

GUILHERME RODRIGUES PEREIRA

**RECOMENDAÇÃO DE GENÓTIPOS DE SOJA UTILIZANDO MODELO
MULTIPLICATIVO MISTO**

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

Orientador: Luiz Antônio dos Santos Dias

Coorientadores: Felipe Lopes da Silva
Kaio Olímpio das Graças Dias

**VIÇOSA – MINAS GERAIS
2023**

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

T

P436r
2023
Pereira, Guilherme Rodrigues, 1997-
Recomendação de genótipos de soja utilizando modelo
multiplicativo misto / Guilherme Rodrigues Pereira. – Viçosa,
MG, 2023.

1 dissertação eletrônica (60 f.): il. (algumas color.).

Inclui apêndice.

Orientador: Luiz Antônio dos Santos Dias.

Dissertação (mestrado) - Universidade Federal de Viçosa,
Departamento de Agronomia, 2023.

Referências bibliográficas: f. 52-58.

DOI: <https://doi.org/10.47328/ufvbbt.2023.260>

Modo de acesso: World Wide Web.

1. Soja - Melhoramento genético - Mato Grosso do Sul.
2. Melhoramento Genético - Métodos estatísticos. 3. Interação
genótipo-ambiente. I. Dias, Luiz Antônio dos Santos, 1957-.
II. Universidade Federal de Viçosa. Departamento de
Agronomia. Programa de Pós-Graduação em Genética e
Melhoramento. III. Título.

CDD 22. ed. 631.5233


GUILHERME RODRIGUES PEREIRA

**RECOMENDAÇÃO DE GENÓTIPOS DE SOJA UTILIZANDO MODELO
MULTIPLICATIVO MISTO**


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

APROVADA: 24 de fevereiro de 2023.

Assentimento:

Documento assinado digitalmente
 GUILHERME RODRIGUES PEREIRA
Data: 10/05/2023 16:23:22-0300
Verifique em <https://validar.iti.gov.br>

Guilherme Rodrigues Pereira
Autor

Documento assinado digitalmente
 LUIZ ANTONIO DOS SANTOS DIAS
Data: 11/05/2023 08:54:57-0300
Verifique em <https://validar.iti.gov.br>

Luiz Antônio dos Santos Dias
Orientador

*Aos meus pais Meire Ivone e Edilson,
e ao meu irmão Jhonathan*

AGRADECIMENTOS

Agradeço a Deus, pela minha vida, pela saúde e por abençoar todos os meus projetos pessoais.

À minha mãe Meire Ivone, pelo carinho, amor, apoio e incentivo a formação pessoal.

Ao meu pai Edilson, pelo companheirismo, amor e apoio.

Ao meu irmão Jhonathan, pelo apoio, amizade e momentos de descontração.

À minha namorada, Leidiane Duarte, por todo amor, companheirismo, amizade e apoio emocional incondicional.

À Universidade Federal de Viçosa e ao Departamento de Agronomia, pela oportunidade de aliar o trabalho com a realização do mestrado. Ao programa de Pós-Graduação em Genética e Melhoramento, pela oportunidade de formação.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

A todos os professores da Universidade Federal de Viçosa que contribuíram com a minha formação profissional. Em especial, ao professor Luiz Antônio dos Santos Dias, pelos ensinamentos, conselhos, confiança, orientação e amizade durante todos esses anos. E aos professores Kaio Olimpo e Felipe Lopes, pela coorientação, ensinamentos e apoio durante o período do mestrado.

Aos amigos do Laboratório de Agroenergia e do Programa Soja, pela amizade, companheirismo, troca de conhecimentos, por momentos compartilhados, colaboração e pelo trabalho em equipe durante a execução deste trabalho. Em especial, ao Saulo e Maurício, pela transmissão do conhecimento, paciência e ajuda na realização das minhas análises.

À Fundação MS pela parceria e confiança, especialmente ao André Ricardo, por sempre ser muito solícito, pela disponibilidade e contribuições para a melhora deste trabalho.

A todos aqueles que contribuíram, direto ou indiretamente, para minha formação pessoal e profissional.

Muito obrigado!

*“Consagre ao Senhor tudo o que
você faz, e os seus planos serão
bem-sucedidos”.*

Provérbios 16:3 (Bíblia)

RESUMO

PEREIRA, Guilherme Rodrigues, M.Sc., Universidade Federal de Viçosa, fevereiro de 2023. **Recomendação de genótipos de soja utilizando modelo multiplicativo misto.** Orientador: Luiz Antônio dos Santos Dias. Coorientadores: Felipe Lopes da Silva e Kaio Olímpio das Graças Dias.

A interação genótipos por ambientes (GxA) representa a resposta diferencial dos genótipos aos diferentes ambientes em que são cultivados, e ignorá-la por ocasião da seleção não é o melhor caminho. Entretanto, antes de realizar a análise conjunta para estudar a GxA, é essencial avaliar a existência de tendências espaciais em cada ensaio, a fim de assegurar a integridade dos dados. Desse modo, objetivou-se verificar e, quando necessário, realizar o ajuste espacial de cada ambiente, a partir de um ensaio de competição conduzido em múltiplos ambientes, no estado do Mato Grosso do Sul, buscando selecionar genótipos superiores de soja. Utilizou-se uma abordagem analítica em dois estágios. No primeiro, realizou-se o ajuste espacial de cada experimento, avaliado na população de ambientes-alvo, com a metodologia 2D P-splines via modelos mistos, chamada de *Spatial Analysis of Field Trials with Splines* (SpATS), para suavização das variações espaciais de campo em duas dimensões. No segundo, empregou-se o *Factor Analytic Mixed Models* (FAMM), um modelo multiplicativo misto com estrutura fator analítico, para compreender a dinâmica da interação GxA em contexto de múltiplos ambientes. As ferramentas FAMM foram usadas para identificar os genótipos de melhor desempenho e estabilidade geral e adaptação específica a ambientes. SpATS e FAMM foram empregados para avaliação de 104 genótipos em 16 ambientes, representados pela combinação de locais, com épocas de plantio distintas, instalados em blocos completos casualizados, com três repetições. O ajuste espacial pela SpATS melhorou 12,5% dos ambientes avaliados, onde o delineamento experimental não se apresentou como método incontestável no controle da variabilidade espacial. Os genótipos com maior adaptação geral foram aqueles que apresentaram alto desempenho (*OP*) e estabilidade geral (*RMSD*), sendo selecionados com base em um índice que tende a favorecer genótipos com maior desempenho. Os valores de *OP* e *RMSD* ficaram entre -152,10 (G087) e 158,42 (G016) e 554,97 (G003) e 3,74 (G024), respectivamente. Foram recomendados 15 genótipos com adaptação geral e, como adaptação específica, o genótipo com maior desempenho.

Palavras-chave: *Glycine Max* [L.] Merrill. Ensaios em múltiplos ambientes. Interação genótipos por ambientes. Seleção de genótipos. FAMM. FAST. SpATS.

ABSTRACT

PEREIRA, Guilherme Rodrigues, M.Sc., Universidade Federal de Viçosa, February 2023. **Recommendation of soybean genotypes using mixed multiplicative model.** Advisor: Luiz Antônio dos Santos Dias. Co-advisors: Felipe Lopes da Silva and Kaio Olímpio das Graças Dias.

The genotype-by-environment interaction (GxE) represents the differential response of genotypes to the different environments in which they are cultivated, and ignoring it during selection is not the best way. However, before performing the pooled analysis to study GxE, it is essential to assess the existence of spatial trends within each run to ensure data integrity. Thus, the objective was to verify and, when necessary, carry out the spatial adjustment of each environment, based on a competition test in multiple environments, in the state of Mato Grosso do Sul, seeking to select superior soybean genotypes. A two-stage analytical approach was used. In the first stage, the spatial adjustment for each experiment evaluated in the population of target environments was performed using the 2D P-splines methodology via mixed models, called Spatial Analysis of Field Trials (SpATS) with Splines, to smooth field trends in two dimensions. In the second stage, the Factor Analytic Mixed Models (FAMM) was used, a mixed multiplicative model with an analytical factor structure for understanding dynamics of the GxE interaction in the context of multi-environment trials. FAMM tools were used to identify the genotypes with the best performance, overall stability, and specific adaptation to environments. SpATS and FAMM were used to evaluate 104 genotypes in 16 environments represented by planting time and location combinations which were installed in complete randomized blocks with three replications. The spatial adjustment by SpATS resulted in an improvement of 12.5% in the environments, where the experimental design did not appear to be an undisputed method for controlling spatial variability. The genotypes with the highest adaptation were those that showed high overall performance (OP) and root mean square deviation (RMSD) that were selected based on an index that tends to genotypes with the overall. The OP and RMSD values were between -152.10 (G087) and 158.42 (G016), and 554.97 (G003) and 3.74 (G024), respectively. Fifteen genotypes with general adaptation were recommended and, as specific adaptation, the genotype with the highest performance was selected.

Keywords: *Glycine Max* [L.] Merrill. Multiple environments trials. Genotype-by-environment interaction. Genotype selection. FAMM. FAST. SpATS.

SUMÁRIO

1. INTRODUÇÃO	10
2. REFERENCIAL TEÓRICO	13
2.1 A cultura da soja (<i>Glycine Max</i> [L.] Merrill)	13
2.2 Análise Espacial	15
2.2.1 <i>Spatial Analysis of Field Trials with Splines</i> - SpATS	15
2.3 Interação Genótipos por Ambientes	17
2.3.1 Modelos Lineares Mistos	18
2.3.2 Fator Analítico Multiplicativo Misto	21
3. MATERIAL E MÉTODOS	24
3.1 Caracterização dos dados	24
3.2 Modelos Utilizados para Análise de Dados	26
3.3.1 Modelo Base sem Análise Espacial com Erros Independentes – Modelo Linear Misto.....	27
3.3.2 Análise Espacial dos Ambientes Utilizando Modelagem Via Spline – Modelo SpATS	27
3.3.3 Seleção de Modelos para Efeito Residual	28
3.3.4 Coeficiente de Variação Genotípica (<i>CVg</i>) e Ambiental (<i>CVe</i>)	28
3.3.5 Herdabilidade	29
3.3.6 Correlação entre os BLUP's.....	30
3.4 Análise em Dois Estágios	31
3.5 Modelos FAMM	31
3.5.1 Escolha do Número de Fatores.....	33
3.5.3 Parâmetros Utilizados.....	34
3.5.4 Natureza da interação GxA	35
3.5.5 <i>Factor Analytic Selection Tools</i> – FAST	35
4. RESULTADOS	37
4.1 Resultado modelagem SpATS	37
4.2 Modelo Fator Analítico Multiplicativo Misto.....	41
5. DISCUSSÃO	48

6. CONCLUSÃO.....	51
6. REFERÊNCIAS	52
APÊNDICE A – Tabela discriminação dos genótipos.....	59

1. INTRODUÇÃO

De acordo com a Organização das Nações Unidas (ONU), a estimativa é de que a população mundial cresça em cerca de 2 bilhões de pessoas até o ano de 2050, onde é esperado alcançar 9,7 bilhões de pessoas, superando os atuais 7,7 bilhões (ONU, 2019). Neste cenário, a produção de alimentos deve aumentar cerca de 60% até 2050, e em até 100% nos países em desenvolvimento (FAO, 2015). Neste contexto, a soja (*Glycine Max* [L.] Merrill) destaca-se pela sua importância socioeconômica, garantindo renda aos produtores e segurança alimentar aos consumidores. O farelo de soja é a principal fonte de farelo proteico ofertado no mundo. Ao todo, na safra 2021/22, foram produzidos 246,57 milhões de toneladas, representando aproximadamente 70% de todo farelo proteico produzido (USDA, 2023a). A produção mundial de soja na safra de 2021/2022 atingiu 358,10 milhões de toneladas, restringindo-se praticamente a três países: Argentina, Estados Unidos e Brasil, em ordem crescente de produção (USDA, 2023b). O Brasil lidera a produção mundial, com 125,55 milhões de toneladas (CONAB, 2023), dos quais 78% foram exportados, na forma de grão, farelo e óleo. O estado do Mato Grosso do Sul contribuiu com 8,83 milhões de toneladas na safra 2021/22, com produtividade média de 2.513 kg/ha (CONAB, 2023).

Uma forma sustentável para atender a segurança alimentar mundial é a utilização de genótipos cada vez mais produtivos. Porém, a obtenção desses genótipos depende da confiabilidade nas estimativas das médias e dos parâmetros genéticos, estando diretamente relacionadas com a precisão experimental. Frequentemente, a modelagem da variância residual dos dados de forma a contabilizar a possível dependência espacial entre as unidades experimentais (parcelas) proporciona uma melhora na acurácia das estimativas. De acordo com Rodríguez -Álvarez et al. (2018), os efeitos espaciais combinados podem gerar diferentes tipos de microambientes que irão incidir de forma diferencial sobre as parcelas, influenciando na produtividade e em outras características. Adotar análises estatísticas robustas que modelam *a posteriori* essas variações, como a análise espacial, é uma maneira segura de mitigar os seus efeitos (GRONDONA et al., 1996; ANDRADE et al., 2020). A análise espacial não gera gasto adicional ou modificação na disposição dos tratamentos no campo. De fato, o único custo envolvido com este procedimento é computacional.

Existem dois modos de se realizar a modelagem espacial. O primeiro é fundamentado em estruturas espaciais de covariâncias, e o segundo em técnicas de suavização. Neste estudo, o método adotado é baseado na proposta de Rodríguez-Álvarez et al. (2018), fundamentado em

modelo 2D P-splines via modelos mistos para a suavização das tendências de campo em duas dimensões. Essa metodologia foi denominada *Spatial Analysis of Field Trials with Splines* (SpATS). O modelo se baseia no uso de P-splines que abordam a suavização como uma regressão penalizada para obter a curva ajustada, onde B-splines são penalizados nos coeficientes de regressão e há uma regressão sobre as funções de base B-splines, utilizando nós igualmente espaçados (MENDOZA, 2017; RODRÍGUEZ-ÁLVAREZ et al., 2018).

O comportamento diferencial dos genótipos a ambientes diferentes é conhecido como interação Genótipos por Ambientes (GxA). Essa interação é um efeito universal que se relaciona a todos os organismos vivos, como humanos, plantas e bactérias (KANG, 1997). A interação GxA é considerada um fator crítico na identificação dos melhores genótipos nos programas de melhoramento para serem recomendados aos agricultores (ENYEW et al., 2021). Assim sendo, o sucesso do programa de melhoramento depende da capacidade em recomendar genótipos com desempenho assegurado, em quantidade e/ou qualidade do caráter fenotípico (MALOSETTI; RIBAUT; EEUWIJK, 2013).

Os ensaios em múltiplos ambientes (*MET*, do inglês *multi-environment trials*) em arranjos de parcelas experimentais ajudam o melhorista a avaliar a magnitude e a significância da interação GxA, estimando com maior precisão os rendimentos dos genótipos e assim, possibilitando maior entendimento dos genótipos nos vários ambientes. Isso permite selecionar as melhores alternativas destinadas ao cultivo (CROSSA, 1990). As análises de adaptabilidade e estabilidade tornam possível a obtenção de genótipos com elevados desempenhos e estabilidade para serem recomendados em determinado conjunto *MET* (RODRÍGUEZ-GONZÁLEZ et al., 2011). Atualmente, há mais de uma dezena de metodologias que vem sendo desenvolvidas e aperfeiçoadas que visam a avaliação do desempenho e da estabilidade dos genótipos. As metodologias utilizadas são diversas, tais como aquelas baseadas em técnicas convencionais de análise de variância (ANOVA), regressão linear e não linear, análise multivariada e estatística não paramétrica (CROSSA, 1990). A escolha do método mais adequado é um pré-requisito para predizer com segurança os desempenhos dos genótipos (MALOSETTI; RIBAUT; EEUWIJK, 2013).

A utilização dos modelos mistos com a estrutura fator analítico foi proposta por Piepho (1997), sendo atualmente chamada de FAMM (Factor Analytic Mixed Models). FAMM é uma técnica que permite o agrupamento de variáveis observadas em variáveis latentes, chamadas “fatores”, utilizando a covariância genética entre ambientes e permitindo um estudo detalhado do efeito da interação GxA. Isto é possível mesmo na presença de heterogeneidade de variâncias e desbalanceamento dos dados (SMITH; CULLIS; THOMPSON, 2001, 2005). O FAMM é

conhecido por fornecerem uma abordagem parcimoniosa, confiável e holística (CULLIS et al., 2014). Ainda que o FAMM proporcione informações precisas do comportamento dos genótipos e da interação dos genótipos com os ambientes (SMITH; CULLIS; THOMPSON, 2001), a utilização de ferramentas auxiliares se faz necessária para facilitar o entendimento dos resultados e garantir rápida tomada de decisão nos programas de melhoramento.

Cullis et al. (2014), apresentaram gráficos de regressão latente como uma tentativa de simplificar os resultados providos pelo FAMM. Para cada genótipo, há vários gráficos para os k Fatores, sendo que, no eixo X estão as cargas dos ambientes para o fator k e no eixo Y os valores genotípicos, onde as inclinações das retas fornecem desempenho e estabilidade. Uma abordagem mais direta é apresentada por Smith e Cullis (2018), a qual explora a estrutura do FAMM, utilizando dos efeitos da interação GxA para recomendar genótipos com adaptação ampla e específica, chamada de *Factor Analytic Selection Tools* (FAST). Esse método sintetiza o resultado do FAMM em estimativas de desempenho geral e estabilidade entre todo o conjunto de ambientes para cada genótipo, facilitando a seleção dos genótipos.

Em virtude do que foi mencionado, objetivou-se verificar e, quando necessário, realizar o ajuste espacial de cada experimento avaliado na população de ambientes-alvo, com a metodologia SpATS no primeiro estágio. Tal tentativa de aumento da eficiência experimental foi conduzida ao segundo estágio, onde foi realizada uma análise conjunta com objetivo de selecionar genótipos com maior desempenho e estabilidade geral e, adaptação específica a ambientes, empregando a metodologia FAST para o Mato Grosso do Sul. Estudos que empregam a metodologia SpATS para correção espacial e a FAST para recomendação de genótipos de soja ainda são escassos.

