

**FELIPE VICENTINO SALVADOR**

**MODELOS ESTATÍSTICOS ESPACIAIS NA ANÁLISE DE DADOS DE  
EXPERIMENTOS DE CAMPO DO MELHORAMENTO DO FEIJOEIRO**

Tese apresentada à Universidade Federal de Viçosa,  
como parte das exigências do Programa de Pós-  
Graduação em Genética e Melhoramento, para  
obtenção do título de *Doctor Scientiae*.

Orientador: Pedro Crescêncio Souza Carneiro

Coorientadores: José Eustáquio de Souza Carneiro  
Antônio Policarpo Souza Carneiro

**VIÇOSA – MINAS GERAIS  
2022**

Ficha catalográfica elaborada pela Biblioteca Central da Universidade  
Federal de Viçosa - Campus Viçosa

T

S182m  
2022

Salvador, Felipe Vicentino, 1992-

Modelos estatísticos espaciais na análise de dados de experimentos de campo do melhoramento do feijoeiro. / Felipe Vicentino Salvador. – Viçosa, MG, 2022.

1 tese eletrônica (75 f.): il. (algumas color.).

Inclui apêndices.

Orientador: Pedro Crescêncio Souza Carneiro.

Tese (doutorado) - Universidade Federal de Viçosa, Departamento de Biologia Geral, 2022.

Inclui bibliografia.

DOI: <https://doi.org/10.47328/ufvbbt.2022.512>

Modo de acesso: World Wide Web.

1. Feijão - Melhoramento genético - Métodos estatísticos.  
2. Análise espacial (Estatística). I. Carneiro, Pedro Crescêncio Souza, 1966-. II. Universidade Federal de Viçosa. Departamento de Biologia Geral. Programa de Pós-Graduação em Genética e Melhoramento. III. Título.

CDD 22. ed. 631.520727

Bibliotecário(a) responsável: Euzébio Luiz Pinto CRB-6/3317

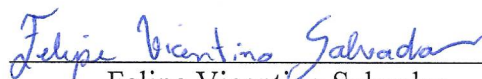
FELIPE VICENTINO SALVADOR

**MODELOS ESTATÍSTICOS ESPACIAIS NA ANÁLISE DE DADOS DE  
EXPERIMENTOS DE CAMPO DO MELHORAMENTO DO FEIJOEIRO**

Tese apresentada à Universidade Federal de Viçosa,  
como parte das exigências do Programa de Pós-  
Graduação em Genética e Melhoramento, para  
obtenção do título de *Doctor Scientiae*.

APROVADA: 19 de julho de 2022.

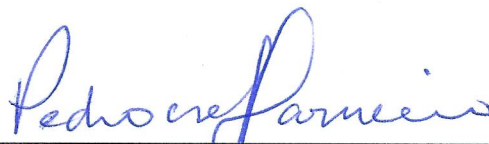
Assentimento:



---

Felipe Vicentino Salvador

Autor



---

Pedro Crescêncio Souza Carneiro

Orientador

## AGRADECIMENTOS

Agradeço primeiramente a Deus, por sua infinita graça e por seu agir ao longo da minha vida.

À minha amada esposa Natália, por todo o seu apoio, amor e carinho incondicionais, por sempre me incentivar a acreditar em mim e nos meus sonhos.

Aos meus pais Luís e Ângela, meus irmãos Gabriel e Renata por todo seu amor, carinho, conselhos e ensinamentos em todos os momentos.

A todos os professores que contribuíram para a minha formação desde a infância até este momento.

À Universidade Federal de Viçosa por me abrir portas desde a graduação.

Ao Programa de Pós-Graduação em Genética e Melhoramento (PPGGM), por toda sua estrutura, corpo docente e pela oportunidade de realizar o doutorado.

Ao meu orientador, professor Pedro Crescêncio Souza Carneiro, por sua excelente orientação, por seu apoio, amizade e confiança.

Aos meus coorientadores, professor José Eustáquio de Souza Carneiro e professor Antônio Policarpo Souza Carneiro pela orientação e apoio.

Aos colegas do Programa Feijão, por toda ajuda na condução dos experimentos, coleta de dados, elaboração de croquis das áreas experimentais, além da amizade, convivência e oportunidade de aprender com todos.

Ao laboratório de Biometria, por toda a acolhida, suporte e amizade.

Aos membros das bancas de qualificação e defesa de tese por suas contribuições ao trabalho e disponibilidade.

Aos funcionários do laboratório do programa Feijão e da estação experimental de Coimbra, Gilberto, Pereira, Sebastião, Beto, Carlinhos, Potoca, Nilson, Fonseca e Douglas, por todo auxílio na realização dos experimentos.

Aos funcionários da secretaria do PPGGM, Marco Tulio e Odilon, por toda a disponibilidade e assistência.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

A todos que, de alguma forma, contribuíram para a realização deste trabalho.

Muito obrigado!

## RESUMO

SALVADOR, Felipe Vicentino, D.Sc., Universidade Federal de Viçosa, julho de 2022. **Modelos estatísticos espaciais na análise de dados de experimentos de campo do melhoramento do feijoeiro**. Orientador: Pedro Crescêncio Souza Carneiro. Coorientadores: José Eustáquio de Souza Carneiro e Antônio Policarpo Souza Carneiro.

Experimentos de campo são realizados com grande frequência em programas de melhoramento de espécies como o feijão-comum (*Phaseolus vulgaris* L.) visando a obtenção acurada dos valores genotípicos. Modelos estatísticos que consideram a informação de posição das parcelas junto à utilização dos delineamentos experimentais podem elevar a eficiência da análise de dados. Assim, o objetivo deste trabalho foi avaliar a eficiência de modelos espaciais baseados na estrutura autorregressiva dos resíduos e do modelo SpATS (análise espacial de experimentos de campo com splines) para captar tendências espaciais em experimentos de feijão-comum, melhorar a predição e a classificação genotípica. O capítulo I foi publicado no periódico Crop Science (Corrigindo dados experimentais para tendências espaciais em um programa de melhoramento de feijão-comum, Salvador, F.V. et al., Crop Science 62/2, Copyright © 2022 The Authors. Crop Science © 2022 Crop Science Society of America, Wiley Periodicals LLC. DOI: <https://doi.org/10.1002/csc2.20703>). Neste capítulo, o modelo tradicional de análise com efeitos dos delineamentos experimentais e erros independentes é comparado a outros modelos que incorporam matrizes de (co)variância residual autorregressiva de primeira ordem no sentido de linhas e colunas para captar tendências locais, assim como efeitos fixos lineares e aleatórios para linhas e colunas visando captar tendências globais e o erro de mensuração. No capítulo II, o modelo tradicional de análise é, desta vez, comparado ao modelo SpATS, que assume erros independentes, porém incorpora uma superfície suavizada capaz de captar tendências globais e locais em um único passo. Em ambas as abordagens estudadas, a análise espacial mostrou-se mais eficiente em relação à forma tradicional de análise, captando tendências espaciais existentes na maioria dos experimentos considerados. Como principais consequências, foram frequentemente observados aumentos nas estimativas de herdabilidade e acurácia ou, pelo menos, reduções na variância do erro de predição, assim como modificações na classificação genotípica. O modelo SpATS apresentou maior praticidade de análise e, além disto, foi possível captar tendências com este modelo em experimentos para os quais não se obteve sucesso com a abordagem autorregressiva. Assim, conclui-se que o uso de modelos espaciais na análise de dados experimentais é um procedimento eficiente e passível de ser utilizado frequentemente no programa de melhoramento da cultura.

Palavras-chave: *Phaseolus vulgaris* L.. Delineamentos experimentais. Análise espacial.  
Estrutura autorregressiva. SpATS.

## ABSTRACT

SALVADOR, Felipe Vicentino, D.Sc., Universidade Federal de Viçosa, July, 2022. **Spatial statistical models in analysis of data from common bean breeding trials**. Adviser: Pedro Crescêncio Souza Carneiro. Co-advisers: José Eustáquio de Souza Carneiro and Antônio Policarpo Souza Carneiro.

Field experiments are frequently carried out in breeding programs for species such as common bean (*Phaseolus vulgaris* L.) to obtain accurate genotypic values. Statistical models that consider the information of plot position jointly with the use of experimental designs may increase the efficiency of data analysis. Thus, the objective of this work was to evaluate the efficiency of spatial models based on the autoregressive residual structure and the SpATS model (Spatial Analysis of Field Experiments with Splines) to capture spatial trends in common bean field trials, improve prediction and genotypic classification. The chapter I was published in the Crop Science journal (Correcting experimental data for spatial trends in a common bean breeding program, Salvador, F.V. et al., Crop Science 62/2, Copyright © 2022 The Authors. Crop Science © 2022 Crop Science Society of America, Wiley Periodicals LLC. DOI: <https://doi.org/10.1002/csc2.20703>). In this chapter, the traditional model of analysis with effects of the experimental design and independent errors is compared to other models that include first-order autoregressive residual (co)variance matrices along rows and columns to capture local trends as well as linear fixed and random effects based on rows and columns to capture global trends and the measurement error. In chapter II, the traditional analysis model is, this time, compared to the SpATS model, which assumes independent errors, but incorporates a smoothing surface able to capture global and local trends in a single step. In both studied approaches, the spatial analysis proved to be more efficient in relation to the traditional form of analysis, capturing spatial trends existing in most of the considered experiments. As main consequences, increases in heritability and accuracy estimates or, at least reductions in prediction error variance, as well as changes in genotypic classification, were frequently observed. The SpATS model presented greater practicality for analysis, and, in addition, it was possible to capture trends with this model in experiments for which the autoregressive approach was not successful. Thus, the use of spatial models in the analysis of experimental data is an efficient procedure to be frequently used in the common bean breeding program.

Keywords: *Phaseolus vulgaris* L.. Experimental designs. Spatial analysis. Autoregressive structure. SpATS.

## SUMÁRIO

INTRODUÇÃO GERAL .....	8
REFERÊNCIAS .....	11
CAPÍTULO I: CORRIGINDO DADOS EXPERIMENTAIS PARA TENDÊNCIAS ESPACIAIS EM UM PROGRAMA DE MELHORAMENTO DE FEIJÃO-COMUM .....	13
Resumo .....	14
1 Introdução .....	15
2 Material e Métodos .....	16
4 Resultados .....	23
5 Discussão .....	31
6 Referências .....	35
7 Material suplementar .....	40
CAPÍTULO II: ANÁLISE ESTATÍSTICA ESPACIAL DE EXPERIMENTOS DE CAMPO NO MELHORAMENTO DO FEIJOEIRO – MODELO SPATS .....	46
Resumo .....	47
1 Introdução .....	47
2 Material e Métodos .....	49
3 Resultados e Discussão .....	53
4 Referências .....	61
5 Material suplementar .....	64
CONCLUSÕES GERAIS .....	66
APÊNDICES .....	68

## INTRODUÇÃO GERAL

O feijão-comum (*Phaseolus vulgaris* L.) é uma leguminosa de grande importância e que faz parte da dieta básica da população brasileira. Diante da alta demanda, os melhoristas têm empenhado constantes esforços no desenvolvimento de novas linhagens que superem agronomicamente àquelas atualmente cultivadas pelos agricultores (RAMALHO; ABREU, 2015). Os trabalhos voltados para o melhoramento genético da cultura são realizados principalmente por instituições públicas como a EMBRAPA, UFV, UFLA, EPAMIG, IAC, IDR-Paraná, IPA, EMCAPA, dentre outros. Neste cenário, é crescente a exigência por materiais que, além de altamente produtivos, apresentem resistência aos principais patógenos que assolam a cultura, com plantas mais eretas visando a colheita mecanizada e grãos que atendam às especificidades do mercado consumidor quanto a cor, tamanho, formato e aspectos culinários (RAMALHO; ABREU, 2015; ASSEFA et al., 2019).

Assim como em qualquer atividade de pesquisa agrícola, as tomadas de decisão nas diferentes etapas de um programa de melhoramento genético vegetal são baseadas na experimentação. Com seus experimentos, os melhoristas esperam separar as variações de origem genética daquelas que ocorrem devido a efeitos ambientais ou ao acaso (RAMALHO; FERREIRA; OLIVEIRA, 2012). Isto se deve pois, segundo Pimentel-Gomes (2009), a influência da variação devido a fatores não controlados, mesmo que passíveis de controle, dificulta a inferência acerca de uma hipótese e, em consequência, pode alterar completamente a interpretação dos resultados obtidos a partir da análise dos dados.

Visando estimar fidedignamente a variação ao acaso nos experimentos e permitir que as inferências se baseiem apenas nos efeitos dos tratamentos comparados, Fisher (1926) definiu o que chamamos hoje de princípios básicos da experimentação. A repetição dos tratamentos possibilita mensurar a variação devida ao acaso; a casualização dá a qualquer parcela a mesma probabilidade de receber qualquer tratamento e valida a variação devida ao acaso estimada pela repetição, uma vez que evita a alocação sistemática de tratamentos a unidades experimentais mais ou menos favoráveis; e, por último, o controle local na casualização, utilizado quando as unidades experimentais não estão sob condições homogêneas devido a fatores controláveis pelo pesquisador e que, quando adotado, visa reduzir a estimativa da variância devida ao acaso (PIMENTEL GOMES, 2009; CASLER, 2015).

Nas diferentes etapas do melhoramento do feijão são conduzidos vários experimentos de campo cujos objetivos incluem a comparação de populações segregantes, progênies endogâmicas ou linhagens, além de estimar parâmetros genéticos para avaliar o potencial dos

materiais para a seleção. Na fase inicial, após obter as populações segregantes a partir de cruzamentos, os parâmetros genéticos necessários para a escolha das populações de maior potencial são obtidos após a avaliação destas populações em experimentos que, por muitas vezes, são conduzidos com o uso do delineamento em blocos casualizados (FERREIRA et al., 2018; MOURA et al., 2016, 2018). Na condução das populações pelo método *bulk* dentro de progênies ou em seleção recorrente, experimentos são realizados para a avaliação de progênies em sucessivas gerações ( $F_{2:3}$ ,  $F_{2:4}$  e  $F_{2:5}$ , por exemplo) utilizando os delineamentos de blocos incompletos como o látice quadrado (BATISTA et al., 2017; MELO et al., 2020; SALVADOR et al., 2021). O delineamento de blocos aumentados também é uma opção neste estágio do programa, seja pelo grande número de genótipos a serem avaliados ou pela escassez de sementes disponíveis após a obtenção de progênies ou linhagens. Estes mesmos delineamentos são também preferencialmente utilizados nas gerações avançadas de endogamia ( $F_6$  ou  $F_7$ ), quando é derivado um grande número de linhagens ( $F_{6:7}$  ou  $F_{7:8}$ ) para avaliações preliminares, visando indicar genótipos promissores a serem incluídos nos ensaios de VCU.

Apesar de experimentos de campo bem planejados e que utilizem adequadamente os delineamentos experimentais serem, em geral, eficientes para captar a heterogeneidade de fertilidade do solo ao longo da área, existem variações de pequena escala causadas por práticas de manejo e condições do ambiente que não são captadas pelas estratégias de controle local impostas por estes delineamentos (RESENDE; STURION, 2003; BURGUEÑO, 2018). A existência de heterogeneidade espacial devido a fatores como a fertilidade e estrutura do solo, umidade, interceptação de luz etc. (RESENDE; STURION, 2003), contribuem para inflar a variação residual e, conseqüentemente, reduzir a acurácia seletiva, sendo importante fazer uso de metodologias capazes de contabilizar estes efeitos na etapa de análise dos dados (RESENDE; SILVA; AZEVEDO, 2014).

Neste sentido, diversas metodologias foram propostas visando a captação destas heterogeneidades ao longo da área experimental e a correção dos efeitos de tratamentos. Dentre as pioneiras, podem ser citadas o método de Papadakis (PAPADAKIS, 1937) que consiste na correção da observação fenotípica de cada parcela por uma covariável, dada pela média do erro estimado nas parcelas vizinhas. Em contrapartida, no método da média móvel, um raciocínio parecido é utilizado, porém os erros não são utilizados para obter a covariável de correção, mas sim os próprios valores fenotípicos das parcelas (RICKEY, 1924).

Com o desenvolvimento das equações de modelos mistos (HENDERSON, 1959) outras técnicas de análise espacial foram propostas. Gilmour et al. (1997) propôs que tendências espaciais locais existentes em um experimento poderiam ser modeladas por meio dos resíduos

com uma estrutura autorregressiva de primeira ordem no sentido de linhas e colunas (CULLIS e GLEESON, 1991). Além disto, neste processo, a estrutura do modelo linear misto é utilizada em múltiplos passos para incorporar diversos efeitos fixos e aleatórios baseados na posição das parcelas a fim de modelar tendências espaciais (GILMOUR et al., 1997). Este método se tornou a metodologia padrão de análise espacial ao longo dos anos.

Mais recentemente, Rodríguez-Álvarez et al. (2018), propuseram uma metodologia baseada em técnica de suavização (*smoothing*) com P-splines bidimensionais que adiciona ao modelo linear misto uma superfície suavizada ajustada capaz de acomodar as tendências espaciais da área experimental. A principal funcionalidade deste método é a capacidade de modelar tendências locais e globais em um único passo, sendo que os autores desenvolveram um pacote em denominado SpATS para a sua utilização (RODRÍGUEZ-ÁLVAREZ et al., 2018).

Assim, o objetivo deste trabalho foi avaliar a eficiência dos modelos estatísticos com estrutura autorregressiva e SpATS para análise espacial de experimentos do melhoramento do feijoeiro.

## REFERÊNCIAS

- ASSEFA, T. et al. A review of breeding objectives, genomic resources, and marker-assisted methods in common bean (*Phaseolus vulgaris* L.). **Molecular Breeding**, v. 39, n. 20, p. 1-23, fev. 2019. DOI: <https://doi.org/10.1007/s11032-018-0920-0>
- BATISTA, L. G. et al. Multigeneration index in the selection of common bean inbred families. **Crop Science**, v. 57, n. 5, p. 2354–2360, set./out. 2017. DOI: <https://doi.org/10.2135/cropsci2016.08.0665>
- BURGUENÑO, J. Spatial Analysis of Field Experiments. In: GLAZ, B.; YEATER, K. M. **Applied Statistics in Agricultural, Biological, and Environmental Sciences**. Madison, WI: ASA, CSSA, and SSSA Books, 2018. p. 319-344. DOI: <https://doi.org/10.2134/appliedstatistics.2016.0011.c12>
- CASLER, M. D. Fundamentals of Experimental Design: Guidelines for Designing Successful Experiments. **Agronomy Journal**, v. 107, p. 692-705. 2015. DOI: <https://doi.org/10.2134/agronj2013.0114>
- CULLIS, B. R.; GLEESON, A. C. Spatial Analysis of Field Experiments-An Extension to Two Dimensions. **Biometrics**, v. 47, n. 4, p. 1449, dez. 1991.
- FERREIRA, L. U. et al. Combining ability as a strategy for selecting common bean parents and populations resistant to white mold. **Crop Breeding and Applied Biotechnology**, v. 18, n. 3, p. 276–283, jan. 2018. DOI: <https://doi.org/10.1590/1984-70332018v18n3a41>
- FISHER, R.A. The arrangement of field experiments. **Journal of the Ministry of Agriculture, Great Britain**, v. 33, p. 503–513. 1926. DOI: <https://doi.org/10.23637/rothamsted.8v61q>
- GILMOUR, A. R.; CULLIS, B. R.; VERBYLA, A. P. Accounting for Natural and Extraneous Variation in the Analysis of Field Experiments. **Journal of Agricultural, Biological, and Environmental Statistics**, v. 2, n. 3, p. 269–293, set. 1997. DOI: <https://doi.org/10.2307/1400446>
- HENDERSON, C. R. et al. The estimation of environmental and genetic trends from records subject to culling. **Biometrics**, v. 15, 192-218. 1959.
- MELO, V. L. DE et al. Modeling (co)variance structures for genetic and non-genetic effects in the selection of common bean progenies. **Euphytica**, v. 216, n. 5, abr. 2020. DOI: <https://doi.org/10.1007/s10681-020-02607-9>
- MOURA, L. M. et al. Diallel analysis to choose parents for black bean (*Phaseolus vulgaris* L.) breeding. **Genetics and Molecular Research**, v. 15, n. 3, p. 1–12, ago. 2016. DOI: <https://doi.org/10.4238/gmr.15038574>
- MOURA, L. M. et al. Combining ability of common bean parents in different seasons, locations and generations. **Euphytica**, v. 214, n. 10, set. 2018. DOI: <https://doi.org/10.1007/s10681-018-2259-3>

PAPADAKIS, J. S. **Méthode statistique pour des expériences sur champ**. Thessalonike: Institut d'Amélioration des Plantes à Salonique. 1937. 30p.

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

RAMALHO, M. A. P.; ABREU, A. F. B. Obtenção de cultivares. In: CARNEIRO, J. E. S.; PAULA JÚNIOR, T. J.; BORÉM, A. **Feijão: do plantio à colheita**. Viçosa: Editora UFV, 2015. p. 96-114.

RAMALHO, M. A. P.; FERREIRA, D. F.; OLIVEIRA, A. C. **Experimentação em Genética e Melhoramento de Plantas**. 3. Ed. Lavras: Editora UFLA. 2012. 329p.

RESENDE, M.D.V.; SILVA, F.F.; AZEVEDO, C.F. **Estatística matemática, biométrica e computacional: modelos mistos, multivariados, categóricos e generalizados (REML/BLUP), inferência Bayesiana, regressão aleatória, seleção genômica, QTL-GWAS, estatística espacial e temporal, competição, sobrevivência**. Viçosa, MG: Suprema, 2014. 881p.

RESENDE, M.D.V.; STURION, J.A. Análise estatística espacial de experimentos via modelos mistos individuais com erros modelados por processos ARIMA em duas dimensões. **Rev. Mat. Estat.**, v. 21, n. 1, p. 7-33, jan. 2003.

RODRÍGUEZ-ÁLVAREZ et al. Correcting for spatial heterogeneity in plant breeding experiments with P-splines. **Spatial Statistics**, v. 23, p. 52-71. 2018. DOI: <https://doi.org/10.1016/j.spasta.2017.10.003>

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

SALVADOR F. V. et al. Accounting for heterogeneous genotypic variance of progenies within populations in autogamous plant breeding: An evaluation in common bean. **Crop Science**, v. 6, n. 6, p. 3947-3961. 2021. DOI: [10.1002/csc2.20590](https://doi.org/10.1002/csc2.20590)

SALVADOR, F. V., et al. Correcting experimental data for spatial trends in a common bean breeding program. **Crop Science**, v. 62, n. 2, p. 825-838. 2022. DOI: <https://doi.org/10.1002/csc2.20703>

## CAPÍTULO I

### **CORRIGINDO DADOS EXPERIMENTAIS PARA TENDÊNCIAS ESPACIAIS EM UM PROGRAMA DE MELHORAMENTO DE FEIJÃO-COMUM<sup>1</sup>**

---

<sup>1</sup>Este capítulo está publicado em inglês no seguinte artigo científico:  
Correcting experimental data for spatial trends in a common bean breeding program, Salvador, F.V. et al., *Crop Science* 62/2, Copyright © 2022 The Authors. *Crop Science* © 2022 Crop Science Society of America, Wiley Periodicals LLC. O artigo pode ser acessado em: <https://doi.org/10.1002/csc2.20703>.

## Resumo

No melhoramento genético do feijão-comum (*Phaseolus vulgaris* L.), vários experimentos são realizados em condições de campo para prever os valores genotípicos, mas os delineamentos experimentais podem não ser suficientes para capturar a heterogeneidade na área experimental. O objetivo deste trabalho foi avaliar o potencial de modelos espaciais em corrigir os dados de um programa de melhoramento de feijão-comum para tendências espaciais e melhorar a predição dos valores genotípicos. Utilizamos dados reais de 19 experimentos de campo de um programa melhoramento de feijão-comum e três delineamentos experimentais. O modelo estatístico tradicional com os efeitos dos delineamentos e erros independentes foi ajustado e usado como modelo básico. Em seguida, ajustamos uma sequência de modelos espaciais para incluir diferentes estruturas de (co)variância residual para tendências locais, além de efeitos fixos e aleatórios baseados nas informações da posição das parcelas para capturar tendências globais e extrínsecas. O modelo básico e o modelo espacial de melhor ajuste foram comparados em relação às estimativas de herdabilidade, acurácia, variância de erro de predição e discordância nos genótipos mais bem classificados. Na maioria dos casos, o uso de modelos espaciais melhorou as estimativas de herdabilidade e acurácia ou, pelo menos, reduziu as estimativas de variância de erro de predição. Além disso, mudanças na classificação dos valores genotípicos foram observadas. Como nenhum modelo apresentou o melhor ajuste para todos os experimentos, alguns dos modelos testados foram recomendados para experimentos futuros com base nos padrões das tendências espaciais observadas. Assim, o uso de modelos espaciais ajudou a melhorar os análise de dados e a predição de valores genotípicos capturando a heterogeneidade nos nossos experimentos de campo de feijão-comum.

