



RENAN SIMÕES CUNHA

**FENOTIPAGEM DIGITAL INTEGRADA A MODELOS MISTOS
PARA SELEÇÃO DE GENÓTIPOS EM TRIGO**

LAVRAS - MG

2026

RENAN SIMÕES CUNHA

**FENOTIPAGEM DIGITAL INTEGRADA A MODELOS MISTOS
PARA SELEÇÃO DE GENÓTIPOS EM TRIGO**

Projeto de dissertação apresentado à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-graduação em Genética e Melhoramento de Plantas, Mestrado Profissional, área de concentração em Genética e Melhoramento de Plantas, para a obtenção do título de Mestre.

Orientador: Prof. Dr. José Maria Villela Pádua

Orientador

Dra. Scheila Roberta Guilherme Zambiazzi

Coorientadora

Dr. Filipe Inácio Matias

Coorientador

LAVRAS - MG

2026

**Ficha Catalográfica elaborada pelo Sistema de Geração
de Ficha Catalográfica da Biblioteca Universitária da UFLA, com
dados informados pelo(a) próprio(a) autor(a).**

Cunha, Renan Simões.

Fenotipagem digital integrada a modelos mistos para seleção de genótipos em
trigo / Renan Simões Cunha. - 2026.
38 p.

Orientador: José Maria Villela Pádua

Coorientadora: Scheila Roberta Guilherme Zambiazzi

Coorientador: Filipe Inácio Matias

Dissertação (Mestrado Profissional) - Universidade Federal de Lavras, 2026.
Bibliografia.

1. Fenotipagem de alto rendimento. 2. Sensoriamento remoto por drones. 3.
Melhoramento genético de trigo. 4. Modelos mistos. 5. Fenótipos digitais. I. Pádua,
José Maria Villela. II. Zambiazzi, Scheila Roberta Guilherme. III. Matias, Filipe
Inácio. IV. Universidade Federal de Lavras. V. Título.

RENAN SIMÕES CUNHA

**FENOTIPAGEM DIGITAL INTEGRADA A MODELOS MISTOS
PARA SELEÇÃO DE GENÓTIPOS EM TRIGO**

**DIGITAL PHENOTYPING INTEGRATED WITH MIXED
MODELS FOR WHEAT GENOTYPE SELECTION**

Projeto de dissertação apresentado à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-graduação em Genética e Melhoramento de Plantas, Mestrado Profissional, área de concentração em Genética e Melhoramento de Plantas, para a obtenção do título de Mestre.

APROVADO em 28 de março de 2026.

Prof. Dr. José Maria Villela Pádua (UFLA)

Dra. Scheila Roberta Guilherme Zambiazzi (Syngenta)

Dr. Pedro Henrique Araújo Diniz Santos (TMG)

Prof. Dr. José Maria Villela Pádua
Coorientador

LAVRAS - MG

2026

AGRADECIMENTOS

Agradeço a Deus. À Ele por ter me dado a oportunidade de realizar tamanho feito, por ter me suportado com uma família maravilhosa que esteve sempre ao meu lado e por ter colocado pessoas incríveis que me deram inúmeros ensinamentos e muito suporte.

Gabriela, obrigado pelo seu amor, encorajamento e pelas doses diárias de resiliência. Você, Maria Julia e Maria Luiza são o meu mundo.

Um imenso obrigado aos meus orientadores José Mária Pádua, Scheila Zambiazzi, Filipe Matias e aos meus amigos Pedro Diniz, Filipe Bertagna e Tiago Corazza. Sem vocês nada disso seria possível.

Benhur, Gustavo, Milena, Gabriela e Giovana, agradeço por todos os dias de aulas compartilhados, por muitos apuros e muitas risadas. Guardo com carinho esse período, vocês são muito especiais!

Obrigado às instituições UFLA e CAPES por tornar possível essa conquista.

RESUMO

A fenotipagem de alto rendimento (High Throughput Phenotyping – HTP), baseada em imagens obtidas por veículos aéreos não tripulados (VANTs), tem se destacado como uma ferramenta promissora para complementar a fenotipagem convencional, permitindo a coleta rápida, precisa e não destrutiva de características relacionadas ao desenvolvimento das plantas. No melhoramento genético do trigo (*Triticum aestivum* L.), a integração entre fenótipos digitais e métodos estatísticos robustos pode aumentar a eficiência da seleção de genótipos superiores. Nesse contexto, o presente estudo teve como objetivo avaliar o potencial de uso de fenótipos digitais derivados de imagens multiespectrais obtidas por drones como variáveis auxiliares em métodos de seleção para produtividade de grãos em trigo, sob diferentes ajustes estatísticos. Foram utilizados dados públicos provenientes de um experimento conduzido com 106 linhagens do programa de melhoramento de trigo da North Dakota State University (NDSU), avaliadas em delineamento em blocos casualizados na safra de 2021. Durante o ciclo da cultura foram realizados 15 voos com drone equipado com sensor multiespectral, permitindo a obtenção de ortomosaicos RGB, multiespectrais e modelos de superfície digital. As análises estatísticas foram realizadas no ambiente R, incluindo estimativas de herdabilidade, correlações entre fenótipos digitais e características agrônômicas e comparação de diferentes métodos de seleção baseados em médias observadas, BLUE (Best Linear Unbiased Estimator) e BLUP (Best Linear Unbiased Prediction), com e sem inclusão de variáveis secundárias oriundas da fenotipagem digital. Os resultados evidenciaram altas herdabilidades para alguns fenótipos digitais, especialmente características associadas à altura estimada da planta e bandas individuais. As correlações entre os diferentes modelos de seleção variaram de 0,47 a 0,99, indicando concordância na classificação dos genótipos. Conclui-se que os modelos avaliados apresentaram alta consistência no ranqueamento das linhagens, sugerindo que abordagens estatísticas mais simples podem ser utilizadas sem perda relevante de informação na seleção de genótipos promissores.

Palavras-chave: Fenotipagem de alto rendimento; Sensoriamento remoto por drones; Melhoramento genético de trigo; Modelos mistos; Seleção genotípica; Fenótipos digitais.

ABSTRACT

High Throughput Phenotyping (HTP), based on images acquired by unmanned aerial vehicles (UAVs), has emerged as a promising tool to complement conventional phenotyping by enabling rapid, precise, and non-destructive data collection on plant development. In wheat (*Triticum aestivum* L.) breeding programs, the integration of digital phenotypes with robust statistical methods may increase the efficiency of selecting superior genotypes. In this context, the objective of this study was to evaluate the potential use of digital phenotypes derived from multispectral drone imagery as auxiliary variables in selection methods for grain yield in wheat under different statistical modeling approaches. Public data from a wheat breeding experiment conducted by the North Dakota State University (NDSU) breeding program were used. The dataset comprised 106 wheat lines evaluated in a randomized complete block design during the 2021 growing season. Throughout the crop cycle, 15 UAV flights equipped with a multispectral sensor were performed, generating RGB orthomosaics, multispectral images, and digital surface models. Statistical analyses were conducted in the R environment, including the estimation of heritability, correlations between digital phenotypes and agronomic traits, and the comparison of different selection methods based on observed means, Best Linear Unbiased Estimators (BLUE), and Best Linear Unbiased Predictions (BLUP), with and without the inclusion of secondary variables derived from digital phenotyping. The results revealed high heritability for some digital phenotypes, particularly traits associated with estimated plant height and individual bands. Correlations among the different selection models ranged from 0.47 to 0.99, indicating agreement in genotype ranking. Overall, the evaluated models showed high consistency in identifying superior lines, suggesting that simpler statistical approaches may be applied without substantial loss of information in genotype selection.

Keywords: High-throughput Phenotyping; UAV-based remote sensing; Wheat breeding; Mixed models; Genotypic selection; Digital phenotypes.