2. REFERENCIAL TEÓRICO

2.1 A cultura da soja (*Glycine Max* [L.] Merrill)

Em várias partes do mundo a soja não é conhecida como alimento básico, diferentemente dos cereais de maneira geral. Todavia, ela é uma das culturas mais importantes, devido às suas características nutricionais (SILVA et al., 2022). O grão de soja é utilizado principalmente na agroindústria, com foco na produção de óleo, gêneros alimentícios com altos teores de proteínas e também para ração animal (HUNGRIA; NOGUEIRA; ARAUJO, 2013; ZHI et al., 2020). Por esses motivos, os programas de melhoramento genético buscam, entre os seus principais objetivos, aumentar a produtividade de grãos.

A semente é composta de proteínas, óleo, carboidratos, isoflavonas e minerais que determinam a sua composição nutricional (BELLALLOUI et al., 2011). O teor de proteína pode variar entre 30% e 53%, sendo que as cultivares brasileiras têm um teor médio de 40%. Já o teor de óleo encontra-se entre 13% e 28%, com média de 20%. A cultura se tornou importante fonte proteica devido a sua qualidade e baixo custo de produção, principalmente para os países em desenvolvimento (SILVA et al., 2022). De acordo com o USDA (2023a), o farelo de soja é o principal farelo proteico ofertado no mundo. Ao todo, na safra 2021/22, foram produzidas 246,57 milhões de toneladas, representando aproximadamente 70% de todo farelo proteico produzido. No Brasil, entre as safras de 2000/2001 e 2017/2018, a produção de soja teve um aumento constante ao longo dos anos, em cerca de 6,2% ao ano, partindo de uma produção de 38,4 milhões para 119,3 milhões de toneladas ao ano. Esse crescimento está ligado à dois elementos importantes: o crescimento da área de cultivo (4,6% a.a.) e o incremento na produtividade (1,5% a.a.) (SEIXAS et al., 2020).

A soja é uma planta herbácea da família Fabaceae. O possível centro de domesticação da espécie é a região nordeste da China, que posteriormente se expandiu para outras regiões do mundo. A partir dos anos 1940, a cultura da soja encontrou condições favoráveis de cultivo nos trópicos brasileiros, graças a um clima adequado e topografia favorável, aliados à presença de agricultores visionários. Esse período marcou um significativo crescimento na produção de soja no país (GAZZONI; DALL'AGNOL, 2018). A adaptação da soja é influenciada pelo grau de maturidade relativa de cada genótipo, sendo esse definido em função da latitude, das macrorregiões sojícolas e das regiões edafoclimáticas. Dessa forma, a medida em que o cultivo se desloca em direção ao sul ou ao norte, a época de semeadura e o ciclo da cultura serão diferentes (EMBRAPA, 2019).

As raízes podem ser divididas em: (a) primária, conhecida como principal, é pivotante e forma-se a partir das sementes como radícula; (b) laterais secundárias, emergindo-se a partir da raiz primária; e (c) laterais terciárias que se originam das secundárias (LERSTEN; CARLSON, 2004). As raízes são ricas em nódulos de bactérias do gênero *Bradyrhizobium*, que capturam o nitrogênio atmosférico e o convertem a nutriente para a planta (NOGUEIRA et al., 2009). O sistema radicular possui plasticidade, devido às características morfológicas, arquitetônicas e fisiológicas da espécie, onde em condições de estresses, podem levar a um aumento na profundidade das raízes, para auxiliar a tolerância ao déficit hídrico, ou a uma arquitetura menos profunda que facilita a absorção de nutrientes na camada superficial, é capaz de aumentar a quantidade de exsudados radiculares, como prótons e ácidos orgânicos. Por conta dessas características adaptativas, as plantas possuem capacidade de tolerar o déficit hídrico e/ou deficiência de nutrientes em diversos ambientes (ZHAO et al., 2004).

O caule é herbáceo, ereto e ramificado, desenvolvendo-se a partir do eixo embrionário (MÜLLER, 1981). A cultura possui três tipos de crescimento entre os genótipos cultivados: determinado, semideterminado e indeterminado, em que cada tipo diferencia entre si em relação ao florescimento e a estatura final atingida pós florescimento. Em relação à inclinação dos ramos laterais, os genótipos podem ser classificados como eretas, semieretas e prostradas (SILVA et al., 2022). Os programas de melhoramento genético têm buscado aumentar o diâmetro do caule, para aumentar a tolerância ao acamamento de novos genótipos (DIAS et al., 2021).

As folhas podem ser de quatro tipos: i) cotiledonares; ii) simples ou unifolioladas; iii) trifolioladas ou compostas; iv) prófilos ou brácteas. As unifolioladas localizam-se no primeiro nó acima das folhas cotiledonares. A soja possui um par de cotilédones que são utilizados como reserva nutricional nos estágios iniciais de desenvolvimento. As trifolioladas arranjam-se alternativamente sobre a haste, com diferentes variações das formas dos folíolos de acordo com o genótipo. Os prófilos estão presentes em pares na base de cada ramo lateral. As flores são completas, possuindo todas as estruturas femininas e masculinas na mesma planta, protegidas por um conjunto de pétalas (VERNETTI; GASTAL, 1979; NOGUEIRA et al., 2009; MATSUO; FERREIRA; SEDIYAMA, 2015). A coloração do hipocótiolo tem grande importância no melhoramento genético, pois a pigmentação antocianínica é utilizada na hibridação artificial e como descritor de plantas (SILVA et al., 2022).

2.2 Análise Espacial

Os ensaios de melhoramento de plantas, como todo experimento biológico, estão sujeitos à variabilidade espacial causada por fatores imprevisíveis (CASLER, 2018). A dependência espacial pode estar relacionada com os tratos culturais desuniformes como variações criadas pelo modo de plantio e colheita (STEFANOVA; SMITH; CULLIS, 2009), e fatores abióticos inerentes às áreas experimentais de cultivo (GRONDONA et al., 1996). Uma vez que há heterogeneidade do solo, a estimativa do erro experimental está sendo inflacionada. Dessa forma, a alta variabilidade no campo causada pelos fatores não-controláveis reduz a capacidade do melhorista em discriminar os genótipos (VAN ES; VAN ES, 1993).

De maneira geral, os delineamentos experimentais que visam o controle local, atenuam de maneira *a priori* a variação residual. No entanto, eles podem ser pouco eficientes em controlar efeitos gerados após o delineamento e instalação do experimento no campo (GRONDONA et al., 1996). Devido ao grande número de genótipos com que se trabalha nas extensas áreas experimentais, muitas das vezes não é possível garantir a homogeneidade nos blocos (ANDRADE et al., 2020).

Brownie, Bowman e Burton (1993) sugerem que esse método estatístico pode aumentar os ganhos na precisão experimental, quando comparados com as análises convencionais. Ressaltam que essa metodologia pode ser utilizada em vários contextos, nos mais variados ambientes e em diferentes tipos de delineamentos. Porém para aplicação dessa metodologia é preciso conhecer as posições das parcelas para estimar e corrigir a variação espacial nas observações. Segundo Grondona et al. (1996), para correção da variabilidade espacial, a maioria das metodologias adota a existência de dois efeitos de tendências centrais. O primeiro sendo um componente com tendência global, com uma variação em larga-escala, e outro com tendência local, afetando em pequena escala.

2.2.1 *Spatial Analysis of Field Trials with Splines - SpATS*

Na análise espacial, considera-se que há a presença de efeitos de dependência espacial entre as parcelas. Os métodos que se fundamentam em estruturas espaciais de covariância, as parcelas localizadas mais próximas são consideradas como mais semelhantes, em função de um raio de influência. Quanto menor essa distância entre parcelas, maior a influência, conseqüentemente, parcelas mais distantes umas das outras são menos similares e assim, tendo uma correlação espacial entre os resíduos menor, chamado de autocorrelação espacial

(BROWNIE; BOWMAN; BURTON, 1993; VAN ES; VAN ES, 1993; GRONDONA et al., 1996; RESENDE; STURION, 2003; ANDRADE et al., 2020).

Existe outra abordagem com potencial de ajuste semelhante, como os métodos de análise de superfície de tendência, baseados em superfícies de resposta, como polinomiais ou splines, que são utilizados como técnicas de suavização que podem ser aplicados para controle da variação global, através da inclusão desse efeito no sentido das linhas e/ou colunas (RESENDE; STURION, 2003). A vantagem de se utilizar essa metodologia é a facilidade na visualização dos efeitos de tendências espaciais, o bom ajuste do modelo e por ajustar perfeitamente às análises de experimentos de campo, conhecida pela sua grande flexibilidade e versatilidade (VERBYLA et al., 1999; DURBÁN; CURRIE, 2003; PIEPHO; BOER; WILLIAMS, 2022).

Em diversos trabalhos, as principais utilizações das funções polinomiais e de *splines* se limitam em suavizar o efeito de tendência unidimensional, ao longo das linhas e colunas no experimento. Dessa forma, sofrem com a limitação de nem sempre conseguirem modelar a tendência local. Portanto, utilizar suavizadores *splines* e polinômios são mais apropriados para modelar a tendência global. Caso o interesse seja em controlar tendências locais, há de se considerar a inclusão de componentes espacialmente correlacionados (GILMOUR; CULLIS; VERBYLA, 1997; RODRÍGUEZ-ÁLVAREZ et al., 2018).

As funções *splines* têm como princípio a interpolação de um polinômio que passe por todos os pontos do plano, e entre eles, pode-se interpolar infinitos polinômios para se obter um melhor formato que se deseja traçar dos dados. Os pontos de ruptura são chamados de nós e servem como pontes de ligação entre dois polinômios. Os tipos de *splines* diferem entre si em relação ao grau do polinômio, ao método utilizado para determinar a quantidade de nós, à posição dos nós e ao tipo de método de suavização das curvas. Existem diversos tipos de *splines*, dentre eles: *Spline* Linear, *Spline* Quadrático, *Spline* Cúbico, *B-Spline* e *P-Spline* (SOARES, 2019).

Como alternativa aos métodos unidimensionais de suavização, Rodríguez-Álvarez et al. (2018) propuseram um novo modelo 2D *P-splines* via modelos mistos para a suavização das tendências de campo em linhas e colunas. A metodologia SpATS se baseia no uso de *P-splines* que abordam a suavização como uma regressão penalizada para obter a curva ajustada, onde *B-splines* são penalizados nos coeficientes de regressão e há uma regressão sobre as funções de base *B-splines*, utilizando nós igualmente espaçados (MENDOZA, 2017; RODRÍGUEZ-ÁLVAREZ et al., 2018).

No modelo SpATS, as posições das parcelas são coletadas em vetores de coordenadas de linhas (\mathbf{r}) e colunas (\mathbf{c}), onde as tendências de campo são suavizadas por uma função bivariada das coordenadas espaciais $f(\mathbf{r}, \mathbf{c})$, representando a função 2D *P-splines*. A partir da função bivariada, derivam-se dois termos para estimar a superfície espacial. O primeiro refere-se a parte de natureza fixa e não penalizado, representando às tendências lineares ao longo das linhas, colunas e da interação entre elas; e o segundo, a parte de natureza aleatória, que é penalizada correspondendo às tendências de suavização ao longo das linhas, colunas e da interação entre elas, com cinco componentes espaciais aleatórios (VELAZCO et al., 2017).

A representação dos *P-splines* como modelos mistos foi alcançado utilizando *B-splines* cúbicos e com a penalidade de segunda ordem, sendo as configurações mais comumente usadas na estrutura *P-splines* (MENDOZA, 2017; VELAZCO et al., 2017), fornecendo um modelo para ser utilizado em ensaios de campo. Esse modelo oferece a flexibilidade de incorporar outros componentes fixos e aleatórios, como efeitos de correção de linhas e colunas (RODRÍGUEZ-ÁLVAREZ et al., 2018).

SpATS tem se mostrado uma importante técnica para o controle da dependência espacial nos ensaios de melhoramento de plantas. Velazco et al. (2017) demonstraram que a SpATS é uma alternativa flexível e confiável para controle da variabilidade espacial em ensaios de melhoramento de sorgo. Andrade et al. (2020) observaram que a metodologia SpATS é equivalente com modelos baseados em processos autorregressivos para o controle da dependência espacial, resultando em maiores ganhos com a seleção devido ao aumento da herdabilidade e diminuição do resíduo em ensaios de melhoramento de batata.

2.3 Interação Genótipos por Ambientes

Os genótipos se diferem entre si em relação a eficiência, adequação e capacidade de resposta frente às condições ambientais, determinadas pelo seu conjunto alélico. Os ambientes por sua vez, diferem-se em relação à quantidade e qualidade dos fatores abióticos fornecidos às plantas (MALOSETTI; RIBAUT; EEUWIJK, 2013). A GxA é um fenômeno biológico importante no melhoramento de plantas. Frente às mudanças ambientais, pode ocorrer um comportamento diferencial dos genótipos em relação a diversas características. Devido a esse fenômeno, pode-se levar a uma escolha incorreta do melhor genótipo frente às condições locais de cultivo.

O efeito da interação GxA quando grande, o desempenho de um conjunto de genótipos em determinado ambiente pode não ser informativo para compreender o comportamento desse

genótipo em outros ambientes (MALOSETTI; RIBAUT; EEUWIJK, 2013). Por ser um componente da variância fenotípica total, esse fator influencia diretamente no ganho de seleção, afetando negativamente a herdabilidade da característica (KANG, 1997). A herdabilidade representa a proporção herdável da variabilidade total, tendo um papel fundamental no avanço genético. Além disso, a interação GxA tem enormes implicações na alocação de recursos em programas de melhoramento. Quando intensa, exige a instalação de muitos ensaios experimentais, de modo a ampliar a amostragem de ambientes na região-alvo do programa de melhoramento. Com isto, aumenta-se exigência de mão-de-obra e custos de implantação e manutenção de experimentos (KANG, 1997).

Os ensaios *MET* fornecem informações referentes ao desempenho dos genótipos na região de destinação dos genótipos (SMITH et al., 2015; VOLPATO et al., 2019). A avaliação da interação GxA é fundamental para se conhecer a previsibilidade do desempenho do genótipo em diferentes locais/anos/épocas (YAU, 1995; KANG, 1997; EEUWIJK; BUSTOS-KORTS; MALOSETTI, 2016).

Tradicionalmente, as análises em *MET* eram feitas através da análise conjunta de variância (ANOVA) e técnicas de regressão linear (RESENDE; THOMPSON, 2004). No entanto, pode ser ineficiente utilizar essas análises tradicionais para avaliar o desempenho e estabilidade dos genótipos quando há desbalanceamento genético e estatístico dos dados (KRAUSE et al., 2022).

2.3.1 Modelos Lineares Mistos

A metodologia de modelos lineares mistos (BLUP), foi apresentada para ser utilizada no melhoramento animal, visando obter estimadores imparciais dos parâmetros genéticos e ambientais em bovinos (HENDERSON, 1949). Segundo Resende (2002), devido às dificuldades inerentes ao procedimento estatístico, essa metodologia não foi amplamente disseminada na época. Porém, com os avanços das técnicas computacionais das últimas décadas esses problemas foram superados e assim seu uso intensificado a partir de 1980.

Os modelos matemáticos descrevem o comportamento de uma variável aleatória dependente, em função de outras variáveis preditoras e, só se define como modelo linear quando os parâmetros presentes nos modelos são simples coeficientes das variáveis classificatórias ou explanatórias. Os modelos mistos são modelos estatísticos lineares com efeitos fixos, excetuando-se o intercepto, e efeitos aleatórios, sem contar com o erro experimental que é

sempre aleatório (PEREIRA; GRAVINA; THIÉBAUT, 2012; RESENDE; SILVA; AZEVEDO, 2014).

A natureza desses efeitos dos parâmetros do modelo depende da inferência estatística a ser feita sobre o conjunto dos dados. Considerando como efeito fixo, o interesse está no estudo da diferença média que existe entre os níveis das variáveis fixas, aplicadas apenas àqueles dados, em virtude de a amostra ser a própria população. Para efeitos aleatórios, visa-se entender a variabilidade que existe entre os níveis das variáveis aleatórias, sendo amostras representativas da população, e assim, podendo extrapolar as estimativas dos parâmetros obtidas para a população (EISENHART, 1947).

Smith, Cullis e Thompson (2005), defendem que ao se considerar o efeito de genótipos como aleatório, permite que os agricultores possam, acima de tudo, tomarem as suas próprias escolhas a respeito dos melhores genótipos para seus ambientes. Uma vez que, ao considerar o efeito de genótipos como aleatório pode tornar as estimativas de ganho de seleção mais realistas, em comparação com a abordagem que considera o efeito de genótipo como fixo, minimizando assim os erros de seleção dos genótipos. Essa abordagem também é útil para validar análises de dados entre anos combinados no estágio de seleção.

A predição confiável dos valores genotípicos dos diversos genótipos candidatos é um dos principais desafios do melhoramento genético, que depende sobretudo da adoção de acurados procedimentos de seleção (RESENDE, 2002, 2004). Henderson (1975) apresentou métodos simplificados para estimar a “melhor estimativa linear não viciada”, denominados de BLUE. Também descreveu a “melhor predição linear não viciada”, o BLUP. Esses resultados estão relacionados aos efeitos fixos e aleatórios das funções lineares estimáveis, respectivamente. A forma matricial da equação de modelo linear misto é:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e},$$

onde $\mathbf{y}^{(n \times 1)}$ é o vetor de observação, $\mathbf{X}^{(n \times f)}$ é a matriz de incidência conhecida para os efeitos fixos, $\boldsymbol{\beta}$ é o vetor de efeitos fixos desconhecido, $\mathbf{Z}^{(n \times i)}$ é a matriz de incidência conhecida para os efeitos aleatórios, \mathbf{u} é o vetor de efeitos aleatórios e \mathbf{e} é o vetor dos resíduos, não observáveis com média nula e

$$\text{Var}\begin{pmatrix} \mathbf{u} \\ \mathbf{e} \end{pmatrix} = \begin{pmatrix} \mathbf{G} & \mathbf{0} \\ \mathbf{0} & \mathbf{R} \end{pmatrix}$$

em que \mathbf{G} , sendo a matriz de variância-covariância de \mathbf{u} e \mathbf{R} , a matriz de variância-covariância residual, ambas matrizes não singulares, sendo $\mathbf{G} = \mathbf{I}\sigma_u^2$ e o $\mathbf{R} = \mathbf{I}\sigma_e^2$.