## 1 Introdução

O feijão-comum (*Phaseolus vulgaris* L.) é uma cultura econômica e socialmente importante. É considerada um alimento básico em vários países em desenvolvimento (GEPTS et al., 2008) e uma importante fonte de proteínas, carboidratos, fibra dietética (LOS et al., 2018) e minerais essenciais, como ferro e zinco (CARVALHO et al., 2012; IZQUIERDO et al., 2018). O Brasil é o maior produtor desta leguminosa e o consumo médio per capita em 2013 foi de 16 kg (BANSAL et al., 2019). Dada a sua relevância, os melhoristas de feijão têm incessantemente empenhado esforços para desenvolver novos genótipos com desempenhos superiores para rendimento de grãos e outros atributos agrônômicos, resistência/tolerância a estresses bióticos e abióticos e qualidade nutricional e de cozimento (ASSEFA et al., 2019).

O processo de melhoramento é realizado com base na avaliação de campo de populações, progênies, linhagens etc. Isso é necessário para predizer os valores genotípicos dos candidatos à seleção, o que deve ser realizado com a maior precisão possível (PIEPHO et al., 2008; RESENDE e ALVES, 2020; RESENDE e DUARTE, 2007). Para isso, delineamentos experimentais são usados para estimar e validar o erro experimental (CASLER, 2015). Um exemplo é o delineamento de blocos completos casualizados (DBC), que é usado especialmente quando o número total de tratamentos (genótipos e cultivares testemunhas) é pequeno e quando heterogeneidade no campo permite a alocação de todos os tratamentos em blocos homogêneos. Para a maioria das situações, como aquelas referentes a avaliações de muitas progênies ou linhagens, esta condição é provavelmente impossível de se atender, de modo que os melhoristas preferem utilizar delineamentos com blocos incompletos (RAMALHO et al., 2012; ZYSTRO et al., 2018). Nesses casos, os experimentos de campo podem ocorrer com ou sem repetições devido à escassez de sementes logo após a extração de progênies ou linhagens. Assim, alguns delineamentos comumente utilizados são o látice quadrado (DLQ) (YATES, 1936) e o de blocos aumentados (DBA) (FEDERER, 1956).

Apesar da capacidade dos delineamentos experimentais de lidar com a heterogeneidade do solo existente na área, geralmente há variações de pequena escala causadas por práticas experimentais e condições ambientais que não são capturadas pelo controle local na casualização desses delineamentos (BURGUEÑO, 2018; RESENDE e STURION, 2003). Assim, a inclusão de informações de posição das parcelas para modelar a estrutura de (co)variância autorregressiva de primeira ordem (AR1) ao longo de linhas e colunas no grid experimental (CULLIS e GLEESON, 1991; GILMOUR et al., 1997; GLEESON e CULLIS, 1987) tem se mostrado eficiente para capturar tendências locais. Além disso, as tendências espaciais globais e variações extrínsecas, como as causadas por operações de campo, podem

ser capturadas com a inclusão de efeitos fixos e aleatórios baseados nas posições das parcelas no grid experimental (GILMOUR et al., 1997). Os modelos espaciais podem ser úteis para levar em conta a provável correlação entre parcelas adjacentes de cada experimento, para melhorar a acurácia de predição dos efeitos genotípicos e reduzir o viés na seleção dos genótipos (DUARTE e VENCOVSKY, 2005).

Assim, a modelagem de tendências espaciais com base em estrutura de (co)variância autorregressiva de primeira ordem em conjunto com o uso de delineamentos experimentais tem sido uma estratégia vantajosa para controlar a heterogeneidade espacial na análise de dados de experimentos de campo para várias espécies, como a cevada (*Hordeum vulgare* L.) (SMITH, CULLIS, THOMPSON, 2001), algodão (*Gossypium hirsutum* L.) (LIU et al., 2015), tremoço (*Lupinus* spp.) (STEFANOVA et al., 2009), milho (*Zea mays* L.) (ARIEF et al., 2019), ervilha (*Pisum sativum* L.) (COWLING et al., 2015), batata (*Solanum tuberosum* L.) (ANDRADE et al., 2020), soja [*Glycine max* (L.) Merr.] (BERNARDELI et al., 2021) e trigo (*Triticum aestivum* L.) (BORGES et al., 2019; STEFANOVA et al., 2009). No entanto, há necessidade de mais estudos sobre seu uso no melhoramento do feijão-comum. Assim, o objetivo deste trabalho foi avaliar o potencial desses modelos espaciais para corrigir dados de experimentos de campo para tendências espaciais e melhorar a predição de valores genotípicos em um programa de melhoramento de feijão-comum.

## **2 Material e Métodos**

### **2.1 Experimentos de campo**

Utilizamos os dados fenotípicos de 19 experimentos de campo do programa de melhoramento de feijão-comum da Universidade Federal de Viçosa, Brasil. Uma descrição completa de cada experimento é mostrada na Tabela 1. A principal estratégia de melhoramento deste programa de melhoramento genético é a seleção recorrente e os 19 experimentos são representativos de diferentes objetivos de melhoramento, ciclos de seleção recorrente ou gerações endogâmicas. Portanto, esses experimentos não estão conectados, exceto por algumas cultivares testemunhas comuns que podem ocorrer em alguns deles. A maioria dos experimentos se referiam a avaliações de progênies com o objetivo de selecionar as superiores para recombinação ou extração de linhagens endogâmicas (experimentos J225, G45, G62, D400, E225, F550, A45, F100, B100 e B500). Outros experimentos foram avaliações preliminares de linhagens endogâmicas extraídas após ciclos de seleção recorrente (experimentos C400, P400, J64 e H650). Quatro experimentos (VI18, VI21, VA18 e VA21)

foram ensaios para valor de cultivo e uso, e um experimento (S112) foi uma avaliação de todas as cultivares de feijão-comum lançadas no Brasil nos últimos 50 anos.

Os 19 experimentos ocorreram em cinco diferentes safras: inverno/2017 (experimentos C400 e P400), inverno/2019 (experimento S112), águas/2019 (experimento J225), inverno/2020 (experimentos G45, G62, J64, VI18, VI21, D400, E225, H650 e F550), e águas/2020 (experimentos A45, F100, VA18, VA21, B100 e B500). Todos eles foram realizados na estação experimental pertencente ao Departamento de Agronomia da Universidade de Viçosa localizada em Coimbra, Minas Gerais, Brasil (20°83' S, 42°73' W, 720 m de altitude). O delineamento experimental (DBC, DLQ ou DBA) usado para cada experimento foi escolhido de acordo com o número de tratamentos a serem avaliados, informações prévias sobre a heterogeneidade da área e disponibilidade de sementes para repetir os tratamentos. Os tamanhos das parcelas variaram entre os experimentos (Tabela 1), mas os sulcos no solo foram preparados a cada 0,5 m de distância, e uma densidade de sementes de 15 sementes  $m^{-1}$  foi utilizada para semear os experimentos.

Coletamos as informações das posições das parcelas de acordo com um grid retangular dado pelo número de linhas (N<sub>Lin</sub>) e colunas (N<sub>Col</sub>) para todos os experimentos (Tabela 1). Cada linha do grid era representada por um conjunto de parcelas paralelas aos sulcos do solo. Cada coluna foi representada por um grupo de parcelas perpendiculares aos sulcos do solo. As práticas culturais e de manejo ocorreram de acordo com às recomendações técnicas para o cultivo do feijão-comum na região (CARNEIRO et al., 2015). Cada parcela foi colhida individualmente, e a produtividade de grãos (PROD) foi computada ( $g\text{ parcela}^{-1}$ ) e convertida em  $kg\text{ ha}^{-1}$ .

**Tabela 1** - Descrição dos 19 experimentos de campo de feijão-comum de acordo com seus delineamentos experimentais, tipo de material genético avaliado, número total de tratamentos (NTrat) subdividido em genótipos (NGen) e testemunhas (Ntest), número de repetições (NRep), linhas (NLin), colunas (NCol) e o tamanho das parcelas.

Experimento	Delineamento <sup>a</sup>	Material Genético	NTrat	NGen	NTest	NRep <sup>b</sup>	NLin	NCol	Tamanho de parcela <sup>c</sup>
S112	DBC	Cultivares	112	112	0	3	16	21	2 linhas de 3 m
G45	DBC	Progênes F <sub>2:8</sub>	45	38	7	3	21	7	2 linhas de 2 m
G62	DBC	Progênes F <sub>2:6</sub>	62	58	4	3	21	9	2 linhas de 1 m
J64	DBC	Linhagens F <sub>6:10</sub>	64	59	5	3	21	10	2 linhas de 1 m
VI18	DBC	Linhagens e multilinhas	18	13	5	3	18	3	4 linhas de 4 m
VI21	DBC	Linhagens e multilinhas	21	19	2	3	21	3	4 linhas de 4 m
A45	DBC	Progênes F <sub>2:9</sub>	45	38	7	3	18	8	2 linhas de 2 m
F100	DBC	Progênes F <sub>2:6</sub>	100	96	4	3	45	7	1 linha de 2 m
VA18	DBC	Linhagens e multilinhas	18	13	5	3	18	3	4 linhas de 4 m
VA21	DBC	Linhagens e multilinhas	21	19	2	3	21	3	4 linhas de 4 m
C400	DLQ	Linhagens F <sub>6:7</sub>	400	380	20	2	20	40	1 linha de 1 m
P400	DLQ	Linhagens F <sub>5:6</sub>	400	395	5	2	40	20	1 linha de 1 m
J225	DLQ	Progênes F <sub>2:4</sub>	225	220	5	3	21	33	2 linhas de 2 m
D400	DLQ	Progênes F <sub>3:5</sub>	400	385	15	2	20	40	2 linhas de 1 m
E225	DLQ	Progênes F <sub>3:7</sub> e F <sub>1:5</sub>	225	220	5	2	21	22	2 linhas de 1 m
B100	DLQ	Progênes F <sub>4:8</sub>	100	95	5	3	45	7	1 linha de 2 m
F550	DBA	Progênes F <sub>2:3</sub>	555	550	5	22	30	22	1 linha de 1 m
H650	DBA	Linhagens F <sub>5:6</sub>	654	650	4	13	39	18	1 linha de 1 m
B500	DBA	Progênes F <sub>2:6</sub>	504	500	4	10	45	12	1 linha de 1 m

<sup>a</sup>DBA, Delineamento em blocos aumentados; DBC, Delineamento em blocos casualizados; DLQ, Delineamento látice quadrado.

<sup>b</sup>O número de repetições (NRep) diz respeito a todos os tratamentos nos delineamentos DBC e DLQ, e somente às testemunhas no delineamento DBA.

<sup>c</sup>As linhas de cada parcela foram espaçadas em 0,5 m e a densidade de sementes no plantio foi de 15 sementes por metro.

### 3.2 Análises estatísticas

Utilizamos a metodologia de modelos mistos (HENDERSON et al., 1959) para prever os valores genotípicos com base nos dados fenotípicos coletados em cada experimento de campo.

#### 3.2.1 Modelo básico

A análise dos dados para cada experimento começou a partir de um modelo básico, que representa o modelo de análise tradicional com os efeitos associados ao respectivo delineamento experimental, resíduos independentes e sem correlação entre parcelas adjacentes.

Para os experimentos instalados no delineamento DBC, o modelo básico (modelo 1) utilizado é mostrado abaixo (Equação 1):

$$\mathbf{y} = \mathbf{X}\boldsymbol{\tau} + \mathbf{Z}_g\mathbf{u}_g + \mathbf{e} \quad (1)$$

em que  $\mathbf{y}$  é o vetor de observações fenotípicas,  $\boldsymbol{\tau}$  é o vetor de efeitos fixos de repetições somados à média geral, com matriz de incidência  $\mathbf{X}$ ;  $\mathbf{u}_g$  é o vetor de efeitos aleatórios de genótipos (linhagens, progênies, cultivares etc.) incluindo as testemunhas, com matriz de incidência  $\mathbf{Z}_g$  e  $\mathbf{e}$  é o vetor de resíduos aleatórios.

Para os experimentos instalados em delineamentos DLQ e DBA, o modelo básico (modelo 1) foi alterado para o seguinte (Equação 2):

$$\mathbf{y} = \mathbf{X}\boldsymbol{\tau} + \mathbf{Z}_g\mathbf{u}_g + \mathbf{Z}_b\mathbf{u}_b + \mathbf{e} \quad (2)$$

Em que  $\mathbf{y}$  é o vetor de observações fenotípicas,  $\boldsymbol{\tau}$  é o vetor de efeitos fixos de repetições somadas à média geral, com matriz de incidência  $\mathbf{X}$ ;  $\mathbf{u}_g$  é o vetor de efeitos aleatórios de genótipos (linhagens, progênies, cultivares etc.) incluindo as testemunhas, com matriz de incidência  $\mathbf{Z}_g$ ;  $\mathbf{u}_b$  é o vetor de efeitos aleatórios de blocos incompletos, com matriz de incidência  $\mathbf{Z}_b$  e  $\mathbf{e}$  é o vetor de resíduos aleatórios.

As seguintes distribuições foram assumidas para os efeitos aleatórios nos modelos básicos:  $\mathbf{u}_g \sim N(\mathbf{0}, \mathbf{G})$ ,  $\mathbf{u}_b \sim N(\mathbf{0}, \mathbf{B})$  e  $\mathbf{e} \sim N(\mathbf{0}, \mathbf{R})$ , em que  $\mathbf{G} = \sigma_g^2 \mathbf{I}_g$ ,  $\mathbf{B} = \sigma_b^2 \mathbf{I}_b$  e  $\mathbf{R} = \sigma^2 [\mathbf{I}_r \otimes \mathbf{I}_c]$ .  $\sigma_g^2$  é a variância genotípica,  $\sigma_b^2$  é a variância entre blocos incompletos e  $\sigma^2$  é a variância residual.  $\mathbf{I}_g$ ,  $\mathbf{I}_b$ ,  $\mathbf{I}_r$ , e  $\mathbf{I}_c$  são matrizes identidade cuja ordem equivale ao número de genótipos, blocos, linhas (índice  $r$  – *rows*) (NLin) e colunas (índice  $c$  – *columns*) (NCol) no grid experimental (Tabela 1), respectivamente.

### 3.2.2 Modelos espaciais

Para investigar as tendências espaciais existentes, depois do modelo básico, os dados fenotípicos de cada experimento foram analisados utilizando uma sequência de modelos espaciais dividida em três etapas, conforme apresentado na tabela 2. Na etapa I, além do modelo básico (modelo 1), ajustamos os modelos 2, 3 e 4 (Tabela 2) que diferem do modelo anterior pois os resíduos são assumidos como correlacionados ( $\mathbf{e} = \boldsymbol{\xi}$ ) com uma estrutura de (co)variância residual ( $\mathbf{R}$ ) para acomodar tendências espaciais locais [ $\boldsymbol{\xi} \sim N(\mathbf{0}, \mathbf{R})$ ] (GILMOUR et al., 1997). No modelo 2, os resíduos são assumidos correlacionados por uma estrutura de (co)variância autorregressiva de primeira ordem no sentido de linhas ( $\mathbf{R} = \sigma_{\xi}^2 \left[ \sum_r (\rho_r) \otimes \mathbf{I}_c \right]$ ); no modelo 3, no sentido de colunas ( $\mathbf{R} = \sigma_{\xi}^2 \left[ \mathbf{I}_r \otimes \sum_c (\rho_c) \right]$ ) e no modelo 4, no sentido de linhas e colunas simultaneamente ( $\mathbf{R} = \sigma_{\xi}^2 \left[ \sum_r (\rho_r) \otimes \sum_c (\rho_c) \right]$ ), onde  $\sigma_{\xi}^2$  é a variância de resíduos correlacionados,  $\sum_r (\rho_r)$  e  $\sum_c (\rho_c)$  são as matrizes de correlação autorregressiva de primeira ordem, com  $\rho_r$  e  $\rho_c$  como parâmetros de autocorrelação nas direções espaciais de linhas e colunas, respectivamente. O modelo com o melhor ajuste na etapa I também apresentou a estrutura de (co)variância residual mais adequada. Portanto, esta estrutura residual foi selecionada para ser utilizada nos modelos ajustados nas próximas etapas (Tabela 2).

Na etapa II, os modelos 5 a 19 incluíram a estrutura residual selecionada na etapa I, mas também termos para capturar tendências espaciais globais e variações extrínsecas ao longo de linhas e colunas (Tabela 2). Esses termos foram representados pelos efeitos fixos de regressão linear da produtividade sobre o índice das linhas [lin(linhas)] e sobre o índice das colunas [lin(colunas)] e também pelos efeitos aleatórios de linhas e colunas.

A etapa III foi somente necessária se o modelo com o melhor ajuste na etapa I apresentasse o vetor de resíduos correlacionados ( $\boldsymbol{\xi}$ ), isto é, os modelos 2, 3 ou 4. Neste caso, os modelos 20 a 35 foram ajustados (Tabela 2). Estes modelos apresentam os mesmos efeitos fixos e aleatórios que o modelo escolhido na etapa I e que os modelos 5 a 19, mas também incluíram o vetor de resíduos independentes ( $\boldsymbol{\eta}$ ), comumente denominado como efeito nugget. Portanto, nos modelos da etapa III, os resíduos são modelados em duas partes: os resíduos correlacionados e os independentes ( $\mathbf{e} = \boldsymbol{\xi} + \boldsymbol{\eta}$ ) (GILMOUR et al, 1997).

**Tabela 2** - Sequência dos modelos básico e espaciais ajustados para a análise de dados dos experimentos de campo do melhoramento do feijoeiro e seus componentes como efeitos fixos e aleatórios e estrutura de (co)variância para os resíduos.

Etapa	Modelo	Tendências espaciais				
		Globais/Extrínsecas		Locais		
		Efeitos fixos <sup>a</sup>	Efeitos aleatórios <sup>b</sup>	Estrutura residual <sup>c</sup>		
I	1			$\sigma^2[\mathbf{I}_r \otimes \mathbf{I}_c]$		
	2			$\sigma^2_{\xi}[\mathbf{\Sigma}_r(\rho_r) \otimes \mathbf{I}_c]$		
	3			$\sigma^2_{\xi}[\mathbf{I}_r \otimes \mathbf{\Sigma}_c(\rho_c)]$		
	4			$\sigma^2_{\xi}[\mathbf{\Sigma}_r(\rho_r) \otimes \mathbf{\Sigma}_c(\rho_c)]$		
II	5	lin(linhas)				
	6	lin(colunas)				
	7		linhas			
	8		colunas			
	9	lin(linhas), lin(colunas)				
	10	lin(linhas)	linhas			
	11	lin(linhas)	colunas	Escolhida na etapa I.		
	12	lin(colunas)	linhas			
	13	lin(colunas)	colunas			
	14		linhas, colunas			
	15	lin(linhas), lin(colunas)	linhas			
	16	lin(linhas), lin(colunas)	colunas			
	17	lin(linhas)	linhas, colunas			
	18	lin(colunas)	linhas, colunas			
	19	lin(linhas), lin(colunas)	linhas, colunas			
	III	20				
		21	lin(linhas)			
22		lin(colunas)				
23			linhas			
24			colunas			
25		lin(linhas), lin(colunas)				
26		lin(linhas)	linhas			
27		lin(linhas)	colunas		Escolhida na etapa I + $\sigma^2_{\eta}[\mathbf{I}_r \otimes \mathbf{I}_c]$	
28		lin(colunas)	linhas			
29		lin(colunas)	colunas			
30			linhas, colunas			
31		lin(linhas), lin(colunas)	linhas			
32		lin(linhas), lin(colunas)	colunas			
33		lin(linhas)	linhas, colunas			
34		lin(colunas)	linhas, colunas			
35		lin(linhas), lin(colunas)	linhas, colunas			

<sup>a</sup> lin(linhas) e lin(colunas): efeito de regressão linear da produtividade nos índices de linha e nos índices de coluna, respectivamente.

<sup>b</sup> linhas and colunas: efeitos aleatórios de linhas e colunas, respectivamente.

<sup>c</sup>  $\sigma^2[\mathbf{I}_r \otimes \mathbf{I}_c]$ : estrutura de (co)variância assumindo resíduos independentes;  $\sigma^2_{\xi}[\mathbf{\Sigma}_r(\rho_r) \otimes \mathbf{I}_c]$ : estrutura de (co)variância residual assumindo correlação autorregressiva de primeira ordem no sentido de linhas;  $\sigma^2_{\xi}[\mathbf{I}_r \otimes \mathbf{\Sigma}_c(\rho_c)]$ : estrutura de (co)variância residual assumindo correlação autorregressiva de primeira ordem no sentido de colunas;  $\sigma^2_{\xi}[\mathbf{\Sigma}_r(\rho_r) \otimes \mathbf{\Sigma}_c(\rho_c)]$ : estrutura de (co)variância residual assumindo correlação autorregressiva de primeira ordem no sentido de linhas e colunas;  $\sigma^2_{\eta}[\mathbf{I}_r \otimes \mathbf{I}_c]$ : estrutura de (co)variância residual para a parte independente dos resíduos (efeito nugget) no modelo espacial.  $\sigma^2$  é a variância residual,  $\sigma^2_{\xi}$  é a variância para a parte correlacionada dos resíduos e  $\sigma^2_{\eta}$  é a variância para parte independente dos resíduos.

A significância dos efeitos fixos baseados nas coordenadas de linhas e colunas foi testada utilizando o teste de Wald. Os efeitos aleatórios de genótipos, linhas e colunas foram testados utilizando o teste de razão de verossimilhança (WILKS, 1938). Os componentes de variância para os efeitos aleatórios e os parâmetros de autocorrelação foram estimados pelo método máxima verossimilhança residual (PATTERSON e THOMPSON, 1971).

Todos os modelos foram ajustados e os efeitos fixos e aleatórios testados utilizando o Echidna Mixed Model Software (GILMOUR, 2020).

### 3.3 Comparações entre modelos estatísticos

O critério de informação de Akaike (AKAIKE, 1974) foi usado para escolha o modelo com o melhor ajuste para cada experimento. Uma vez que os modelos estudados apresentam diferentes efeitos fixos e as verossimilhanças restritas não são comparáveis, utilizamos as estimativas corrigidas do critério de informação de Akaike ( $AIC_c$ ) propostas por Verbyla (2019), que podem ser calculadas da seguinte forma (Equação 3):

$$AIC_c = -2\log L_{Full} + 2(p + q) \quad (3)$$

Em que  $\log L_{Full}$  é o logaritmo da função completa de máxima verossimilhança para as estimativas de máxima verossimilhança residual,  $p$  é o número de parâmetros fixos no modelo e  $q$  é o número de componentes de variância no modelo.

O modelo com melhor ajuste foi aquele com a estimativa mínima de  $AIC_c$ . Para aquelas situações em que a diferença entre os dois menores valores de  $AIC_c$  foi inferior ou igual a duas unidades, escolhemos o modelo com menor número de parâmetros, conforme os critérios relatados em Cavanaugh e Neath (2019).

### 3.4 Estimação da herdabilidade e da acurácia de seleção

Para cada experimento, computamos as estimativas da herdabilidade e de acurácia de seleção para o modelo básico e os modelos espaciais. A herdabilidade ( $h_g^2$ ) foi estimada utilizando o seguinte estimador generalizado (Equação 4) proposto por Cullis et al. (2006).

$$h_g^2 = 1 - \frac{\bar{v}_{BLUP}}{2\sigma_g^2} \quad (4)$$

Em que  $\bar{v}_{BLUP}$  é a média da variância do erro de predição par a par entre duas melhores predições lineares não viesadas (BLUP) dos efeitos de genótipo e  $\sigma_g^2$  é a variância genotípica.

A acurácia de seleção ( $r_{\hat{g}g}$ ) foi calculada de acordo com Resende et al. (2014) conforme a seguir (Equação 5):

$$r_{\hat{g}g} = \sqrt{1 - \frac{PEV}{\sigma_g^2}} \quad (5)$$

Em que  $PEV$  é a média da variância do erro de predição e  $\sigma_g^2$  é a variância genotípica.

### 3.5 Efeito na classificação dos valores genotípicos

Para os experimentos em que houve um modelo especial com melhor ajuste comparado ao modelo básico, os valores genotípicos obtidos destes dois modelos foram classificados e os genótipos mais bem classificados nas faixas de 5%, 10%, 20% e 30% foram comparados. A frequência de genótipos coincidentes (FC) para cada faixa considerada utilizando os valores genotípicos dos modelos básico e espacial foi computada utilizando o índice de coincidência proposto por Hamblin and Zimmerman (1986) (Equação 6):

$$FC(\%) = \left( \frac{A - C}{B - C} \right) \times 100 \quad (6)$$

Em que  $A$  é o número de genótipos coincidentes entre aqueles classificados usando o modelo básico e usando o modelo espacial;  $B$  é o número total de genótipos selecionados considerando a faixa de genótipos mais bem classificados (5%, 10%, 20%, ou 30%), e  $C$  é o número de genótipos coincidentes devido ao acaso, que é dado pelo produto de  $B$  e a faixa de genótipos mais bem classificados (5%, 10%, 20%, ou 30%).