INDICADORES DE IMPACTOS

Os resultados desta pesquisa apresentam impactos potenciais relevantes nos âmbitos tecnológico, econômico e social, especialmente no contexto da agricultura moderna e dos programas de melhoramento genético de culturas. O estudo avaliou a integração entre dados de fenotipagem digital obtidos por veículos aéreos não tripulados (VANTs) e modelos estatísticos baseados em modelos mistos (BLUE e BLUP) para apoiar a seleção de genótipos superiores de trigo. Essa abordagem contribui para o avanço metodológico no melhoramento vegetal ao demonstrar a viabilidade de integrar dados fenômicos de alta resolução com ferramentas estatísticas robustas para otimização do processo de seleção genética. Do ponto de vista tecnológico, a pesquisa contribui para o desenvolvimento e validação de metodologias aplicáveis à agricultura digital e à fenotipagem de alto rendimento. A utilização de sensores multiespectrais embarcados em drones possibilita a coleta rápida e não destrutiva de grandes volumes de dados fenotípicos ao longo do ciclo da cultura. No estudo, foram analisados dados provenientes de 15 voos realizados durante o desenvolvimento da cultura, permitindo identificar características com elevada herdabilidade e potencial para uso em seleção indireta. Esses resultados indicam que variáveis derivadas de imagens podem auxiliar na identificação de genótipos superiores, aumentando a eficiência dos programas de melhoramento genético e reduzindo custos associados à fenotipagem tradicional. Em termos econômicos e sociais, a aplicação dessas metodologias pode contribuir para tornar os programas de melhoramento mais eficientes, reduzindo o tempo necessário para o desenvolvimento de novas cultivares e aumentando o ganho genético por ciclo de seleção. Esse avanço pode favorecer o aumento da produtividade agrícola e a competitividade do setor produtivo. Os resultados possuem potencial de aplicação em diferentes contextos produtivos, podendo beneficiar pesquisadores, melhoristas de plantas, empresas de sementes e produtores rurais. Os impactos desta pesquisa podem ser classificados principalmente nas áreas temáticas de Tecnologia e Produção e Educação, conforme a Política Nacional de Extensão, além de estarem alinhados aos Objetivos de Desenvolvimento Sustentável (ODS) da ONU, especialmente ODS 2 (Fome Zero e Agricultura Sustentável), ODS 9 (Indústria, Inovação e Infraestrutura), ODS 12 (Consumo e Produção Responsáveis) e ODS 13 (Ação Climática). De modo geral, os impactos são predominantemente potenciais,

pois fornecem bases metodológicas para o desenvolvimento de tecnologias mais eficientes no melhoramento genético de culturas agrícolas.

IMPACT INDICATORS

The results of this research present relevant potential impacts in the technological, economic, and social spheres, particularly within the context of modern agriculture and plant breeding programs. The study evaluated the integration of digital phenotyping data obtained from unmanned aerial vehicles (UAVs) with statistical approaches based on mixed models (BLUE and BLUP) to support the selection of superior wheat genotypes. This approach contributes to methodological advances in plant breeding by demonstrating the feasibility of integrating high-resolution phenomic data with robust statistical tools to optimize the genetic selection process. From a technological perspective, the research contributes to the development and validation of methodologies applicable to digital agriculture and high-throughput phenotyping. The use of multispectral sensors mounted on drones enables rapid, non-destructive collection of large volumes of phenotypic data throughout the crop cycle. In this study, data obtained from 15 flights conducted during crop development were analyzed, allowing the identification of traits with high heritability and potential use for indirect selection. These results indicate that variables derived from image analysis can assist in identifying superior genotypes, increasing the efficiency of breeding programs and reducing costs associated with traditional phenotyping. In economic and social terms, the application of these methodologies can contribute to making breeding programs more efficient by reducing the time required for the development of new cultivars and increasing genetic gain per selection cycle. This advancement may favor increased agricultural productivity and strengthen the competitiveness of the agricultural sector. The results have potential applicability in different production contexts and may benefit researchers, plant breeders, seed companies, and farmers. The impacts of this research can be classified mainly within the thematic areas of Technology and Production and Education, according to the Brazilian National Extension Policy, and are also aligned with the United Nations Sustainable Development Goals (SDGs), particularly SDG 2 (Zero Hunger and Sustainable Agriculture), SDG 9 (Industry, Innovation and Infrastructure), SDG 12 (Responsible Consumption and Production), and SDG 13 (Climate Action). Overall, the impacts are predominantly potential, as the study provides methodological foundations for the development of more efficient technologies in the genetic improvement of agricultural crops.

SUMÁRIO

1	INTRODUÇÃO.....	10
2	REFERENCIAL TEÓRICO	11
2.1	Importância da cultura do trigo	11
2.2	Melhoramento de trigo	12
2.3	Fenotipagem de alto rendimento.....	13
2.4	Modelos mistos e definição de BLUE e BLUP	14
2.5	Seleção indireta.....	16
2.6	AIC e BIC.....	17
3	MATERIAL E MÉTODOS	18
3.1	Detalhamento da coleta de dados fenotípicos	18
3.2	Análises estatísticas.....	19
3.3	Ajuste dos modelos ou métodos de seleção	20
3.4	Ordenamento (ranking) e coincidência de seleção.....	22
3.5	Cálculo do ganho de seleção.....	22
4	RESULTADOS E DISCUSSÃO.....	23
5	CONCLUSÕES.....	30

1 INTRODUÇÃO

O trigo (*Triticum aestivum* L.) é uma das culturas mais importantes para a humanidade, tendo sido domesticado há mais de dez mil anos na região da Mesopotâmia e desempenhado papel central na transição das sociedades nômades para sistemas agrícolas sedentários (PIANA; CARVALHO, 2008). Atualmente, o trigo comum representa cerca de 90% da produção mundial, enquanto o trigo duro (*T. turgidum* var. *durum*) ocupa a maior parte da fração restante. Além de constituir a base alimentar de grande parcela da população global, contribuindo com aproximadamente 20% da ingestão calórica e proteica humana, o trigo também possui relevância estratégica no comércio internacional e na segurança alimentar (HAWKESFORD et al., 2013; SHEWRY, 2009). Sua ampla adaptação, cultivado desde altas latitudes do hemisfério norte até regiões do hemisfério sul, reflete a notável plasticidade genômica associada à sua natureza alopoliploide (DUBCOVSKY; DVORAK, 2007; FELDMAN, 1995).

O sucesso dessa cultura ao longo da história está intimamente relacionado ao melhoramento genético, entendido como a aplicação de princípios genéticos para o desenvolvimento de cultivares superiores (ALLARD, 2023). O trigo hexaploide ($2n=6x=42$; AABBDD) originou-se da hibridização entre espécies previamente cultivadas, o que contribuiu para sua elevada variabilidade e adaptabilidade (MARCUSSEN et al., 2014; PETERSEN et al., 2006; SHEWRY, 2009). Em programas de melhoramento, a eficiência da seleção depende tanto da variabilidade genética disponível quanto da qualidade das avaliações fenotípicas e das metodologias estatísticas empregadas (BERNARDO, 2020). Nesse contexto, decisões relacionadas ao número de genótipos avaliados, ambientes de teste e critérios de seleção são determinantes para o ganho genético ao longo dos ciclos (PEREIRA, 2018).

A fenotipagem de alto rendimento (High Throughput Phenotyping – HTP) surge como uma ferramenta estratégica para ampliar a eficiência desses programas. Por meio do uso de sensores multiespectrais, câmeras RGB e veículos aéreos não tripulados, é possível coletar grande volume de dados não destrutivos em diferentes estádios fenológicos, capturando características complexas com rapidez e precisão (COBB et al., 2013; SHAKOOR et al., 2017; MATIAS et al., 2022). Essa abordagem permite aumentar a escala de avaliação e reduzir custos operacionais, além de viabilizar análises mais robustas em múltiplos ambientes, aspecto essencial para

culturas amplamente cultivadas como o trigo (YANG et al., 2020; SILVEIRA et al., 2020). Assim, a integração entre fenômica, inteligência artificial e melhoramento tende a acelerar o desenvolvimento de cultivares mais produtivas e adaptadas às mudanças climáticas (MOLIN et al., 2015).