A obtenção dos BLUE's e BLUP's passa pela solução da equação de modelos mistos (HENDERSON, 1975):

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

em que a obtenção de $\hat{\beta}$ e \hat{u} implica na inversão da matriz dos coeficientes pelo método dos quadrados mínimos generalizados. Porém, algumas suposições devem ser tomadas, como a suposição de aditividade dos efeitos genéticos e ambientais, distribuição normal dos resíduos, independência entre as observações e homogeneidade da variância dos erros (RESENDE, 2002; CARVALHO; FRITSCH NETO; GERALDI, 2008).

O método de Máxima Verossimilhança Residual (REML) foi proposto por Patterson e Thompson (1971) como uma alternativa para contornar as pressuposições dos modelos mistos. Desde então, tem sido reconhecido como um procedimento altamente recomendado para estimar os componentes de variância em situações de desbalanceamento de dados (RESENDE; SILVA; AZEVEDO, 2014; BERNARDO, 2020). Significa que, os componentes de variância nas matrizes G e R na equação de modelos mistos precisam ser substituídos pelas estimativas obtidas por REML (PIEPHO et al., 2008), garantindo que a predição seja realizada com componentes de variância obtidos de forma precisa (RESENDE, 2004).

Sob certas condições de balanceamento de dados, os resultados da metodologia e análise de variância (ANOVA) são equivalentes. No entanto, em cenários de desbalanceamento ou em desregramento de certas suposições da ANOVA, a metodologia de modelos lineares mistos tem propriedades que a tornam mais apropriada (BERNARDO, 2020; ISIK; HOLLAND; MALTECCA, 2017). Desse modo, o método mais eficiente para predição dos valores genotípicos é o BLUP, enquanto que o REML é o procedimento mais adequado para estimação dos componentes de variância (RESENDE, 2004).

No melhoramento de plantas, a metodologia BLUP é frequentemente empregada para identificar as melhores progênies de soja em situações de interação GxA (BIANCHI et al., 2020), na predição do valor genético para produtividade de grãos, visando a recomendação de linhagens mais adaptadas e estáveis (ALBUQUERQUE et al., 2022), em modelagem de experimentos que corrigem efeitos de tendências espaciais (BERNARDELI et al., 2020), e em procedimentos que consideram a estrutura de correlação entre caracteres avaliados, visando uma seleção mais precisa de progênies segregantes de soja (VOLPATO et al., 2019).

2.3.2 Fator Analítico Multiplicativo Misto

As metodologias baseadas em ANOVA falham em dois pontos principais: i) a obtenção do efeito da interação GxA só fornece uma visão geral da magnitude desse efeito, não sendo empregado diretamente para avaliar os padrões de comportamento ao longo dos ambientes; ii) As pressuposições da estrutura de variância uniforme, podem não ser satisfatórias, devido a presença de variância e covariância heterogêneas (SMITH; CULLIS; THOMPSON, 2005).

As melhores metodologias para abordagem de dados em *MET* são as que combinam os modelos mistos lineares com uma estrutura multiplicativa para realização das análises, onde se utiliza de toda informação de cada ambiente. *Additive Main Effects and Multiplicative Interaction Analysis* (AMMI) (GAUCH, 1988, 1992) e FAMM (PIEPHO, 1997, 1998; SMITH; CULLIS; THOMPSON, 2001) fornecem uma abordagem para acomodar a heterogeneidade da variância dos efeitos GxA em contexto *MET*, por meio do relaxamento da suposição relativa de independência desses efeitos (RESENDE; THOMPSON, 2004).

A pressuposição de variâncias e covariâncias heterogêneas dos dados em contexto *MET*, garantem que a variação do erro e da dependência espacial não seja comum entre os ambientes. Da mesma forma, cada ambiente dispõe da sua própria variância genética, refletindo na real manifestação genotípica ao longo dos ambientes. Bem como, a covariância genética entre os pares de ambientes, assegurando a particularidade entre os pares (SMITH; CULLIS; THOMPSON, 2001).

A combinação de análise de fatores com modelos lineares mistos, proporciona uma grande eficiência na análise de dados *MET* (NUVUNGA, 2014). Utilizar dos modelos lineares mistos em contexto *MET*, no melhoramento de plantas, inclui vantagens como, facilidade em lidar com o desbalanceamento de dados, capacidade de incluir a modelagem espacial para controle da variação do erro, e possibilidade de assumir os efeitos como fixo ou aleatório (SMITH; CULLIS; THOMPSON, 2005).

Smith, Cullis e Thompson (2001), demonstraram que a metodologia FAMM pode ser utilizada para acomodar todas as fontes de variações heterogêneas em contexto *MET*. Tendo potencial para ser utilizada em análises de interação GxA, por fornecer informações precisas do comportamento dos genótipos e da interação dos genótipos com os ambientes. Inúmeros trabalhos surgiram a partir de então (KELLY et al., 2007; CULLIS et al., 2010, 2014; SMITH; CULLIS, 2018; DIAS et al., 2018; KRAUSE et al., 2020; CHAVES et al., 2023) sendo a

metodologia preferível para analisar dados *MET* nos programas Australianos de melhoramento de plantas (KELLY et al., 2007).

FAMM é uma técnica que reuni um conjunto de métodos estatísticos, relacionando o comportamento de variáveis (como ambientes) utilizando um conjunto reduzido de variáveis latentes, conhecidos como fatores, permitindo que cada ambiente seja explicado por fatores comuns (cargas fatoriais) e por uma parte não explicada por esses fatores, chamada de variância específica. Esse agrupamento é realizado por meio da correlação entre os ambientes, onde os mais correlacionados estão fortemente ligados entre si em um fator, e mais fracamente em outro (MURAKAMI, 2001; BAKARE et al., 2022).

O ponto principal da metodologia FAMM é utilizar da covariância genética entre ambientes para captar o efeito da interação GxA, utilizando de um número reduzido de fatores (SMITH; CULLIS; THOMPSON, 2001). Esse modelo fornece previsões confiáveis do comportamento dos genótipos em cada ambiente (SMITH; CULLIS, 2018). Em razão da complexidade da abordagem, à falta de um software livre ou até mesmo devido à alta demanda computacional que a análise exige, esta abordagem estava sendo pouco utilizada (FIGUEIREDO et al., 2015), aliado a falta de ferramentas que sintetizam os resultados obtidos do FAMM que sejam estatisticamente ou biologicamente aceitáveis sem perder informações da variância da interação GxA (SMITH; CULLIS, 2018).

Cullis et al. (2010) com o intuito de desenvolver ferramentas auxiliares adequadas para o FAMM em contexto MET, desenvolveram os mapas de calor que exploram a interação GxA. Uma vez que a heterogeneidade nas correlações genéticas entre os pares de ambientes, representam a interação GxA do tipo complexa, que afeta o ranqueamento dos melhores genótipos. Por isso, os autores sugeriram que os mapas de calor sejam utilizados para compreender os padrões de interação GxA.

Cullis et al. (2014), aproveitando das propriedades da FAMM que a ligam com as análises de regressão múltipla, construiu gráficos para facilitar o entendimento dos resultados e garantir rápida tomada de decisão nos programas de melhoramento. Cullis et al. (2014), para melhor visualização da resposta genotípica, utilizaram gráficos de regressão para cada covariável ambiental latente (cargas fatoriais rotacionadas). Ou seja, para determinado indivíduo, há uma série de gráficos para k fatores, onde no eixo X estão as cargas do fator k e no eixo Y os valores genotípicos. As inclinações das retas representam o desempenho e a estabilidade do genótipo para cada covariável ambiental latente.

Ainda que a análise de regressão múltipla seja uma técnica muito intuitiva no estudo da interação GxA, apresenta limitação quando se trabalha com grande número de genótipos. As

curvas de cada genótipos nos k fatores ficam sobrepostas, dificultando a visualização do desempenho *per se*. Segundo Smith e Cullis (2018) essa metodologia fornece apenas uma análise informal do desempenho geral e da estabilidade dos genótipos.

A metodologia FAST sintetiza os resultados do FAMM em estimativas de desempenho geral e estabilidade, facilitando a seleção dos genótipos. Smith e Cullis (2018) consideraram que, uma vez que a maioria das cargas rotacionadas estimadas para o primeiro fator são positivas, os genótipos podem ser recomendados explorando as cargas e escores fatoriais para obter as estimativas de Desempenho Geral (OP_i), que é uma medida de desempenho geral; e Raiz do Quadrado Médio dos Desvio ($RMSD_i$), medida de estabilidade geral. A Capacidade de Resposta (RE_i) também pode ser utilizado como uma medida de estabilidade específica para cada fator k, em função do sinal das cargas no fator k. O melhor genótipo será aquele que apresenta alto desempenho e estabilidade geral. Para que isso ocorra, deve-se obter alto valor de OP_i , e $RMSD_i$ e RE_i próximos de zero, para cada fator k.

FAST tem sido utilizada para facilitar as análises em contextos MET. Sjoberg et al. (2021) demonstraram que FAST pode ser aplicado à outras características, além da produtividade, para ranquear genótipos com melhores qualidades nutricionais em contexto MET em trigo. No Brasil, Chaves et al., (2023) selecionaram híbridos milho com elevado desempenho e estabilidade em contexto MET utilizando FAST.

3. MATERIAL E MÉTODOS

3.1 Caracterização dos dados

Os dados analisados são referentes a ensaios de competição de genótipos de soja do ano agrícola 2021/22 em 10 municípios, totalizando 16 ambientes. Neste trabalho os ambientes são a combinação de locais e épocas de plantio distintas, sendo o E08 irrigado via pivô central (Tabela 1). Os ambientes abrangeram duas macrorregiões sojícolas no Mato Grosso do Sul, em três microrregiões edafoclimáticas distintas para recomendação de genótipos de soja (EMBRAPA, 2019). O estado na classificação de Köppen possui quatro tipos climáticos: Af (clima equatorial), Am (clima de monção); Aw (clima de savana); e Cfa (clima subtropical úmido) (ALVARES et al., 2013). A discriminação dos municípios e do número de ambientes em cada foram representados no mapa do estado (Figura 1). Por conta da limitação de área, recursos e genótipos mais recomendados para determinados municípios, nem todos os genótipos foram avaliados em todos os ambientes, ocorrendo um desbalanceamento dos dados, com um valor aproximado de 56,55% (Figura 2).

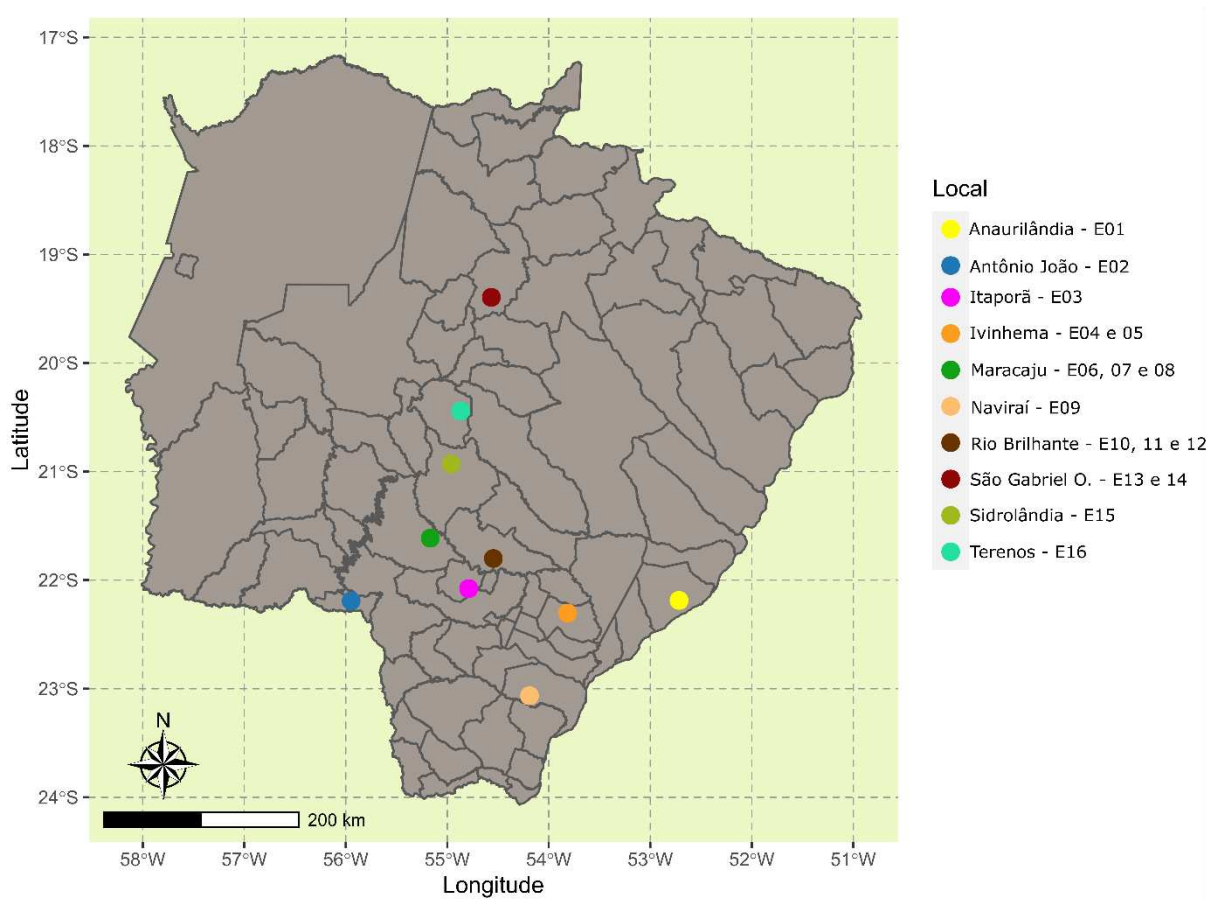
Tabela 1. Códigos dos ambientes, épocas de plantio, número de genótipos avaliadas, e número de linhas, colunas e produtividade média de cada ambiente.

Ambientes ^a	Município	Épocas ^b	Nº de genótipos avaliados	Linhas	Colunas	Média PROD (kg/ha)
E01	Anaurilândia	-	58	58	3	1638,042
E02	Antônio João	-	52	52	3	2180,458
E03	Itaporã	-	33	33	3	1707,784
E04	Ivinhema	1 ^a	58	58	3	1510,182
E05	Ivinhema	2 ^a	60	20	3	2015,447
E06	Maracaju	1 ^a	58	58	3	3664,107
E07	Maracaju	2 ^a	62	62	3	3313,456
E08	Maracaju Irrigado	-	30	15	6	4658,906
E09	Naviraí	-	41	41	3	1374,386
E10	Rio Brilhante	1 ^a	40	40	3	3436,198
E11	Rio Brilhante	2 ^a	59	59	3	3830,484
E12	Rio Brilhante	3 ^a	42	42	3	1927,656
E13	São Gabriel do Oeste	1 ^a	46	46	3	4400,474
E14	São Gabriel do Oeste	2 ^a	42	42	3	4262,903
E15	Sidrolândia	-	60	60	3	1628,289
E16	Terenos	-	40	40	3	1223,645

^a Número do ambiente precedido pela letra E.

^b Ambientes com uma única época são denominados de -. Os demais, são divididos em uma ou mais épocas de plantio.

Figura 1. Discriminação dos municípios com ensaios de competição de genótipos de soja na safra agrícola 2021/22, com seus respectivos números de ambientes (E), no estado do Mato Grosso do Sul, Brasil.



Fonte: O autor (2023)

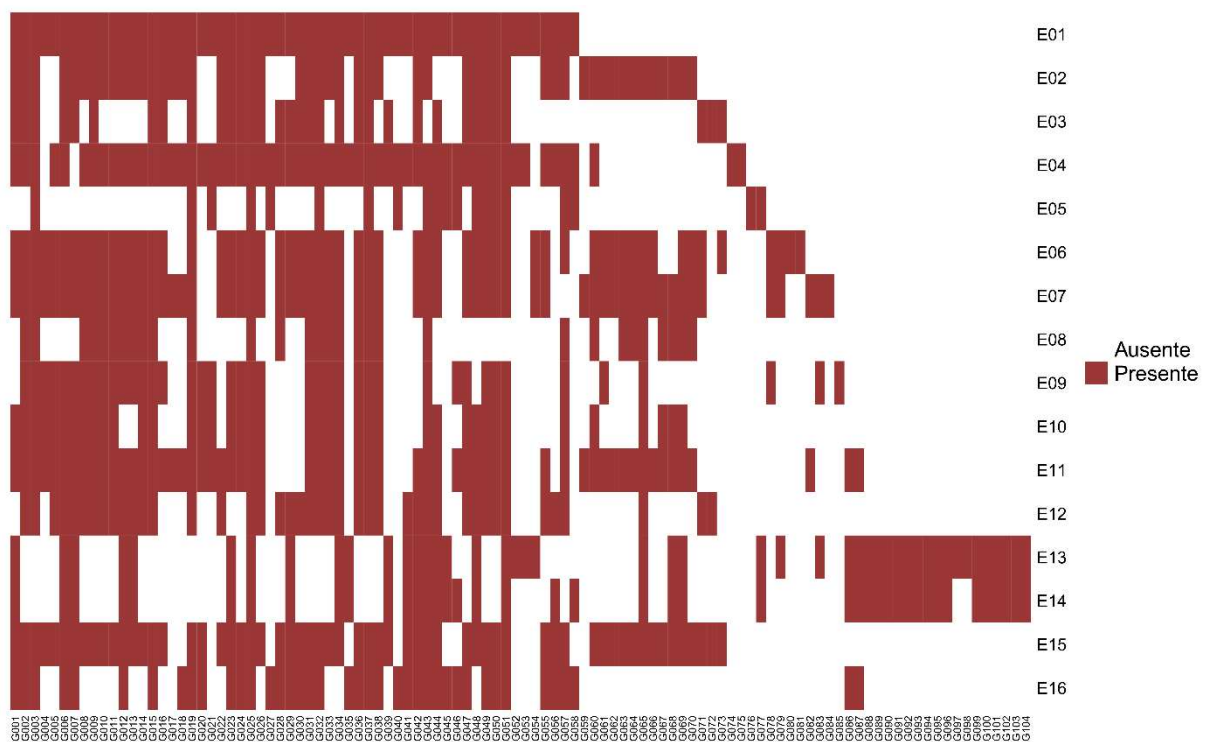
Nesses ensaios as parcelas consistiam de cinco linhas de 12 m, com espaçamento entre linhas de 0,5 m, totalizando 30 m². A área útil da parcela foi formada a partir da colheita das três linhas centrais de 11 m cada, desprezando-se 0,5 m de cada extremidade, e representando 18 m². O delineamento experimental foi blocos completos casualizados (DBC), com três repetições.