Consequentemente, a frequência de genótipos não coincidentes (FNC) foi computada conforme a seguir (Equação 7):

$$FNC(\%) = 100 - FC(\%) \quad (7)$$

## 4 Resultados

### 4.1 Seleção de modelos e estruturas de covariância residual

Não houve um único modelo que forneceu o melhor ajuste aos dados fenotípicos de todos os experimentos de campo estudados (Tabela 3). Considerando o critério AICc, para apenas 5 experimentos (S112, VI18, VI21, A45 e VA18), todos realizados sob delineamento DBC, o modelo básico foi o modelo com melhor ajuste (Tabela 3), indicando que o delineamento experimental foi suficiente para capturar os níveis de heterogeneidade espacial na área utilizada para esses experimentos. Além disso, a suposição de independência entre os resíduos de parcelas adjacentes foi adequada para esses experimentos.

**Tabela 3** - Comparação entre o modelo básico e o modelo especial selecionado (em caso de melhor ajuste que o modelo básico) com relação à estrutura de (co)variância residual, número de parâmetros fixos ( $p$ ), número de componentes de variância ( $q$ ), logaritmo da função completa de máxima verossimilhança ( $\log L_{Full}$ ) e valores de AIC corrigidos de acordo com Verbyla (2019) ( $AIC_c$ ) para cada experimento de campo de feijão-comum.

Delin <sup>a</sup> .	Exp.	Modelo	Estrutura residual <sup>b</sup>	$p$	$q$	$\log L_{Full}$	$AIC_c$
S112	Básico	1	$\sigma^2[\mathbf{I}_r \otimes \mathbf{I}_c]$	3	2	-2392,61	4795,21
	Espacial						
G45	Básico	1	$\sigma^2[\mathbf{I}_r \otimes \mathbf{I}_c]$	3	2	-932,09	1874,19
	Espacial	3	$\sigma^2_{\xi}[\mathbf{I}_r \otimes \Sigma_c(\rho_c)]$	3	3	-927,51	1867,03
G62	Básico	1	$\sigma^2[\mathbf{I}_r \otimes \mathbf{I}_c]$	3	2	-1301,29	2612,58
	Espacial	3	$\sigma^2_{\xi}[\mathbf{I}_r \otimes \Sigma_c(\rho_c)]$	3	3	-1298,07	2608,14
J64	Básico	1	$\sigma^2[\mathbf{I}_r \otimes \mathbf{I}_c]$	3	2	-1378,81	2767,62
	Espacial	5	$\sigma^2_{\xi}[\Sigma_r(\rho_r) \otimes \mathbf{I}_c]$	4	3	-1371,79	2757,57
VII8	Básico	1	$\sigma^2[\mathbf{I}_r \otimes \mathbf{I}_c]$	3	2	-359,48	728,96
	Espacial						
VI21	Básico	1	$\sigma^2[\mathbf{I}_r \otimes \mathbf{I}_c]$	3	2	-444,71	899,42
	Espacial						
A45	Básico	1	$\sigma^2[\mathbf{I}_r \otimes \mathbf{I}_c]$	3	2	-899,97	1809,94
	Espacial						
F100	Básico	1	$\sigma^2[\mathbf{I}_r \otimes \mathbf{I}_c]$	3	2	-1974,93	3959,85
	Espacial	7	$\sigma^2_{\xi}[\Sigma_r(\rho_r) \otimes \Sigma_c(\rho_c)]$	3	5	-1960,79	3937,59
VA18	Básico	1	$\sigma^2[\mathbf{I}_r \otimes \mathbf{I}_c]$	3	2	-345,83	701,66
	Espacial						
VA21	Básico	1	$\sigma^2[\mathbf{I}_r \otimes \mathbf{I}_c]$	3	2	-424,16	858,33
	Espacial	5	$\sigma^2_{\xi}[\mathbf{I}_r \otimes \Sigma_c(\rho_c)]$	4	3	-417,89	849,78
C400	Básico	1	$\sigma^2[\mathbf{I}_r \otimes \mathbf{I}_c]$	2	3	-5753,22	11516,44
	Espacial	7	$\sigma^2[\mathbf{I}_r \otimes \mathbf{I}_c]$	2	4	-5736,55	11485,10
P400	Básico	1	$\sigma^2[\mathbf{I}_r \otimes \mathbf{I}_c]$	2	3	-5678,69	11367,38
	Espacial	25	$\sigma^2_{\xi}[\mathbf{I}_r \otimes \Sigma_c(\rho_c)] + \sigma^2_{\eta}[\mathbf{I}_r \otimes \mathbf{I}_c]$	4	5	-5645,42	11308,83
J225	Básico	1	$\sigma^2[\mathbf{I}_r \otimes \mathbf{I}_c]$	3	3	-4619,53	9251,07
	Espacial	22	$\sigma^2_{\xi}[\Sigma_r(\rho_r) \otimes \Sigma_c(\rho_c)] + \sigma^2_{\eta}[\mathbf{I}_r \otimes \mathbf{I}_c]$	4	6	-4592,93	9205,86
D400	Básico	1	$\sigma^2[\mathbf{I}_r \otimes \mathbf{I}_c]$	2	3	-5731,39	11472,79
	Espacial	20	$\sigma^2_{\xi}[\Sigma_r(\rho_r) \otimes \Sigma_c(\rho_c)] + \sigma^2_{\eta}[\mathbf{I}_r \otimes \mathbf{I}_c]$	2	6	-5706,89	11429,79
E225	Básico	1	$\sigma^2[\mathbf{I}_r \otimes \mathbf{I}_c]$	2	3	-3251,75	6513,51
	Espacial	5	$\sigma^2_{\xi}[\mathbf{I}_r \otimes \Sigma_c(\rho_c)]$	3	4	-3236,10	6486,19
B100	Básico	1	$\sigma^2[\mathbf{I}_r \otimes \mathbf{I}_c]$	3	3	-2137,09	4286,19
	Espacial	7	$\sigma^2_{\xi}[\mathbf{I}_r \otimes \Sigma_c(\rho_c)]$	3	5	-2128,12	4272,23
F550	Básico	1	$\sigma^2[\mathbf{I}_r \otimes \mathbf{I}_c]$	1	3	-5017,06	10042,13
	Espacial	9	$\sigma^2[\mathbf{I}_r \otimes \mathbf{I}_c]$	3	3	-5009,94	10018,90
H650	Básico	1	$\sigma^2[\mathbf{I}_r \otimes \mathbf{I}_c]$	1	3	-5164,09	10336,19
	Espacial	20	$\sigma^2_{\xi}[\Sigma_r(\rho_r) \otimes \Sigma_c(\rho_c)] + \sigma^2_{\eta}[\mathbf{I}_r \otimes \mathbf{I}_c]$	1	6	-5114,37	10242,74
B500	Básico	1	$\sigma^2[\mathbf{I}_r \otimes \mathbf{I}_c]$	1	3	-3859,31	7726,62
	Espacial	5	$\sigma^2_{\xi}[\Sigma_r(\rho_r) \otimes \Sigma_c(\rho_c)]$	2	5	-3846,91	7707,81

<sup>a</sup>DBA, Delineamento em blocos aumentados; DBC, Delineamento em blocos casualizados; DLQ, Delineamento látice quadrado.

<sup>b</sup> $\sigma^2[\mathbf{I}_r \otimes \mathbf{I}_c]$ : estrutura de (co)variância assumindo resíduos independentes;  $\sigma^2_{\xi}[\Sigma_r(\rho_r) \otimes \mathbf{I}_c]$ : estrutura de (co)variância residual assumindo correlação autorregressiva de primeira ordem no sentido de linhas;  $\sigma^2_{\xi}[\mathbf{I}_r \otimes \Sigma_c(\rho_c)]$ : estrutura de (co)variância residual assumindo correlação autorregressiva de primeira ordem no sentido de colunas;  $\sigma^2_{\xi}[\Sigma_r(\rho_r) \otimes \Sigma_c(\rho_c)]$ : estrutura de (co)variância residual assumindo correlação autorregressiva de primeira ordem no sentido de linhas e colunas;  $\sigma^2_{\eta}[\mathbf{I}_r \otimes \mathbf{I}_c]$ : estrutura de (co)variância residual para a parte independente dos resíduos (efeito nugget) no modelo espacial.  $\sigma^2$  é a variância residual,  $\sigma^2_{\xi}$  é a variância para a parte correlacionada dos resíduos e  $\sigma^2_{\eta}$  é a variância para parte independente dos resíduos.

Para os outros 14 experimentos, um modelo espacial apresentou melhor ajuste que o modelo básico (Tabela 3). A contribuição das tendências espaciais globais e locais descritas abaixo para as observações fenotípicas em cada parcela foi ilustrada nas Figuras Suplementares S1 (experimento G45), S2 (experimento G62), S3 (experimento J64), S4 (experimento F100), S5 (experimento VA21), S6 (experimento C400), S7 (experimento P400), S8 (experimento J225), S9 (experimento D400), S10 (experimento E225), S11 (experimento B100), S12 (experimento F550), S13 (experimento H650) e S14 (experimento B500).

Para os demais experimentos sob delineamento DBC (experimentos G45, G62, J64, F100 e VA21), houve um modelo espacial com estrutura residual AR1 com melhor ajuste que o modelo básico (Tabela 3). A estrutura residual AR1 foi necessária ao longo das linhas para o experimento J64 (modelo 5); ao longo das colunas para os experimentos G45 (modelo 3), G62 (modelo 3) e VA21 (modelo 5); e ao longo de ambas as coordenadas espaciais para o experimento F100 (modelo 7).

Para cada um dos experimentos instalados sob os delineamentos DLQ e DBA, houve um modelo espacial específico com ajuste superior ao modelo básico (ou seja, estimativas de AICc inferiores) (Tabela 3). No experimento C400, embora os resíduos fossem considerados independentes, o modelo espacial selecionado capturou a variabilidade entre as linhas (modelo 7). Resíduos independentes também foram incluídos no modelo espacial para o experimento F550, mas tendências espaciais globais foram modeladas usando os efeitos fixos de regressão linear de PROD nos índices de linha e coluna (modelo 9). A estrutura residual AR1 foi exigida ao longo das colunas para os experimentos P400 (modelo 25), E225 (modelo 5) e B100 (modelo 7); e para ambas as direções para os experimentos J225 (modelo 22), D400 (modelo 20), H650 (modelo 20) e B500 (modelo 5). Os modelos espaciais selecionados para os experimentos P400, J225, D400 e H650 também incluíram o efeito nugget ou erro de mensuração (Tabela 3).

## **4.2 Tendências espaciais globais e estimativas de autocorrelação**

Com os modelos espaciais selecionados, detectamos significância para as tendências globais pelos efeitos fixos da regressão linear de PROD ao longo das linhas (Tabela Suplementar S1) para os experimentos J64 (modelo 5), VA21 (modelo 5), P400 (modelo 25), E225 (modelo 5), F550 (modelo 9) e B500 (modelo 5). Os mesmos efeitos ao longo das colunas (Tabela Suplementar S1) foram significativos para os experimentos P400 (modelo 25), J225 (modelo 22) e F550 (modelo 9). Estes resultados indicam a presença de gradientes de variabilidade espacial ao longo da área nestes experimentos (Figuras Suplementares S3, S5, S7, S8, S10, S12 e S14).

Outras tendências globais foram detectadas pelos efeitos aleatórios de linhas significativos quando o modelo espacial 7 foi usado nos experimentos F100, C400 e B100 (Tabela 4 e Figuras Suplementares S4, S6 e S11). Para os experimentos em que a estrutura de (co)variância residual AR1 proporcionou um ajuste superior ao modelo espacial, os parâmetros de autocorrelação foram estimados (Tabela 4). Essas estimativas foram baixas para os experimentos conduzidos sob delineamento DBC, com valores variando de 0,18 (experimento J64) a 0,35 (experimento F100) ao longo das linhas e de 0,08 (experimento F100) a 0,40 (experimento G45) ao longo das colunas. Para os experimentos realizados no delineamento DLQ, as estimativas de autocorrelação foram entre 0,90 (experimento J225) e 0,91 (experimento D400) ao longo das linhas e entre 0,05 (experimento B100) e 0,91 (experimento P400) ao longo das colunas. Para os experimentos sob delineamento DBA, essas estimativas foram de 0,41 (experimento B500) e 0,95 (experimento H650) ao longo das linhas e 0,43 (experimento B500) e 0,91 (experimento H650) ao longo das colunas. Na maioria dos casos, altas estimativas para a autocorrelação ( $\geq 0,90$ ) foram observadas quando o efeito nugget foi incluído nos modelos espaciais (Tabela 4).

### **4.3 Variabilidade genotípica, herdabilidade e estimativas de precisão de seleção**

Foi detectada significância para o efeito de genótipos usando o modelo básico e o modelo espacial selecionado para todos os experimentos (Tabela 4).

Os cinco experimentos em que o modelo básico foi o modelo selecionado (experimentos S112, VI18, VI21, A45 e VA18) apresentaram estimativas de herdabilidade de 0,90, 0,84, 0,86, 0,83 e 0,91 e estimativas de acurácia de 0,94, 0,89, 0,90, 0,90, 0,93, respectivamente.

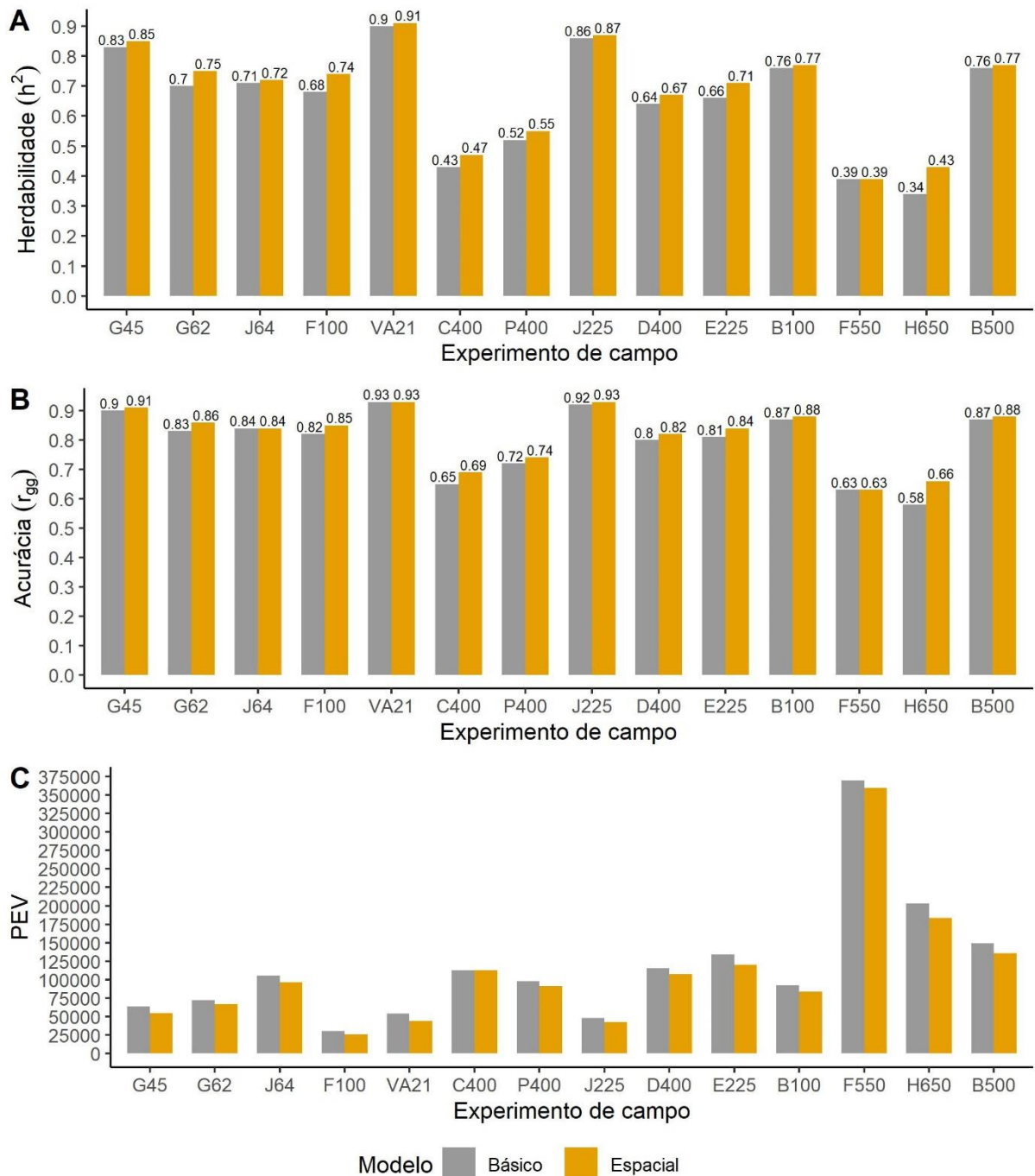
**Tabela 4.** Estimativas de componentes de variância para os efeitos de genótipos ( $\hat{\sigma}_g^2$ ), blocos ( $\hat{\sigma}_b^2$ ), linhas ( $\hat{\sigma}_{lin}^2$ ), colunas ( $\hat{\sigma}_{col}^2$ ), resíduos ( $\hat{\sigma}^2$ ), resíduos correlacionados ( $\hat{\sigma}_\xi^2$ ), e resíduos independentes ( $\hat{\sigma}_\eta^2$ ); estimativas de autocorrelação ao longo de linhas ( $\hat{\rho}_r$ ) e colunas ( $\hat{\rho}_c$ ) para os experimentos de campo de feijão-comum

Delin. <sup>a</sup>	Exp.	Modelo		$\hat{\sigma}_g^2$	$\hat{\sigma}_b^2$	$\hat{\sigma}_{lin}^2$	$\hat{\sigma}^2$	$\hat{\sigma}_\xi^2$	$\hat{\sigma}_\eta^2$	$\hat{\rho}_r$	$\hat{\rho}_c$
DBC	S112	Básico	1	765844**			267588				
		Espacial									
	G45	Básico	1	330646**			208447				
		Espacial	3	327354**					213575		0,40
	G62	Básico	1	233728**			298142				
		Espacial	3	259750**					296151		0,32
	J64	Básico	1	348762**			427496				
		Espacial	5	331570**					407335		0,18
	VI18	Básico	1	285969**			163792				
		Espacial									
	VI21	Básico	1	537747**			274486				
		Espacial									
A45	Básico	1	258588**			155675					
	Espacial										
F100	Básico	1	94829**			132450					
	Espacial	7	95410**		20782*			113481		0,35	0,08
VA18	Básico	1	275097**			79964					
	Espacial										
VA21	Básico	1	390286**			124572					
	Espacial	5	338042**					108625		0,35	
DLQ	C400	Básico	1	196407**	35358		505554				
		Espacial	7	213802**	39082	51118**	441689				
	P400	Básico	1	205360**	86547		352207				
		Espacial	25	202887**	45263			90477	264906		0,91
	J225	Básico	1	326931**	35754		153748				
		Espacial	22	320649**	3			79824	118382	0,90	0,66
	D400	Básico	1	322649**	66959		338801				
		Espacial	20	324919**	5199			153208	277259	0,91	0,85
	E225	Básico	1	396.360**	22835		388487				
		Espacial	5	414166**	3222			384561			0,34
	B100	Básico	1	377909**	6622		348570				
		Espacial	7	359174**	6422	79461**		276752			0,05
DBA	F550	Básico	1	606477**	87244		912348				
		Espacial	9	593310**	50622		911562				
	H650	Básico	1	307463**	92740		581762				
		Espacial	20	322291**	6			721357	353288	0,95	0,91
	B500	Básico	1	634600**	12172		180869				
		Espacial	5	598523**	10098			205408		0,41	0,43

<sup>a</sup>DBA, Delineamento em blocos aumentados; DBC, Delineamento em blocos casualizados; DLQ, Delineamento látice quadrado.

\*Significativo a 5% de significância; \*\*\*Significativo a 1% de significância pelo teste de razão de verossimilhança (LRT).

Em geral, a maioria dos 14 ensaios em que um modelo espacial específico foi superior ao modelo básico em ajuste apresentaram altas estimativas de herdabilidade (Figura 1a) e acurácia de seleção (Figura 1b). Essas estimativas foram menores para os experimentos C400, P400, F550 e H650 devido à grande variação residual ou efeito nugget nesses experimentos (Figuras Suplementares S6, S7, S12 e S13, respectivamente). No entanto, na maioria dos casos, os modelos espaciais resultaram em maiores estimativas de herdabilidade (Figura 1a) e acurácia de seleção (Figura 1b). As maiores mudanças na herdabilidade para PROD (Figura 1a) foram cerca de 26,5% para o experimento H650 (de 0,34 a 0,43), 9,3% para o experimento C400 (de 0,43 a 0,47), 8,8% para o experimento F100 (0,68 a 0,74), 7,6% para o experimento E225 (de 0,66 a 0,71), 7,1% para o experimento G62 (de 0,70 a 0,75), 5,8% para o experimento P400 (de 0,52 a 0,55), 4,7% para o experimento D400 (de 0,64 para 0,67) e 2,4% para o experimento G45 (de 0,83 a 0,85). Esses experimentos também foram os que obtiveram os maiores ganhos na acurácia da seleção quando os modelos espaciais foram usados em vez do modelo básico (Figura 1b): 13,8% para o experimento H650 (de 0,58 a 0,66), 6,2% para o experimento C400 (de 0,65 a 0,69), 3,7% para o experimento E225 (de 0,81 a 0,84), 3,7% para o experimento F100 (de 0,82 a 0,85), 3,6% para o experimento G62 (de 0,83 a 0,86), 2,8% para o experimento P400 (de 0,72 a 0,74), 2,5% para o experimento D400 (de 0,8 a 0,82) e 1,1% para os experimentos J225 (de 0,92 a 0,93) e G45 (de 0,90 a 0,91). Embora para alguns experimentos (J64, VA21 e F550) não houve melhora na acurácia da seleção, observamos reduções nas estimativas do PEV (Figura 1c) com o uso dos modelos espaciais.

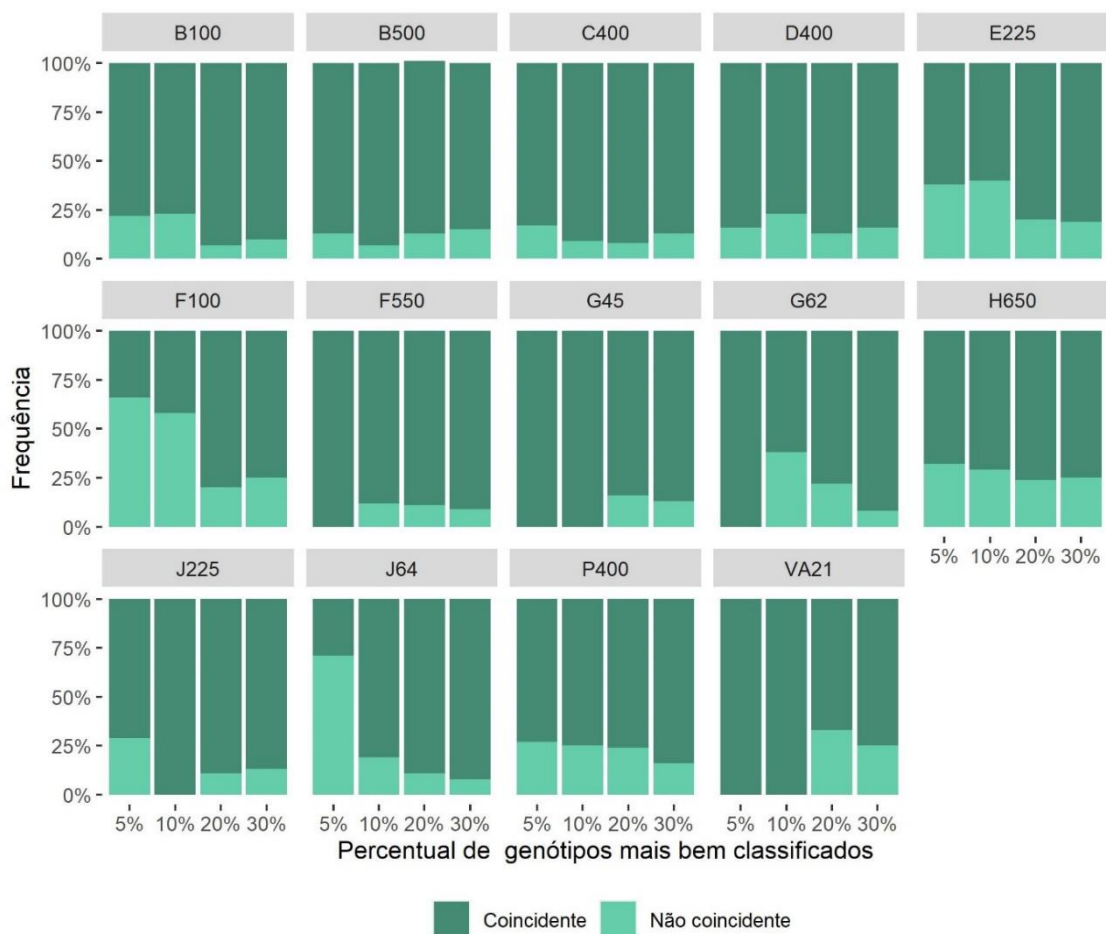


**Figura 1.** Estimativas de herdabilidade [a], acurácia [b] e variância do erro de predição (PEV) [c] obtidas com o modelo básico e o modelo espacial selecionado, para os experimentos de campo de feijão-comum nos quais houve um modelo especial com melhor ajuste que o modelo básico.