Entretanto, o potencial da HTP só é plenamente explorado quando aliado a métodos estatísticos capazes de modelar adequadamente a variabilidade genética e ambiental. Os modelos lineares mistos representam uma das principais ferramentas para análise de dados em experimentação agrícola, permitindo a estimação de efeitos fixos (BLUE) e a predição de efeitos aleatórios (BLUP) com maior acurácia (HENDERSON, 1975; THOMPSON, 2008; PIEPHO et al., 2008). A utilização de procedimentos de máxima verossimilhança contribui para estimativas menos viesadas e maior precisão na predição do mérito genético (ROCHA, 2025; RESENDE, 2016). Dessa forma, os modelos mistos possibilitam melhor aproveitamento das informações disponíveis, mesmo em experimentos desbalanceados ou conduzidos em múltiplos ambientes (NUNES; RAMALHO; FERREIRA, 2008).

Adicionalmente, a integração entre HTP e modelos mistos amplia as oportunidades para seleção indireta, especialmente para caracteres complexos e de baixa herdabilidade, como produtividade de grãos (RAMALHO et al., 2012; NOGUEIRA et al., 2012). A eficiência dessa estratégia depende da magnitude das correlações genéticas entre caracteres secundários e primários, bem como da herdabilidade relativa entre eles (CRUZ; REGAZZI; CARNEIRO, 2014; NETO et al., 2018). Assim, a utilização de variáveis derivadas de imagens como caracteres auxiliares, associadas à predição via BLUP, pode aumentar a acurácia seletiva e acelerar o ganho genético em trigo, consolidando uma abordagem integrada entre fenotipagem digital e modelagem estatística no contexto do melhoramento moderno.

Neste sentido, este trabalho tem como objetivo avaliar como o uso de dados oriundos de imagens de drones influenciam no ganho genético de trigo sob diferentes modelos estatísticos.

2 REFERENCIAL TEÓRICO

2.1 Importância da cultura do trigo

Originário do sudoeste da Ásia, na região da Mesopotâmia, estima-se que o trigo tenha sido domesticado há mais de dez mil anos onde desempenhou um papel fundamental no desenvolvimento das primeiras civilizações. Ele foi crucial na transição da humanidade de sociedades nômades para sociedades agrícolas sedentárias. Atualmente, existem várias espécies de trigo, tanto silvestres quanto cultivadas, mas apenas duas têm grande importância econômica: o trigo comum ou trigo para pão (*Triticum aestivum* sin. *T. vulgare*) e o trigo duro ou trigo para macarrão (*T. turgidum* var. *durum*). O trigo comum representa cerca de 90% de todo o trigo cultivado no mundo, e o restante é quase todo ocupado pelo trigo duro (PIANA; CARVALHO, 2008).

Este cereal é a base da dieta de uma parte expressiva da população, sendo uma fonte importante de energia (cerca de 20% da demanda populacional) e proteína (também próximo de 20%), além de vitaminas e outros compostos benéficos. Ele é amplamente utilizado também para alimentação animal (HAWKESFORD *et al.*, 2013; SHEWRY, 2009).

A produção e o comércio de trigo são e continuam sendo essenciais para a economia global. Ele é cultivado em muitos países e tem um papel significativo no comércio internacional de alimentos. O trigo é a espécie cultivada que ocupa a maior parte da área agrícola do planeta, sendo cultivado do 67° Norte, nos países Escandinavos e Rússia, ao 45° Sul, na Argentina, abrangendo regiões de altitudes variadas, embora seja menos cultivado em regiões tropicais (FELDMAN, 1995).

2.2 Melhoramento de trigo

O melhoramento vegetal consiste na aplicação de princípios genéticos para o desenvolvimento de plantas que tenham maior utilidade para as necessidades humanas. Isso é obtido através da escolha de plantas consideradas vantajosas do ponto de vista do melhorista, que leva em consideração valores econômicos, ambientais e/ou estéticos, começando pelo cruzamento de indivíduos selecionados e, posteriormente, selecionando seus melhores descendentes (ALLARD, 2023).

O trigo é uma espécie hexaploide ($2n=6x=42$, genoma AABBDD), resultante da junção de 3 inter-relacionados genomas diploides (EL BAIDOURI *et al.*, 2017; MARCUSSEN *et al.*, 2014; PETERSEN *et al.*, 2006; PONT; SALSE, 2017).

Diferentemente de outras espécies cultivadas, o trigo hexaploide não foi selecionado a partir de espécies silvestres, mas surgiu a partir da hibridização entre espécies já cultivadas pelo homem (*T. turgidum* e *Ae. tauschii*), assim, é possível dizer que jamais existiu *T. aestivum* na forma silvestre (SHEWRY, 2009).

As razões pelas quais este cereal foi tão amplamente adotado pelo homem incluem a elevada adaptabilidade, dado possivelmente pela sua natureza alopoliploide que tem lhe conferido a chamada “plasticidade genômica”, as suas qualidades como alimento (fonte de carboidratos, proteínas, vitaminas), com especial destaque à singular propriedade elástica do seu glúten, que tem dado a sua farinha os mais diversos usos desde os primórdios da humanidade e ainda outras, como a possibilidade de armazenar os seus grãos, sob determinadas condições, por tempo indefinido (DUBCOVSKY; DVORAK, 2007; SHEWRY, 2009).

2.3 Fenotipagem de alto rendimento

O sucesso dos programas de melhoramento genético, no Brasil e no mundo, resulta da implementação de estratégias bem elaboradas e executadas. Isso é particularmente evidente quando consideramos as decisões relacionadas aos métodos de condução, número de progênies/linhagens avaliadas e as seleções das melhores linhagens para avanços (PEREIRA, 2018). A etapa de seleção de linhagens promissoras se apoiará na assertividade da coleta dos dados obtidos nos experimentos e suas metodologias de análise (BERNARDO, 2020).

Nesse contexto, torna-se fácil compreender que o melhoramento genético requer a atenta observação de diversas características das plantas durante todo seu ciclo de desenvolvimento. Em muitos casos, essa coleta de dados demanda a participação de uma grande equipe para avaliar essas características de maneira consistente e precisa. Atualmente, existem formas eficientes e econômicas para coleta desses dados, permitindo a obtenção de maior quantidade e qualidade das informações pertinentes para o processo de seleção. A implementação de novas tecnologias e abordagens como a fenômica, possibilitam ao melhorista obter dados assertivos para características sensíveis à cultura, como por exemplo, a maturidade fisiológica. Para complementar o trabalho de fenotipagem visual do avaliador, veículos aéreos não tripulados equipados com sensores são cada vez mais utilizados para

coletar grandes quantidades de dados multidimensionais para capturar características fenotípicas das plantas (MATIAS *et al.*, 2022).

High Throughput Phenotyping (HTP), em português “Fenotipagem de Alto Rendimento”, é uma alternativa tecnológica que vem revolucionando o melhoramento de plantas visto sua capacidade veloz de avaliação e precisão dos dados em larga escala. A HTP utiliza de imagens multiespectrais por sensores avançados e uma alta capacidade de análise de dados para capturar características fenotípicas complexas em grandes populações de plantas, o que demonstra o avanço das ferramentas de automação de fenotipagem para avaliações não destrutivas e em larga escala aplicáveis a diversas culturas agrônomicas (SHAKOOR *et al.*, 2017; SILVEIRA *et al.*, 2020).

Uma das principais vantagens de se utilizar a HTP é justamente na viabilidade de acelerar o processo de seleção de genótipos, visto sua capacidade de fornecer uma gama vasta de avaliações em um curto espaço de tempo. Essa capacidade permite aumentar a eficiência do trabalho do melhorista, viabilizando um potencial aumento no número de genótipos testados ou mesmo reduzindo custos excessivos com genótipos mal avaliados, algo valioso para seleção de características complexas como tolerância a seca e eficiência no uso de nutrientes. Além disso, o melhorista se vê mais suportado para aumentar o número de locais avaliados devido a maior precisão e velocidade de avaliação da HTP, o que contribuirá para seleção de variedades mais adaptáveis a diferentes locais (COBB *et al.*, 2013; YANG *et al.*, 2020).