As sementes empregadas nos plantios foram tratadas com o fungicida e inseticida (Standak® Top), na dose de 200 mL do produto comercial para cada 50 kg de sementes. Foi realizada a coinoculação com *Bradyrhizobium* (4 doses) mais *Azospirillum* (1 dose), por meio de via líquida, aplicada em sulco de plantio. Utilizou-se uma adubação de manutenção de 330 kg ha⁻¹ com o adubo formulado (02-20-18). Também, foram feitas aplicações de defensivos para

prevenção e controle de pragas e doenças, de acordo com o estabelecido na região de cultivo, respeitando a rotação de moléculas, visando diminuir a seleção de indivíduos resistentes.

Os genótipos tiveram sua produtividade de grãos (PROD) em kg por hectare aferida, com o objetivo de recomendar as melhores para o estado. O sistema de colheita foi mecanizado e a data dela variou de acordo com a maturação dos genótipos, embora todas tenham sido colhidas no estágio de maturação plena (R8) (RITCHIE et al., 1982).

Figura 2. Discriminação da presença/ausência dos genótipos não avaliados (no eixo X de G001 a G104) nos ambientes (no eixo Y de E01 a E16).



Fonte: O autor (2023)

3.2 Modelos Utilizados para Análise de Dados

Os ensaios de competição são compostos por 104 genótipos de soja (Apêndice A). Aprioristicamente foi estabelecido que o conjunto de dados *MET* compreende de algumas notações matemáticas para melhor abordagem dos métodos nesta dissertação, onde g representa o número total de genótipos ($i = 1, 2, \dots, 104$), r é o número de linhas, c é o número de colunas e b é o número de blocos ($f = 1, 2, \dots, 3$) e a é o número de ambientes ($j = 1, 2, \dots, 16$). Os dados de PROD foram analisados utilizando o modelo base (sem análise espacial) e SpATS, em cada ambiente de maneira individual. No modelo conjunto, utilizou-se a estrutura do modelo Fator

Analítico. REML foi utilizado para estimar os componentes de variância e o BLUP para obtenção dos BLUE's e BLUP's em todas as análises (PATTERSON, THOMPSON, 1971; HENDERSON, 1975).

3.3.1 Modelo Base sem Análise Espacial com Erros Independentes – Modelo Linear Misto

Primeiramente, para todos os ambientes, foram preditos os valores genotípicos com base em um modelo sem análise espacial. Considerou-se o modelo linear misto assumindo a independência dos erros ($R = \sigma_e^2 \mathbf{I}_e$), considerando somente os efeitos inerentes ao delineamento experimental:

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{X}_b\boldsymbol{\tau}_b + \mathbf{Z}_g\mathbf{u}_g + \mathbf{e}$$

onde $\mathbf{y}^{(n \times 1)}$ refere-se ao vetor de observação dos efeitos fenotípicos para cada ambiente; μ é o intercepto de cada ambiente multiplicado pelo vetor de uns $\mathbf{1}^{(n \times 1)}$; $\boldsymbol{\tau}_b^{(b \times 1)}$ é o vetor de efeitos fixos de blocos com uma matriz de incidência $\mathbf{X}_b^{(n \times b)}$; $\mathbf{u}_g^{(g \times 1)}$ é o vetor de efeitos aleatórios de genótipos, com $\mathbf{u}_g \sim N(0, \sigma_g^2 \mathbf{I}_g)$ e matriz de incidência associada $\mathbf{Z}_g^{(n \times g)}$; $\mathbf{e}^{(n \times 1)}$ é o vetor de resíduos aleatórios, com $\mathbf{e} \sim N(0, \sigma_e^2 \mathbf{I}_e)$; σ_g^2 e σ_e^2 são os componentes de variância dos efeitos \mathbf{u}_g e \mathbf{e} , respectivamente.

3.3.2 Análise Espacial dos Ambientes Utilizando Modelagem Via Spline – Modelo SpATS

A fim de capturar a heterogeneidade ambiental presente nos ambientes, foi empregado a metodologia SpATS, que se baseia nas coordenadas espaciais das parcelas expressas em vetores de coordenadas de linhas (\mathbf{r}) e colunas (\mathbf{c}). Essa técnica utiliza a metodologia 2D P-splines para suavizar as tendências espaciais ao longo da área experimental, por meio de uma função bivariada das coordenadas espaciais $f(\mathbf{r}, \mathbf{c})$, representando a função de suavização espacial nos sentidos das linhas e colunas, controlada pelo parâmetro de suavização $\lambda_{S_k} = \hat{\sigma}_e^2 / \hat{\sigma}_{S_k}^2$, para modelar as tendências espaciais e obter estimativas confiáveis dos efeitos fixos e aleatórios. Os parâmetros de suavização são estimados via REML, a partir de um processo anisotrópico que procura modelar a variabilidade espacial em todas as direções possíveis, a fim de minimizar o erro de predição (RODRÍGUEZ-ÁLVAREZ et al., 2018), sendo:

$$\mathbf{y} = \mathbf{1}\mu + f(\mathbf{r}, \mathbf{c}) + \mathbf{Z}_d\mathbf{u}_d + \mathbf{Z}_g\mathbf{u}_g + \mathbf{e}$$

em que $f(\mathbf{r}, \mathbf{c})$ é a função de suavização bivariada das coordenadas espaciais, sendo igual à soma dos termos $\mathbf{X}_s\boldsymbol{\tau}_s$ e $\mathbf{Z}_s\mathbf{u}_s$, onde $\mathbf{X}_s\boldsymbol{\tau}_s$ é a parte de natureza fixa e não penalizado, representando às tendências lineares ao longo das linhas ($\beta_{s1}\mathbf{r}$), colunas ($\beta_{s2}\mathbf{c}$) e da interação entre elas ($\beta_{s3}\mathbf{r}\mathbf{c}$); e o $\mathbf{Z}_s\mathbf{u}_s$, a parte de natureza aleatória, que é penalizada correspondendo às tendências lineares de suavização ao longo das linhas ($f_1(\mathbf{r})$), colunas ($f_2(\mathbf{c})$), às tendências suavizadas devido a inclinação da tendência linear ao longo das linhas ($h_3(\mathbf{r})\mathbf{c}$), colunas ($rh_4(\mathbf{c})$) e a interação entre a tendência suavizada nos sentidos das linhas e colunas ($f_5(\mathbf{r}, \mathbf{c})$) (VELAZCO et al., 2017). O vetor \mathbf{u}_d é o vetor aleatórios de linhas e colunas, contabilizando variação de campo descontínua com uma matriz de incidência \mathbf{Z}_d .

A superfície suavizada foi obtida com nós igualmente espaçados, com B-splines cúbicos, sendo o número interno de nós igual ao número de linhas e de colunas, e com a penalidade de segunda ordem. O modelo SpATS é um modelo misto espacial que utiliza da metodologia P-spline ANOVA (PS-ANOVA) desenvolvida por Lee, Durbán e Eilers (2013), para representar o processo 2D P-spline com a estrutura de modelos mistos. Para melhor compreensão das matrizes \mathbf{X}_s , \mathbf{Z}_s e matriz de covariância \mathbf{S} , que compõem o modelo, são apresentadas por Lee, Durbán e Eilers (2013) e Rodríguez-Álvarez et al. (2018).

3.3.3 Seleção de Modelos para Efeito Residual

Para obter o melhor modelo ajustado, comparou-se o modelo SpATS com o modelo base sem considerar a análise espacial, em termos de parâmetros utilizados no melhoramento de plantas, que impactam de forma prática na seleção/recomendação dos melhores genótipos, sendo eles:

3.3.4 Coeficiente de Variação Genotípica (CV_g) e Ambiental (CV_e)

Através do modelo SpATS e do modelo base não espacial, foi possível quantificar a variabilidade genética entre os genótipos e a variação experimental em cada ambiente. Esses estimadores são úteis e comparáveis por serem adimensionais. O coeficiente de variação genética (CV_g) é a razão, em termos percentuais, entre a estimativa do desvio-padrão genotípico

e a média, demonstrando a extensão da variabilidade em relação à estimativa da média dos genótipos, sendo:

$$CV_g = \frac{\sqrt{\hat{\sigma}_g^2}}{\hat{\mu}} \times 100$$

O coeficiente de variação ambiental (CV_e) é a razão, em termos percentuais, entre a estimativa do desvio-padrão residual e a média, onde estima-se a influência do ambiente na produtividade dos genótipos, sendo:

$$CV_e = \frac{\sqrt{\hat{\sigma}_e^2}}{\hat{\mu}} \times 100$$

onde:

$\hat{\sigma}_g^2$ = estimativa do componente de variância genética entre os genótipos;

$\hat{\sigma}_e^2$ = estimativa do componente de variância residual experimental;

$\hat{\mu}$ = estimativa da média do ambiente de acordo com cada modelo.

3.3.5 Herdabilidade

A herdabilidade representa a proporção herdável da variabilidade total, medindo a relação entre o fenótipo e o valor genético. A comparação entre herdabilidades parte do princípio que, se houve maior expressão da variância devido aos efeitos genéticos quando utilizado determinado modelo, este modelo possui menor variância residual e, portanto, proporciona inferências mais confiáveis. Desta forma, modelos com maior herdabilidade foram escolhidos.

A herdabilidade estimada para o modelo base (H^2) foi calculada utilizando a equação proposta por Cullis, Smith e Coombes (2006):

$$H^2 = 1 - \frac{\bar{V}(\Delta)}{2\hat{\sigma}_g^2}$$

onde:

$\bar{V}(\Delta)$ = variância média do erro de predição;

Para o modelo SpATS, a herdabilidade (H^2) foi calculada utilizando a metodologia proposta de Rodríguez-Álvarez et al. (2018):

$$H^2 = \frac{ED_g}{n_g}$$

onde:

ED_g = dimensão efetiva associada aos efeitos genotípicos;

n_g = número total de genótipos.

Esses dois estimadores da herdabilidade são equivalentes, e assim podem ser utilizados para comparação direta entre a herdabilidade do modelo SpATS e do modelo base sem análise espacial (VELAZCO et al., 2017), de tal forma que:

$$H^2 = 1 - \frac{\bar{V}(\Delta)}{2\hat{\sigma}_g^2} = \frac{ED_g}{n_g}$$

Em ambientes com valores idênticos de H^2 , optou-se pelos BLUE's do modelo base não espacial, tendo em vista que a dependência residual está sendo captada pelo delineamento em blocos.

3.3.6 Correlação entre os BLUP's

Visando avaliar o impacto dos métodos espaciais na seleção dos melhores genótipos foi processada a correlação de Spearman entre o modelo SpATS e o modelo base sem análise espacial dos BLUP's nos diferentes ambientes, podendo servir como uma medida comparativa. Essa correlação é uma medida não paramétrica da dependência estatística baseada em classificação entre duas variáveis, sendo:

$$\rho_j = \frac{\Sigma(x' - BLUP_{x'}) (y_i' - BLUP_{y'})}{\sqrt{\Sigma(x' - BLUP_{x'})^2 (y_i' - BLUP_{y'})^2}}$$

onde:

x' = ranque do genótipo x no ambiente j para o modelo SpATS;

y_i' = ranque do genótipo x no ambiente j para o modelo base sem análise espacial;

$BLUP_{x'}$ = BLUP do genótipo x no ambiente j para o modelo SpATS;

$BLUP_{y'}$ = BLUP do genótipo x no ambiente j para o modelo base sem análise espacial.

3.4 Análise em Dois Estágios

Usualmente, em contexto de MET, os dados são analisados em dois estágios (two-stage). No primeiro estágio os ensaios são analisados com modelos que inclui o efeito de delineamento e o de tendência espacial. Em seguida, a partir das análises dos ensaios individuais, os BLUE's e os pesos ajustados, que normalmente são as variâncias dos BLUE's, são carregados para o segundo estágio. No segundo estágio a saída dos dados da primeira fase são combinados para formar a análise conjunta dos MET (SMITH; CULLIS; THOMPSON, 2005; MALOSETTI; RIBAUT; EEUWIJK, 2013). Neste estudo, os BLUE's e os pesos ajustados de cada modelo selecionado no primeiro estágio foram transportados para o modelo conjunto.

3.5 Modelos FAMM

Os modelos FAMM, propostos por Piepho (1997, 1998) e descritos Smith, Cullis e Thompson (2001), fornecem informações confiáveis e precisas do comportamento *per se* de cada genótipo, em cada ambiente alvo, levando em consideração a interação GxA. O modelo conjunto que utiliza a estrutura fator analítico para modelar os efeitos genéticos foi definido por:

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{X}_a\boldsymbol{\tau}_a + \mathbf{Z}_g\mathbf{u}_g + \mathbf{e}$$

onde: $\mathbf{y}^{(n \times 1)}$ é o vetor de efeitos fenotípicos nos j ambiente, sendo $n = \sum_{j=1}^a n_j$, em que n_j é o número de parcelas em a ambientes; $\boldsymbol{\tau}_a^{(a \times 1)}$ é o vetor de efeitos fixos de ambiente com uma matriz de incidência $\mathbf{X}_a^{(n \times a)}$; $\mathbf{u}_g^{(ga \times 1)}$ é o vetor de efeitos aleatórios de genótipos nos a ambientes, com $\mathbf{u}_g \sim N(0, \sigma_g^2 \mathbf{I}_g)$ e matriz de incidência associada $\mathbf{Z}_g^{(n \times ga)}$; $\mathbf{1}^{(n \times 1)}$ é um vetor que na sua diagonal contém número uns e μ é a média de cada ambiente.

O modelo ajustado no contexto FAMM para os efeitos de genótipos foi dado por:

$$g_{ij} = \sum_{k=1}^r \lambda_{jk} f_{ik} + \delta_{ij}$$

onde:

g_{ij} = efeito do genótipo i para o ambiente j ;

λ_{jk} = carga fatorial no fator k para o ambiente j ;

f_{ik} = escore fatorial do genótipo i para o fator k ;

δ_{ij} = erro residual, denominado como fator específico, devido à falta de ajuste para o modelo.

Em termos de denotação de vetores, de maneira compacta é:

$$\mathbf{g} = (\mathbf{\Lambda} \otimes \mathbf{I}_g) \mathbf{f} + \boldsymbol{\delta}$$

onde:

$\mathbf{\Lambda}^{(j \times k)}$ = matriz de cargas fatoriais (λ_{jk}), sendo $[\lambda_1 \dots \lambda_k]$;

$\mathbf{f}^{(i \times 1)}$ = vetor de escores fatoriais (f_{ik}), sendo $[f_1 \dots f_k]$;

$\boldsymbol{\delta}^{(ij \times 1)}$ = vetor de efeitos específicos (δ).

A distribuição simultânea de \mathbf{f} e $\boldsymbol{\delta}$ é expressa da seguinte forma:

$$\begin{pmatrix} \mathbf{f} \\ \boldsymbol{\delta} \end{pmatrix} \sim N \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \mathbf{G}_f \otimes \mathbf{I}_g & 0 \\ 0 & \boldsymbol{\Psi} \otimes \mathbf{I}_g \end{pmatrix} \right]$$

onde:

$\boldsymbol{\Psi}$ = matriz das variâncias específicas, sendo $\text{diag}[\Psi_1, \dots, \Psi_j]$;

Ψ_j = variância específica para o j -ésimo ambiente.

Os escores fatoriais são assumidos como independentes, com variância unitária, sendo $\mathbf{G}_f = \mathbf{I}_g$. Dessa forma, a matriz de variância para os efeitos dos genótipos é expressa da seguinte forma:

$$\begin{aligned}\text{var}(\mathbf{g}) &= (\mathbf{\Lambda} \otimes \mathbf{I}_g) \text{var}(\mathbf{f}) (\mathbf{\Lambda}' \otimes \mathbf{I}_g) + \text{var}(\boldsymbol{\delta}) \\ &= (\mathbf{\Lambda} \mathbf{\Lambda}' + \boldsymbol{\Psi}) \otimes \mathbf{I}_g\end{aligned}$$

Com isso, a matriz de covariâncias genéticas do modelo FAMM tem a estrutura:

$$\mathbf{G}_h = \mathbf{\Lambda} \mathbf{\Lambda}' + \boldsymbol{\Psi}$$

Para obtenção de fatores ortogonais, as cargas fatoriais foram rotacionadas. A rotação foi feita com base em Cullis et al. (2010), chamada de decomposição por valores singulares, obtendo-se $\mathbf{\Lambda}^*$ e \mathbf{f}^* , cargas e escores fatoriais rotacionados. Em posse desses valores, foram estimados as covariâncias genéticas e os valores genotípicos preditos.

3.5.1 Escolha do Número de Fatores

O número de fatores foi definido considerando o equilíbrio entre variância explicada e a parcimônia. O critério de informação de Akaike (*AIC*), proposto por Akaike (1974), foi utilizado, sendo:

$$AIC = -2 \log L + 2p$$

onde L é ponto máximo da função de verossimilhança e o p é o número de parâmetros estimados.

De acordo com Smith et al. (2015), o critério de AIC pode levar a inconsistências na escolha do modelo, por indicar que modelos de ordem inferior são superiores àqueles com maior variância percentual geral para o conjunto dos fatores. Dessa forma, além do AIC é necessário utilizar da porcentagem da variância explicada pelo conjunto de fatores como parâmetro auxiliar na seleção do modelo, dada por:

$$\bar{v} = \frac{\text{tr}(\mathbf{\Lambda} \mathbf{\Lambda}')}{\text{tr}(\mathbf{G}_h)}$$

$$v_{jk} = \frac{\sum_1^k \lambda_{jk}^2}{\sum_1^k \lambda_{jk}^2 + \Psi_j}$$

onde \bar{v} é a variância explicada para todos os k fatores e v_{jk} para cada fator no j -ésimo ambiente.