#### 4.4 Impacto na classificação dos valores genotípicos

A consequência mais prática de se levar em conta as tendências espaciais nos modelos estatísticos é a melhoria da predição dos valores genotípicos e, conseqüentemente, mudanças na sua classificação. Para a maioria das combinações de experimentos e porcentagens de

genótipos mais bem classificados, houve discordâncias entre o modelo básico e o espacial (Figura 2), mesmo naqueles experimentos para os quais foram observadas poucas mudanças na acurácia. Para os 5% mais bem classificados, não houve discordâncias para os experimentos G45, G62, VA21 e F550, mas houve discordâncias para todos os outros experimentos, com frequências de genótipos não coincidentes variando de 13% no experimento B500 a 71% no experimento J64. Para os 10% superiores, não foram observadas discordâncias para os experimentos G45, J225 e VA21. No entanto, para os outros experimentos, a frequência de discordância variou de 7% no experimento B500 a 58% no experimento F100. Considerando os genótipos 20% e 30% mais bem classificados, foram observadas discordâncias entre o modelo básico e o espacial para todos os experimentos. A frequência de discordâncias variou de 7% no experimento B100 a 33% no experimento VA21 considerando os 20% mais bem classificados, e de 8% nos experimentos G62 e J64 a 25% nos experimentos VA21, F100 e H650 quando os 30% mais bem classificados dos genótipos foram considerados (Figura 2).



**Figura 2.** Frequência de genótipos coincidentes e não coincidentes dentre os 5, 10, 20 e 30% mais bem classificados considerando os valores genotípicos preditos com o modelo básico e o modelo espacial de melhor ajuste para os experimentos de feijão-comum estudados.

## 5 Discussão

A análise de dados de experimentos de campo é fundamental no melhoramento de plantas para permitir a predição acurada dos valores genotípicos dos candidatos à seleção (RESENDE e ALVES, 2020). Como as informações coletadas nesses experimentos são fenotípicas, é importante controlar as fontes de variação que podem existir ao longo da área experimental e que não se refiram aos efeitos genotípicos. A estratégia usual é usar delineamentos experimentais baseados nos princípios de repetição, casualização e controle local (FISHER, 1926).

No entanto, observou-se que, para apenas cinco dos 19 ensaios do programa de melhoramento de feijoeiro relatados neste estudo, o modelo básico, que incluiu apenas os efeitos referentes aos delineamentos experimentais e assumiu independência para os resíduos, foi suficiente para capturar a heterogeneidade espacial existente no campo. A maioria desses cinco experimentos possui poucos tratamentos e todos eles foram conduzidos sob o delineamento DBC. Esse delineamento é amplamente utilizado em pesquisas agrícolas e, em geral, os melhoristas o utilizam em situações em que há conhecimento prévio sobre fatores ambientais ou práticas de manejo que afetam de forma diferenciada a área experimental, mas ainda é possível definir blocos homogêneos. A restrição é que esses blocos devem ser completos, ou seja, os blocos precisam acomodar todos os tratamentos (CASLER, 2015). Para os outros cinco experimentos conduzidos no DBC, os modelos espaciais que consideram a correlação residual e a os índices de posição das parcelas (linha e coluna) foram os modelos de melhor ajuste, aumentando a acurácia de seleção e as estimativas de herdabilidade ou, pelo menos, reduzindo a PEV. Utilizando experimentos com delineamento DBC para avaliar genótipos de soja quanto às características de composição de sementes, Bernardeli et al. (2021) observaram que os modelos espaciais que consideram a correlação residual pela estrutura de (co)variância AR1 foram predominantemente superiores em termos de ajuste, herdabilidade de sentido amplo e acurácia em comparação com o modelo DBC padrão. Liu et al. (2015) avaliaram genótipos de algodão quanto à produção de pluma e observaram a superioridade do modelo com uma estrutura de (co)variância bidimensional AR1 para levar em conta as variações globais e locais em comparação com o modelo básico para DBC.

Se o número de tratamentos a serem avaliados for muito grande no melhoramento de plantas, delineamentos com blocos incompletos, como DLQ e DBA, são preferidos em relação ao delineamento DBC para reduzir a heterogeneidade ambiental dentro dos blocos, já que blocos grandes implicam em heterogeneidade dentro dos blocos (RAMALHO et al., 2012; ZYSTRO et al., 2018). Mesmo com essa estratégia de controle da variação de campo,

observamos superioridade dos modelos espaciais para os seis experimentos estudados que foram instalados sob o delineamento DLQ em relação ao modelo básico em termos de aumento da acurácia e das estimativas de herdabilidade ou, pelo menos, de redução da PEV. Esses resultados revelam que o uso conjunto do delineamento experimental e a posterior modelagem de tendências espaciais é importante para melhorar ainda mais o controle da heterogeneidade espacial. Conforme enfatizado por Grondona et al. (1996), o uso de delineamentos mais complexos, como o de blocos incompletos, não é garantia de eficiência no controle da variabilidade espacial, principalmente quando ocorre variabilidade em pequena escala no campo.

Alguns dos experimentos realizados no delineamento DBA relatados neste trabalho apresentaram alta variância residual e as menores estimativas de herdabilidade, resultando em estimativas de acurácia moderada de acordo com Resende e Alves (2020). No caso do experimento F550, não foi observada melhora com o uso dos modelos espaciais. Muito disso porque, neste delineamento, o erro experimental é estimado apenas com base nas cultivares testemunhas repetidas no experimento e casualizadas dentro dos blocos (FEDERER, 1956; RAMALHO et al., 2012). De acordo com Müller et al. (2010), quando os tratamentos são completamente casualizados dentro dos blocos, como neste estudo, a distribuição espacial das testemunhas no experimento pode não otimizar a precisão da estimação dos resíduos e parâmetros espaciais e, conseqüentemente, a comparação dos tratamentos. Assim, esses autores observaram vantagens em alocar as parcelas de testemunhas com diferentes arranjos dentro dos blocos, o que poderia possibilitar melhorias na precisão experimental quando são utilizados modelos espaciais (MÜLLER et al., 2010).

Além disso, em ensaios com delineamento DBA, o erro é estimado com base nas cultivares testemunhas e assume-se que este seja representativo para todos os genótipos não repetidos. Essa suposição pode não refletir a realidade, principalmente para os experimentos F500 e B500, em que as cultivares testemunhas são linhagens comerciais e os tratamentos não repetidos são progênies obtidas em gerações precoces, que podem apresentar variabilidade genotípica dentro (RAMALHO et al., 2012). Para reduzir esse viés para a estimação de erro e aumentar a precisão experimental, uma alternativa seria utilizar os delineamentos p-rep propostos por Cullis et al. (2006), em que uma parte dos genótipos candidatos (progênies, linhagens etc.) são repetidos e substituem as cultivares padrão nas parcelas de testemunhas (CULLIS et al., 2006).

Não encontramos um único modelo espacial capaz de capturar a heterogeneidade espacial e com o melhor ajuste aos dados de todos os experimentos estudados. No entanto, ficou

claro que para muitos desses experimentos, a estrutura AR1 era mais adequada para capturar tendências espaciais locais, seja na direção de linhas ou colunas ou mesmo em ambas as direções. Como enfatizam Resende e Sturion (2003), mesmo com a utilização de delineamentos experimentais, que se baseiam no princípio casualização dos tratamentos às unidades experimentais para evitar a correlação espacial, as parcelas vizinhas podem apresentar certo grau de correlação devido a efeitos ambientais na área experimental. Nesse sentido, parcelas adjacentes tendem a ser mais correlacionadas fenotipicamente do que aquelas mais distantes umas das outras. A comparação dos tratamentos é afetada por essa correlação se ela não for considerada corretamente no modelo estatístico. A comparação de tratamentos localizados a curtas distâncias é mais propensa ao erro tipo II onde diferentes desempenhos médios são considerados iguais, enquanto a comparação de tratamentos localizados mais distantes é mais propensa ao erro tipo I onde desempenhos médios idênticos são considerados diferentes (VAN ES e VAN ES, 1993).

Para os experimentos em que os modelos espaciais assumiram a estrutura AR1, as estimativas de autocorrelação foram todas positivas, o que indica uma associação direta da observação de PROD e a similaridade ambiental, bem como a ausência de competição entre parcelas vizinhas (RESENDE e STURION, 2003). Os coeficientes de autocorrelação, os coeficientes de regressão linear fixos e os efeitos aleatórios baseados nos índices das linhas indicaram a existência de tendências espaciais ao longo das colunas e paralelas aos blocos e, portanto, ajudam a capturar as heterogeneidades existentes entre as parcelas dentro dos blocos. Por outro lado, em alguns experimentos, coeficientes de autocorrelação e coeficientes de regressão linear usando índices de coluna foram obtidos e capturaram as tendências espaciais ao longo das linhas e perpendiculares aos blocos. Essas tendências ocorrem como manchas locais ou gradientes globais nos experimentos, causados por fatores intrínsecos e extrínsecos ao experimento que podem incluir a heterogeneidade de fertilidade, umidade e aspectos físicos do solo, topografia, a direção em que as práticas de manejo são realizadas (GILMOUR et al., 1997; GEZAN et al., 2010) e a ocorrência de doenças e estresses abióticos (ANDRADE et al., 2020). A inclusão do erro independente ou de mensuração (nugget) foi necessária para apenas quatro experimentos, o que aumentou as estimativas de autocorrelação devido à redução da variância residual correlacionada, conforme também relatado por Resende e Sturion (2003).

Apesar de capturar as tendências espaciais existentes nos experimentos, a adição de informações relacionadas à posição das parcelas nos modelos espaciais não consumiu a variabilidade entre blocos para alguns experimentos realizados nos delineamentos DLQ e DBA. Isso enfatiza a importância de manter as informações provenientes do delineamento

experimental. Essa importância do uso da informação espacial aliada ao delineamento experimental também foi relatada em outros estudos (ANDRADE et al., 2020; BORGES et al., 2019; HOEFLER et al., 2020).

Em termos práticos, para a maioria dos experimentos, a escolha de modelos espaciais ao invés do modelo básico tradicional e a correção dos dados fenotípicos para tendências espaciais causaram mudanças consideráveis na classificação dos valores genotípicos. Essas mudanças se devem ao controle adicional da heterogeneidade ambiental proporcionado pelos modelos espaciais selecionados. Isso resultou em estimativas de acurácia de seleção mais altas ou, pelo menos, estimativas mais baixas para a variância do erro de predição (PEV). De acordo com Resende e Alves (2020), a acurácia mede o grau de ordenação correta dos genótipos, pois reflete a correlação entre os valores genotípicos verdadeiros e preditos, enquanto valores menores de PEV indicam maior precisão na predição dos valores genotípicos (RESENDE e ALVES, 2020). Outra perspectiva relevante é que os modelos espaciais aumentaram as estimativas de herdabilidade para PROD em vários experimentos. Se parâmetros como intensidade de seleção, variabilidade genotípica e tempo para completar um ciclo de seleção permanecerem constantes, o uso de estimativas de herdabilidade mais altas pode resultar em aumentos nos ganhos com a seleção ao considerar as informações espaciais no modelo estatístico. Esta é uma forma mais simples e barata de melhorar a eficiência de um programa de melhoramento de feijoeiro em comparação com o aumento da herdabilidade com mais genótipos avaliados ou incluindo mais repetições nos experimentos.

A análise espacial relatada neste estudo foi composta por uma sequência de 35 modelos possíveis. Esta foi uma estratégia útil para investigar e compreender a variabilidade espacial que ocorre em um número representativo de experimentos de nosso programa de melhoramento de feijão-comum. Não houve um modelo específico com o ajuste adequado para todos os experimentos. No entanto, com base em nossos resultados e nos padrões de variabilidade espacial observados, a análise de dados para experimentos futuros pode ser resumida em duas etapas: uma primeira etapa para selecionar a melhor estrutura residual: erros independentes (modelo 1), ou estrutura AR1 ao longo das linhas, colunas ou ambas as direções (modelos 2, 3 e 4, respectivamente), pois todas ocorreram. Um segundo passo seria ajustar alguns modelos para testar a significância dos tipos mais frequentes de tendências globais, como os efeitos fixos [lin(linhas) e lin(colunas)], e os efeitos aleatórios de linhas (modelos 5, 6, 7, 9 e 15). Dependendo da estrutura residual escolhida, o efeito nugget também pode ser adicionado ao modelo selecionado para avaliar sua relevância. Devido à possível existência de interação genótipo x ambiente, é fortemente recomendado que a seleção de genótipos superiores seja

baseada na análise de dados de experimentos multi-ambientes. Assim, para permitir uma seleção confiável, os genótipos avaliados em cada um dos experimentos relatados neste estudo precisarão ser avaliados em mais safras, anos e locais. Embora a análise dos dados de experimentos multi-ambientes seja comumente realizada em uma análise em dois estágios, o primeiro estágio é crucial, pois permite a modelagem do delineamento experimental e das tendências espaciais existentes dentro do ambiente com modelos espaciais, corrigindo os efeitos genotípicos para essas tendências antes do segundo estágio (ARIEF et al., 2019; SMITH; CULLIS; GILMOUR, 2001).

Assim, concluímos que os modelos estatísticos espaciais associados ao uso de delineamentos experimentais ajudaram a capturar a heterogeneidade espacial existente para corrigir os dados fenotípicos de experimentos de campo individuais para tendências espaciais e melhorar a predição de valores genotípicos. Um uso mais frequente desses modelos na análise de dados aumentaria a eficiência do programa de melhoramento de feijoeiro estudado.

## 6 Referências

AKAIKE, H. A new look at the statistical model identification. **IEEE Transaction on Automatic Control**, v. 19, n. 6, p. 716-723, dez. 1974. DOI: <https://doi.org/10.1109/TAC.1974.1100705>

ANDRADE, M. H. M. L. et al. Accounting for spatial trends to increase the selection efficiency in potato breeding. **Crop Science**. v. 60, p. 2354–2372, mai. 2020. DOI: <https://doi.org/10.1002/csc2.20226>

ARIEF, V. N. et al. Utilization of Multiyear Plant Breeding Data to Better Predict Genotype Performance. **Crop Science**, v. 59, p. 480-490. 2019. DOI: <https://doi.org/10.2135/cropsci2018.03.0182>

ASSEFA, T. et al. A review of breeding objectives, genomic resources, and marker-assisted methods in common bean (*Phaseolus vulgaris* L.). **Molecular Breeding**, v. 39, n. 20, p. 1-23, fev. 2019. DOI: <https://doi.org/10.1007/s11032-018-0920-0>

BANSAL, P.; RAWAL, V.; BANSAL, V. Pulses of *Phaseolus* and *Vigna* Genera. In V. Rawal, & D. K. Navarro (Ed), **The Global Economy of Pulses**. p. 39-70. FAO. 2019. DOI: <https://doi.org/10.4060/I7108EN>

BERNARDELI, A. et al. Modeling spatial trends and enhancing genetic selection: An approach to soybean seed composition breeding. **Crop Science**. v. 61, p. 976-988. 2021. DOI: <https://doi.org/10.1002/csc2.20364>

BORGES, A. et al. Can Spatial Modeling Substitute for Experimental Design in Agricultural Experiments? **Crop Science**. v. 59, p. 44-53. 2019. DOI: <https://doi.org/10.2135/cropsci2018.03.0177>

BURGUEÑO, J. Spatial Analysis of Field Experiments. In: GLAZ, B.; YEATER, K. M. **Applied Statistics in Agricultural, Biological, and Environmental Sciences**. Madison, WI: ASA, CSSA, and SSSA Books, 2018. p. 319-344. DOI: <https://doi.org/10.2134/appliedstatistics.2016.0011.c12>

CARNEIRO, J. E. S.; PAULA JÚNIOR, T. J.; BORÉM, A. **Feijão: do plantio à colheita**. Viçosa: Editora UFV. 2015. 384p.

CARVALHO, L. M. J. et al. Iron and zinc retention in common beans (*Phaseolus vulgaris* L.) after home cooking. **Food and Nutrition Research**, v. 56, p. 1–6. 2012. DOI: <https://doi.org/10.3402/2Ffnr.v56i0.15618>

CASLER, M. D. Fundamentals of Experimental Design: Guidelines for Designing Successful Experiments. **Agronomy Journal**, v. 107, p. 692-705. 2015. DOI: <https://doi.org/10.2134/agronj2013.0114>

CAVANAUGH, J. E.; NEATH, A. A. The Akaike information criterion: Background, derivation, properties, application, interpretation, and refinements. **WIREs Computational Statistics**, v. 11, n. 3. 2019. DOI: <https://doi.org/10.1002/wics.1460>

COWLING, W. A. et al. Using the animal model to accelerate response to selection in a self-pollinating crop. **G3: Genes, Genomes, Genetics**, v. 5, p. 1419–1428. 2015. DOI: <https://doi.org/10.1534/g3.115.018838>

CULLIS, B. R.; GLEESON, A. C. Spatial Analysis of Field Experiments-An Extension to Two Dimensions. **Biometrics**, v. 47, n. 4, p. 1449, dez. 1991.

CULLIS, B. R.; SMITH, A. B.; COOMBES, N. E. On the design of early generation variety trials with correlated data. **Journal of Agricultural, Biological, and Environmental Statistics**. v. 11, p. 381–393. 2006. DOI: <https://doi.org/10.1198/108571106x154443>

DUARTE, J. B., VENCOVSKY, R. Spatial statistical analysis and selection of genotypes in plant breeding. **Pesquisa Agropecuária Brasileira**, 40, 107-114. 2005. DOI: <https://doi.org/10.1590/S0100-204x2005000200002>

FEDERER, W. T. Augmented (or hoonuiaku) designs. **Biometrics Unit Technical Reports**. 1956. Disponível em: <https://hdl.handle.net/1813/32841>

FISHER, R.A. The arrangement of field experiments. **Journal of the Ministry of Agriculture, Great Britain**, v. 33, p. 503–513. 1926. DOI: <https://doi.org/10.23637/rothamsted.8v61q>

GEPTS, P. et al. Genomics of *Phaseolus* Beans, a Major Source of Dietary Protein and Micronutrients in the Tropics. In: P. H. Moore & R. Ming (Ed.), **Genomics of Tropical Crop Plants** (pp. 113-143). Springer. 2008. [https://doi.org/10.1007/978-0-387-71219-2\\_5](https://doi.org/10.1007/978-0-387-71219-2_5)

GEZAN, S. A.; WHITE, T. L.; HUBER, D. A. Accounting for Spatial Variability in Breeding Trials: A Simulation Study. **Agronomy Journal**. v. 102, p. 1562-1571, set. 2010. DOI: <https://doi.org/10.2134/agronj2010.0196>

GILMOUR, A. R.; CULLIS, B. R.; VERBYLA, A. P. Accounting for Natural and Extraneous Variation in the Analysis of Field Experiments. **Journal of Agricultural, Biological, and Environmental Statistics**, v. 2, n. 3, p. 269–293, set. 1997. DOI: <https://doi.org/10.2307/1400446>

GILMOUR, A.R. (2020). **Echidna Mixed Model Software**. 2020. Disponível em: <http://www.EchidnaMMS.org>

GLEESON, A. C.; CULLIS, B. R. Residual Maximum Likelihood (REML) Estimation of a Neighbour Model for Field Experiments. **Biometrics**, v. 43, p. 277–287. 1987. DOI: <https://doi.org/10.2307/2531812>

GRONDONA, M. O.; CROSSA, J.; FOX, P. N.; PFEIFFER, W. H. Analysis of Variety Yield Trials Using Two-Dimensional Separable ARIMA Processes. **Biometrics**, v. 52, p. 763-770, jun. 1996.

HAMBLIN, J.; ZIMMERMAN, M. J. O. Breeding common bean for yield mixtures. **Plant Breeding Reviews**, v. 4, p. 245–272. 1986. DOI: <https://doi.org/10.1002/9781118061015.ch8>

HENDERSON, C. R. et al. The estimation of environmental and genetic trends from records subject to culling. **Biometrics**, v. 15, n. 2, p. 192-218, 1959.

IZQUIERDO, P. et al. Meta-QTL analysis of seed iron and zinc concentration and content in common bean (*Phaseolus vulgaris* L.). **Theoretical and Applied Genetics**, v. 131, p. 1645–1658, mai. 2018.

HOEFLER, R. et al. Do Spatial Designs Outperform Classic Experimental Designs? **Journal of Agricultural, Biological, and Environmental Statistics**, v. 25, p. 523–552, ago. 2020. DOI: <https://doi.org/10.1007/s13253-020-00406-2>

LIU, S. M. et al. Benefit of spatial analysis for furrow irrigated cotton breeding trials. **Euphytica**, v. 201, p. 253–264. 2015. DOI: <https://doi.org/10.1007/s10681-014-1205-2>

LOS, F. G. B. et al. Beans (*Phaseolus vulgaris* L.): whole seeds with complex chemical composition. **Current Opinion in Food Science**, v. 19, p. 63–71, jan. 2018. DOI: <https://doi.org/10.1016/j.cofs.2018.01.010>

MÜLLER, B. U., SCHÜTZENMEISTER, A., PIEPHO, H. P. (2010). Arrangement of check plots in augmented block designs when spatial analysis is used. **Plant Breeding**, v. 129, p. 581-589. 2010. DOI: <https://doi.org/10.1111/j.1439-0523.2010.01803.x>

PATTERSON, H. D.; THOMPSON, R. Recovery of inter-block information when block sizes are unequal. **Biometrika**, v. 58, p. 545–554. 1971. DOI: <https://doi.org/10.1093/biomet/58.3.545>

PIEPHO, H. P. et al. BLUP for phenotypic selection in plant breeding and variety testing. **Euphytica**, v. 161, p. 209–228, jul. 2008. DOI: <https://doi.org/10.1007/s10681-007-9449-8>

RAMALHO, M. A. P.; FERREIRA, D. F.; OLIVEIRA, A. C. **Experimentação em Genética e Melhoramento de Plantas**. 3. Ed. Lavras: Editora UFLA. 2012. 329p.

RESENDE, M. D. V.; ALVES, R. S. Linear, generalized, hierarchical, bayesian and random regression mixed models in genetics/genomics in plant breeding. **Functional Plant Breeding Journal**. v. 2, p. 1-31, jun. 2020. DOI: <https://doi.org/10.35418/2526-4117/v2n2a1>

RESENDE, M. D. V.; DUARTE, J. B. Precisão e controle de qualidade em experimentos de avaliação de cultivares. **Pesquisa Agropecuária Tropical**. v.37, p. 182–194, set. 2007.

RESENDE, M. D. V.; SILVA, F.F.; AZEVEDO, C.F. **Estatística matemática, biométrica e computacional**: modelos mistos, multivariados, categóricos e generalizados (REML/BLUP), inferência Bayesiana, regressão aleatória, seleção genômica, QTL-GWAS, estatística espacial e temporal, competição, sobrevivência. Viçosa, MG: Suprema, 2014. 881p.

RESENDE, M. D. V.; STURION, J. A. Análise estatística espacial de experimentos via modelos mistos individuais com erros modelados por processos ARIMA em duas dimensões. **Rev. Mat. Estat.**, v. 21, n. 1, p. 7-33, jan. 2003.