O futuro que se propõe unindo a HTP e o melhoramento de trigo deverá ser promissor, visto o desenvolvimento contínuo de novas tecnologias e métodos de análise de dados. Soma-se ainda o crescimento da inteligência artificial e *machine learning*, ambos temas com crescente aumento de pesquisas para interpretação de dados fenotípicos e previsões mais precisas de desempenho de culturas. É possível esperar que a HTP obtenha um papel relevante na abordagem de desafios globais, como a adaptação de culturas às mudanças climáticas, através do auxílio no desenvolvimento mais rápido e assertivo de variedades de trigo cada vez mais produtivas e mais resistentes à estresses bióticos e abióticos (MOLIN *et al.*, 2015).

2.4 Modelos mistos e definição de BLUE e BLUP

No que tange à avaliação dos dados obtidos em programas de melhoramento de plantas, os modelos mistos são uma ferramenta fundamental que permite a análise de dados complexos com estruturas de variância-covariância heterogêneas. Em 1930, os modelos mistos já eram utilizados na experimentação agrícola afim de corrigir delineamentos experimentais desequilibrados, mas obteve destaque em 1950 quando aplicado os modelos para avaliar desempenho genético de touros com base na produção de leite de duas filhas (THOMPSON, 1976).

A utilização desses modelos adiciona efeitos aleatórios aos fixos, onde permitem uma abordagem mais flexível para lidar com variabilidade genética e ambiental inerente aos experimentos de melhoramento. A acurácia de seleção depende do uso de métodos estatísticos que retirem o máximo de proveito das informações disponíveis, uma vez alcançado por meio de uma experimentação adequada (NUNES; RAMALHO; FERREIRA, 2008).

Uma vez avançada a acessibilidade a tecnologia e aprimorada as ferramentas estatísticas, procedimentos de análises possibilitaram ajustar e aplicar diferentes metodologias com ganho prático de velocidade e garantindo assertividade nos resultados. A aplicação da metodologia de máxima verossimilhança na estimativa dos modelos foi uma delas, corrigindo estimativas enviesadas obtidas pelo método de mínimos quadrados (ROCHA, 2025). Dessa forma, com tal metodologia obtêm-se o BLUE (Best Linear Unbiased Estimator), para os efeitos fixos, e BLUP (Best Linear Unbiased Prediction) para efeitos aleatórios do ajuste do modelo (THOMPSON, 2008).

O BLUP é aplicado para predizer efeitos aleatórios, como valores genéticos aditivos, de dominância e genotípicos a depender da situação (HENDERSON, 1975). Se considerar a estrutura de parentesco entre indivíduos e a herdabilidade de caracteres, o BLUP pode ser uma ferramenta valiosa para o melhorista permitindo a predição de efeitos genéticos aditivos e possibilitando a correção de padrões de fluxo de informação genética de geração em geração, mesmo na existência de desequilíbrio causado a partir de seleções. Assim, possibilitando o melhorista obter predições mais acuradas do mérito genético o que, conseqüentemente, será relevante para seleção de genótipos (PIEPHO *et al.*, 2008; ISIK; HOLLAND; MALTECCA, 2017; RESENDE, 2016).

2.5 Seleção indireta

A seleção indireta é uma estratégia fundamental em um programa de melhoramento de plantas. Isso se faz necessário devido à complexidade de muitos caracteres de interesse do melhorista estarem relacionados a heranças quantitativas (NOGUEIRA *et al.*, 2012).

A estratégia de se selecionar muitos caracteres muitas vezes se torna inviável devido às diferentes correlações entre elas, o que pode ser benéfico ou não ao interesse do melhorista. Dessa forma, a seleção indireta se torna uma alternativa valiosa quando a característica de interesse é difícil de medir, tem baixa herdabilidade, ou é expressa tardiamente no ciclo de vida da planta (RAMALHO *et al.*, 2012).

A eficácia da seleção indireta depende diretamente da correlação genética entre o caráter secundário e o caráter primário. Uma correlação genética forte e positiva indica que a seleção para o caráter secundário resultará em ganhos genéticos no caráter primário. Assim, se faz necessário a herdabilidade do caráter secundário ser maior do que o caráter primário para que a seleção indireta seja vantajosa (CRUZ; REGAZZI; CARNEIRO, 2014).

Em programas de melhoramento genético, a seleção indireta é amplamente utilizada para caracteres complexos, como produtividade de grãos, peso de mil grãos e número de grãos por espiga, com o objetivo de maximizar o ganho em produtividade total. Nesse contexto, é fundamental que o melhorista compreenda as relações entre os caracteres, uma vez que as correlações influenciam diretamente a eficiência da seleção.

As correlações entre caracteres podem ser classificadas em fenotípicas, genotípicas e ambientais. A correlação fenotípica é estimada a partir das observações dos caracteres e resulta da combinação dos efeitos genéticos e ambientais. Já a correlação genotípica expressa a associação entre os valores genéticos dos caracteres, sendo independente do ambiente. Essa correlação pode ser decorrente de pleiotropia, quando um mesmo gene afeta simultaneamente dois ou mais caracteres, ou de ligação gênica, quando genes distintos que controlam diferentes caracteres estão fisicamente próximos no cromossomo e tendem a ser herdados em conjunto (NETO *et al.*, 2018; CRUZ; REGAZZI; CARNEIRO, 2014). Por sua vez, a correlação ambiental refere-se à associação entre caracteres causada

exclusivamente por fatores ambientais, refletindo a influência conjunta das condições de cultivo sobre múltiplas variáveis fenotípicas.

Com o avanço de tecnologias de fenotipagem de alto rendimento, novas oportunidades para seleção indireta têm surgido. Características secundárias podem ser medidas com maior rapidez e precisão, auxiliando o melhorista no processo de seleção de linhagens de interesse em um programa de melhoramento.

2.6 AIC e BIC

Os critérios de informação são ferramentas estatísticas relevantes na seleção de modelos no melhoramento de plantas. Com o objetivo de selecionar modelos, os critérios ajudam a encontrar um equilíbrio entre o quão bem um modelo se ajusta aos dados e o quão complexo ele é, permitindo que os melhoristas selecionem a opção mais simples e eficiente para seus objetivos. São frequentemente utilizados critérios de informação como o Critério de Informação de Akaike (AIC), o Critério de Informação de Akaike corrigido (AICc) e o Critério de Informação Bayesiano (BIC). Sendo a seleção do modelo baseada no que possuir o menor valor de AIC, AICc ou BIC (EMILIANO, 2013).

O Critério de Informação de Akaike (AIC) foi desenvolvido e denominado por Hirotugu Akaike em 1974 como “um critério de informação” (AKAIKE, 1974). Dessa forma, o AIC é fundamentado no conceito de informação e estima a quantidade relativa de informação perdida por um determinado modelo. No contexto do melhoramento de plantas, este critério pode ser utilizado para comparar modelos concorrentes, como diferentes estruturas de variância-covariância em análises de múltiplos ambientes ou na modelagem de interações genótipo-ambiente (EMILIANO, 2013). Além deste critério, uma relevante derivação foi o Critério de Akaike corrigido (AICc), que surge na intenção de corrigir desempenhos ruins do AIC quando existirem muitos parâmetros em comparação com o tamanho da amostra (SUGIURA, 1978).

O Critério de Informação Bayesiano (BIC), foi proposto por Schwarz (1978) e incorpora uma penalidade mais forte para a complexidade do modelo em comparação ao AIC. Também chamado de Critério de Schwarz, ele consiste na avaliação dos modelos definidos em termos de probabilidade a posteriori, no qual foi chamado dessa forma pois o autor utilizou um argumento bayesiano para prová-lo (EMILIANO, 2013).

Assim, para um programa de melhoramento de plantas, o melhorista deverá escolher entre AIC e BIC a depender do objetivo de sua análise. Sendo AIC um critério de maior auxílio a modelos mais complexos e utilizado em modelos de predição, enquanto o BIC tende a ser mais propício para modelos mais simples e usado com objetivo de identificar o melhor modelo gerador de dados.