3.5.3 Parâmetros Utilizados

A acurácia foi calculada com base em Mrode (2014), sendo:

$$r = \sqrt{1 - \frac{PEV}{\hat{\sigma}_g^2}}$$

onde a PEV é a variância do erro de predição, obtida a partir da diagonal da inversa da matriz dos coeficientes. A confiabilidade de cada genótipo também é dada por:

$$r_i^2 = 1 - \frac{PEV_i}{\hat{\sigma}_g^2}$$

A herdabilidade estimada é a mesma utilizada para o modelo base sem análise espacial, dada por Cullis, Smith e Coombes (2006):

$$H^2 = 1 - \frac{\bar{V}(\Delta)}{2\hat{\sigma}_g^2}$$

As correlações genéticas entre os ambientes foram calculadas com base no critério descrito por Cullis et al. (2010), dada por:

$$\rho_{g_{jj'}} = \frac{\sum_{i=1}^k \lambda_{jk}^2 \lambda_{j'k}^2}{\sqrt{\hat{\sigma}_{gj}^2 \hat{\sigma}_{gj'}^2}} = \mathbf{D}\mathbf{G}_h\mathbf{D}$$

onde $\hat{\sigma}_{gj}^2$ e $\hat{\sigma}_{gj'}^2$ são os componentes de variância genotípica nos ambientes j e j' e a matriz \mathbf{D} é uma matriz diagonal com os elementos invertidos das raízes quadradas da diagonal dos elementos de \mathbf{G}_h . A correlação genética foi representada utilizando um mapa de calor, associado a um dendrograma que agrupou os ambientes pelo método completo. O pacote

ComplexHeatmap, versão 2.14.0 (GU; EILS; SCHLESNER, 2016) foi utilizado para gerar o mapa.

3.5.4 Natureza da interação GxA

A natureza do tipo de interação GxA em contexto MET, foi decomposta pela metodologia dada por Cooper e DeLacy (1994). Os autores consideram medidas de similaridade e dissimilaridade em análises METs. A interação GxA foi ajustada pelo modelo de simetria composta, onde as variâncias em cada ambiente são homogêneas ($\hat{\sigma}_g^2 + \hat{\sigma}_{ga}^2$) e, entre qualquer par de ambiente são uniforme ($\hat{\sigma}_g^2$). A interação do tipo simples ($\hat{\sigma}_{ga}^2$) foi particionada do tipo complexa a partir da variância genética dentro de cada ambiente, subtraindo-a pelo covariância genética entre os ambientes, estimando a parte simples.

$$\hat{\sigma}_{ga}^2 = \frac{\sum_{j < j', j'} (\hat{\sigma}_{gj}^2 + \hat{\sigma}_{gj'}^2 - 2\hat{\sigma}_{gjj'})}{a(a-1)} = \frac{\sum_{j < j', j'} \frac{g}{g-1} D_{jj'}}{a(a-1)}$$

onde $\hat{\sigma}_{gjj'}$ é a covariância genética entre os pares de ambientes j e j' . $D_{jj'}$ é a diferença no desempenho genotípico nos ambientes j e j' .

3.5.5 Factor Analytic Selection Tools – FAST

As ferramentas de seleção propostas por Smith e Cullis (2018), utiliza da estrutura do FAMM para obter medidas formais de desempenho geral e estabilidade. Vale lembrar que os efeitos comuns previsíveis de genótipos ao longo dos ambientes, a partir da rotação das cargas, é dado por: $\mathbf{g} = (\mathbf{\Lambda}^* \otimes \mathbf{I}_m) \mathbf{f}^*$, sendo $g_{ij} = \lambda_{j1}^* f_{i1}^* + \lambda_{j2}^* f_{i2}^* + \lambda_{j3}^* f_{i3}^* + \dots + \lambda_{jk}^* f_{ik}^*$ (CULLIS et al., 2014). No FAST, esses efeitos comuns são convertidos em efeitos associados ao primeiro fator e o restante, e dessa forma $g_{ij} = \lambda_{j1}^* f_{i1}^* + e_{ij}$, sendo $e_{ij} = \lambda_{j2}^* f_{i2}^* + \lambda_{j3}^* f_{i3}^* + \dots + \lambda_{jk}^* f_{ik}^*$. A partir disso, foi possível estimar os parâmetros recomendados por Smith e Cullis (2018), sendo eles a Desempenho Geral (OP_i), que é uma medida de desempenho geral; Raiz do Quadrado Médio dos Desvio ($RMSD_i$) medida de estabilidade geral e Capacidade de Resposta (RE_i), dadas por:

$$OP_i = \frac{1}{a} \sum_{j=1}^a \lambda_{j1}^* f_{i1}^*$$

$$RMSD_i = \sqrt{\frac{1}{a} \sum_{j=1}^a e_{ij}^2}$$

$$RE_i = (\bar{\lambda}_{k+}^* - \bar{\lambda}_{k-}^*) f_{ik}^*$$

onde $\bar{\lambda}_{k+}^*$ é a média das cargas positivas e $\bar{\lambda}_{k-}^*$ média das cargas negativas do k -ésimo fator. De acordo com Cullis et al. (2014), quando se obtém as cargas rotacionadas e, por conseguinte os fatores rotacionados, é possível estudar a estabilidade específica dos genótipos em todos os fatores, excetuando-se o primeiro fator que indica desempenho e o restante estabilidade. Os plots para a análise FAST foram elaborados utilizando o pacote ggplot2, versão 3.4.0 (WICKHAM, 2009).

Foram selecionados os 15 melhores genótipos em relação a OP e $RMSD$, sendo que, os mais produtivos têm uma maior vantagem no critério de seleção. O seguinte índice de seleção (IS_i) foi utilizado para os identifica-los:

$$IS_i = 3 \times \left(\frac{OP_i - \overline{OP}}{\sqrt{V(OP)}} \right) - \left(\frac{RMSD_i - \overline{RMSD}}{\sqrt{V(RMSD)}} \right)$$

Todos os procedimentos analíticos foram conduzidos utilizando o software R, versão 4.2.2 (RCORETEAM, 2022), utilizando os pacotes de ASReml-R, versão 4.1.0.176 (BUTLER et al., 2018) e SpATS, versão 1.0-18 (RODRÍGUEZ-ÁLVAREZ et al., 2018).

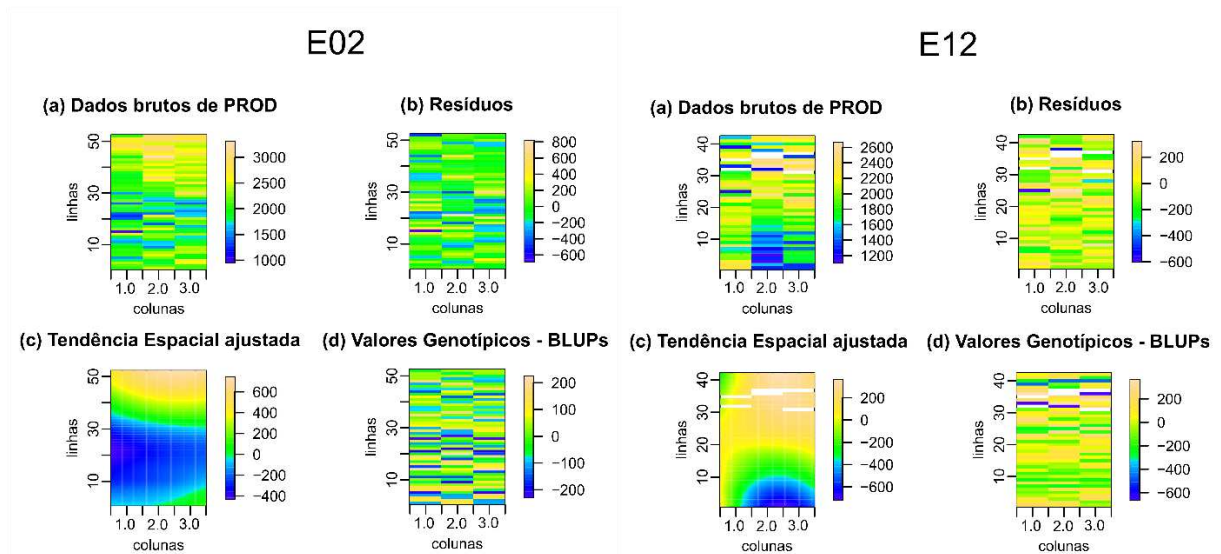
4. RESULTADOS

4.1 Resultado modelagem SpATS

Os resultados para o modelo SpATS (Figura 3), onde optou-se por escolher os dois ambientes mais contrastantes em relação a correção espacial realizada pela SpATS. O E12 foi melhor corrigido pela SpATS tendo uma maior diferença de herdabilidade corrigida em relação ao modelo base não espacial (dif. de 0,10), enquanto o E02 não precisou de correção (dif. de 0,33).

Nos dois ambientes, sugere-se que o padrão espacial dos resíduos não foi totalmente capturado pelo delineamento experimental (Figura 3b), existindo diferentes padrões residuais nos sentidos das linhas e das colunas. Indicando que o uso de metodologias que corrigem tendências espaciais *a posteriori* pode ser efetivo. Por meio da tendência espacial ajustada, é possível visualizar os gradientes de tendência espacial ajustados em cada ambiente (Figura 3c). Os valores genotípicos, considerando o efeito de genótipos como aleatório, são corrigidos e plotados (Figura 3d).

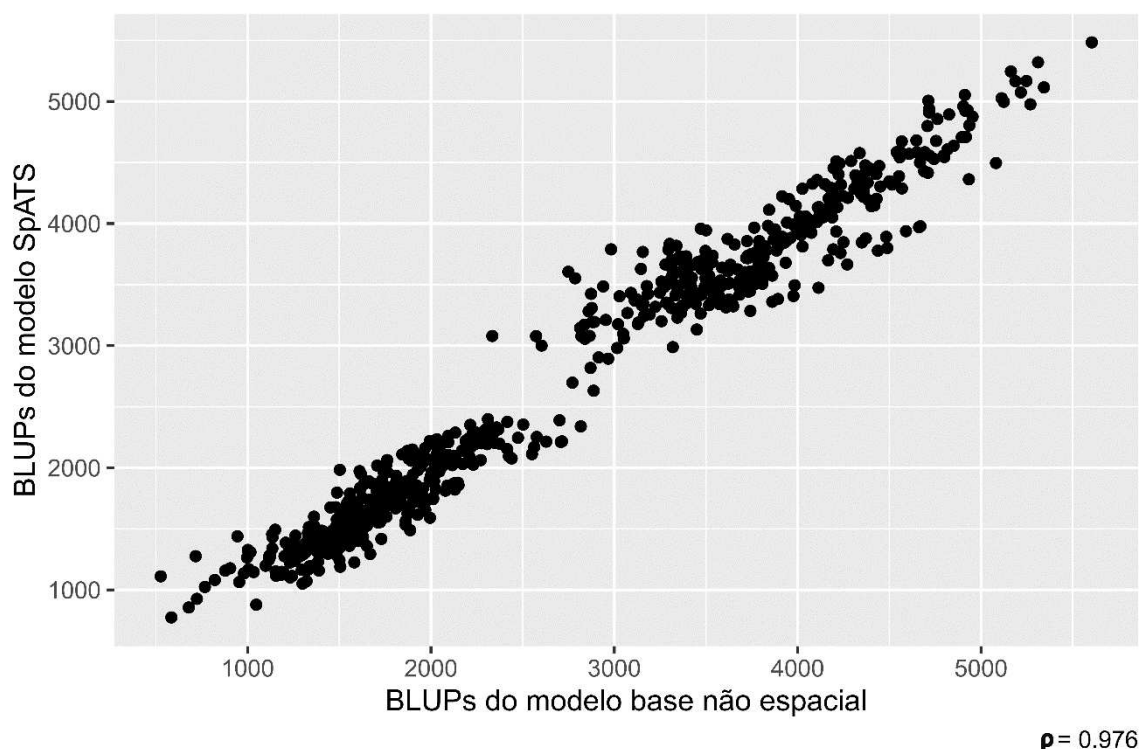
Figura 3. Plots resultados SpATS em dois ambientes (E02 e E12): (a) dados brutos de produtividade. Observa-se que há um contraste entre os dois grupos de ambientes, em relação a quantidade de linhas, com número igual de colunas, delimitando a repetição dos genótipos; (b) distribuição dos resíduos nos ensaios; (c) tendência espacial ajustada; (d) valores genotípicos pós modelagem. As parcelas estão representadas por diferentes cores, sendo que em branco simbolizam parcelas perdidas.



Fonte: SpATS (RODRÍGUEZ-ÁLVAREZ et al., 2018)

Os resultados obtidos comparando a metodologia da SpATS com o modelo base sem espacial foram similares. A mudança no ranqueamento, calculada por meio da correlação de Spearman ($\rho = 0,976$), apresentou-se muito baixa, com pouca variação entre os modelos. É possível observar uma tendência quase linear, onde os BLUP's dos genótipos menos e mais produtivos continuam sendo os menores e os maiores valores nos dois modelos (Figura 4).

Figura 4. Distribuição dos BLUP's para o modelo SpATS vs modelo base não espacial para a característica de PROD.

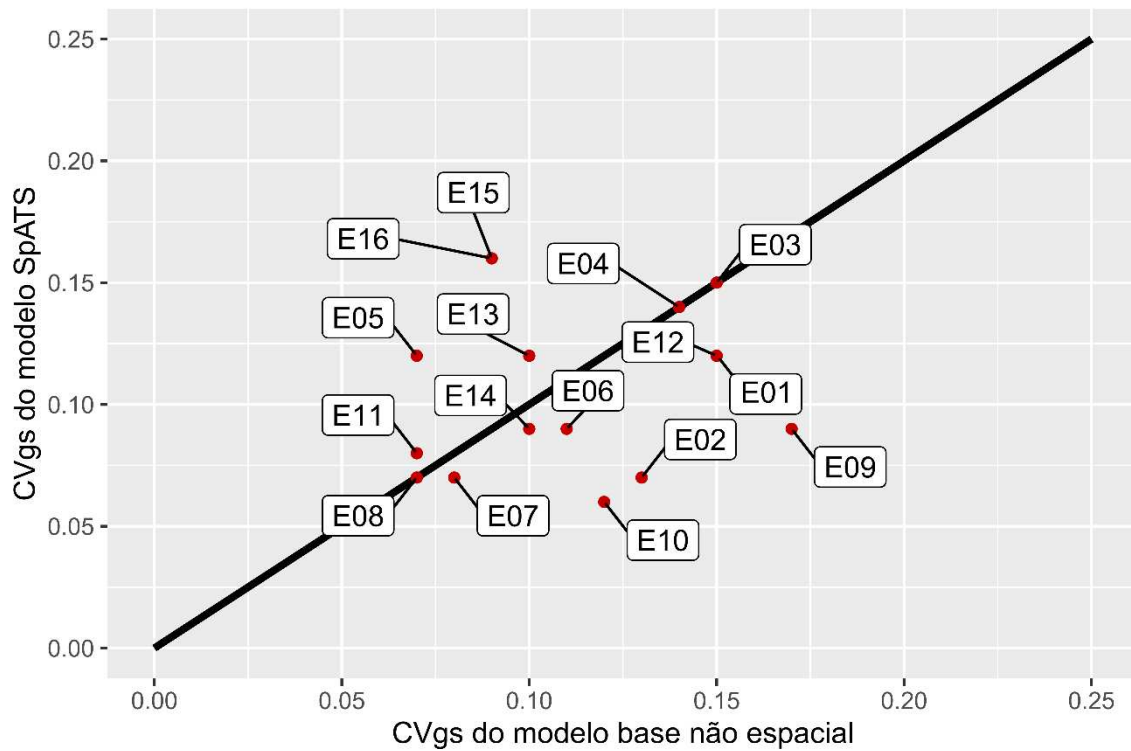


Fonte: O autor (2023)

Os coeficientes de variação genotípica foram divergentes entre o modelo SpATS e o sem espacial (Figura 5). Foi possível identificar três grupos distintos de ambientes: 1) Grupo de ambientes que apresentaram maior variabilidade genética quando utilizado SpATS; 2) Grupo de ambientes que apresentaram maior variabilidade genética quando utilizado o modelo sem espacial; 3) Ambientes que mantiveram a variabilidade genética similar entre os dois modelos, localizados no segmento da reta.

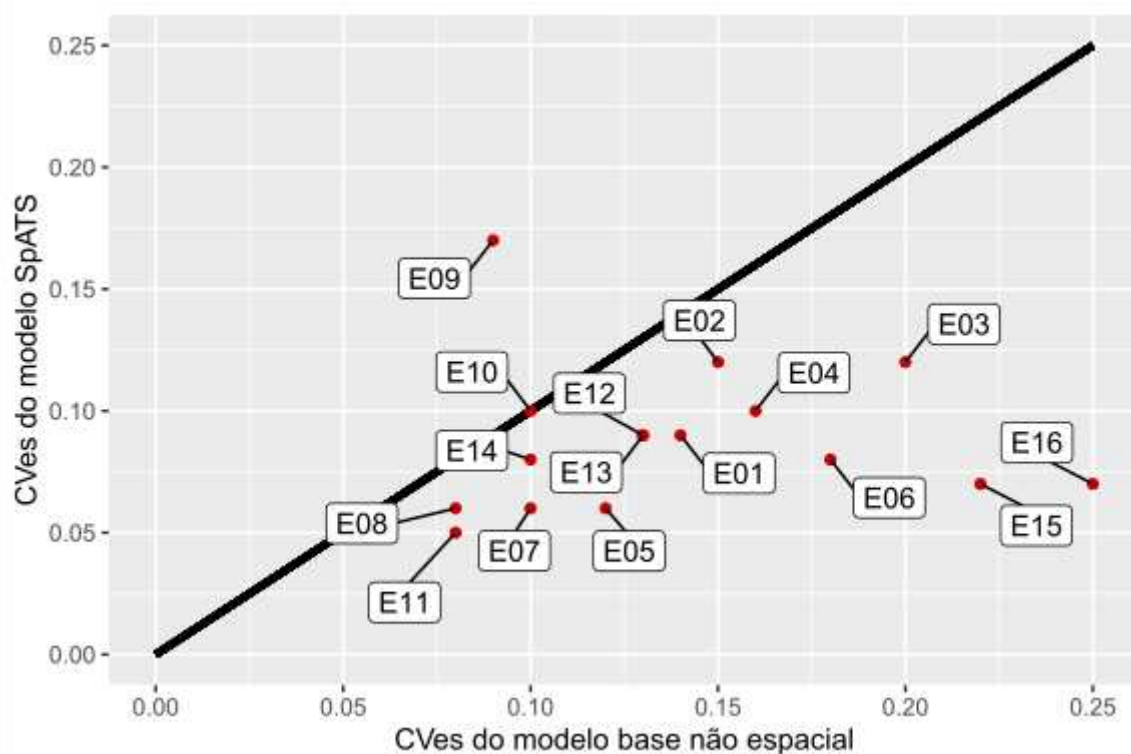
Os coeficientes de variação ambiental (Figura 6), apresentaram o mesmo padrão de grupos de ambientes descritos no parágrafo anterior, porém uma menor variabilidade ambiental é encontrada no modelo SpATS. O E09 foi o único ambiente com maior variação ambiental na SpATS quando comparado com o modelo sem espacial.