SALVADOR, F. V., et al. Correcting experimental data for spatial trends in a common bean breeding program. **Crop Science**, v. 62, n. 2, p. 825-838. 2022. DOI: <https://doi.org/10.1002/csc2.20703>

SMITH, A. B.; CULLIS, B. R.; GILMOUR, A. The analysis of crop variety evaluation data in Australia. **Australian and New Zealand Journal of Statistics**. v. 43, p. 129–145. 2001. DOI: <https://doi.org/10.1111/1467-842X.00163>

SMITH, A.; CULLIS, B.; THOMPSON, R. Analyzing variety by environment data using multiplicative mixed models and adjustments for spatial field trend. **Biometrics**, v. 57, n. 4, p. 1138–1147, dez. 2001. DOI: <https://doi.org/10.1111/j.0006-341X.2001.01138.x>

STEFANOVA, K. T.; SMITH, A. B.; CULLIS, B. R. Enhanced diagnostics for the spatial analysis of field trials. **Journal of Agricultural, Biological, and Environmental Statistics**, v. 14, n. 4, p. 392–410, dez. 2009. DOI: <https://doi.org/10.1198/jabes.2009.07098>

VAN ES, H. M.; VAN ES, C. L. Spatial Nature of Randomization and Its Effect on the Outcome of Field Experiments. **Agronomy Journal**. v. 85, p. 420-428, mar./abr. 1993. DOI: <https://doi.org/10.2134/agronj1993.00021962008500020046x>

VERBYLA, A. P. A note on model selection using information criteria for general linear models estimated using REML. **Australian & New Zealand Journal of Statistics**. v. 61, p. 39-50, mar. 2019. DOI: <https://doi.org/10.1111/anzs.12254>

WILKS, S. S. The large-sample distribution of the likelihood ratio for testing composite hypotheses. **The Annals of Mathematical Statistics**. v. 9, p. 60–62. 1938. DOI: <https://doi.org/10.1214/aoms/1177732360>

YATES, F. A new method of arranging variety trials involving a large number of varieties. **Journal of Agricultural Science**, v. 26, p. 424–455, 1936. DOI: <https://doi.org/10.1017/S0021859600022760>

ZYSTRO, J.; COLLEY, M.; DAWSON, J. Alternative Experimental Designs for Plant Breeding. In: GOLDMAN, I. **Plant Breeding Reviews - Volume 42**. Hoboken: John Wiley & Sons, Inc., 2018. p. 87-117. DOI: <https://doi.org/10.1002/9781119521358.ch3>

## 7 Material suplementar

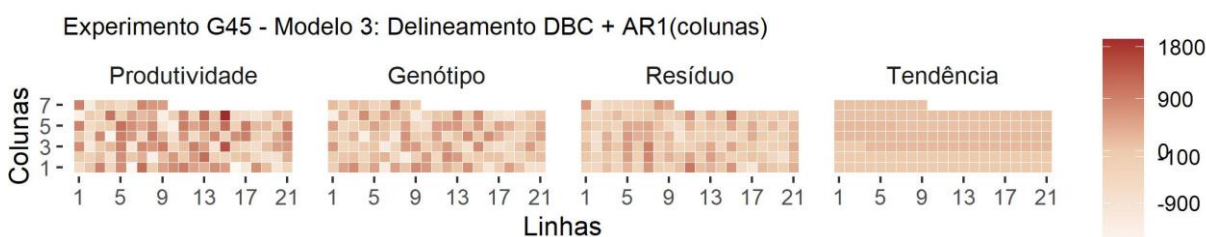
**Tabela S1.** Estimativas, erros-padrão (EP) e p-valor para os efeitos fixos de regressão linear de produtividade no sentido de linhas e colunas dos modelos espaciais selecionados para capturar a heterogeneidade especial em experimentos de campo de feijoeiro.

Delimitação <sup>a</sup>	Experimento	Modelo Espacial	Efeito fixo <sup>b</sup>	Estimativa	EP
DBC	J64	5	lin(linhas)	30,39**	9,90
DBC	VA21	5	lin(linhas)	-27,96**	9,58
DLQ	P400	25	lin(linhas)	14,57**	4,34
DLQ	P400	25	lin(colunas)	44,39**	14,64
DLQ	J225	22	lin(colunas)	-36,91**	12,44
DLQ	E225	5	lin(linhas)	-20,75**	7,63
DBA	F550	9	lin(linhas)	13,28*	5,29
DBA	F550	9	lin(colunas)	31,12**	10,41
DBA	B500	5	lin(linhas)	-10,20*	4,25

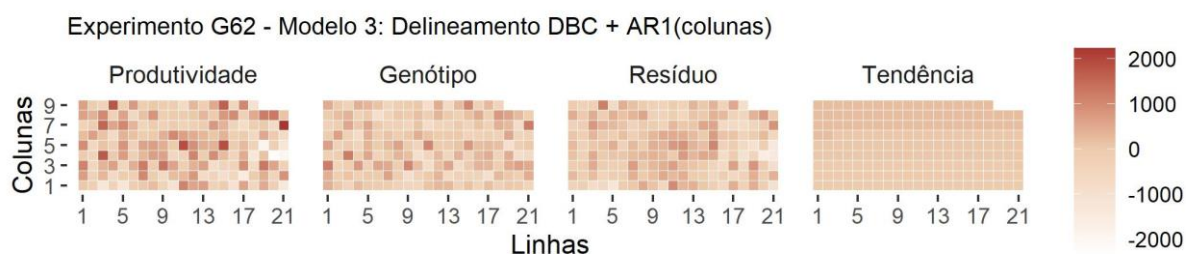
<sup>a</sup>DBA, Delimitação em blocos aumentados; DBC, Delimitação em blocos casualizados; DLQ, Delimitação látice quadrado.

<sup>b</sup>lin(linhas) e lin(colunas): efeito de regressão linear da produtividade nos índices de linha e nos índices de coluna, respectivamente.

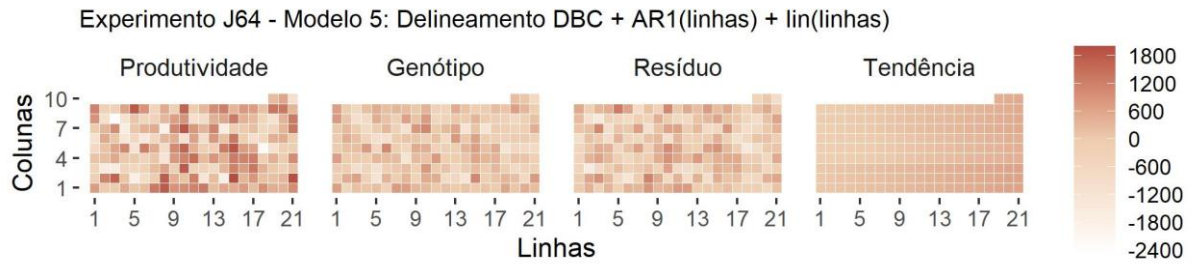
\*Significativo a 5% de significância; \*\*Significativo a 1% de significância pelo teste Wald.



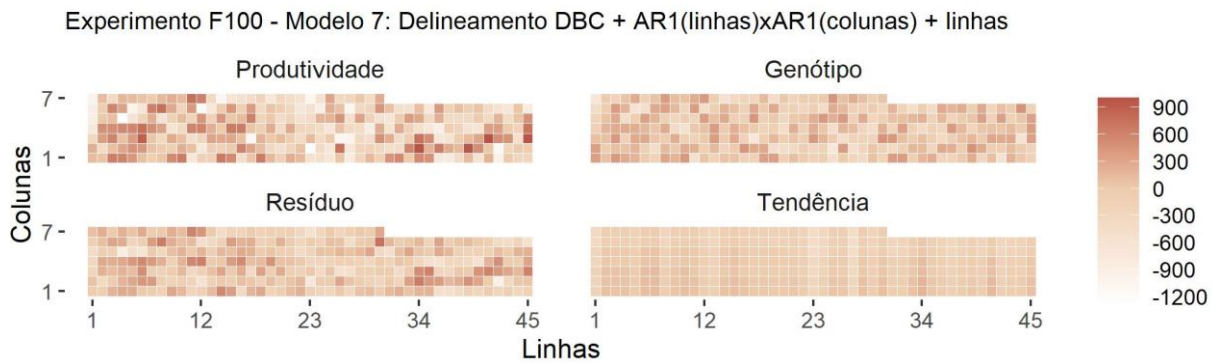
**Figura S1.** Mapas de calor em nível de parcela para o experimento G45 referentes aos dados fenotípicos centrados ( $\text{kg ha}^{-1}$ ), valores genotípicos (BLUP), efeitos de resíduo e efeitos de tendência representados pelos efeitos de repetição (bloco).



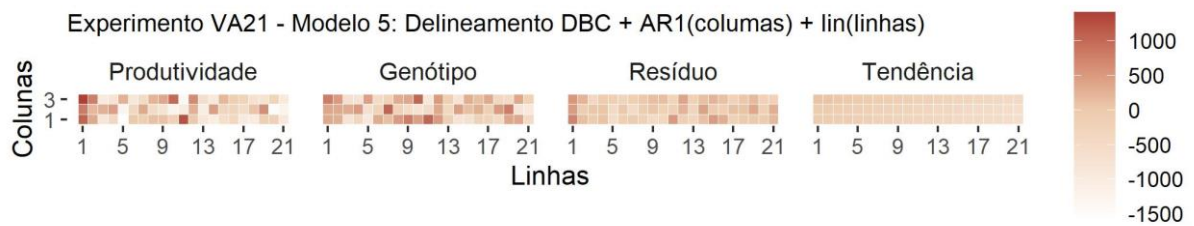
**Figura S2.** Mapas de calor em nível de parcela para o experimento G62 referentes aos dados fenotípicos centrados ( $\text{kg ha}^{-1}$ ), valores genotípicos (BLUP), efeitos de resíduo e efeitos de tendência representados pelos efeitos de repetição (bloco).



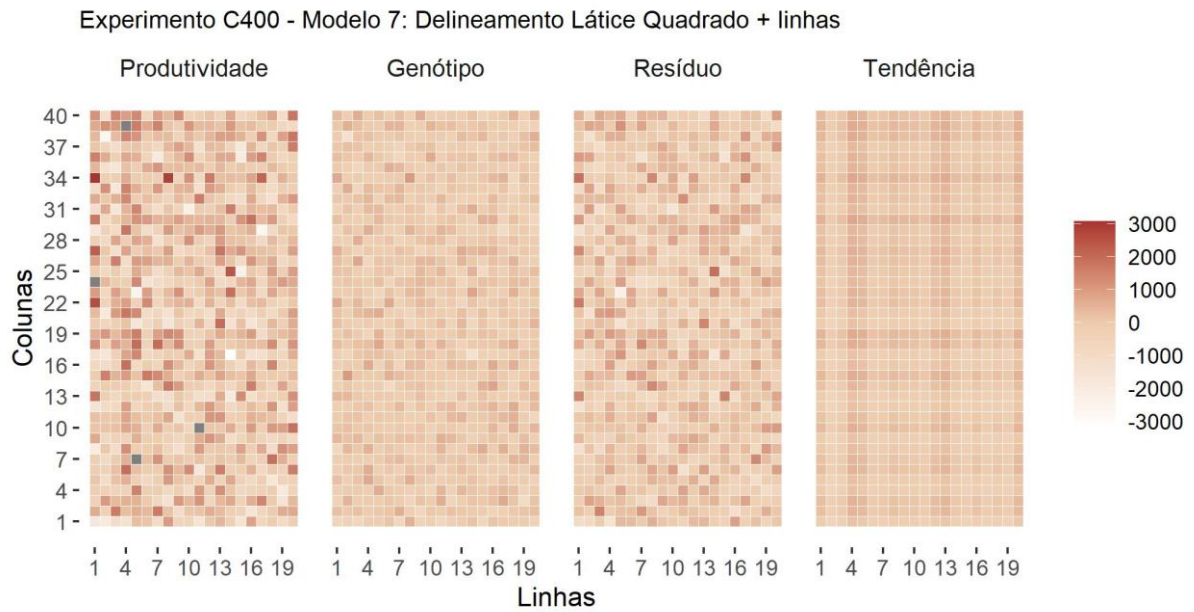
**Figura S3.** Mapas de calor em nível de parcela para o experimento J64 referentes aos dados fenotípicos centrados ( $\text{kg ha}^{-1}$ ), valores genotípicos (BLUP), efeitos de resíduo e efeitos de tendência representados pela soma dos efeitos de repetição (blocos) e efeito linear de linhas.



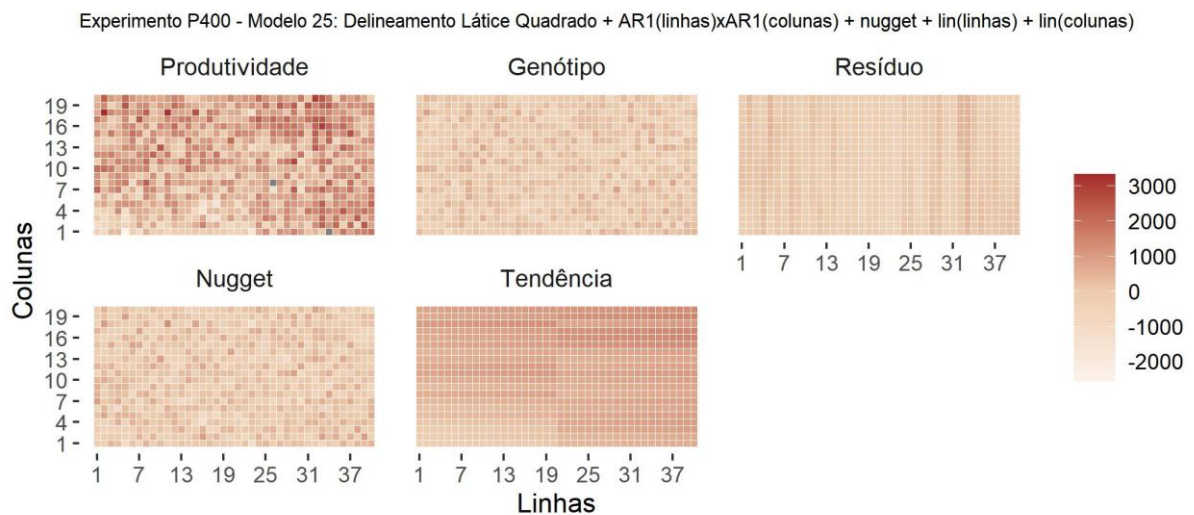
**Figura S4.** Mapas de calor em nível de parcela para o experimento F100 referentes aos dados fenotípicos centrados ( $\text{kg ha}^{-1}$ ), valores genotípicos (BLUP), efeitos de resíduo e efeitos de tendência representados pela soma dos efeitos de repetição (blocos) e efeitos aleatório de linhas.



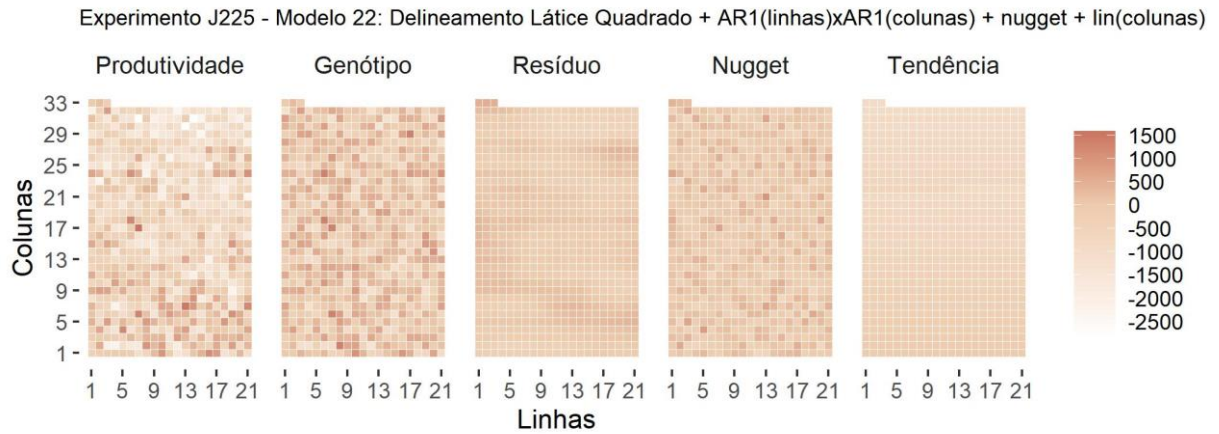
**Figura S5.** Mapas de calor em nível de parcela para o experimento VA21 referentes aos dados fenotípicos centrados ( $\text{kg ha}^{-1}$ ), valores genotípicos (BLUP), efeitos de resíduo e efeitos de tendência representados pela soma dos efeitos de repetição (blocos) e efeito linear de linhas.



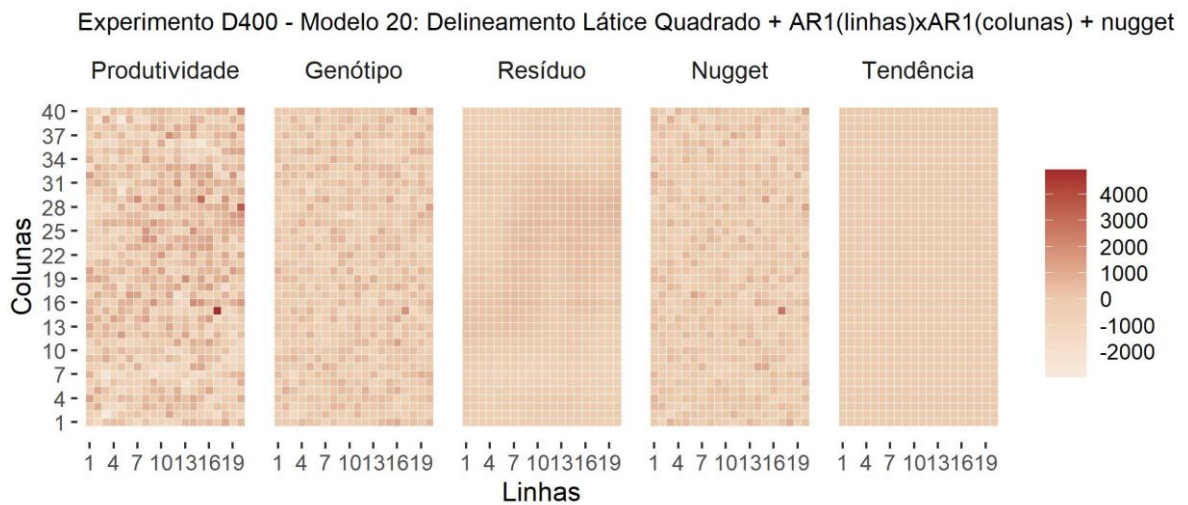
**Figura S6.** Mapas de calor em nível de parcela para o experimento C400 referentes aos dados fenotípicos centrados ( $\text{kg ha}^{-1}$ ), valores genotípicos (BLUP), efeitos de resíduo e efeitos de tendência representados pela soma dos efeitos de repetição, blocos e efeitos aleatórios de linha. Quadrados em cinza escuro referem-se a parcelas perdidas.



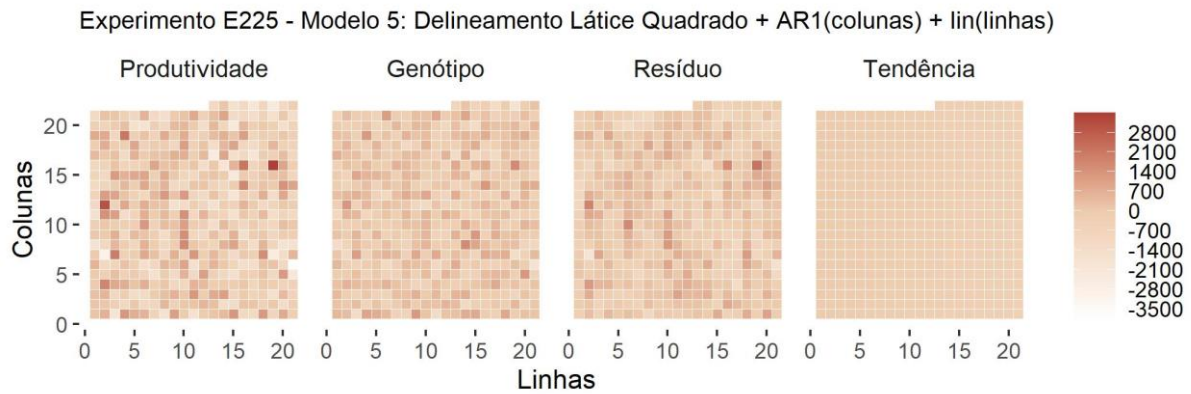
**Figura S7.** Mapas de calor em nível de parcela para o experimento P400 referentes aos dados fenotípicos centrados ( $\text{kg ha}^{-1}$ ), valores genotípicos (BLUP), efeitos de resíduo, efeitos nugget e efeitos de tendência representados pela soma dos efeitos de repetição, blocos, efeito linear de linhas e efeito linear de colunas. Quadrados em cinza escuro referem-se a parcelas perdidas.



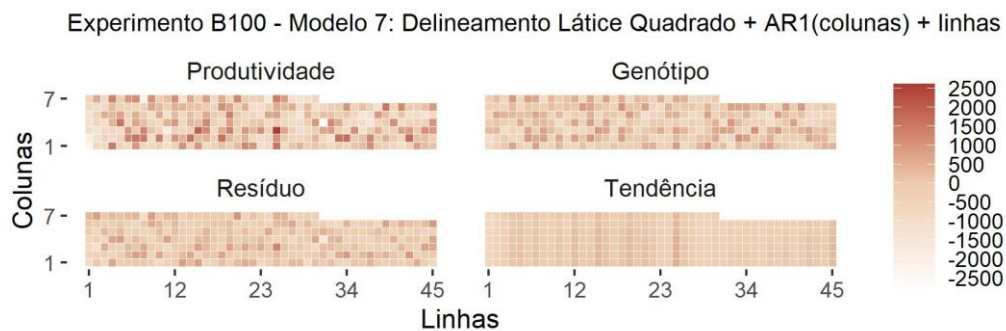
**Figura S8.** Mapas de calor em nível de parcela para o experimento J225 referentes aos dados fenotípicos centrados ( $\text{kg ha}^{-1}$ ), valores genotípicos (BLUP), efeitos de resíduo, efeitos nugget e efeitos de tendência representados pela soma dos efeitos de repetição, blocos e efeito linear de colunas.



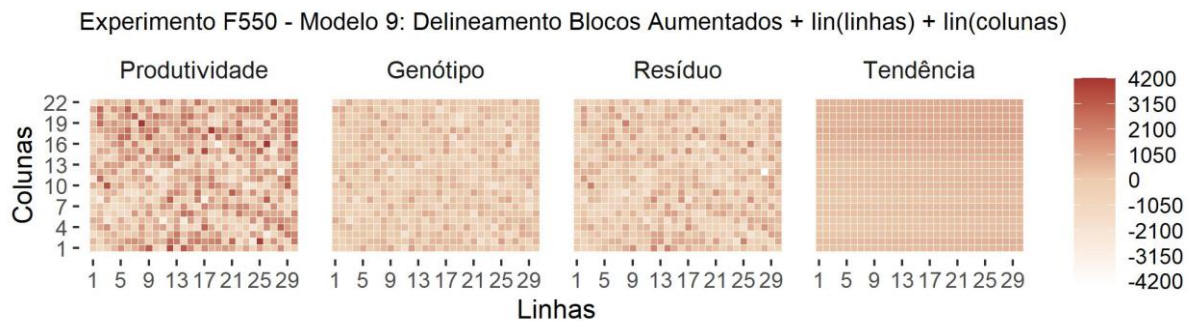
**Figura S9.** Mapas de calor em nível de parcela para o experimento D400 referentes aos dados fenotípicos centrados ( $\text{kg ha}^{-1}$ ), valores genotípicos (BLUP), efeitos de resíduo, efeitos nugget, e efeitos de tendência representados pela soma dos efeitos de repetição e blocos.



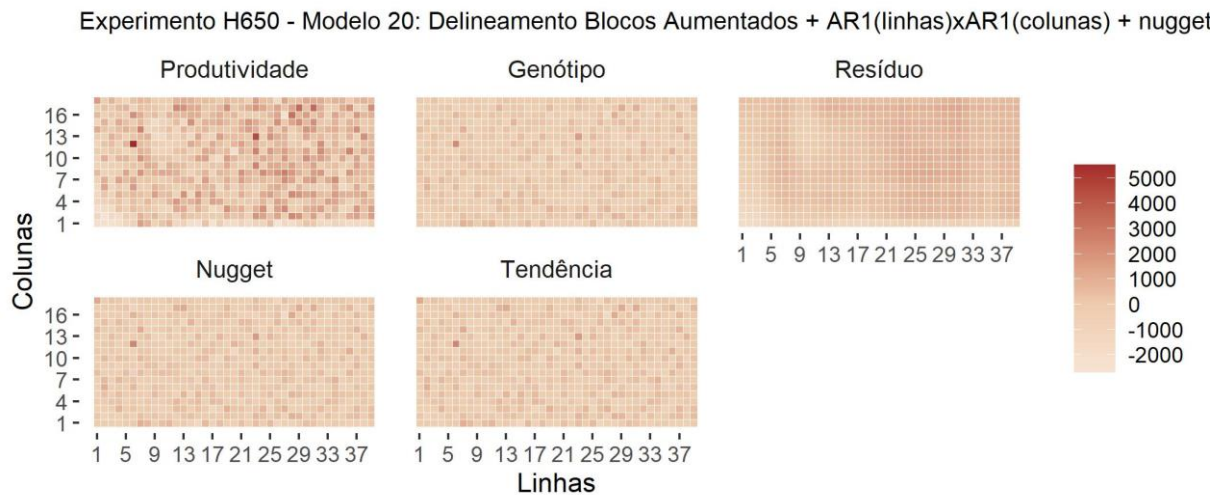
**Figura S10.** Mapas de calor em nível de parcela para o experimento E225 referentes aos dados fenotípicos centrados ( $\text{kg ha}^{-1}$ ), valores genotípicos (BLUP), efeitos de resíduo e efeitos de tendência representados pela soma dos efeitos de repetição, blocos e efeito linear de linha.