3 MATERIAL E MÉTODOS

Os genótipos e os detalhes do experimento de campo são descritos por Matias *et al.* (2021) e estão sumarizados a seguir. O experimento foi conduzido durante a safra de trigo de 2021, na fazenda experimental da North Dakota State University (NDSU), localizada em Casselton, Dakota do Norte, EUA (46.88051067579893, -97.2370382539691).

Foram utilizados dados públicos provenientes do programa de melhoramento de trigo da NDSU, compostos por 106 linhagens em fase de teste de produtividade no segundo ano de avaliação. O delineamento experimental adotado foi em blocos casualizados (DBC), conforme descrito por Matias *et al.* (2021). Todas as linhagens foram avaliadas com pelo menos duas repetições, enquanto algumas cultivares testemunhas apresentaram até três repetições, como ALPINE, BARLOW, ELGIN-ND, FALLER, GLENN, SYINGMAR, NDSW0932 e NDSW14098.

As parcelas experimentais e a condução do ensaio seguiram os protocolos estabelecidos pelo programa de melhoramento da instituição, assegurando condições adequadas para avaliação do desempenho agrônomo das linhagens em teste de rendimento. Esse conjunto de dados foi utilizado como base para o desenvolvimento e aplicação das metodologias de análise, incluindo abordagens fundamentadas em modelos mistos e integração com dados de fenotipagem.

3.1 Detalhamento da coleta de dados fenotípicos

Os conjuntos de dados dos drones foram obtidos por um equipamento DJI Inspire 2 com um sensor multiespectral Sentera 6x. Ao longo do experimento, foram obtidos um conjunto de dados de 15 voos realizados por volta do meio-dia, em dias sem nuvens, nas datas de: (1) 11/maio, (2) 23/maio, (3) 26/maio, (4) 01/junho, (5)

07/junho, (6) 10/junho, (7) 17/junho, (8) 22/junho, (9) 25/junho, (10) 28/junho, (11) 01/julho, (12) 13/julho, (13) 19/julho, (14) 20/julho, e (15) 26/julho. Os parâmetros de voos foram definidos para 75% de sobreposição, 15 mph e altura de voo de 200 pés. Como resultado dos voos, foram obtidos 15 ortomosaicos RGB (Red, Green and Blue), 15 ortomosaicos multiespectrais e 15 ortomosaicos DSM (Modelos de Superfície Digital).

Com os dados gerados por imagens de drones, foram obtidos os seguintes caracteres:

- Bandas individuais: "Blue"; "Green"; "Red"; "RedEdge"; "NIR";
- Índices vegetativos: "NGRDI"; "BGI"; "GLI"; "NDVI"; "NDRE"; "CIG"; "CIRE";
- Cobertura de dossel: "Canopy"
- Altura estimada da planta, onde os valores de 0 (mínimo) a 100 (máximo) referem-se aos quantis na distribuição de dados DSM para altura do pixel: "Height_0"; "Height_10"; "Height_25"; "Height_50"; "Height_75"; "Height_90"; "Height_100"

Além dos citados, para auxiliar nos estudos gerados por Matias *et al.* (2021) foram avaliados os seguintes caracteres agrônômicos: Dia de maturação (MAT_DAY) no qual se considerou a maturação quanto 50% dos pedúnculos das plantas na parcela estão amarelos. Dias para o florescimento (DH) onde se considerou o DAP (dias após plantio) em que a parcela apresentou 50% das plantas com espiga exposta. Altura da planta (HT) e Acamamento (LODG) levaram em consideração a coleta de imagens da superfície (DSM) do solo antes da emergência das plantas e ao longo do ciclo de crescimento para cálculo da estimativa de altura de planta e dossel e, por fim, a Produtividade das parcelas (YLD).

3.2 Análises estatísticas

Neste estudo, foi utilizado o software R (R Core Team, 2026) para desenvolver as análises estatísticas necessárias a fim de avaliar a influência de imagens de drones no ganho genético de trigo.

Os passos seguidos neste estudo a fim de realizar as análises estatísticas foram: (0) Avaliação dos fenótipos digitais através do cálculo de herdabilidade e

correlação entre fenótipos digitais e características agronômicas (MATIAS *et al.* 2021); (1) Ajuste dos modelos ou métodos de seleção em relação a característica agronômica de produtividade; (2) Ordenamento (ranking) e coincidência de seleção; (3) Cálculo do ganho de seleção; sendo detalhados conforme os pontos a seguir:

Na primeira etapa de análise, serão calculadas as herdabilidades e correlação entre as imagens obtidas e características agronômicas. Os caracteres utilizados para as análises foram apresentados no estudo de Matias *et al.* (2021): Bandas Individuais, Índices Vegetativos, Cobertura de Dossel e Altura Estimada de Planta. Estes caracteres serão correlacionados aos dias após o plantio (DAP) em que foram realizados os voos de drones.

Assim, para se realizar essa avaliação foram utilizados os modelos estatísticos:

I. Herdabilidade das características agronômicas:

$$y_{ij} = \mu + g_i + e_j$$

II. Herdabilidade das características fenômicas:

$$y_{ij} = \mu + g_i + e_j$$

III. Correlação entre agronômicas e fenômicas:

$$y_{ij} = \mu + g_i + e_j$$

Onde:

- y_{ij} : valor fenotípico observado;
- μ : constante associada a todas as observações;
- g_i : efeito aleatório do genótipo;
- e_j : erro experimental aleatório com $e_j \sim N(0, \sigma_e^2)$.

3.3 Ajuste dos modelos ou métodos de seleção

Neste ponto foram avaliados os métodos de seleção levando em consideração a característica agronômica de produtividade (YLD) no estudo realizado por Matias *et al.* (2021). Os métodos e modelos estatísticos para essa avaliação, foram:

I. Média Geral:

$$y_{ij} = \mu$$

II. BLUE com efeito de g_i fixo:

- $y_{ij} = \mu + g_i + r_j + \varepsilon_{ij}$
- III. BLUP com efeito de g_i aleatório:
 - $y_{ij} = \mu + g_i + r_j + \varepsilon_{ij}$
- IV. BLUE com efeito de g_i fixo e uma variável secundária (único voo com maior h^2)
 - $y_{ij} = \mu + g_i + \beta_1 p_{1ij} + r_j + \varepsilon_{ij}$
- V. BLUP com efeito de g_i aleatório e uma variável secundária (único voo com maior h^2)
 - $y_{ij} = \mu + g_i + \beta_1 p_{1ij} + r_j + \varepsilon_{ij}$
- VI. BLUE com efeito de g_i fixo e uma variável secundária (voo com segunda maior h^2)
 - $y_{ij} = \mu + g_i + \beta_2 p_{2ij} + r_j + \varepsilon_{ij}$
- VII. BLUP com efeito de g_i aleatório e uma variável secundária (único voo com segunda maior h^2)
 - $y_{ij} = \mu + g_i + \beta_2 p_{2ij} + r_j + \varepsilon_{ij}$
- VIII. BLUP com efeito de g_i aleatório e duas variáveis secundárias (único voo com as duas maiores h^2)
 - $y_{ij} = \mu + g_i + \beta_1 p_{1ij} + \beta_2 p_{2ij} + r_j + \varepsilon_{ij}$
- IX. BLUE com efeito de g_i fixo e duas variáveis secundárias (único voo com as duas maiores h^2)
 - $y_{ij} = \mu + g_i + \beta_1 p_{1ij} + \beta_2 p_{2ij} + r_j + \varepsilon_{ij}$

Onde:

- y_{ij} : valor observado de YLD do genótipo i na repetição j
- μ : média geral
- g_i : efeito fixo do genótipo i
- p_{1ij} : covariável 1
- β_1 : coeficiente de regressão associado à covariável p_1
- p_{2ij} : covariável 2
- β_2 : coeficiente de regressão associado à covariável p_2
- $r_j \sim N(0, \sigma_r^2)$: efeito aleatório da repetição j
- $\varepsilon_{ij} \sim N(0, \sigma_e^2)$: erro residual.