Figura 5. Comparação entre os coeficientes de variação genotípica no modelo SpATS vs modelo não espacial.



Fonte: O autor (2023)

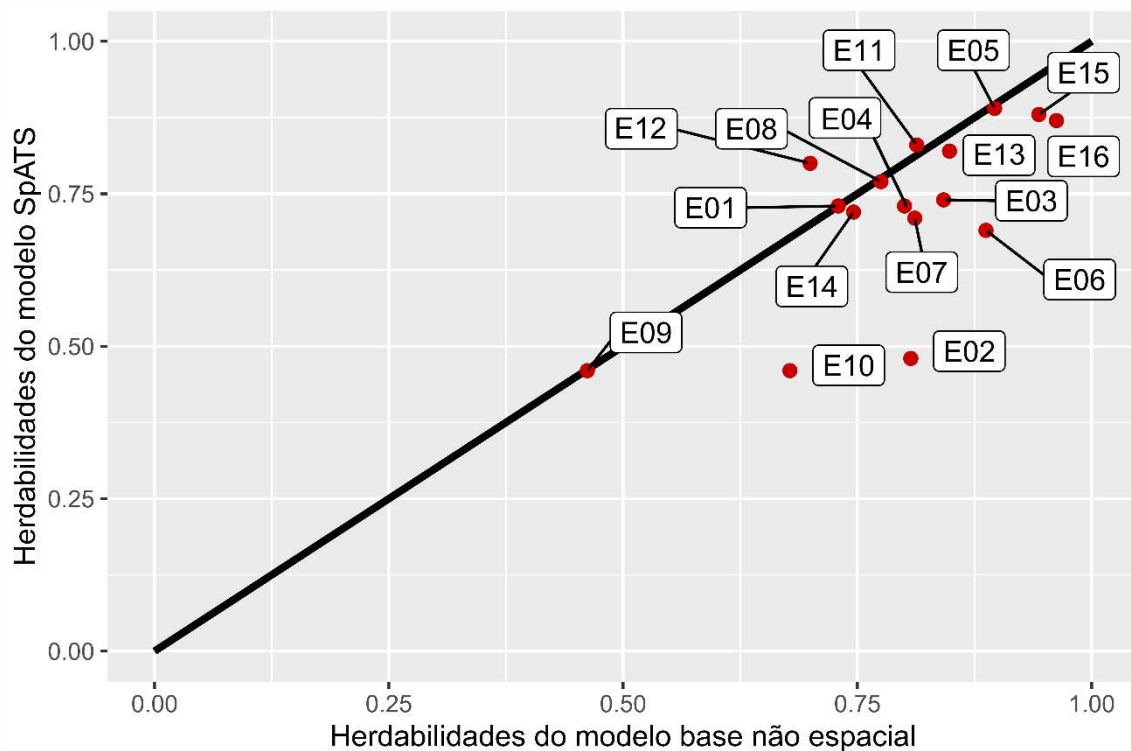
Figura 6. Comparação entre os coeficientes de variação ambiental no modelo SpATS vs modelo não espacial.



Fonte: O autor (2023)

Ainda que os resultados dos coeficientes de variação ambiental indicassem que um melhor ajuste dos efeitos de tendência seria propiciado pela SpATS, devido a redução da variabilidade residual quando comparado com o modelo sem espacial, isso não se refletiu no aumento da herdabilidade (Figura 7). Observa-se que, houve um padrão de valores similares (12,50%) e maiores valores obtidos pelo modelo não espacial (75,00%), sendo a SpATS bem-sucedida em ajustar os efeitos espaciais em dois ambientes (12,50%).

Figura 7. Comparação entre os valores de herdabilidade no modelo SpATS vs modelo não espacial.



Fonte: O autor (2023)

Esses resultados são promissores no sentido de demonstrar que a modelagem SpATS pode ser utilizada como ferramenta auxiliar em ambientes que o delineamento experimental não consegue controlar todos os efeitos residuais presentes nos ensaios. Essa ferramenta não tem nenhum custo extra experimental, não necessitando de nenhuma mudança no posicionamento ou disposição das parcelas no campo, podendo ser utilizada nos mais diversos contextos e em diferentes tipos de delineamentos experimentais.

4.2 Modelo Fator Analítico Multiplicativo Misto

Os resultados do FAMM foram obtidos considerando a análise em dois estágios, sendo os BLUE's e os pesos dos ambientes E11 e E12 obtidos pela SpATS e o restante pelo modelo base sem espacial. Os BLUE's e os pesos foram arranjados, para posteriormente serem utilizados na análise FAMM.

O modelo escolhido foi definido com base no \bar{v} mínimo de 80%, sugerido por Smith et al. (2015), e pelo menor valor de *AIC*. Nota-se que na classificação do *AIC*, os modelos FA3 e FA4 tiveram menores valores de 8989,38 e 8990,00, respectivamente (Tabela 2). FA3 foi o escolhido, por ser um modelo menos parametrizado, tendo menor valor de *AIC* e sendo o primeiro modelo FA com um de \bar{v} acima de 80%, levando a um ajuste adequado dos dados.

Tabela 2. Parâmetros utilizados para seleção do melhor modelo de fator analítico em conjunto de dados MET.

Modelo	Parâmetros ^a	<i>AIC</i>	<i>r</i>	<i>H</i> ²	\bar{v}
FA1	32	9058,68	0,69	0,50	39,62
FA2	48	9026,50	0,75	0,59	62,91
FA3	64	8989,38	0,77	0,62	82,45
FA4	80	8990,00	0,78	0,63	86,74
FA5	96	8994,25	0,79	0,63	92,37

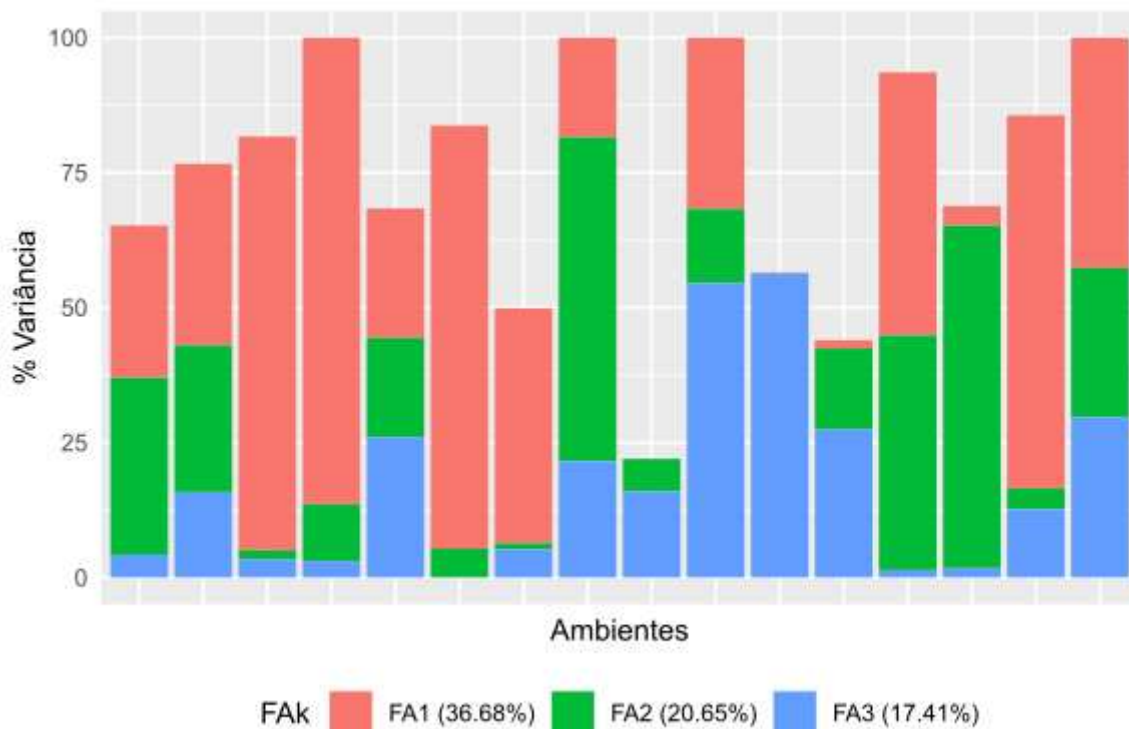
^a Número de parâmetros estimados é $kj + j$. Está em negrito o modelo FAMM escolhido.

Tabela 3. Valores para o modelo FA3 de cargas fatoriais, variância específica (Ψ), variância genética (σ_g^2).

Ambientes	Cargas fatoriais ^a			Variância	
	FA1	FA2	FA3	Psi (Ψ)	σ_g^2
E01	215,30	-232,05	83,01	57136,88	164228,24
E02	-192,91	173,49	132,14	25864,15	110638,29
E03	297,89	44,86	61,84	21229,56	115804,77
E04	247,92	86,48	46,59	0,00	71111,74
E05	128,22	-112,15	133,35	21670,71	68471,01
E06	640,10	-166,27	14,11	85091,96	522668,55
E07	241,01	-38,83	83,18	66903,83	133416,40
E08	-173,83	313,55	187,70	0,00	163761,05
E09	0,61	-32,18	52,34	13406,73	17182,35
E10	-206,91	135,90	271,04	0,00	134743,42
E11	14,92	5,18	244,11	45977,58	105815,58
E12	30,46	-94,24	127,29	33156,04	59167,47
E13	-612,11	-578,86	100,96	49673,55	769627,36
E14	88,33	-369,41	-60,58	67028,62	214962,10
E15	334,14	-79,25	142,77	23259,48	161572,76
E16	196,14	-157,98	163,49	0,00	90156,72

^a Cargas em negrito no fator, representam a maior variância explicada para o ambiente.

Figura 8. Variância explicada para cada fator no j-ésimo ambiente.



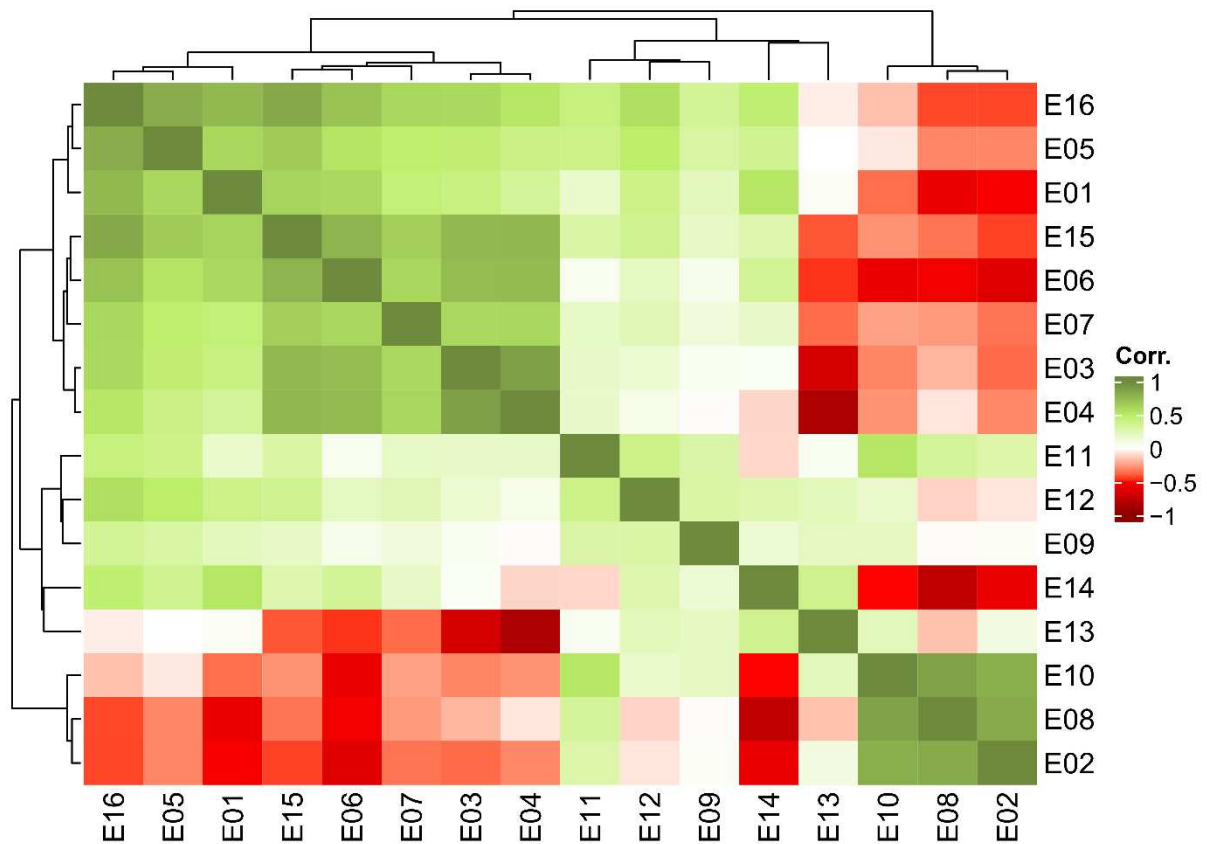
Fonte: O autor (2023)

As cargas fatoriais estimadas pelo modelo FA3 para cada ambiente, foram rotacionadas, a fim de garantir a sua ortogonalidade (Tabela 3). Note que no primeiro fator a maioria das cargas são positivas, indicando que os efeitos da interação GxA não são cruzados neste fator, ou seja, o somatório dos efeitos de cada ambiente para o Fator 1 não se anulam, sendo possível estimar o desempenho dos genótipos com o *OP* e o *RMSD* (SMITH, CULLIS, 2018).

A variância explicada para cada fator entre os ambientes (v_{jk}) foi estimada pelo modelo FA3 (Figura 8). Note que os ambientes têm uma grande variação entre eles. Ambientes que tiveram uma baixa variância explicada pelos fatores comuns, está relacionado com a porcentagem de variância não explicada pelo modelo. Os ambientes apresentaram aproximadamente 30% de interação do tipo simples e 70% do tipo complexa, demonstrando a importância de se utilizar a estrutura Fator Analítico para estudar a previsibilidade do comportamento dos genótipos frente as mudanças ambientais.

A correlação genética ($\rho_{g_{jj'}}$), foi calculada para cada par de ambientes, sendo obtido em sua maioria correlações positivas e baixas, com valores extremos de -0,8425% (E04 vs E13) e 0,8883% (E03 vs E04) (Figura 9). As correlações heterogêneas entre os ambientes corroboram com a natureza da interação GxA do tipo complexa. Dado que a presença de ambientes com correlações negativas e nulas (ex. E04 vs E13 e E09 vs E02), acarretam em mudanças no ranqueamento dos genótipos. Ambientes que mesmo não inseridos geograficamente no mesmo local possuem um alto valor de correlação entre eles (Figura 9). Também, o inverso é verdadeiro, podendo ser notado pela ausência (próxima a nulidade) e/ou pelas correlações negativamente entre ambientes distantes. Ambiente que se diferem em relação à época são similares de acordo com a correlação genética.

Figura 9. Mapa de calor das correlações genéticas nos 16 ambientes par a par utilizando o modelo FA3. A intensidade da cor está relacionada com a magnitude da correlação.



Fonte: O autor (2023)

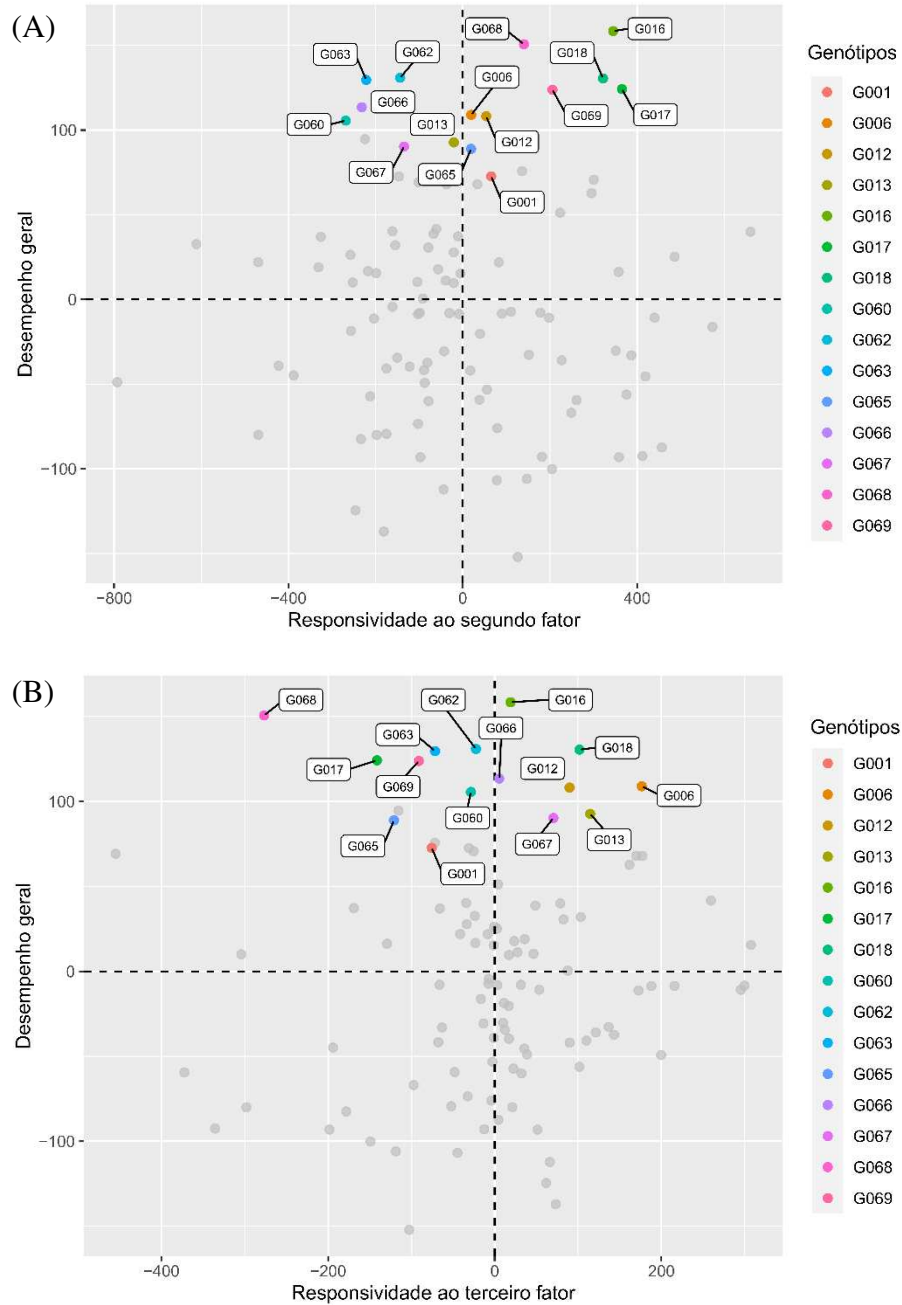
Os genótipos são apresentados no biplot OP vs $RMSD$ (Figura 10), os selecionados sendo destacados no gráfico. Esse gráfico proporciona uma facilidade na visualização dos melhores genótipos no contexto MET. O eixo Y representa o desempenho dos genótipos, sendo que, o desempenho geral classifica os genótipos de acordo com sua produtividade, então quanto maior o valor de OP_i , mais produtivo é o genótipo. O eixo X representa a estabilidade geral dos genótipos, sendo que, genótipos com valores mais próximos de zero, possuem uma maior estabilidade, tendo como interpretação oposta ao OP_i .