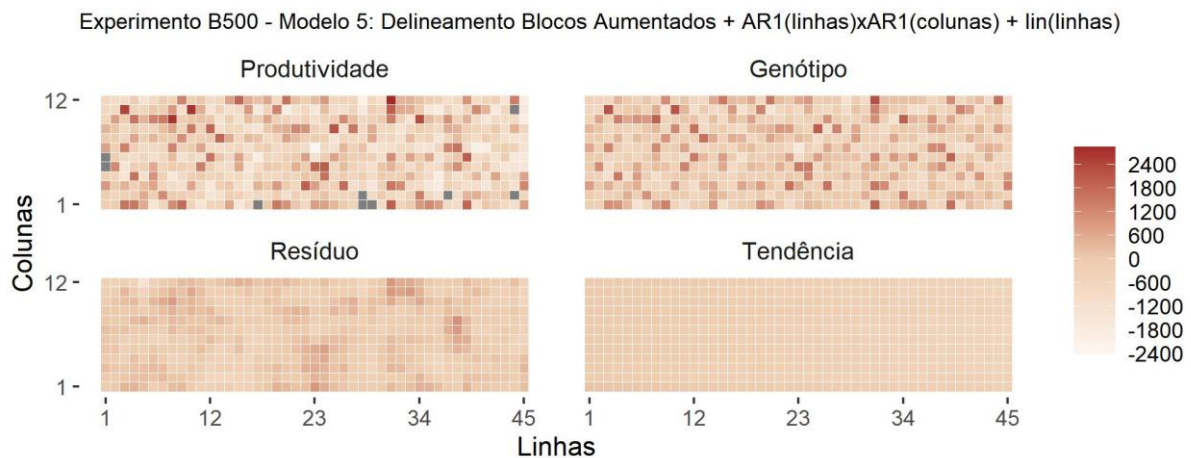
**Figura S11.** Mapas de calor em nível de parcela para o experimento B100 referentes aos dados fenotípicos centrados ( $\text{kg ha}^{-1}$ ), valores genotípicos (BLUP), efeitos de resíduo e efeitos de tendência representados pela soma dos efeitos de repetição, blocos e efeitos aleatórios de linha.



**Figura S12.** Mapas de calor em nível de parcela para o experimento F550 referentes aos dados fenotípicos centrados ( $\text{kg ha}^{-1}$ ), valores genotípicos (BLUP), efeitos de resíduo e efeitos de tendência representados pela soma dos efeitos de blocos, efeito linear de linhas e efeito linear de colunas.



**Figura S13.** Mapas de calor em nível de parcela para o experimento H650 referentes aos dados fenotípicos centrados ( $\text{kg ha}^{-1}$ ), valores genotípicos (BLUP), efeitos de resíduo, efeitos nugget e efeitos de tendência representados pelos efeitos de blocos.



**Figura S14.** Mapas de calor em nível de parcela para o experimento B500 referentes aos dados fenotípicos centrados ( $\text{kg ha}^{-1}$ ), valores genotípicos (BLUP), efeitos de resíduo e efeitos de tendência representados pela soma dos efeitos blocos e efeito linear de linhas. Quadrados em cinza escuro referem-se a parcelas perdidas.

## **CAPÍTULO II**

### **ANÁLISE ESTATÍSTICA ESPACIAL DE EXPERIMENTOS DE CAMPO NO MELHORAMENTO DO FEIJOEIRO – MODELO SpATS**

## **Resumo**

O uso de delineamentos experimentais nem sempre garante a captação da heterogeneidade espacial ao longo da área de experimentos como os realizados no melhoramento do feijoeiro. Recentemente, uma abordagem de análise espacial denominada SpATS baseada no ajuste de uma superfície de suavização penalizada foi proposta. Assim, o objetivo deste trabalho foi avaliar a eficiência do modelo SpATS em captar as tendências espaciais existentes em experimentos com feijoeiro, assim como a praticidade de sua utilização. Foram utilizados dados fenotípicos de 19 experimentos de um programa de melhoramento de feijoeiro para os quais o modelo tradicional de análise e o modelo SpATS foram ajustados. Os modelos foram comparados com base nas estimativas de herdabilidade, acurácia e variância do erro de predição (PEV), assim como nas discordâncias entre os genótipos mais bem classificados. O modelo SpATS captou tendências espaciais através da superfície suavizada e dos efeitos aleatórios de linha e coluna para vários experimentos, o que pode ser constatado pelas estimativas de dimensão efetiva. Aumentos nas estimativas de herdabilidade, acurácia e diminuição da PEV em razão da captação das tendências espaciais pelo modelo SpATS foram observados para a maioria dos experimentos. Isto resultou em discordâncias entre os genótipos mais bem classificados ao considerar os modelos básico e SpATS. Em termos de praticidade, o modelo SpATS mostrou-se muito atrativo, pois permitiu a captação de todas as tendências em um único passo. Assim, concluiu-se o modelo SpATS contribuiu com praticidade e eficiência para a análise de dados experimentais do melhoramento do feijoeiro, sendo uma excelente ferramenta para ser incorporada à rotina de análise de dados do programa de melhoramento.

## **1 Introdução**

Programas de melhoramento genético vegetal utilizam diferentes estratégias para a obtenção, avaliação e seleção de genótipos candidatos, com o intuito de desenvolver novas

cultivares que atendam às principais demandas de produtores, indústria e consumidores (BORÉM e MIRANDA., 2013). Dentro deste processo, a fenotipagem dos materiais desenvolvidos em experimentos é a atividade que detém a maior parcela do orçamento destes programas (RAMALHO, FERREIRA, OLIVEIRA, 2012; COBB et al., 2019; YAN, 2021). Para a cultura do feijão-comum (*Phaseolus vulgaris* L.), os programas de melhoramento no Brasil são conduzidos especialmente por instituições públicas com limitação de recursos disponíveis. Embora esses programas tenham obtido crescente progresso genético ao longo dos anos para vários caracteres (FARIA et al., 2013; BARILI et al., 2016; ZEFFA et al., 2020), estratégias que aumentem a precisão experimental destes programas sem aumentar os custos são importantes de serem adotadas.

Neste sentido, a correta modelagem estatística dos dados experimentais é fundamental para a eficiência das etapas de seleção, uma vez que promove a acurada predição dos valores genotípicos (RESENDE e ALVES, 2020). No caso da avaliação de caracteres quantitativos, dos quais tem-se como o principal exemplo a produtividade de grãos, a variação fenotípica é contínua devido a contribuição de muitos genes, mas também devido ao efeito de fatores ambientais (FALCONER e MACKAY, 1991; BERNARDO, 2010). Por sua vez, na área experimental a heterogeneidade de fatores como fertilidade e estrutura do solo, umidade, interceptação de luz podem ocorrer em grande ou mesmo em pequena escala e devem ser captadas para reduzir o erro experimental (RESENDE e STURION, 2003; GEZAN et al. 2010).

Assim, a utilização de delineamentos experimentais têm sido procedimento padrão para controlar *a priori* a heterogeneidade da área experimental e fazer a estimação e validação do erro experimental (RAMALHO, FERREIRA, OLIVEIRA, 2012; CASLER et al., 2015). No entanto, devido a incapacidade destes delineamentos de captar toda a heterogeneidade ao longo da área, sua utilização associada a informações da posição das parcelas em busca de captar as tendências da área tem sido recomendada.

Com desenvolvimento das equações de modelos mistos (HENDERSON et al., 1959), Gilmour et al. (1997) propôs a modelagem de tendências locais pela flexibilização da matriz de covariância dos resíduos, substituindo a relação de independência por uma que assume autocorrelação de primeira ordem (AR1) no sentido de linhas e colunas, e as tendências globais sendo modeladas com outros efeitos inseridos no modelo e dependentes das coordenadas espaciais. No entanto, Velazco et al. (2018) relata que algumas desvantagens deste procedimento estão relacionadas ao processo laborioso de modelagem das tendências espaciais e dependente de múltiplos passos podendo envolver vários modelos, o que dificulta a utilização

rotineira para analisar vários experimentos, além de possibilidades de superparametrização e problemas de convergência.

Diante disso, recentemente Rodríguez-Álvarez et al. (2018) propuseram um procedimento baseado em superfície de suavização penalizada (P-splines) bidimensionais. Essa superfície de suavização modela a heterogeneidade espacial de larga e pequena escala em um único passo por meio de modelo misto denominado SpATS (Análise espacial de experimentos de campo com P-Splines) que também inclui os efeitos de genótipos e aqueles associados aos delineamentos experimentais (repetições, blocos). Este procedimento tem sido utilizado com a finalidade de corrigir os valores fenotípicos para as tendências espaciais em diversos trabalhos apresentando resultados satisfatórios (VELAZCO et al. 2017; ANDRADE et al., 2020; FAYE et al., 2021; PÉREZ-VALENCIA et al., 2022; GRUNEBERG et al., 2022). Sua utilização tem sido reportada inclusive para a cultura do feijão (ARIZA-SUAREZ et al., 2022; KELLER et al., 2020).

Assim, o objetivo deste trabalho foi a avaliar a eficiência do modelo SpATS em captar a heterogeneidade espacial existente em experimentos do melhoramento do feijoeiro, para conferir maior eficiência à análise de dados e acurácia à predição dos valores genotípicos, além de sua praticidade visando à sua adoção em futuros experimentos do programa de melhoramento.

## **2 Material e Métodos**

### **3.1 Experimentos de campo**

Para a avaliação do método SpATS, foram utilizados os dados de produtividade de grãos ( $\text{kg ha}^{-1}$ ) por parcela obtidos nos mesmos 19 experimentos (Tabela 1) utilizados para avaliar a os modelos com estrutura autorregressiva, no capítulo anterior. De forma semelhante, além dos dados fenotípicos de produtividade foram coletadas as informações de posição das parcelas (linha e coluna) no grid experimental.

## **2.2 Análises estatística dos dados**

### **2.2.1 Modelo básico de análise**

O modelo de análise tradicional considerando cada tipo de delineamento experimental foi definido como modelo básico. Assim, para os experimentos conduzidos sob o delineamento em blocos casualizados (DBC), o modelo misto abaixo foi utilizado:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\tau} + \mathbf{Z}_g\mathbf{u}_g + \mathbf{e}$$

Em que:  $\mathbf{y}$  é o vetor de observações;  $\boldsymbol{\tau}$  é o vetor de efeitos fixos de repetições (blocos completos) somados a média geral, associado a uma matriz de incidência  $\mathbf{X}$ ;  $\mathbf{u}_g$  é o vetor de efeitos aleatórios de genótipos (linhagens, progênies etc.) que inclui as testemunhas, associado a uma matriz de incidência  $\mathbf{Z}_g$ ;  $\mathbf{e}$  é o vetor de resíduos aleatórios independentes.

Ao se tratar de experimentos sob delineamento látice quadrado (DLQ) e em blocos aumentados (DBA), um vetor referente aos efeitos aleatórios de blocos incompletos e sua matriz de incidência correspondente são adicionados ao modelo, conforme a seguir:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\tau} + \mathbf{Z}_g\mathbf{u}_g + \mathbf{Z}_b\mathbf{u}_b + \mathbf{e}$$

Em que:  $\boldsymbol{\tau}$  é um vetor de efeitos fixos que contempla os efeitos de repetições somados à média geral (para o delineamento DLQ) e somente a média geral (para o delineamento DBA).  $\mathbf{u}_b$  é o vetor de efeitos aleatórios de blocos incompletos, cuja matriz de incidência é  $\mathbf{Z}_b$ . Os vetores  $\mathbf{y}$ ,  $\mathbf{u}_g$  e  $\mathbf{e}$ , assim como as matrizes  $\mathbf{X}$  e  $\mathbf{Z}_g$  são definidos conforme anteriormente.

Assumiu-se independência entre os efeitos aleatórios de genótipo [ $\mathbf{u}_g \sim N(0, \mathbf{G})$ ], de blocos incompletos (quando presentes no delineamento) [ $\mathbf{u}_b \sim N(0, \mathbf{B})$ ], residuais [ $\mathbf{e} \sim N(0, \mathbf{R})$ ], com matrizes de covariância  $\mathbf{G} = \mathbf{I}_g\sigma_g^2$ ,  $\mathbf{B} = \mathbf{I}_b\sigma_b^2$ , e  $\mathbf{R} = \sigma^2[\mathbf{I}_r \otimes \mathbf{I}_c]$ , respectivamente. Em que:  $\sigma_g^2$  é a variância genotípica,  $\sigma_b^2$  é a variância entre blocos incompletos e  $\sigma^2$  é a variância residual.  $\mathbf{I}_g$ ,  $\mathbf{I}_b$ ,  $\mathbf{I}_r$  e  $\mathbf{I}_c$  são matrizes identidade com ordem correspondente ao número de genótipos, blocos incompletos, linhas e colunas de cada experimento, respectivamente.

### 2.2.2 Modelo SpATS

O método SpATS (Spatial Analysis of field Trials with Splines), desenvolvido e apresentado por Rodríguez-Álvarez et al. (2018) foi utilizado. Esta metodologia baseia-se na inclusão de uma função bivariada [ $f(r, c)$ ] das coordenadas de cada parcela dos experimentos (linha – r; coluna – c) no modelo estatístico, com a objetivo de capturar as tendências espaciais. O ajuste desta função resulta em uma superfície suavizada que representa as tendências ao longo da área experimental, e permite a captura de todos os tipos de variação (locais e globais) em um único passo (RODRÍGUEZ-ÁLVAREZ et al., 2018).

Assim, o modelo SpATS ajustado para experimentos sob delineamento DBC, pode ser definido conforme a seguir:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\tau} + \mathbf{X}_s\boldsymbol{\beta}_s + \mathbf{Z}_s\mathbf{s}_s + \mathbf{Z}_d\mathbf{u}_d + \mathbf{Z}_g\mathbf{u}_g + \mathbf{e}$$

Em que:  $\mathbf{y}$  é o vetor de observações;  $\boldsymbol{\tau}$  é o vetor de efeitos fixos de repetições somados a média geral, associado a uma matriz de incidência  $\mathbf{X}$ ;  $\mathbf{u}_d$  é o vetor que inclui os efeitos aleatórios de

linha e coluna, com matriz de incidência  $\mathbf{Z}_d$ , em que  $\mathbf{Z}_d = [\mathbf{Z}_{linha} | \mathbf{Z}_{coluna}]$ ;  $\mathbf{u}_g$  é o vetor de efeitos aleatórios de genótipos (linhagens, progênies etc.) incluindo as testemunhas, associado a uma matriz de incidência  $\mathbf{Z}_g$ ;  $\mathbf{e}$  é o vetor de resíduos aleatórios independentes.

No caso dos experimentos sob delineamento DLQ e DBA, o modelo SpATS ajustados é modificado para o seguinte:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\tau} + \mathbf{X}_s\boldsymbol{\beta}_s + \mathbf{Z}_s\mathbf{s}_s + \mathbf{Z}_d\mathbf{u}_d + \mathbf{Z}_g\mathbf{u}_g + \mathbf{Z}_b\mathbf{u}_b + \mathbf{e}$$

Em que:  $\mathbf{y}$  é o vetor de observações;  $\boldsymbol{\tau}$  é o vetor de efeitos fixos de repetições somados a média geral (para experimentos em DLQ) ou com somente a média geral (para experimentos em DBA), associado a uma matriz de incidência  $\mathbf{X}$ .  $\mathbf{u}_b$  é o vetor de efeitos aleatórios de blocos incompletos, com matriz de incidência  $\mathbf{Z}_b$ . Os demais componentes são definidos conforme para o modelo SpATS para experimentos em DBC.

Assim como no modelo básico, assumiu-se independência entre os efeitos aleatórios de genótipos, blocos incompletos (quando presentes) e residuais. Além disto, os efeitos aleatórios de linha e coluna também foram assumidos como independentes, ou seja  $\mathbf{u}_d \sim N(0, \mathbf{D})$  com matriz de covariância  $\mathbf{D} = \text{diag}(\mathbf{I}_r\sigma_{lin}^2, \mathbf{I}_r\sigma_{col}^2)$ .

Os termos do modelo SpATS, que não foram definidos acima, são os responsáveis por definir a função bivariada que ajusta a superfície suavizada e que capta tendências espaciais, ou seja:  $f(r, c) = \mathbf{X}_s\boldsymbol{\beta}_s + \mathbf{Z}_s\mathbf{s}_s$ . A primeira parte desta expressão ( $\mathbf{X}_s\boldsymbol{\beta}_s$ ) corresponde aos efeitos fixos no modelo misto e não são penalizados, enquanto a segunda parte ( $\mathbf{Z}_s\mathbf{s}_s$ ) é aleatória e penalizada por parâmetros suavizadores que controlam o balanço entre suavização da superfície e o ajuste aos dados (VELAZCO et al, 2017). Para a estimação destes parâmetros, é fundamental que a função bivariada possa ser representada conforme a estrutura de um modelo misto. Assim, o método SpATS faz uso da representação PS-ANOVA (P-Spline ANOVA) proposta por Lee et al. (2013), que permite decompor a função suavizada em:

$$\mathbf{X}_s\boldsymbol{\beta}_s = \beta_{s1}r + \beta_{s2}c + \beta_{s3}rc \quad e$$

$$\mathbf{Z}_s\mathbf{s}_s = f_1(r) + f_2(c) + h_3(r)c + rh_4(c) + f_5(r, c)$$

Em que:  $\beta_{s1}$ ,  $\beta_{s2}$  são coeficientes para tendências lineares ao longo das linhas e ao longo de colunas, respectivamente e  $\beta_{s3}$  para tendências de interação linear ao longo de linhas e colunas.  $f_1(r)$  e  $f_2(c)$  são os efeitos de suavização (*smooth effects*) unidimensionais ao longo de linhas e colunas;  $h_3(r)c$  e  $rh_4(c)$  são tendências de interação linear x suave (*linear-by-smooth*) ao longo de linhas e colunas, respectivamente;  $f_5(r, c)$  são tendências devido à interação suave x suave (*smooth-by-smooth*) ao longo de linhas e colunas (RODRÍGUEZ-ÁLVAREZ et al., 2018). Assim, o vetor  $\mathbf{s}_s$  é composto por 5 sub-vetores referentes aos 5 efeitos mutuamente

independentes da decomposição acima [ $f_1(r)$ ,  $f_2(c)$ ,  $h_3(r)c$ ,  $rh_4(c)$  e  $f_5(r, c)$ ], os quais são dependentes de um parâmetro específico de suavização  $\lambda_k$  em que  $k = 1, 2, \dots, 5$  e que é função do componente de variância associado a cada um destes efeitos, sendo calculados da seguinte forma:  $\lambda_k = \sigma^2 / \sigma_k^2$  (VELAZCO et al., 2017). Assim, por se tratar de uma penalização baseada em vários parâmetros de suavização, o método SpATS é dito como um método de suavização anisotrópica (VELAZCO et al., 2017; RODRÍGUEZ-ÁLVAREZ et al., 2018). A obtenção dos parâmetros de suavização é possível pois eles são função dos componentes de variância associados aos respectivos efeitos aleatórios, obtidos pelo método da máxima verossimilhança restrita – REML (PATTERSON e THOMPSON, 1977).

Ao ajustar o modelo SpATS, foram obtidas as estimativas de dimensão efetiva (DE) do modelo estatístico associadas aos efeitos aleatórios de genótipos, resíduos, linhas, colunas, e cada um dos cinco efeitos constituintes da parte penalizada da curva suavizadora. As estimativas de dimensão efetiva são importantes pois permitem avaliar a complexidade associadas a cada um desses efeitos nos experimentos.

O modelo foi ajustado aos dados fenotípicos de cada experimento utilizando o pacote SpATS (RODRÍGUEZ-ÁLVAREZ et al., 2018) no software R (R CORE TEAM, 2021). Na definição do modelo, foram utilizadas B-splines cúbicas ( $q = 3$ ) com penalizações de segunda ordem para cada marginal. O número de linhas e de colunas de cada experimento foi adotado como número de interior de nós igualmente espaçados em cada uma destas direções.

### 2.3 Comparação entre os modelos

O modelo básico foi comparado ao modelo SpATS com base nos componentes de variância genotípica e residual estimados, assim como nas estimativas de herdabilidade, acurácia seletiva e variância do erro de predição (PEV).

Para o modelo básico, a herdabilidade foi computada utilizando o estimador generalizado, conforme sugerido por Cullis et al. (2006):

$$h_g^2 = 1 - \frac{\bar{v}_{BLUP}}{2\sigma_g^2}$$

Em que  $\bar{v}_{BLUP}$  é a média da variância do erro de predição par a par entre os BLUPs (melhor predição linear não viesada) dos efeitos de genótipo e  $\sigma_g^2$  é a variância genotípica.

Para o modelo SpATS, a herdabilidade foi calculada com base nos conceitos de dimensão efetiva em modelos de regressão com splines, conforme Velazco et al. (2017):

$$h_g^2 = \frac{DE_g}{n_g}$$

Em que:  $DE_g$  refere-se à dimensão efetiva associada aos efeitos genotípicos e  $n_g$  é o número de tratamentos no experimento (Tabela 1) que, neste trabalho, equivale a soma do número de genótipos avaliados e testemunhas. Segundo estes autores, para o caso do presente trabalho que assume independência entre os efeitos genótipos (ausência de pedigree ou parentesco genômico) avaliados dentro de cada experimento, a equação acima equivale à expressão generalizada proposta por Cullis et al. (2006) e utilizada tradicionalmente para modelos mistos. Assim, torna-se possível a comparação direta das estimativas de herdabilidade entre os dois modelos (VELAZCO et al., 2017).

A acurácia de seleção ( $r_{\hat{g}g}$ ) foi calculada da forma tradicional (Resende et al.; 2014, p. 190):

$$r_{\hat{g}g} = \sqrt{1 - \frac{PEV}{\sigma_g^2}}$$

Em que  $PEV$  é a média da variância do erro de predição e  $\sigma_g^2$  é a variância genotípica.

Assim como no capítulo 1, o impacto na seleção dos genótipos mais promissores foi avaliado por meio da concordância e discordância entre os genótipos selecionados quando os valores genotípicos obtidos com os modelos básico e SpATS foram utilizados. Para isto, utilizou-se novamente a metodologia proposta por Hamblin and Zimmerman (1986) considerando os grupos de 5% (A) e 10% (B) mais bem classificados (para experimentos com mais de 45 tratamentos), e considerando os grupos 20% (C) e 30% (D) mais bem classificados (para todos os experimentos).

### 3 Resultados e Discussão

Ao analisar os valores estimados de dimensão efetiva associados aos efeitos do modelo SpATS (Tabela 5), pode-se observar que as variações devido às tendências espaciais captadas pelo modelo (superfície suavizada e efeitos aleatórios de linhas e colunas) nos experimentos avaliados são de baixa proporção em relação às fontes de variação genotípica e residual. Estes resultados indicam que a influência de tendências espaciais existentes ao longo dos experimentos tende a ser leve. Ainda assim, são esperadas melhorias na acurácia e herdabilidade ao modelar estas tendências pois é fundamental que os modelos utilizados sejam capazes de estimar e separar as frações genéticas daquelas devido às tendências espaciais

existentes na área visando a predição acurada dos valores genotípicos (ROBBINS; BACKLUND e SCHNELLE, 2012).

No caso de metodologias baseadas em regressões spline como o modelo SpATS, as estimativas de dimensão efetiva (DE) são primordiais na interpretação da importância de cada componente do modelo SpATS, uma vez que estas informam sobre complexidade associada a cada um deles (VELAZCO et al., 2017; RODRÍGUEZ-ALVARÉZ et al., 2018). Com relação a superfície suavizada ( $\mathbf{Z}_s\mathbf{s}_s$ ) utilizada para captar as tendências espaciais no modelo SpATS, os maiores valores de dimensão efetiva (Tabela 5) foram observados para experimentos cujo modelo espacial escolhido no capítulo I (SALVADOR et al., 2022) apresentou algum tipo de estrutura autorregressiva para os resíduos (G45, G62, F100, VA21, P400, J225, D400, E225, B100, H650 e B500), evidenciando a associação entre complexidade das tendências espaciais existente na área e a dimensão efetiva. Duas exceções, são os experimentos S112 e F550, cujos modelos ajustados no capítulo I (Salvador et al., 2022) não necessitaram de estrutura autorregressiva para a modelagem dos resíduos, porém, com o modelo SpATS apresentaram-se dentre os experimentos com maiores valores de dimensão efetiva para a função suavizada (Tabela 5).

É importante salientar que houve predominância dos efeitos de interação do tipo suave x suave [ $f_5(r, c)$ ] nos experimentos S112, G45, VI21, F100, C400, J225, E225 e B500 (Tabela 5). Os efeitos de tendências de interação do tipo linear x suave ao longo de linhas ou colunas [ $h_3(r)c$  ou  $rh_4(c)$ , respectivamente] foram importantes para os experimentos G62, J225, D400 e H650 (Tabela 5). Já os efeitos de suavização unidimensionais ao longo de linhas [ $f_1(r)$ ] apresentaram importância relativa na superfície suavizada para os experimentos VA21, P400, B100 e F500; enquanto, no sentido de colunas [ $f_2(c)$ ], para os experimentos F550 e H650 (Tabela 5). Observou-se que para vários experimentos, houve efeitos cujas estimativas de dimensão efetiva foram iguais ou próximas a zero, o que significa que estes efeitos não contribuem ou não têm impacto na variabilidade espacial destes experimentos (RODRÍGUEZ-ALVARÉZ et al., 2018).

1 **Tabela 5.** Dimensões efetivas para os componentes espaciais da parte aleatória da superfície suavizada ( $\mathbf{Z}_s\mathbf{s}_s$ ), dos efeitos aleatórios de linha ( $\mathbf{u}_r$ ) e  
 2 coluna ( $\mathbf{u}_c$ ), genótipos ( $\mathbf{u}_g$ ) e resíduos ( $\mathbf{e}$ ) considerando o modelo SpATS ajustado aos dados de produtividade coletados em experimentos de feijão.