Para fins de auxílio no texto a ser apresentado nas figuras de análise, foram aplicadas as legendas para referenciar os modelos conforme a apresenta a tabela 1.

Tabela 1 – Legenda dos modelos analisados

Modelo	Legenda
Média Geral	RANK_YLD
BLUE com efeito de g_i fixo	RANK_YLD_BLUE_mod_2
BLUP com efeito de g_i aleatório	RANK_YLD_BLUP_mod_3
BLUE com efeito de g_i fixo e uma variável secundária (único voo com maior h^2)	RANK_YLD_BLUE_mod_4
BLUP com efeito de g_i aleatório e uma variável secundária (único voo com maior h^2)	RANK_YLD_BLUP_mod_5
BLUE com efeito de g_i fixo e uma variável secundária (voo com segunda maior h^2)	RANK_YLD_BLUE_mod_6
BLUP com efeito de g_i aleatório e uma variável secundária (único voo com segunda maior h^2)	RANK_YLD_BLUP_mod_7
BLUP com efeito de g_i aleatório e duas variável secundárias (único voo com as duas maiores h^2)	RANK_YLD_BLUP_mod_8
BLUE com efeito de g_i fixo e duas variável secundária (único voo com as duas maiores h^2)	RANK_YLD_BLUE_mod_9

Fonte: elaboração própria

3.4 Ordenamento (ranking) e coincidência de seleção

Após os ajustes dos modelos no ponto (1), foi realizada a ordenação (ranking) de seleção entre as metodologias, de forma a avaliar a comparação da coincidência de seleção entre os métodos. Os modelos foram comparados pela correlação de Spearman, entre eles.

3.5 Cálculo do ganho de genético

Com a informação dos 20% de indivíduos superiores selecionados em cada metodologia, foi calculado o ganho de genético pela equação abaixo:

$$\Delta G = h^2 \times S$$

Onde:

- ΔG : Ganho de seleção;
- h^2 : Herdabilidade;
- S: Diferencial de seleção.

4 RESULTADOS E DISCUSSÃO

Foram obtidas as análises de herdabilidade dos índices dentro de cada época para avaliação das duas características com maiores valores a fim de se continuar os estudos propostos. Como demonstrado na Figura 1, as maiores herdabilidades se encontram nos caracteres de Altura Estimada da Planta “Height_25” (Quantil de 25%), com herdabilidade de 0,796 na avaliação com 73 dias após o plantio (DAP) e no caractere de Banda Individual “RedEdge”, com herdabilidade de 0,778 na avaliação com 74 dias após o plantio (DAP). Para fins deste estudo, optou-se pelos cálculos com DAP 73, visto a maior herdabilidade estar disposta nessa avaliação.

Figura 1 – Herdabilidade dos caracteres avaliados por data após plantio

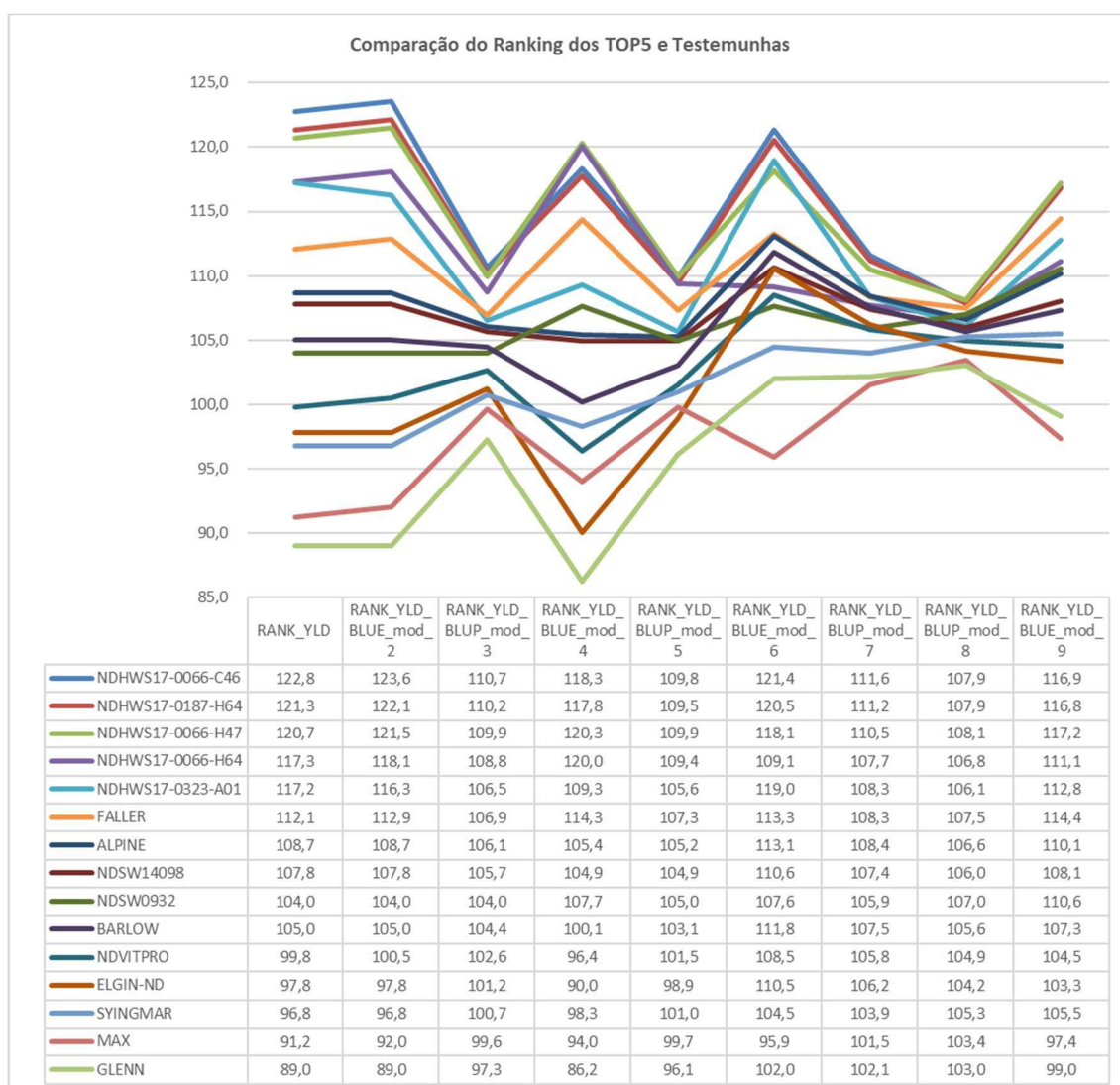
Trait	DAP													
	16	19	25	31	34	41	46	49	52	55	67	73	74	80
BGI	0,459	0,620	0,549	0,298	0,350	0,420	0,694	0,643	0,599	0,608	0,619	0,622	0,643	0,714
Blue	0,052	0,016	0,275	0,278	0,185	0,087	0,287	0,253	0,200	0,237	0,596	0,664	0,753	0,707
Canopy	0,645	0,672	0,435	0,172	0,203	0,417	0,637	0,236	0,320	0,397	0,396	0,512	0,489	0,679
CIG	0,386	0,523	0,498	0,304	0,367	0,430	0,658	0,611	0,571	0,518	0,665	0,711	0,735	0,752
CIRE	0,464	0,549	0,509	0,282	0,351	0,442	0,681	0,669	0,580	0,517	0,680	0,719	0,745	0,739
GLI	0,598	0,649	0,567	0,338	0,410	0,448	0,676	0,643	0,616	0,630	0,595	0,598	0,612	0,714
Green	0,006	0,000	0,000	0,467	0,253	0,150	0,254	0,343	0,388	0,409	0,581	0,666	0,768	0,682
Height_0	0,153	0,300	0,298	0,086	0,113	0,059	0,066	0,133	0,000	0,007	0,026	0,114	0,575	0,570
Height_10	0,137	0,257	0,211	0,068	0,110	0,254	0,405	0,330	0,521	0,552	0,071	0,671	0,613	0,641
Height_100	0,120	0,177	0,053	0,094	0,150	0,340	0,426	0,661	0,504	0,639	0,126	0,681	0,567	0,549
Height_25	0,131	0,255	0,203	0,071	0,123	0,296	0,471	0,423	0,569	0,601	0,246	0,796	0,678	0,673
Height_50	0,125	0,229	0,172	0,074	0,116	0,318	0,487	0,531	0,593	0,590	0,330	0,738	0,696	0,556
Height_75	0,114	0,211	0,155	0,072	0,114	0,426	0,549	0,639	0,608	0,630	0,281	0,705	0,677	0,550
Height_90	0,115	0,198	0,127	0,079	0,149	0,410	0,500	0,660	0,584	0,620	0,194	0,673	0,673	0,565
NDRE	0,446	0,531	0,487	0,244	0,329	0,392	0,655	0,643	0,445	0,448	0,625	0,677	0,711	0,734
NDVI	0,387	0,499	0,452	0,232	0,352	0,349	0,594	0,405	0,365	0,419	0,577	0,596	0,593	0,734
NGRDI	0,431	0,542	0,547	0,351	0,487	0,531	0,675	0,677	0,734	0,761	0,761	0,585	0,652	0,743
NIR	0,055	0,120	0,420	0,352	0,431	0,439	0,669	0,427	0,678	0,688	0,669	0,618	0,645	0,605
Red	0,054	0,019	0,204	0,313	0,234	0,160	0,212	0,304	0,281	0,302	0,485	0,567	0,727	0,694
RedEdge	0,000	0,000	0,074	0,413	0,269	0,158	0,375	0,325	0,435	0,465	0,633	0,711	0,778	0,657