Os valores de OP e $RMSD$ ficaram entre -152,10 (G087) e 158,42 (G016) e 554,97 (G003) e 3,74 (G024), respectivamente. O genótipo G016 com maior OP , teve um valor médio de $RMSD$ na ordem de 241,62, sendo esses valores em termos de BLUP. O genótipo G024 mais estável foi pouco produtivo, tendo 15,34 de OP . Os genótipos foram selecionados com base no IS_i , onde os 15 melhores genótipos foram escolhidos. Os genótipos destacados em cores representam os selecionados, com suas respectivas intensidades de confiabilidade (r_i^2) (Figura 10).

possuir o maior valor de *OP* (Figura 10), porém, com alto valor de *RMSD*. Assim, essa alta instabilidade pode estar vinculada aos ambientes com cargas opostas às recomendadas para esse genótipo. Por isso, com o objetivo de capitalizar no efeito da interação GxA, optou-se por recomendar esse único genótipo para os ambientes E02, E03, E04, E08, E10 e E11, aprimorando o desempenho para as condições ambientais específicas.

Os genótipos selecionados para a análise geral foram o G062, G016, G068, G063, G012, G069, G006, G018, G066, G013, G065, G067, G060, G017 e G001, onde tiveram valores de confiabilidade entre 0,57 a 0,89, G062 e G012, respectivamente. Também foi proposto a recomendação do G016, com uma confiabilidade de 0,73, para os ambientes com condições ambientais específicas, buscando-se obter um maior desempenho desse genótipo.

Figura 11. Desempenho geral (OP_i) e responsividade (RE_i) para o Fator 2 (A) e Fator 3 (B) dos 104 genótipos de soja, nos 16 ambientes utilizando o modelo FA3. Ambientes com altos valores de carga e sinal negativo, genótipos à esquerda são mais adaptados, e para sinal positivo, optar por genótipos à direita do zero no eixo X em cada fator.



Fonte: O autor (2023)

5. DISCUSSÃO

A utilização da metodologia 2D P-splines SpATS se mostrou eficiente como uma técnica auxiliar para o controle da dependência espacial. Neste trabalho, a principal particularidade observada dessa técnica é a facilidade de se modelar os ambientes sem ter que testar os diversos tipos de efeitos que podem incidir nos ambientes, ganhando em rendimento operacional. De fato, vários estudos na literatura demonstram que esse método de correção espacial é rápido, fácil e flexível para execução em diversos tipos de análises (EEUWIJK et al., 2019; ANDRADE et al., 2020; XU; CANNON; BEAVIS, 2022). Além disso, é estável, já que não costuma demonstrar problemas de convergência, como nos modelos $AR1 \otimes AR1$ em alguns casos (VELAZCO et al., 2017). Igualmente, a metodologia FAST mostrou-se como uma ferramenta importante para o melhorista na seleção de genótipos em contexto MET. Tais ferramentas são capazes de selecionar os melhores genótipos frente às variações ambientais, utilizando a estrutura Fator Analítico com informações concisas, precisas e resumidas na presença de interação GxA. Essa metodologia incluiu medidas *OP*, *RMSD* e *RE* para todos os ambientes e para cada genótipo, que foram incorporadas no IS_I .

O ajuste espacial proporcionado pela metodologia foi comparável para a maioria dos ambientes, onde a correção do efeito ambiental foi realizada pelo delineamento (Figura 7). Esses resultados são apoiados pela invariância do ranqueamento dos genótipos ($\rho = 0,976$) ao longo dos ambientes. Ainda que, o delineamento possa ter controlado as tendências espaciais na maioria dos ambientes, não se apresentou como método incontestável no controle da dependência espacial. Logo, a utilização da SpATS, como técnica auxiliar, foi bem-sucedida em controlar as tendências espaciais onde somente o delineamento não foi suficiente, sucedendo uma melhora em 12,5% dos casos. À vista disso, por se tratar de uma técnica gratuita, os programas de melhoramento podem se valer dela para melhoria da qualidade experimental, levando ao aperfeiçoamento da seleção. Ainda mais que, a precisão da seleção é uma importante ferramenta para avaliar o êxito de um programa de melhoramento genético (RESENDE; ALVES, 2022).

Na análise de fator analítico, o modelo FA3 foi o melhor modelo ajustado para este conjunto de dados (Tabela 2), já que possuiu o menor AIC e variância explicada acima de 80%. Por consequência, o modelo apresentou eficiência computacional por ser mais parcimonioso, e representou grande parte da variação do conjunto *MET*. Isto pode ser validado pelos valores individuais de v_{jk} (Figura 8), os quais demonstram que maior parte dos ambientes (9 de 16),

apresentassem valores individuais superiores a 75%, fornecendo um ajuste adequado aos dados (SMITH et al., 2015).

A magnitude da correlação genética (Figura 9), demonstrou a presença da interação GxA, onde o comportamento par a par dos ambientes com épocas distintas, no mesmo município, foi sempre positivo e, nos ambientes distantes, variou entre padrões de semelhança, dissemelhança e até ausência de quaisquer relações. Conforme Cullis et al. (2010) a heterogeneidade da correlação entre os pares, evidenciam a presença da interação GxA do tipo complexa. A presença da interação GxA neste trabalho foi corroborada pelo método de Cooper e DeLacy (1994), onde a maior parte da interação é de natureza complexa (70%). Os resultados enfatizam a importância da análise de fator analítico em contexto de MET, a qual é bem-sucedida em aproveitar-se da covariância genética entre ambientes para embasar tomadas de decisões mais precisas acerca da população de ambientes-alvo.

A capacidade do modelo em fornecer informações seguras do comportamento dos genótipos no contexto *MET* está diretamente relacionado com a magnitude da covariância entre os ambientes, tornando-se um importante parâmetro a ser considerado (SMITH et al., 2015; KRAUSE et al., 2020). Note que, no dendrograma (Figura 9) há a formação de grupos de ambientes mais correlacionados, onde os ambientes mais próximos têm padrões similares de interação GxA do que aos mais afastados. Isso demonstra a possibilidade de estratificação ambiental ser feita. Contudo, informações mais robustas devem ser adicionadas para obter uma estratificação mais confiável, como informação genômica dos genótipos (KRAUSE et al., 2023) e considerar a diferença nos efeitos abióticos entre os ambientes por meio das covariáveis ambientais (SMITH et al., 2015; OLIVEIRA et al., 2020; CHAVES et al., 2023).

A análise em dois estágios possibilitou que a dependência espacial fosse corrigida, e, conseqüentemente, o viés das estimativas dos efeitos de genótipos fossem dirimidos. Isto propiciou resultados aceitáveis utilizando os modelos FMM. Vários outros estudos relacionados, também tem demonstrado que FMM pode ser utilizado de forma combinada com análises de modelagem espacial, como em milho (TESFAYE et al., 2023), milho (FIGUEIREDO et al., 2015), sorgo (WAGAW et al., 2020) e cevada (TADESE; LAKEW; TAYE, 2019). Entretanto, estudos aplicando a metodologia SpATS para correção espacial e a análise FMM, com destaque na utilização do FAST, para recomendação de genótipos de soja ainda são escassos.

FAST (SMITH, CULLIS, 2018) como ferramenta de análise de dados, fornece informações sucintas dos resultados emitidos pelo FMM, sendo demonstrado por diversos tipos de trabalhos atuais o poder de sua performance (TOLHURST et al., 2019; SJOBERG et

al., 2021; FERRANTE et al., 2021; COAST et al., 2022; CHAVES et al., 2023). Neste trabalho, empregamos o FAMM para compreender a dinâmica da interação GxA, resumindo os resultados nas três estimativas propostas pelo FAST (Figura 10 e Figura 11), utilizando índice de seleção. Logrou-se êxito em recomendar genótipos com adaptação geral e específica, onde o genótipo com maior desempenho no conjunto de dados, foi recomendado para as suas condições de especificidade, buscando melhorar o seu desempenho, aplicando a ferramenta FAST. Isso só foi possível, devido à maioria das cargas serem positivas para o FA1 (Tabela 3). Caso houvesse um aumento na magnitude e nos números de ambientes com cargas negativas para esse fator, elevando a sua proporção, chegaria a um ponto que a interação GxA complexa imperaria, não sendo mais possível utilizar das medidas gerais de seleção de genótipos (SMITH, CULLIS, 2018; TOLHURST et al., 2019).

6. CONCLUSÃO

A implantação da SpATS corrigiu a dependência espacial em ambientes que o delineamento experimental não conseguiu controlar todos os efeitos residuais presentes nos ensaios. O uso da ferramenta FAST apresentou robustez, característica da estrutura fator analítico e dos modelos lineares mistos, e, simultaneamente, flexibilidade por possibilitar a recomendação com base em índices de seleção, seja para desempenho e a estabilidade geral, ou para ambientes específicos, almejando a melhora do desempenho de um genótipo. O G016 foi recomendado para ambientes específicos; G062, G016, G068, G063, G012, G069, G006, G018, G066, G013, G065, G067, G060, G017 e G001 para os ambientes de forma geral, sendo os mais produtivos e estáveis.

6. REFERÊNCIAS

AKAIKE, H. A new look at the statistical model identification. **IEEE Transactions on Automatic Control**, v. 19, n. 6, p. 716–723, 1974.

ALBUQUERQUE, J. R. T. DE et al. Adaptability and stability of soybean (*Glycine max* L.) genotypes in semiarid conditions. **Euphytica**, v. 218, n. 5, p. 61, 2022. <https://doi.org/10.1007/s10681022-03012-0>

ALVARES, C. A. et al. Köppen's climate classification map for Brazil. **Meteorologische Zeitschrift**, v. 22, p. 711–728, 2013. <https://doi.org/10.1127/0941-2948/2013/0507>

ANDRADE, M. H. M. L. et al. Accounting for spatial trends to increase the selection efficiency in potato breeding. **Crop Science**, v. 60, n. 5, p. 2354–2372, 2020. <https://doi.org/10.1002/csc2.20226>

BAKARE, M. A. et al. Parsimonious genotype by environment interaction covariance models for cassava (*Manihot esculenta*). **Frontiers in Plant Science**, v. 13, p. 978248, 2022. <https://doi.org/10.3389/fpls.2022.978248>

BELLALOU, N. et al. **Soybean seed composition and quality: Interactions of environment, genotype, and management practices**. Em: *Soybeans: Cultivation, Uses and Nutrition*. (ed. Maxwell, J. E.), Nova Science Publishers, p. 1–42, 2011.

BERNARDELI, A. et al. Modeling spatial trends and enhancing genetic selection: An approach to soybean seed composition breeding. **Crop Science**, v. 61, n. 2, p. 976–988, 2020. <https://doi.org/10.1002/csc2.20364>

BERNARDO, R. Reinventing quantitative genetics for plant breeding: something old, something new, something borrowed, something BLUE. **Heredity**, v. 125, n. 6, p. 375–385, 2020. <https://doi.org/10.1038/s41437-020-0312-1>

BIANCHI, M. C. et al. Heritability and the genotype × environment interaction in soybean. **AgroSystems, Geosciences & Environment**, v. 3, n. 1, p. e20020, 2020. <https://doi.org/10.1002/agg2.20020>

BROWNIE, C.; BOWMAN, D. T.; BURTON, J. W. Estimating Spatial Variation in Analysis of Data from Yield Trials: A Comparison of Methods. **Agronomy Journal**, v. 85, n. 6, p. 1244–1253, 1993. <https://doi.org/10.2134/agronj1993.00021962008500060028x>

BUTLER, D. G. et al. **ASReml estimates variance components under a general linear**. Hemel Hempstead, UK: VSN International Ltd, 2018.

CARVALHO, A. D. F.; FRISTICHE NETO, R.; GERALDI, I. O. Estimation and prediction of parameters and breeding values in soybean using REML/BLUP and least squares. **Crop Breeding and Applied Biotechnology**, v. 8, n. 3, p. 219–224, 2008.

CASLER, M. D. Blocking Principles for Biological Experiments. Em: **Applied Statistics in Agricultural, Biological, and Environmental Sciences**. [s.l.] John Wiley & Sons, Ltd, 2018. p. 53–72.

CHAVES, S. F. S. et al. Employing factor analytic tools for selecting high-performance and stable tropical maize hybrids. **Crop Science**, p. 1–12, 2023. <https://doi.org/10.1002/csc2.20911>

CONAB. **Acompanhamento de safra brasileira de grãos**. v. 9, safra 2021/22, n. 12 – décimo segundo levantamento. Disponível em: <<http://www.conab.gov.br/info-agro/safras/graos/boletim-da-safra-de-graos>>. Acesso em: 19 jan. 2023.

COOPER, M.; DELACY, I. H. Relationships among analytical methods used to study genotypic variation and genotype-by-environment interaction in plant breeding multi-environment experiments. **Theoretical and Applied Genetics**, v. 88, n. 5, p. 561–572, 1994. <https://doi.org/10.1007/bf01240919>

CROSSA, J. Statistical Analyses of Multilocation Trials. Em: BRADY, N. C. (Ed.). **Advances in Agronomy**. [s.l.] Academic Press, 1990. v. 44p. 55–85. [https://doi.org/10.1016/S0065-2113\(08\)60818-4](https://doi.org/10.1016/S0065-2113(08)60818-4)

CULLIS, B. R. et al. Analysis of yield and oil from a series of canola breeding trials. part ii. exploring variety by environment interaction using factor analysis. **Genome**, 11. v. 53, p. 1002–1016, 2010. <https://doi.org/10.1139/G10-080>

CULLIS, B. R. et al. Factor analytic and reduced animal models for the investigation of additive genotype-by-environment interaction in outcrossing plant species with application to a *Pinus radiata* breeding programme. **Theoretical and Applied Genetics**, v. 127, n. 10, p. 2193–2210, 2014. <https://doi.org/10.1007/s00122-014-2373-0>

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, n. 4, p. 381–393, 2006. <https://doi.org/10.1198/108571106X154443>

DIAS, G. H. DE O. et al. Desenvolvimento de cultivares de soja de crescimento indeterminado após a poda apical. **Research, Society and Development**, v. 10, n. 5, p. e46510513688–e46510513688, 2021. <https://doi.org/10.33448/rsd-v10i5.13688>

DIAS, K. O. D. G. et al. Estimating Genotype × Environment Interaction for and Genetic Correlations among Drought Tolerance Traits in Maize via Factor Analytic Multiplicative Mixed Models. **Crop Science**, v. 58, n. 1, p. 72–83, 2018. <https://doi.org/10.2135/cropsci2016.07.0566>

DURBÁN, M.; CURRIE, I. D. A note on P-spline additive models with correlated errors. **Computational Statistics**, v. 18, n. 2, p. 251–262, 2003. <https://doi.org/10.1007/s001800300143>

EEUWIJK, F. A. V. et al. Modelling strategies for assessing and increasing the effectiveness of new phenotyping techniques in plant breeding. **Plant Science**, The 4th International Plant Phenotyping Symposium. v. 282, p. 23–39, 2019. <https://doi.org/10.1016/j.plantsci.2018.06.018>

EEUWIJK, F. A. V.; BUSTOS-KORTS, D. V.; MALOSETTI, M. What Should Students in Plant Breeding Know About the Statistical Aspects of Genotype × Environment Interactions? **Crop Science**, v. 56, n. 5, p. 2119–2140, 2016. <https://doi.org/10.2135/cropsci2015.06.0375>

EISENHART, C. The assumptions underlying the analysis of variance. **Biometrics**, v. 3, p. 1-21, 1947. <https://doi.org/10.2307/3001534>

EMBRAPA. **CULTIVARES DE SOJA: CENTRO-SUL DO BRASIL E MACRORREGIÕES 1, 2 E 3 E REC 401**. 2019. Disponível em :<<https://www.embrapa.br/en/busca-de-publicacoes/-/publicacao/1116134/brs-cultivares-desoja-centro-sul-do-brasil-macrorregioes-1-2-3-e-rec-401> >. Acesso em: 07/12/2022.

ENYEW, M. et al. Genotype by environment interaction, correlation, AMMI, GGE biplot and cluster analysis for grain yield and other agronomic traits in sorghum (*Sorghum bicolor* L. Moench). **PLOS ONE**, v. 16, n. 10, p. e0258211, 2021. <https://doi.org/10.1371/journal.pone.0258211>

FAO. **Desafios: melhores políticas e mais investimentos, incluindo a adaptação da agricultura às alterações climáticas**. 2015. Disponível em: <<https://www.fao.org/brasil/noticias/detail-events/pt/c/293047/>>. Acesso em: 19 jan. 2023.