Componente	Experimento																		
	DBC						DLQ						DBA						
	S112	G45	G62	J64	VI18	VI21	A45	F100	VA18	VA21	C400	P400	J225	D400	E225	B100	F550	H650	B500
$f_1(r)$	1,8	1,2	0,3	0,9	0,0	0,0	0,4	0,0	0,0	3,1	0,0	3,4	1,8	0,6	0,0	3,6	2,9	1,6	1,0
$f_2(c)$	1,1	0,0	0,0	0,0	0,3	0,0	0,0	1,1	0,0	0,0	0,0	0,0	1,5	3,3	0,0	1,6	2,6	2,9	0,4
$h_3(r)c$	0,0	0,0	4,4	0,0	0,7	0,0	0,4	0,0	2,1	0,7	0,5	1,8	0,9	2,4	0,0	1,5	1,0	4,8	3,2
$rh_4(c)$	0,8	0,0	1,7	0,0	0,0	0,3	0,0	2,7	0,4	0,6	0,0	2,6	6,5	4,2	3,1	0,0	2,2	0,0	1,2
$f_5(r,c)$	5,3	6,0	5,4	0,0	0,0	3,1	0,2	9,9	0,0	0,0	2,8	2,8	10,2	5,0	12,2	0,0	3,6	4,1	10,3
Total ( $\mathbf{Z}_s\mathbf{s}_s$ )	9,0	7,2	11,8	0,9	1,0	3,4	1,0	13,7	2,5	4,4	3,3	10,6	20,9	15,5	15,3	6,7	12,3	13,4	16,1
Linhas	0,0	11,4	8,1	6,2	0,1	0,0	0,2	0,0	0,0	0,0	14,2	17,2	0,1	5,5	10,8	17,6	0,1	20,6	9,4
Colunas	1,1	0,0	0,0	4,2	-	-	4,4	20,9	-	-	-	3,3	15,5	-	1,8	0,0	1,1	9,5	2,8
Total	10,1	18,6	19,9	11,3	1,1	3,4	5,6	34,6	2,5	4,4	17,5	31,1	36,5	21	27,9	24,3	13,5	43,5	28,3
$\mathbf{u}_g$	100,5	37,8	48,2	46,0	15,0	17,2	36,8	71,3	15,8	18,6	190,8	220,9	195	263	162,3	77,8	218,9	302	370,7
$\mathbf{e}$	219,5	72,6	111,9	128,7	32,9	35,5	84,5	188,1	23,5	34,6	562,4	526,1	431,9	493,2	254,1	192	422,5	352,5	125,2

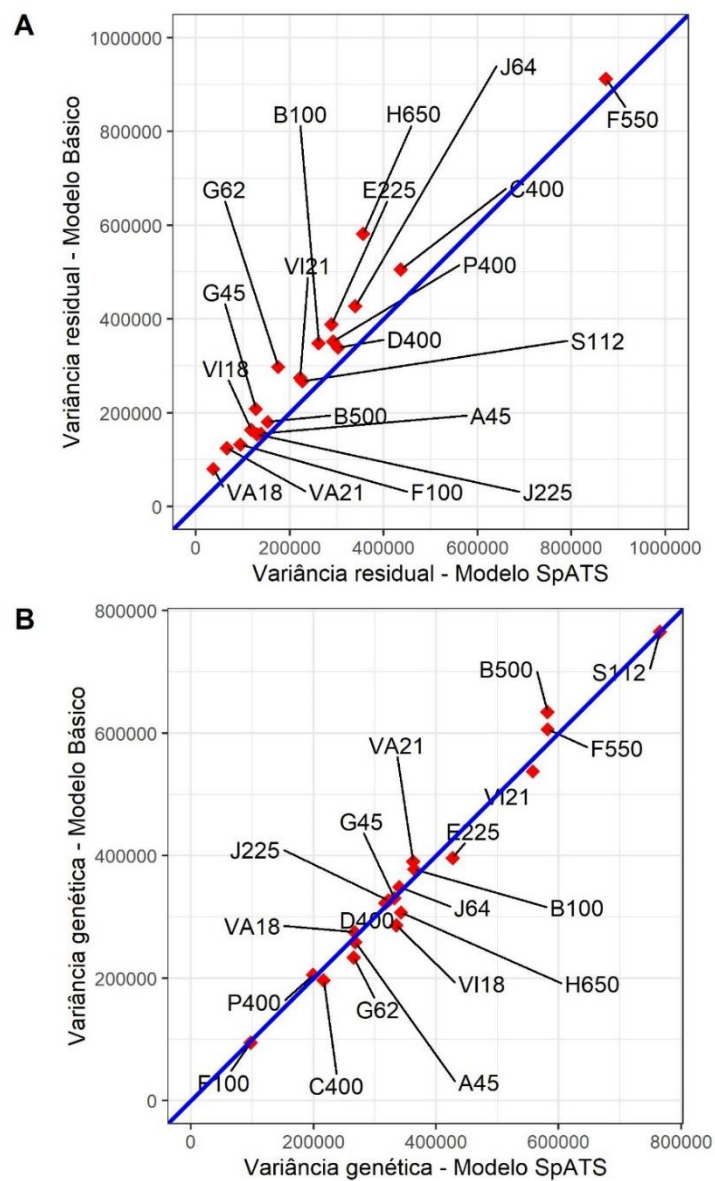
As tendências espaciais captadas ao longo da área experimental pela superfície suavizada nos experimentos sob delineamentos de blocos completos (DBC) e incompletos (DLQ e DBA) podem ser visualizados nos gráficos de calor das figuras suplementares S15 e S16, respectivamente. Por meio destas representações gráficas, observa-se que a heterogeneidade espacial pode se manifestar de diferentes formas ao longo da área experimental, ou seja, de maneira local como manchas, pontos ou em faixas específicas, assim como em gradientes ao longo de toda a área. As causas para a ocorrência desta heterogeneidade ao longo de linhas e colunas de um experimento podem ser inerentes a diversos fatores como variações em aspectos físicos e químicos do solo, topografia, interceptação de luz, práticas de manejo como a irrigação e ocorrência de doenças e estresses que estejam se manifestando de forma não uniforme ao longo da área (GILMOUR et al., 1997; RESENDE e STURION, 2003; GEZAN et al, 2010, ANDRADE et al., 2020).

Além da superfície suavizada, as correções espaciais foram complementadas no modelo SpATS pela captação dos efeitos aleatórios de linhas e colunas (Tabela 5). Com base nos valores de dimensão efetiva, estes efeitos foram importantes para captar tendências no sentido de linhas nos experimentos G45, G62, J64, C400, P400, D400, E225, B100, H650 e B500 e no sentido de colunas nos experimentos J64, A45, F100, J225 e H650 (Tabela 5). Segundo Piepho et al. (2022), a inclusão no modelo de efeitos aleatórios e independentes associados às coordenadas espaciais de linha e coluna é importante para captar quaisquer outras tendências que não foram captadas pela curva suavizada.

As comparações entre as estimativas obtidas pelo modelo básico e pelo modelo SpATS para a variância residual e genotípica podem ser visualizadas na figura 3. Houve redução da variância residual ao modelar as tendências espaciais com o modelo SpATS para todos os experimentos (Figura 3-A), o que indica a eficiência deste modelo em captar as tendências espaciais existentes ao longo da área experimental que antes não eram captadas pelo modelo básico. A heterogeneidade espacial, quando não captada pelo modelo estatístico de análise, infla a variância residual, o que afeta a comparação entre os genótipos avaliados (DUARTE e VENCOVSKY, 2005), além de prejudicar parâmetros importantes para a definição do ganho de seleção, como a herdabilidade e a acurácia (HOARAU et al., 2021; RESENDE, SILVA, AZEVEDO, 2014). A redução da variância residual com o auxílio do modelo SpATS é especialmente importante em experimentos conduzidos sob delineamento DBA (B500, H650 e B500), para os quais a estimação dos resíduos é muitas vezes imprecisa devido à ausência de repetição para a maioria dos genótipos avaliados e pelo uso de tratamentos comuns

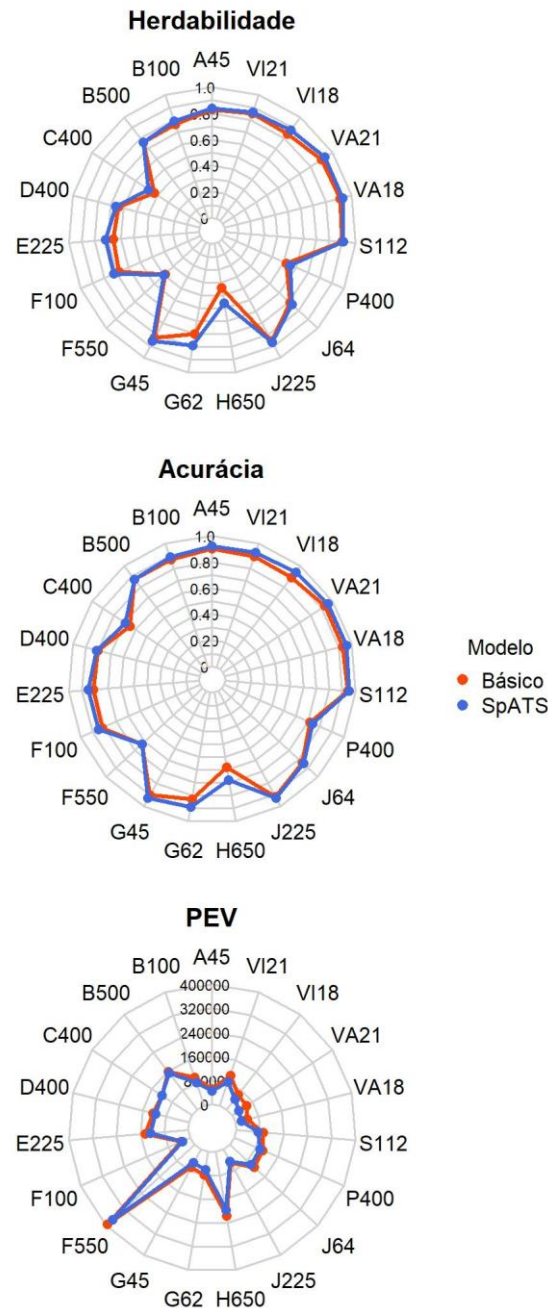
(testemunhas) cujo nível de variação dentro não é compatível com os demais genótipos (RAMALHO, FERREIRA, OLIVEIRA, 2012).

A redução observada na variância residual com o modelo SpATS foi acompanhada por um aumento na variância genotípica nos experimentos VI18, VI21, A45, G62, C400, E225 e H650. Em contrapartida, a variância genotípica foi ligeiramente menor com o modelo SpATS nos experimentos F550, B500, VA21 e VA18 (Figura 3-B). Esta instabilidade no comportamento da variância genotípica ao utilizar os modelos básico e SpATS também foi observada por Andrade et al. (2020) em experimentos no melhoramento de batata.



**Figura 3.** Comparações das estimativas de variância genética (A) e residual (B) obtidas ao utilizar o modelo básico e o modelo SpATS para cada um dos 19 experimentos de feijoeiro. A linha diagonal azul em cada gráfico identifica a posição de valores idênticos.

O ajuste conferido pelo modelo SpATS resultou em aumento nos valores de herdabilidade ( $h_g^2$ ) e acurácia ( $r_{\hat{g}}$ ), para a maioria dos experimentos (Figura 4). As exceções são os experimentos B500, para a herdabilidade e acurácia, e F550 para acurácia, para os quais essas estimativas foram idênticas em ambos os modelos. Além disso, a variância do erro de predição (PEV) foi reduzida com a modelagem espacial em todos os experimentos, com exceção do experimento C400, cuja PEV foi igual em ambos os modelos (Figura 4).

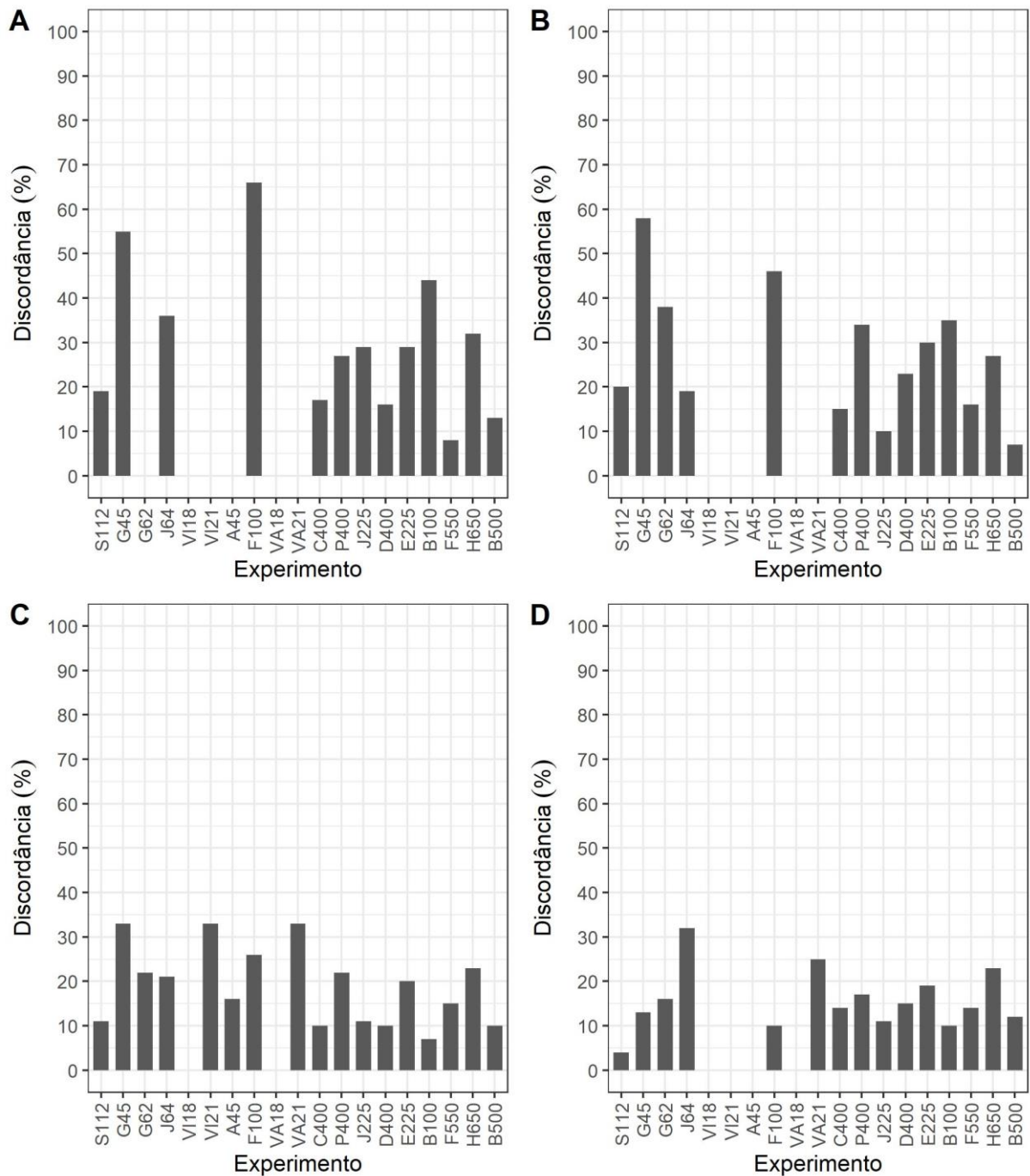


**Figura 4.** Comparação das estimativas de herdabilidade, acurácia e variância do erro de predição (PEV) obtidas ao utilizar o modelo básico e o modelo SpATS para cada um dos 19 experimentos de feijoeiro.

Com relação a herdabilidade, os maiores incrementos ao utilizar o modelo SpATS foram nos experimentos H650 (35%), G62 (13%), C400 (12%) e E225 (9%) (Figura 4-A). Conseqüentemente, esses experimentos também apresentaram aumentos consideráveis nas estimativas de acurácia: H650 (17%), G62 (7%), C400 (6%) e E225 (5%) (Figura 4-B). Utilizando o modelo SpATS, incrementos na herdabilidade de caracteres agrônômicos, morfológicos e de qualidade foram obtidos por Muktar et al. (2022) ao fazer a correção dos valores fenotípicos em experimentos com capim Napier e por Andrade et al. (2022) na maioria de seus experimentos com batata.

Ao observar os genótipos melhor classificados considerando os valores genotípicos obtidos por cada modelo, apenas para o experimento VI18 não foram constatadas discordâncias entre os modelos (Figura 5). Para as demais combinações de experimentos e grupos de genótipos mais bem classificados, a correção para os efeitos espaciais captados pelo modelo SpATS resultaram em discordâncias. Este efeito sobre os genótipos mais bem classificados traz conseqüências positivas à seleção e é coerente com os aumentos na acurácia e as diminuições da PEV, uma vez que estas estatísticas representam a correlação do valor genotípico predito com o verdadeiro e a precisão com que as predições ocorrem, respectivamente (RESENDE e DUARTE, 2007; RESENDE e ALVES, 2020).

Os resultados deste trabalho demonstram que o modelo SpATS, por meio da modelagem de uma função bivariada baseada nas coordenadas espaciais, foi eficiente em captar as tendências espaciais existentes nas áreas dos experimentos avaliados e isso resultou na estimação/predição mais fidedigna de componentes de variância, parâmetros genéticos e valores genotípicos. É importante notar que, ao contrário dos modelos baseados na estrutura autorregressiva dos resíduos (GILMOUR et al., 1997), o método SpATS não presume a existência de correlação entre os resíduos de parcelas vizinhas. Alternativamente, as tendências espaciais são modeladas de forma direta por meio da inclusão da função bivariada no modelo e da estimação de parâmetros de suavização no ajuste aos dados (RODRIGUEZ-ÁLVAREZ et al., 2018), o que facilita muito os procedimentos de análise pois a modelagem de tendências locais e globais é feita em um único passo. Além disto, a decomposição da superfície suavizada utilizada pelo pacote SpATS em componentes lineares e de suavização (LEE et al. 2013) torna a interpretação da importância das tendências espaciais bastante simples pois é feita de forma direta por meio das estimativas de dimensão efetiva.



**Figura 5.** Frequência de genótipos discordantes ao classificar os valores genotípicos obtidos pelos modelos básico e SpATS, considerando os grupos de 5% (A) e 10% (B) mais bem classificados (para experimentos com mais de 45 tratamentos), e considerando os grupos 20% (C) e 30% (D) mais bem classificados (para todos os experimentos).

Diante do exposto, conclui-se que a utilização do modelo SpATS contribuiu para o aumento da eficiência da análise de dados experimentais do melhoramento do feijoeiro. Diante da fácil implementação deste método e do seu potencial para a captação de tendências espaciais existentes nas áreas experimentais, o modelo é uma excelente ferramenta a ser incorporada à rotina de análise de dados do programa de melhoramento.

#### 4 Referências

- ANDRADE, M. H. M. L. et al. Accounting for spatial trends to increase the selection efficiency in potato breeding. **Crop Science**. v. 60, p. 2354–2372, mai. 2020. DOI: <https://doi.org/10.1002/csc2.20226>
- ARIZA-SUAREZ, D. et al. Genetic analysis of resistance to bean leaf crumple virus identifies a candidate LRR-RLK gene. **The Plant Journal**. 2022. DOI: <https://doi.org/10.1111/tpj.15810>
- BARILI, L. D. et al. Genetic progress resulting from forty-three years of breeding of the carioca common bean in Brazil. **Genetics and Molecular Research**, v. 15, n. 3, p. 1–11, 2016. DOI: [10.4238/gmr.15038523](https://doi.org/10.4238/gmr.15038523).
- BERNARDO, R. **Quantitative traits in plants**. 2.ed. Woodbury, MN: Stemma Press, 2010. 390p.
- BORÉM, A.; MIRANDA, G. V. **Melhoramento de Plantas**. 6 ed. Viçosa: Editora UFV: 2013. 525p.
- CASLER, M. D. Fundamentals of Experimental Design: Guidelines for Designing Successful Experiments. **Agronomy Journal**, v. 107, p. 692-705. 2015. DOI: <https://doi.org/10.2134/agronj2013.0114>
- COBB, J. N. et al. Enhancing the rate of genetic gain in public-sector plant breeding programs: lessons from the breeder's equation. **Theoretical and Applied Genetics**, v. 132, p. 627–645. 2019. DOI: <https://doi.org/10.1007/s00122-019-03317-0>
- CULLIS, B. R.; SMITH, A. B.; COOMBES, N. E. On the design of early generation variety trials with correlated data. **Journal of Agricultural, Biological, and Environmental Statistics**. v. 11, p. 381–393. 2006. DOI: <https://doi.org/10.1198/108571106x154443>
- DUARTE, J. B., VENCOVSKY, R. Spatial statistical analysis and selection of genotypes in plant breeding. **Pesquisa Agropecuária Brasileira**, 40, 107-114. 2005. DOI: <https://doi.org/10.1590/S0100-204x2005000200002>
- FALCONER, D. S.; MACKAY, T. F. C. **Introduction to Quantitative Genetics**. 4 ed. Harlow: Longman. 1996. 464p.
- FARIA, L. C. et al. Genetic progress during 22 years of improvement of carioca-type common bean in Brazil. **Field Crops Research**, v. 142, p. 68–74, 2013. DOI: <https://doi.org/10.1016/j.fcr.2012.11.016>
- FAYE, J. M. et al. Quantitative and population genomics suggest a broad role of stay-green loci in the drought adaptation of sorghum. **The Plant Genome**. v. 15, p. 1-17. 2021. DOI: [10.1002/tpg2.20176](https://doi.org/10.1002/tpg2.20176)
- GEZAN, S. A.; WHITE, T. L.; HUBER, D. A. Accounting for Spatial Variability in Breeding Trials: A Simulation Study. **Agronomy Journal**. v. 102, p. 1562-1571, set. 2010. DOI: <https://doi.org/10.2134/agronj2010.0196>

GILMOUR, A. R.; CULLIS, B. R.; VERBYLA, A. P. Accounting for Natural and Extraneous Variation in the Analysis of Field Experiments. **Journal of Agricultural, Biological, and Environmental Statistics**, v. 2, n. 3, p. 269–293, set. 1997.

GRÜNEBERG, W. J. et al. Heterosis and Responses to Selection in Orange-Fleshed Sweetpotato (*Ipomoea batatas* L.) Improved Using Reciprocal Recurrent Selection. **Frontiers in Plant Science**. v.13. p. 1-15. 2022. DOI: 10.3389/fpls.2022.793904

HAMBLIN, J.; ZIMMERMAN, M. J. O. Breeding common bean for yield mixtures. **Plant Breeding Reviews**. v. 4, p. 245–272. 1986. DOI: <https://doi.org/10.1002/9781118061015.ch8>

HENDERSON, C. R. et al. The estimation of environmental and genetic trends from records subject to culling. **Biometrics**, v. 15, 192-218. 1959.

HOARAU, J. Y. et al. Applications of Quantitative Genetics and Statistical Analyses in Sugarcane Breeding. **Sugar Tech**. v. 24, p. 320–340. 2021. DOI: <https://doi.org/10.1007/s12355-021-01012-3>

LEE, D.; DURBÁN, M.; EILERS, P. Efficient two-dimensional smoothing with P spline ANOVA mixed models and nested bases. **Computational Statistics and Data Analysis**. v.61. p. 22-37. 2013. DOI: <https://doi.org/10.1016/j.csda.2012.11.013>

KELLER, B. et al. Genomic Prediction of Agronomic Traits in Common Bean (*Phaseolus vulgaris* L.) Under Environmental Stress. **Frontiers in Plant Science**. v. 11. p. 1-15. 2020. DOI: <https://doi.org/10.3389/fpls.2020.01001>

MUKTAR, M. S. et al. Insights Into the Genetic Architecture of Complex Traits in Napier Grass (*Cenchrus purpureus*) and QTL Regions Governing Forage Biomass Yield, Water Use Efficiency and Feed Quality Traits. **Frontiers in Plant Science**. v. 12. p. 1-22. 2022. DOI: <https://doi.org/10.3389/fpls.2021.678862>

PATTERSON, H. D.; THOMPSON, R. Recovery of inter-block information when block sizes are unequal. **Biometrika**. v. 58, p. 545–554. 1971. DOI: <https://doi.org/10.1093/biomet/58.3.545>

PÉREZ-VALENCIA, D. M. et al. A two-stage approach for the spatio-temporal analysis of high-throughput phenotyping data. **Scientific Reports**. v.12. 2022. DOI: <https://doi.org/10.1038/s41598-022-06935-9>

PIEPHO, H. P.; BOER, M. P.; WILLIAMS, E. R. Two-dimensional P-spline smoothing for spatial analysis of plant breeding trials. **Biometrical Journal**. v.64, n. 5, p. 835-857. 2022. DOI: <https://doi.org/10.1002/bimj.202100212>

R CORE TEAM. **R: A language and environment for statistical computing**. R Found. Stat. Comput., Vienna. 2021.