Fonte: elaboração própria

Após a definição dos caracteres e a aplicação dos modelos, podemos observar o comparativo do ordenamento obtido pelos modelos pela Figura 2, que apresenta as testemunhas e as cinco linhagens com maior média de produção.

É possível observar que em comparação à média geral de produtividade (RANK_YLD), os modelos que não possuem aplicação de variáveis secundárias (RANK_YLD_BLUE_mod_2 e RANK_YLD_BLUP_mod_3) apresentam correlação monotônica. A correlação entre estes modelos com a média geral é apresentada também na Figura 3 que corrobora com os resultados mais altos de correlação com a média geral ("r" acima de 0,97).

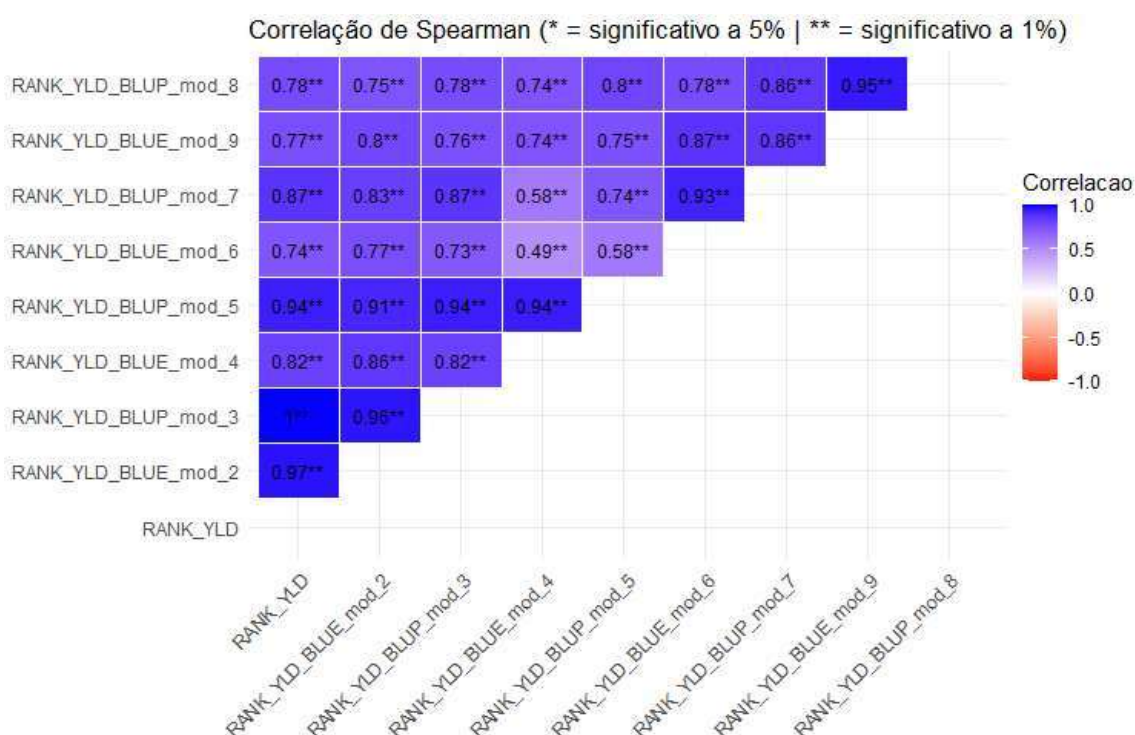
Figura 2 – Comparação do Ranking dos TOP5 e Testemunhas



Fonte: elaboração própria

Corroborando com a Figura 2, a matriz de correlação de Spearman obtida a partir dos dados das avaliações (Figura 3) evidencia a concordância entre os diferentes modelos avaliados, tanto para estimativas baseadas em médias observadas (YLD), quanto para estimativas ajustadas via BLUE e previsões via BLUP. De maneira geral, as correlações entre as estimativas de produtividade (YLD) e aquelas derivadas dos modelos mistos foram superiores a 0,47, atingindo valores próximos de 1,00 em diversos casos.

Figura 3 – Correlação de Spearman entre os ranks dos modelos



Fonte: elaboração própria

Conforme apresentado na Figura 3, observa-se que as correlações quase perfeitas entre os modelos que não possuem aplicação de variáveis secundárias RANK_YLD_BLUE_mod_2 e RANK_YLD_BLUP_mod_3 ($p = 0,96$), entre os modelos com aplicação da variável secundária de maior herdabilidade (Height_25) RANK_YLD_BLUP_mod_4 e RANK_YLD_BLUE_mod_5 ($p = 0,94$), entre os modelos com aplicação da variável secundária de segunda maior herdabilidade (RedEdge) RANK_YLD_BLUP_mod_6 e RANK_YLD_BLUE_mod_7 ($p = 0,93$) e entre os modelos com aplicação das duas variáveis secundárias RANK_YLD_BLUP_mod_8 e RANK_YLD_BLUE_mod_9 ($p = 0,95$). Esses resultados indicam que os ajustes

realizados pelos modelos mistos pouco alteraram a ordenação relativa dos genótipos quando comparados às médias fenotípicas simples. Tal comportamento sugere que, nas condições experimentais avaliadas (DBC com número relativamente equilibrado de repetições), o efeito ambiental intrabloco pode ter sido eficientemente controlado pelo delineamento, reduzindo discrepâncias entre médias ajustadas e observadas.

Entretanto, diferenças tornam-se mais evidentes quando analisado os ranks entre propostas de modelos com variáveis diferentes. As correlações entre modelos que se utilizam de variáveis diferentes entre si, casos como RANK_YLD_BLUE_mod_4 e RANK_YLD_BLUE_mod_6 ($p = 0,492$) e entre RANK_YLD_BLUE_mod_4 e RANK_YLD_BLUP_mod_7 ($p = 0,578$) indicam que determinados ajustes estruturais no modelo influenciaram a classificação relativa das linhagens. Embora essas correlações ainda sejam consideradas relativamente altas, diferenças dessa magnitude podem impactar decisões de avanço em programas de melhoramento, sobretudo quando a seleção é aplicada sobre frações reduzidas da população. Este fator deve ser observado e avaliado com a capacidade operacional da obtenção de dados do experimento em detrimento do estudo com as variáveis secundárias apontadas.