FIGUEIREDO, A. G. et al. Application of mixed models for evaluating stability and adaptability of maize using unbalanced data. **Euphytica**, v. 202, n. 3, p. 393–409, 2015. <https://doi.org/10.1007/s10681-014-1301-3>

GAUCH, H. G. Model Selection and Validation for Yield Trials with Interaction. **Biometrics**, v. 44, n. 3, p. 705–715, 1988. <https://doi.org/10.2307/2531585>

GAUCH, H. G. Statistical analysis of regional yield trials: AMMI analysis of factorial designs. Amsterdam: **Elsevier**, p. 172, 1992. <https://doi.org/10.4236/ajps.2020.116056>

GAZZONI, D. L.; DALL'AGNOL, A. **A saga da soja: de 1050 a.C. a 2050 d.C.** 1. ed. Brasília, DF: Embrapa, p. 199, 2018.

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, 1997. <https://doi.org/10.2307/1400446>

GRONDONA, M. O. et al. Analysis of Variety Yield Trials Using Two-Dimensional Separable ARIMA Processes. **Biometrics**, v. 52, n. 2, p. 763–770, 1996. <https://doi.org/10.2307/2532916>

GU, Z.; EILS, R.; SCHLESNER, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. **Bioinformatics**, v. 32, n. 18, p. 2847–2849, 2016.

HENDERSON, C. R. Best Linear Unbiased Estimation and Prediction under a Selection Model. **Biometrics**, v. 31, n. 2, p. 423–447, 1975. <https://doi.org/10.2307/2529430>

HENDERSON, C. R. Estimation of changes in herb environment. **Journal of Dairy Statistics**, v. 32, p. 706, 1949.

HUNGRIA, M.; NOGUEIRA, M. A.; ARAUJO, R. S. Co-inoculation of soybeans and common beans with rhizobia and azospirilla: strategies to improve sustainability. **Biology and Fertility of Soils**, v. 49, n. 7, p. 791–801, 2013. <https://doi.org/10.1007/s00374-012-0771-5>

ISIK, F.; HOLLAND, J.; MALTECCA, C. **Genetic Data Analysis for Plant and Animal Breeding**. Cham: Springer International Publishing, 2017. <https://doi.org/10.1007/978-3-319-55177-7>

KANG, M. S. Using Genotype-by-Environment Interaction for Crop Cultivar Development. **Advances in Agronomy**, v. 62, p. 199–252, 1997. <https://doi.org/10.2135/cropsci2006.08.0540>

KELLY, A. M. et al. The Accuracy of Varietal Selection Using Factor Analytic Models for Multi-Environment Plant Breeding Trials. **Crop Science**, v. 47, n. 3, p. 1063–1070, 2007.

KRAUSE, M. DALSANTE et al. Boosting predictive ability of tropical maize hybrids via genotype-by-environment interaction under multivariate GBLUP models. **Crop Science**, v. 60, n. 6, p. 3049–3065, 2020.

KRAUSE, M. D. et al. Using soybean historical field trial data to study genotype by environment variation and identify mega-environments with the integration of genetic and non-genetic factors. **bioRxiv**, v. 4, n. 11, e487885, 2023. <https://doi.org/10.1101/2022.04.11.487885>

LEE, D.-J.; DURBÁN, M.; EILERS, P. Efficient two-dimensional smoothing with P-spline ANOVA mixed models and nested bases. **Computational Statistics & Data Analysis**, v. 61, p. 22–37, 2013.

LERSTEN, N. R. et al. “**Vegetative morphology**”. Em: Soybeans: Improvement, Production and Uses. 3. ed. Ed. Madison, WI: American Society of Agronomy, Crop Science Society of America, Soil Science Society of America, p. 15–57, 2004. <https://doi.org/10.2134/agronmonogr16.3ed.c2>

MALOSETTI, M.; RIBAUT, J.-M.; EEUWIJK, F. A. V. The statistical analysis of multi-environment data: modeling genotype-by-environment interaction and its genetic basis. **Frontiers in Physiology**, v. 4, 2013. <https://doi.org/10.3389/fphys.2013.00044>

MATSUO, E. et al. **Importância Econômica**. Em: Soja: do plantio à colheita. Viçosa: Editora UFV, p. 27-53, 2015.

MENDOZA, N. V. R. **Estruturas unidimensionais e bidimensionais utilizando P-splines nos modelos mistos aditivos generalizados com aplicação na produção de cana-de-açúcar**. 2017. p. 85. Tese (Doutorado em Estatística e Experimentação Agronômica) - Escola Superior de Agricultura “Luiz de Queiroz”, Piracicaba, SP.

MRODE, R. A. **Linear Models for the Prediction of Animal Breeding Values**. 3rd. ed. Boston, MA: Cabi, 2014.

MÜLLER, L. **Taxonomia e morfologia**. Em: A soja no Brasil. 1. ed. Campinas: Instituto de Tecnologia de Alimentos, p. 65-104, 1981.

MURAKAMI, D. M. **NOVAS METODOLOGIAS DE ANÁLISE DE INTERAÇÃO GENÓTIPOS X AMBIENTES: ANÁLISE COMBINADA DE ESTRATIFICAÇÃO, ADAPTABILIDADE E ESTABILIDADE E ANÁLISE DE REPRESENTATIVIDADE AMBIENTAL**. 2001. p. 155. Tese (Doutorado em Genética e Melhoramento) – Universidade Federal de Viçosa, Viçosa, MG.

NOGUEIRA, A. P. O. et al. **Morfologia, crescimento e desenvolvimento**. Em: Tecnologias de produção e usos da soja. 1. ed. Ed Mecenas, Londrina, p. 314, 2009.

NUVUNGA, J. J. **ANÁLISE DE FATORES PARA ENSAIOS MULTIAMBIENTES SOB DIFERENTES NÍVEIS DE DESBALANCEAMENTO USANDO MODELOS MISTOS**. 2014. p. 85. Dissertação (Mestrado em Estatística e Experimentação Agropecuária) – Universidade Federal de Lavras, Lavras, MG.

OLIVEIRA, I. C. M. et al. Genotype-by-environment interaction and yield stability analysis of biomass sorghum hybrids using factor analytic models and environmental covariates. **Field Crops Research**, v. 257, p. 107929, 2020. <https://doi.org/10.1016/j.fcr.2020.107929>

ONU. **População mundial deve chegar a 9.7 bilhões de pessoas em 2050, diz relatório da ONU**. 2019. Disponível em: <<https://brasil.un.org/pt-br/83427-popula%C3%A7%C3%A3o-mundial-deve-chegar-97-bilh%C3%B5es-de-pessoas-em-2050-diz-relat%C3%B3rio-da-onu>>. Acesso em: 20 jan. 2023.

PATTERSON, H. D.; THOMPSON, R. Recovery of inter-block information when block sizes are unequal. **Biometrika**, v. 58, n. 3, p. 545–554, 1971. <https://doi.org/10.2307/2334389>

PEREIRA, É. M.; GRAVINA, G. A.; THIÉBAUT, J. T. L. Aplicações dos modelos lineares mistos na pesquisa agropecuária. **Natureza on line**, v. 10, n. 2, p. 52–58, 2012.

PIEPHO, H. P. et al. BLUP for phenotypic selection in plant breeding and variety testing. **Euphytica**, v. 161, n. 1, p. 209–228, 2008. <https://doi.org/10.1007/s10681-007-9449-8>

PIEPHO, H.-P. Analyzing Genotype-Environment Data by Mixed Models with Multiplicative Terms. **Biometrics**, v. 53, n. 2, p. 761–766, 1997. <https://doi.org/10.2307/2533976>

PIEPHO, H.-P. Empirical best linear unbiased prediction in cultivar trials using factor-analytic variance-covariance structures. **Theoretical and Applied Genetics**, v. 97, n. 1, p. 195–201, 1998. <https://doi.org/10.1007/s001220050885>

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. <https://doi.org/10.1002/bimj.202100212>

RCORETEAM. **R: A Language and environment for statistical computing**. Viena, Austria: R Foundation for Statistical Computing, 2022.

RESENDE, M. D. V. **Genética biométrica e estatística no melhoramento de plantas perenes**. Embrapa Informação Tecnológica, Brasília, p. 975, 2002.

RESENDE, M. D. V. **Métodos estatísticos ótimos na análise de experimentos de campo**. 1ª Ed. Colombo: Embrapa Florestas, p. 57, 2004.

RESENDE, M. D. V.; ALVES, R. S. Statistical significance, selection accuracy, and experimental precision in plant breeding. **Crop Breeding and Applied Biotechnology**, v. 22, p. e42712238, 2022. <https://doi.org/10.1590/1984-70332022v22n3a31>

RESENDE, M. D. V.; SILVA, F. F.; AZEVEDO, C. F. **Estatística matemática, biométrica e computacional**. Suprema, Visconde do Rio Branco, p. 881, 2014.

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. **Revista de matemática e estatística**, v. 21, n. 1, p. 7–33, 2003.

RESENDE, M. D. V.; THOMPSON, R. Factor analytic multiplicative mixed models in the analysis of multiple experiments. **Revista de Matemática e Estatística, São Paulo**, v. 22, n. 2, p. 31–52, 2004.

RITCHIE, S. W. et al. How a soybean plant develops. **Ames: Iowa State University of Science and Technology, Cooperative Extension Service**, n. 53, p. 20, 1982.

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

RODRÍGUEZ-GONZÁLEZ, R. E. et al. Interacción genotipo-ambiente para la estabilidad de rendimiento en trigo en la región de Mexicali, B.C., México. **Tropical and subtropical agroecosystems**, v. 14, n. 2, p. 543–558, 2011.

SEIXAS, C. D. S. et al. **Tecnologias de produção de soja**. Londrina: Embrapa Soja, p. 347, 2020. (Sistemas de Produção / Embrapa Soja, ISSN 2176-2902; n. 17).

SILVA, F. et al. **Soja: do plantio à colheita**. 2. ed. São Paulo: Oficina de textos, p. 304, 2022.

SJOBERG, S. M. et al. Application of the factor analytic model to assess wheat falling number performance and stability in multi-environment trials. **Crop Science**, v. 61, n. 1, p. 372–382, 2021. <https://doi.org/10.1002/csc2.20293>

SMITH, A. B. et al. Factor analytic mixed models for the provision of grower information from national crop variety testing programs. **Theoretical and Applied Genetics**, v. 128, n. 1, p. 55–72, 2015. <https://doi.org/10.1007/s00122-014-2412-x>

SMITH, A. B.; CULLIS, B. R. Plant breeding selection tools built on factor analytic mixed models for multi-environment trial data. **Euphytica**, v. 214, n. 143, p. 1–19, 2018. <https://doi.org/10.1007/s10681-018-2220-5>

SMITH, A. B.; CULLIS, B. R.; THOMPSON, R. The analysis of crop cultivar breeding and evaluation trials: an overview of current mixed model approaches. **The Journal of Agricultural Science**, v. 143, n. 6, p. 449–462, 2005. <https://doi.org/10.1017/S0021859605005587>

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, 2001. <https://doi.org/10.1111/j.0006-341X.2001.01138.x>

SOARES, A. R. **Regressão spline de nós livres para modelagem de curvas de crescimento multifásica**. 2019. p. 64. Dissertação (Mestrado em Estatística e Experimentação Agronômica) - Escola Superior de Agricultura “Luiz de Queiroz”, Piracicaba, SP.

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, 2009. <https://doi.org/10.1198/jabes.2009.07098>

TESFAYE, K. et al. Evaluation of finger millet (*Eleusine coracana* (L.) Gaertn.) in multi-environment trials using enhanced statistical models. **PLOS ONE**, e0277499. v. 18, n. 2, p. 1–17, 2023. <https://doi.org/10.1371/journal.pone.0277499>

USDA. **World Agricultural Production**. 2023a. Disponível em: <<https://apps.fas.usda.gov/psdonline/circulars/oilseeds.pdf>>. Acesso em: 19 jan. 2023.

USDA. **World Agricultural Production**. 2023b. Disponível em: <<https://apps.fas.usda.gov/psdonline/circulars/production.pdf>>. Acesso em: 19 jan. 2023.

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, n. 2, p. 420–428, 1993. <https://doi.org/10.2134/agronj1993.00021962008500020046x>

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, n. 7, p. 1375–1392, 2017. <https://doi.org/10.1007/s00122-017-2894-4>

VERBYLA, A. P. et al. The Analysis of Designed Experiments and Longitudinal Data by Using Smoothing Splines. **Journal of the Royal Statistical Society. Series C (Applied Statistics)**, v. 48, n. 3, p. 269–311, 1999. <https://www.jstor.org/stable/2680826>

VERNETTI, F. J. et al. **Descrição botânica da soja**. Pelotas: EMBRAPA UEPAE de Pelotas, p. 15, 1979. (Circular Técnica, 7).

VOLPATO, L. et al. Multi-trait multi-environment models in the genetic selection of segregating soybean progeny. **PLOS ONE**, v. 14, n. 4, p. e0215315, 2019. <https://doi.org/10.1371/journal.pone.0215315>

WAGAW, K. et al. The MET Analysis of Yield Performance of Advanced Sorghum [*Sorghum Bicolor* (L.) Moench] Lines under Moisture Stress Areas Using Spatial Analysis. **American Journal of Plant Sciences**, v. 11, n. 10, p. 1618–1631, 2020. <https://doi.org/10.4236/ajps.2020.1110117>

WICKHAM, H. **ggplot2: Elegant Graphics for Data Analysis**. 2. ed. New York, NY: Springer New York, 2009.

XU, Z.; CANNON, S. B.; BEAVIS, W. D. Applying Spatial Statistical Analysis to Ordinal Data for Soybean Iron Deficiency Chlorosis. **Agronomy**, 2025. v. 12, n. 9, p. 1–28, 2022. <https://doi.org/10.3390/agronomy12092095>

YAU, S. K. Regression and AMMI Analyses of Genotype × Environment Interactions: An Empirical Comparison. **Agronomy Journal**, v. 87, n. 1, p. 121–126, 1995. <https://doi.org/10.2134/agronj1995.00021962008700010021x>

ZHAO, J. et al. Characterization of root architecture in an applied core collection for phosphorus efficiency of soybean germplasm. **Chinese Science Bulletin**, v. 49, n. 15, p. 1611–1620, 2004. <https://doi.org/10.1007/BF03184131>

ZHI, Y. et al. Screening of safe soybean cultivars for cadmium contaminated fields. **Scientific Reports**, 12965. v. 10, n. 1, p. 1–12, 2020. <https://doi.org/10.1038/s41598-020-69803-4>

APÊNDICE A – Tabela discriminação dos genótipos

Apêndice A. Discriminação dos genótipos avaliados nos 10 municípios no Mato Grosso do Sul, Brasil.

Genótipos	ID	Genótipos	ID
96R29-IPRO	G001	HO-TERERE-IPRO	G053
97R50-IPRO	G002	ICS3032-IPRO	G054
AS3680-IPRO	G003	ICS7019-RR	G055
AS3707i2x	G004	ICS-RODEO-RR	G056
AS3730-IPRO	G005	LG60159-IPRO	G057
BRASMAX-BONUS-IPRO	G006	LG-EX60006-IPRO	G058
BRASMAX-COLISEU-I2X	G007	M5947-IPRO	G059
BRASMAX-COMPACTA-IPRO	G008	M6100-XTD	G060
BRASMAX-DESAFIO-RR	G009	M6210-IPRO	G061
BRASMAX-FIBRA-IPRO	G010	M6410-IPRO	G062
BRASMAX-FOCO-IPRO	G011	M6430-XTD	G063
BRASMAX-GARRA-IPRO	G012	M6620-I2X	G064
BRASMAX-GUEPARDO-IPRO	G013	NEO660-IPRO	G065
BRASMAX-LENDARIA-CE	G014	NEO680-IPRO	G066
BRASMAX-LOTUS-IPRO	G015	NEO710-IPRO	G067
BRASMAX-NEXUS-I2X	G016	NEO740-IPRO	G068
BRASMAX-OLIMPO-IPRO	G017	NEO790-IPRO	G069
BRASMAX-POTENCIA-RR	G018	NK6356-IPRO	G070
BRASMAX-TANQUE-I2X	G019	NS5933-IPRO	G071
BRASMAX-TORMENTA-CE	G020	NS6010-IPRO	G072
BRASMAX-VORAZ-IPRO	G021	NS6220-IPRO	G073
BRB201576-IPRO	G022	NS6299-IPRO	G074
BRS1061-IPRO	G023	NS6433-I2X	G075
CSR1623-IPRO	G024	NS6446-I2X	G076
CSR1624-IPRO	G025	NS6700-IPRO	G077
CSR2904-IPRO	G026	NS7300-IPRO	G078
CSR390-RR	G027	POTENT-I2X	G079
CZ26B47-I2X	G028	PP-39JM-IPRO	G080
CZ36B96-I2X	G029	PP-50JM-IPRO	G081
CZ37B39-I2X	G030	PP-MS1107-RR	G082
CZ37B43-IPRO	G031	PP-MS1110-RR	G083
DM64I63-IPRO	G032	RESULT-I2X	G084
DM66i68-RSF-IPRO	G033	RK5519-RR	G085
DM70I71-IPRO	G034	RK6022-IPRO	G086
DM79i81-IPRO	G035	RK6316-IPRO	G087
EXP-WS02676/3-IPRO	G036	RK7518-IPRO	G088

EXP-WS02775/3-IPRO	G037	ST591-I2X	G089
FPS1867-IPRO	G038	ST621-I2X	G090
FTR158-RR	G039	ST622-IPRO	G091
FTR2660-IPRO	G040	ST631-I2X	G092
FTR3165-IPRO	G041	ST644-IPRO	G093
FTR3868-IPRO	G042	ST700-I2X	G094
FTR4262-IPRO	G043	ST783-IPRO	G095
GDMEXP-01	G044	ST794-I2X	G096
HO-AMAMBAY-IPRO	G045	ST797-IPRO	G097
HO-APORE-IPRO	G046	TEC7022-IPRO	G098
HO-CORUMBA-IPRO	G047	TMG2165-IPRO	G099
HO-IGUACU-IPRO	G048	TMG2360-IPRO	G100
HO-MARACAI-IPRO	G049	TMG2374-IPRO	G101
HO-PARAGUACU-I2X	G050	TMG71X57I2X	G102
HO-PIRAPO-IPRO	G051	TMG7362-IPRO-INOX	G103
HO-TAQUARI-IPRO	G052	TMG7368-IPRO	G104