RAMALHO, M. A. P.; FERREIRA, D. F.; OLIVEIRA, A. C. **Experimentação em Genética e Melhoramento de Plantas**. 3. Ed. Lavras: Editora UFLA. 2012. 329p.

RESENDE, M. D. V.; ALVES, R. S. Linear, generalized, hierarchical, bayesian and random regression mixed models in genetics/genomics in plant breeding. **Functional Plant Breeding Journal**. v. 2, p. 1-31, jun. 2020. DOI: <https://doi.org/10.35418/2526-4117/v2n2a1>

RESENDE, M. D. V.; DUARTE, J. B. Precisão e controle de qualidade em experimentos de avaliação de cultivares. **Pesquisa Agropecuária Tropical**. v.37, p. 182–194, set. 2007.

RESENDE, M. D. V.; SILVA, F.F.; AZEVEDO, C.F. **Estatística matemática, biométrica e computacional**: modelos mistos, multivariados, categóricos e generalizados (REML/BLUP), inferência Bayesiana, regressão aleatória, seleção genômica, QTL-GWAS, estatística espacial e temporal, competição, sobrevivência. Viçosa, MG: Suprema, 2014. 881p.

RESENDE, M. D. V.; STURION, J. A. Análise estatística espacial de experimentos via modelos mistos individuais com erros modelados por processos ARIMA em duas dimensões. **Rev. Mat. Estat.**, v. 21, n. 1, p. 7-33, jan. 2003.

ROBBINS, K. R.; BACKLUND, J. E.; SCHNELLE, K. D. Spatial Corrections of Unreplicated Trials using a Two-dimensional Spline. **Crop Science**. v.52. p. 1138-1144. 2012. DOI: <https://doi.org/10.2135/cropsci2011.08.0417>

RODRÍGUEZ-ÁLVAREZ et al. Correcting for spatial heterogeneity in plant breeding experiments with P-splines. **Spatial Statistics**, v. 23, p. 52-71. 2018. DOI: <https://doi.org/10.1016/j.spasta.2017.10.003>

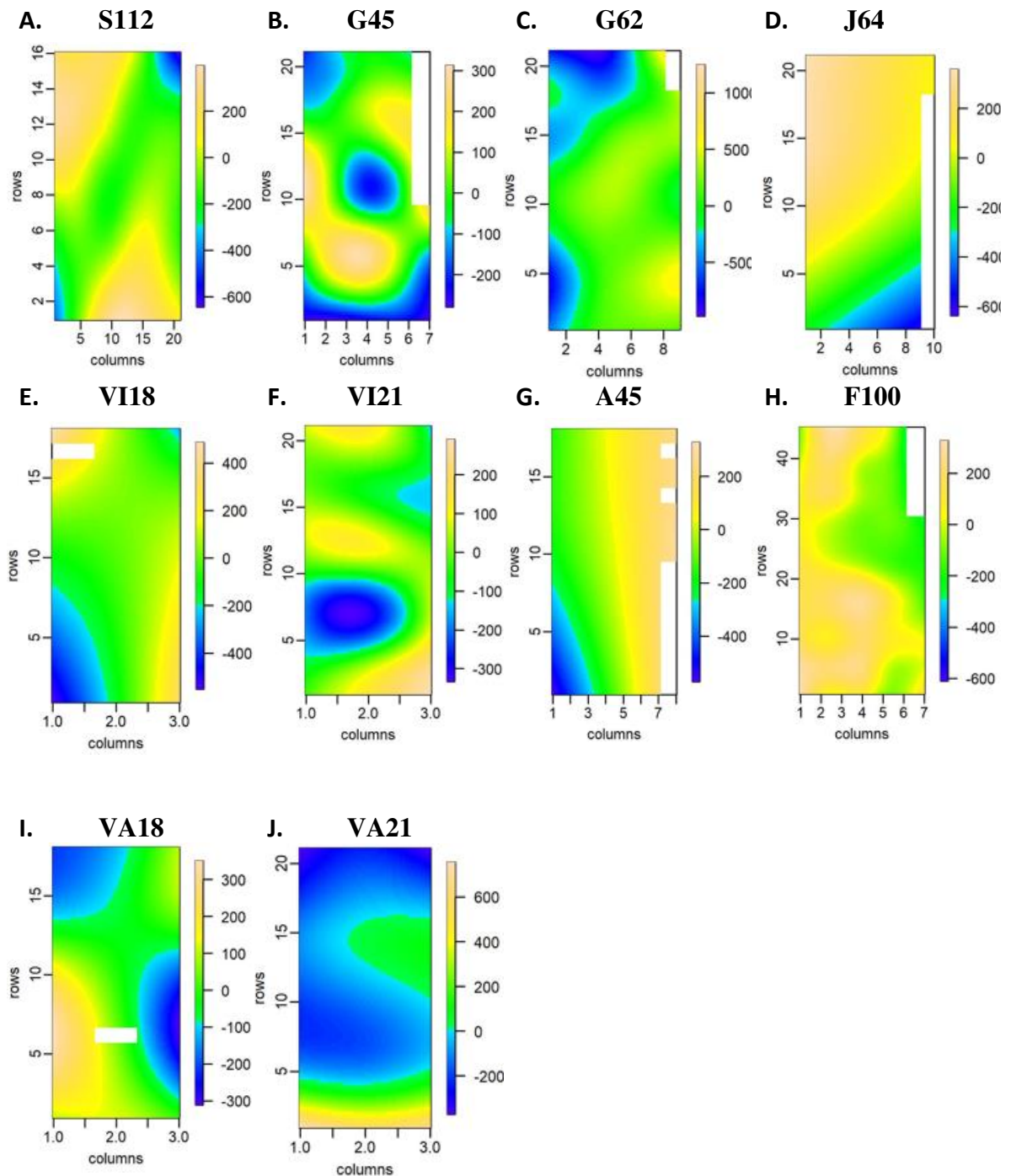
SALVADOR, F. V., et al. Correcting experimental data for spatial trends in a common bean breeding program. **Crop Science**, v. 62, n. 2, p. 825-838. 2022. DOI: <https://doi.org/10.1002/csc2.20703>

VELAZCO, J. G., et al. Modelling spatial trends in sorghum breeding field trials using a two-dimensional P-spline mixed model. **Theoretical and Applied Genetics**, v. 130, p. 1375–1392. 2017. DOI: [10.1007/s00122-017-2894-4](https://doi.org/10.1007/s00122-017-2894-4)

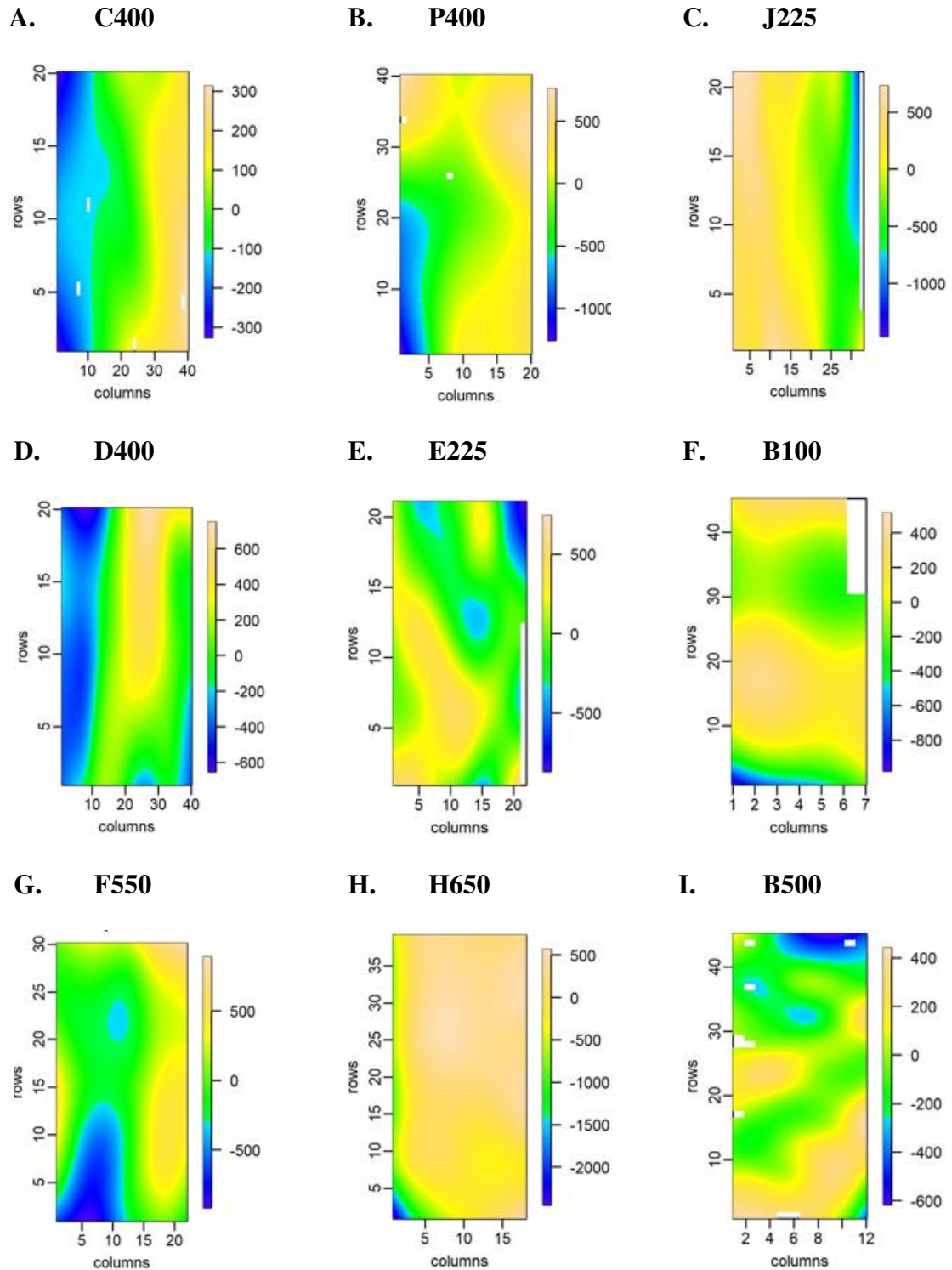
YAN, W. Estimation of the Optimal Number of Replicates in Crop Variety Trials. **Frontiers in Plant Science**, v. 11, p. 1-14. 2021. DOI: [10.3389/fpls.2020.590762](https://doi.org/10.3389/fpls.2020.590762)

ZEFFA, D. M. et al. Genetic Progress of Seed Yield and Nitrogen Use Efficiency of Brazilian carioca Common Bean Cultivars Using Bayesian Approaches. **Frontiers in Plant Science**, v. 11, p. 1-14. 2020. DOI: <https://doi.org/10.3389/fpls.2020.01168>

## 5 Material suplementar



**Figura S15.** Tendência espacial (superfície suavizada) ajustada pelo modelo SpATS para os experimentos de campo de feijoeiro conduzidos sob o delineamento em blocos casualizados (DBC).



**Figura S16.** Tendência espacial (superfície suavizada) ajustada pelo modelo SpATS para os experimentos de campo de feijoeiro conduzidos sob o delineamento látice quadrado (DLQ) e blocos aumentados (DBA).

## CONCLUSÕES GERAIS

Independentemente da abordagem utilizada (estrutura autorregressiva ou SpATS), a análise estatística espacial mostrou-se mais eficiente em relação à forma tradicional de análise (sem considerar informações espaciais) uma vez que, para a maioria dos experimentos com a cultura do feijoeiro considerados, foram detectadas tendências espaciais ao longo das áreas experimentais, resultando em aumentos nas estimativas de herdabilidade e acurácia ou redução na variância do erro de predição (PEV), o que, conseqüentemente, afetou a predição dos valores genotípicos.

Dentre as abordagens de análise espacial, o modelo SpATS apresentou resultados ligeiramente superiores, em termos de acurácia e herdabilidade, à metodologia autorregressiva para a maioria dos experimentos em que ambas as abordagens detectaram tendências espaciais.

Além disto, para os experimentos S112, A45, VA18, VI18 e VI21, foram detectadas tendências espaciais apenas no modelo SpATS, resultando em melhorias nas predições de parâmetros oriundas destes experimentos, o que não foi possível com os modelos autorregressivos. Estes resultados podem estar associados a existência de tendências do tipo spline nestes experimentos, sendo somente captadas pelo modelo SpATS. É importante salientar que efeitos do tipo spline também podem ser adicionados aos modelos espaciais com estrutura autorregressiva, o que poderia aumentar a eficiência destes modelos ao ajustar os dados destes experimentos. No entanto, estes efeitos não foram considerados no presente trabalho pois elevaria muito o número de modelos comparados no capítulo I.

Os modelos com estrutura autorregressiva e o seu processo passo a passo de ajuste são excelentes para entender a natureza das tendências espaciais existentes nos experimentos. No entanto, o processo de ajuste envolve múltiplos passos e a comparação de vários modelos. Mesmo que alguns modelos com tendências mais frequentemente observadas tenham sido recomendados para serem testados em futuros experimentos, é totalmente possível que tendências não observadas anteriormente ocorram, uma vez que a heterogeneidade espacial é inerente a cada experimento. Neste sentido, a praticidade na utilização do modelo SpATS, conferida pela inclusão simultânea de todos os efeitos em um único modelo, é outro ponto a seu favor, em termos de utilização rotineira no programa de melhoramento do feijoeiro.

Assim, conclui-se que a utilização de modelos espaciais na análise de dados experimentais do melhoramento do feijoeiro é um procedimento eficiente e passível de ser utilizado como forma de detecção de tendências espaciais existentes nas áreas experimentais,

assim como na correção destes dados visando maior assertividade nas tomadas de decisão das etapas de seleção e recomendação.

## APÊNDICES

APÊNDICE A – Exemplo da estrutura do arquivo de dados utilizado neste trabalho para as análises com o software Echidna (GILMOUR, 2020) e com o pacote “SpATS” (RODRÍGUEZ-ÁLVAREZ et al., 2018) em R (R Core Team, 2021). As colunas plots, rep, rows, columns, gen e yield são colunas que codificam as parcelas, repetições, linhas, colunas, genótipos e a produtividade de grãos por parcela ( $\text{kg ha}^{-1}$ ).

plots	rep	rows	columns	gen	yield
101	1	1	1	2	3130
102	1	2	1	25	3820
103	1	3	1	5	4680
104	1	4	1	7	3410
105	1	5	1	22	4540
⋮	⋮	⋮	⋮	⋮	⋮
343	3	7	7	35	4630
344	3	8	7	26	4420
345	3	9	7	7	4340

APÊNDICE B – Exemplo de arquivo de parâmetros utilizado para a etapa I da análise de dados do capítulo I no software Echidna (GILMOUR, 2020).

```
!REN !OUT Etapa1_m !ARG 1 2 3 4 !SLN
Analise !DOPART $1
plots *
rep *
rows *
columns *
gen *
yield
```

```
dataF100.txt !SKIP 1 !MVINCLUDE !MAXIT 500 !EXT 3 !VRB
```

```
!PART 1 # Modelo 1 (Básico)
```

```
yield ~ mu rep !r gen !f mv
```

```
Predict gen
```

```
!PART 2 # Modelo 2
```

```
yield ~ mu rep !r gen !f mv
```

```
residual columns.ar1(rows)
```

```
Predict gen
```

```
!PART 3 # Modelo 3
```

```
yield ~ mu rep !r gen !f mv
residual ar1(columns).rows
Predict gen
```

```
!PART 4
```

```
# Modelo 4
```

```
yield ~ mu rep !r gen !f mv
residual ar1(columns).ar1(rows)
Predict gen
```

```
# Referência:
```

```
#GILMOUR, A.R. (2020). Echidna Mixed Model Software. 2020. Disponível em:
#http://www.EchidnaMMS.org
```

APÊNDICE C – Exemplo de arquivo de parâmetros utilizado para a etapa II da análise de dados do capítulo I no software Echidna (GILMOUR, 2020).

```
!REN !OUT Etapa2a_m !ARG 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 !SLN
Analise !DOPART $1
plots *
rep *
rows *
columns *
gen *
yield
```

```
dataF100.txt !SKIP 1 !MVINCLUDE !MAXIT 500 !EXT 3 !VRB
```

#OBSERVAÇÃO: Foi incluída a estrutura residual escolhida na etapa I em todos os modelos da etapa na etapa II.

```
!PART 1
```

```
# Modelo 5
```

```
yield ~ mu rep lin(rows) !r gen !f mv
residual ar1(columns).ar1(rows)
Predict gen
```

```
!PART 2
```

```
# Modelo 6
```

```
yield ~ mu rep lin(columns) !r gen !f mv
residual ar1(columns).ar1(rows)
Predict gen
```

```
!PART 3
```

```
# Modelo 7
```

```
yield ~ mu rep !r gen rows !f mv
residual ar1(columns).ar1(rows)
Predict gen
```

!PART 4 # Modelo 8

```
yield ~ mu rep !r gen columns !f mv
residual ar1(columns).ar1(rows)
Predict gen
```

!PART 5 # Modelo 9

```
yield ~ mu rep lin(rows) lin(columns) !r gen !f mv
residual ar1(columns).ar1(rows)
Predict gen
```

!PART 6 # Modelo 10

```
yield ~ mu rep lin(rows) !r gen rows !f mv
residual ar1(columns).ar1(rows)
Predict gen
```

!PART 7 # Modelo 11

```
yield ~ mu rep lin(rows) !r gen columns !f mv
residual ar1(columns).ar1(rows)
Predict gen
```

!PART 8 # Modelo 12

```
yield ~ mu rep lin(columns) !r gen rows !f mv
residual ar1(columns).ar1(rows)
Predict gen
```

!PART 9 # Modelo 13

```
yield ~ mu rep lin(columns) !r gen columns !f mv
residual ar1(columns).ar1(rows)
Predict gen
```

!PART 10 # Modelo 14

```
yield ~ mu rep !r gen rows columns !f mv
residual ar1(columns).ar1(rows)
Predict gen
```

!PART 11 # Modelo 15

```
yield ~ mu rep lin(rows) lin(columns) !r gen rows !f mv
residual ar1(columns).ar1(rows)
Predict gen
```

!PART 12 # Modelo 16

```
yield ~ mu rep lin(rows) lin(columns) !r gen columns !f mv
residual ar1(columns).ar1(rows)
Predict gen
!PART 13 # Modelo 17
```

```
yield ~ mu rep lin(rows) !r gen rows columns !f mv
residual ar1(columns).ar1(rows)
Predict gen
```

```
!PART 14 # Modelo 18
```

```
yield ~ mu rep lin(columns) !r gen rows columns !f mv
residual ar1(columns).ar1(rows)
Predict gen
```

```
!PART 15 # Modelo 19
```

```
yield ~ mu rep lin(rows) lin(columns) !r gen rows columns !f mv
residual ar1(columns).ar1(rows)
Predict gen
```

# Referência:

#GILMOUR, A.R. (2020). **Echidna Mixed Model Software**. 2020. Disponível em:  
#<http://www.EchidnaMMS.org>

APÊNDICE D – Exemplo de arquivo de parâmetros utilizado para a etapa III da análise de dados do capítulo I no software Echidna (GILMOUR, 2020).

```
!REN !OUT Etapa3_m !ARG 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 !SLN
Analise !DOPART $1
plots *
rep *
rows *
columns *
gen *
yield
```

```
dataF100.txt !SKIP 1 !MVINCLUDE !MAXIT 500 !EXT 3 !VRB
```

#OBSERVAÇÃO: Foi incluída a estrutura residual escolhida na etapa I em todos os modelos da etapa na etapa III.

```
!PART 1 # Modelo 20
```

```
yield ~ mu rep !r gen units !f mv
residual ar1(columns).ar1(rows)
Predict gen
```

!PART 2 # Modelo 21

```
yield ~ mu rep lin(rows) !r gen units !f mv
residual ar1(columns).ar1(rows)
Predict gen
```

!PART 3 # Modelo 22

```
yield ~ mu rep lin(columns) !r gen units !f mv
residual ar1(columns).ar1(rows)
Predict gen
```

!PART 4 # Modelo 23

```
yield ~ mu rep !r gen rows units !f mv
residual ar1(columns).ar1(rows)
Predict gen
```

!PART 5 # Modelo 24

```
yield ~ mu rep !r gen columns units !f mv
residual ar1(columns).ar1(rows)
Predict gen
```

!PART 6 # Modelo 25

```
yield ~ mu rep lin(rows) lin(columns) !r gen units !f mv
residual ar1(columns).ar1(rows)
Predict gen
```

!PART 7 # Modelo 26

```
yield ~ mu rep lin(rows) !r gen rows units !f mv
residual ar1(columns).ar1(rows)
Predict gen
```

!PART 8 # Modelo 27

```
yield ~ mu rep lin(rows) !r gen columns units !f mv
residual ar1(columns).ar1(rows)
Predict gen
```

!PART 9 # Modelo 28

```
yield ~ mu rep lin(columns) !r gen rows units !f mv
residual ar1(columns).ar1(rows)
Predict gen
```

!PART 10 # Modelo 29

```
yield ~ mu rep lin(columns) !r gen columns units !f mv
residual ar1(columns).ar1(rows)
Predict gen
```

```
!PART 11 # Modelo 30
```

```
yield ~ mu rep !r gen rows columns units !f mv
residual ar1(columns).ar1(rows)
Predict gen
```

```
!PART 12 # Modelo 31
```

```
yield ~ mu rep lin(rows) lin(columns) !r gen rows units !f mv
residual ar1(columns).ar1(rows)
Predict gen
```

```
!PART 13 # Modelo 32
```

```
yield ~ mu rep lin(rows) lin(columns) !r gen columns units !f mv
residual ar1(columns).ar1(rows)
Predict gen
```

```
!PART 14 # Modelo 33
```

```
yield ~ mu rep lin(rows) !r gen rows columns units !f mv
residual ar1(columns).ar1(rows)
Predict gen
```

```
!PART 15 # Modelo 34
```

```
yield ~ mu rep lin(columns) !r gen rows columns units !f mv
residual ar1(columns).ar1(rows)
Predict gen
```

```
!PART 16 # Modelo 35
```

```
yield ~ mu rep lin(rows) lin(columns) !r gen rows columns units !f mv
residual ar1(columns).ar1(rows)
Predict gen
```

# Referência:

#GILMOUR, A.R. (2020). **Echidna Mixed Model Software**. 2020. Disponível em:  
 #<http://www.EchidnaMMS.org>

APÊNDICE E – Rotina utilizada para análise espacial dos dados no capítulo II com o software R (R Core Team, 2021) e o pacote SpATS (RODRÍGUEZ-ÁLVAREZ et al., 2018).

```

rm(list = ls())
library(SpATS)           # Carregar o pacote SpATS
setwd("C:\\... ")       # Definir o diretório de trabalho

data <- read.table("data_F100.txt", header = TRUE)   # Carregar os dados
head(data)                                           # Mostrar o início do arquivo de dados

max(data$columns)   # Obter o nº de colunas do experimento
max(data$rows)     # Obter o nº de linhas do experimento

# Transformar variáveis em fatores
data$rep <- as.factor(data$rep)
data$R <- as.factor(data$rows)
data$C <- as.factor(data$columns)
data$gen <- as.factor(data$gen)

# Definir o número de segmentos Spline em cada direção (nseg1 = colunas e nseg2 = linhas).
# Como o número de nós internos em cada direção foi definido, neste trabalho, como o número
# de linhas e colunas do experimento, foram assumidos que nseg1 = nº de colunas + 1 e nseg2 =
# nº de linhas + 1.

nseg1 = 8
nseg2 = 46

#Definindo a função suavizadora

spatial.SpATS <- as.formula(~ PSANOVA(columns, rows, nseg = c(nseg1, nseg2), degree = 3,
pord = 2, nest.div = 2))

#Definindo o modelo SpATS

mod_spats_F100 <- SpATS(response = "yield",
                        spatial = spatial.SpATS,
                        genotype = "gen", genotype.as.random = TRUE, fixed = ~ rep,
                        random = ~ R + C, data = data, control = list(tolerance = 1e-06))

summary(mod_spats_F100, which = "all") # Resumo dos resultados
plot(mod_spats_F100, main = "F100")   #Gráficos gerados pelo pacote

pred_F100 <- predict(mod_spats_F100, which = "gen") #Obter os BLUPs de genótipos
pred_F100

write.table(pred_F100,"C:\\...\\BLUPs_F100.txt")

```

```
# Salvar os resultados
```

```
sink("Resultado_SpATS_F100.txt")  
print(summary(mod_spats_F100, which = "all"))  
sink()
```

```
jpeg(filename = "plots_F100.jpeg",width = 6, height = 6, units = "in", res = 300)  
plot(mod_spats_F100, main = "Trial F100")  
dev.off()
```

```
## Referências
```

```
# RODRÍGUEZ-ÁLVAREZ et al. Correcting for spatial heterogeneity in plant breeding  
#experiments with P-splines. Spatial Statistics, v. 23, p. 52-71. 2018.  
# R CORE TEAM. R: A language and environment for statistical computing. R Found. Stat.  
#Comput., Vienna. 2021.
```