A análise das correlações demonstra que os modelos baseados em RANK_YLD_BLUP apresentaram elevada consistência entre si (“r” entre 0,73 e 0,94), especialmente entre RANK_YLD_BLUP_mod_3, RANK_YLD_BLUP_mod_5 e RANK_YLD_BLUP_mod_7. Isso sugere estabilidade nas predições dos efeitos aleatórios genotípicos, mesmo sob diferentes parametrizações do modelo. Já os modelos baseados em BLUE apresentaram maior diferença entre as correlações dos modelos avaliados (“r” entre 0,49 e 0,87), o que apresenta maior instabilidade em detrimento às avaliações analisadas pelo BLUP.

Do ponto de vista teórico, essa estabilidade é coerente com o arcabouço dos modelos mistos descrito por Henderson (1975) e aprofundado por Piepho *et al.* (2008) e Resende (2016). O BLUP incorpora a estrutura de variância genética e ambiental, promovendo encolhimento (*shrinkage*) das estimativas em direção à média populacional, especialmente para genótipos com menor número de observações. No presente conjunto de dados, como todas as linhagens possuíam ao menos duas repetições e algumas testemunhas até três, o efeito de *shrinkage* provavelmente foi moderado, o que explica a elevada correlação entre BLUPs e médias observadas.

Em contrapartida, os BLUEs, por tratarem genótipos como efeitos fixos, não realizam predição de mérito genético, mas apenas estimativa ajustada de médias. Ainda assim, a alta correlação entre BLUEs e BLUPs (acima de 0,82 na maioria dos casos) indica que, sob as condições experimentais avaliadas, a distinção prática entre estimativa (BLUE) e predição (BLUP) teve impacto relativamente pequeno sobre a classificação global dos genótipos.

Os critérios de informação de Akaike (AIC) e Bayesiano (BIC) foram utilizados para comparação entre modelos conforme demonstra a Figura 4, considerando simultaneamente qualidade de ajuste. Em ambos os critérios, menores valores indicam melhor desempenho como descritos por seus autores Akaike (1974) e Schwarz (1978). Observou-se consistência entre AIC e BIC, com ambos indicando o modelo BLUE_mod_9 como superior. Adicionalmente, todos os modelos do tipo BLUE apresentaram desempenho substancialmente melhor que os modelos BLUP, evidenciado por diferenças superiores a 400 unidades nos critérios, indicando forte evidência a favor dos modelos BLUE.

Figura 4 – Critérios de informação aplicados aos modelos

Modelo	AIC	BIC
BLUE_mod_9	1047,216511	1419,00522
BLUE_mod_6	1060,46165	1428,870462
BLUE_mod_4	1082,017161	1450,425973
BLUE_mod_2	1096,064996	1461,09391
BLUP_mod_8	1526,710354	1546,989738
BLUP_mod_7	1571,323586	1588,223073
BLUP_mod_5	1606,754529	1623,654016
BLUP_mod_3	1623,132605	1636,652194

Fonte: elaboração própria

Entre os modelos BLUE, o BLUE_mod_9 apresentou o menor AIC e BIC, com diferença superior a 10 unidades em relação ao segundo melhor modelo, caracterizando evidência forte de superioridade (BURNHAM e ANDERSON, 2002). A superioridade dos modelos do tipo BLUE em relação aos modelos BLUP, evidenciada pelos menores valores de AIC e BIC, pode ser atribuída à menor complexidade estrutural e à ausência de *shrinkage* dos efeitos genotípicos. Nos modelos BLUP, a inclusão de componentes de variância e a modelagem de efeitos aleatórios aumentam o número efetivo de parâmetros e introduzem penalizações adicionais nos critérios de

informação (VAIDA e BLANCHARD, 2005). Além disso, o processo de *shrinkage*, característico dos BLUPs, tende a reduzir a verossimilhança do modelo ao aproximar os efeitos genotípicos da média, o que impacta negativamente os valores de AIC e BIC (PIEPHO *et al.*, 2008).

Do ponto de vista da seleção indireta, os resultados são particularmente relevantes. Conforme discutido por Ramalho *et al.* (2012) e Cruz, Regazzi e Carneiro (2014), a eficiência da seleção indireta depende da correlação genética entre caracteres e da herdabilidade relativa. No presente estudo, a elevada correlação entre os diferentes modelos sugere que variáveis derivadas de modelagens distintas mantêm associação consistente com a produtividade observada.

Isso implica que características secundárias obtidas por fenotipagem digital, quando integradas a modelos mistos, podem ser utilizadas como critérios auxiliares de seleção sem provocar mudanças drásticas na ordenação dos genótipos superiores. A forte associação observada entre RANK_YLD e os ranks derivados de BLUP (frequentemente “ r ” > 0,78) indica que a utilização de valores genotípicos preditos pode aumentar a robustez da decisão seletiva, principalmente em cenários com maior desbalanceamento experimental ou múltiplos ambientes.

Além disso, pequenas variações nos coeficientes (por exemplo, $r \approx 0,90$ – $0,93$ em alguns modelos BLUE) reforçam que ajustes estatísticos podem modificar posições relativas intermediárias. Em programas de melhoramento, onde a intensidade de seleção é elevada (por exemplo, seleção dos 10–20% superiores), essas mudanças podem resultar na inclusão ou exclusão de determinados genótipos candidatos.

Figura 5 – Resultados de Ganho Genético Estimado

Modelo	Média Geral	Média TOP20	S (Diferencial de Seleção)	h^2	ΔG (Ganho Genético) BLUE: $\Delta G = h^2 \times S$ BLUP: $\Delta G = S$
RANK_YLD	103,70	114,57	10,86	0,21	2,29
RANK_YLD_BLUE_mod_2	104,39	115,25	10,86	0,21	2,29
RANK_YLD_BLUP_mod_3	103,95	107,68	3,73	0,21	3,73
RANK_YLD_BLUE_mod_4	104,50	115,04	10,55	0,23	2,39
RANK_YLD_BLUP_mod_5	103,88	107,68	3,80	0,23	3,80
RANK_YLD_BLUE_mod_6	105,18	116,44	11,26	0,23	2,57
RANK_YLD_BLUP_mod_7	105,28	109,14	3,86	0,23	3,86
RANK_YLD_BLUE_mod_9	105,16	114,44	9,28	0,13	1,25
RANK_YLD_BLUP_mod_8	105,22	107,40	2,18	0,13	2,18

Fonte: elaboração própria

A comparação dos ganhos genéticos estimados, conforme apresentado na Figura 5, evidencia as diferenças estruturais entre os modelos, tanto pela natureza dos efeitos genotípicos fixos em comparação aos aleatórios quanto pela inclusão de uma ou duas variáveis secundárias obtidas pelas imagens de drones.

De forma geral, os modelos baseados em BLUE não apresentam componentes de variância genética e residual, portanto a fim de se calcular o ganho genético estimado para estes modelos foi estimado o valor da herdabilidade do modelo BLUP correspondente às variáveis calculadas. Assim, os modelos BLUP apresentaram os maiores ganhos genéticos estimados (ΔG entre 2,18 e 3,86), enquanto os modelos baseados em BLUE resultaram em ganhos menores (ΔG entre 1,25 e 2,57). Esse padrão é consistente com a teoria de modelos mistos: os BLUEs operam na escala fenotípica ajustada e preservam os valores extremos, elevando o diferencial de seleção (S) e, conseqüentemente, o ganho genético (HENDERSON, 1975; PIEPHO *et al.*, 2008). Em contrapartida, os BLUPs incorporam a estrutura de variância genética e promovem *shrinkage* dos efeitos, reduzindo a influência de extremos e fornecendo estimativas mais conservadoras e robustas do valor genético (RESENDE, 2016; BERNARDO, 2020).

A inclusão de variáveis secundárias impactou positivamente os ganhos. É possível observar que o modelo RANK_YLD_BLUE_mod_6 apresentou o maior ganho genético ($\Delta G = 2,57$), indicando que a inclusão das covariáveis melhora a discriminação entre genótipos e amplia o diferencial de seleção. De forma semelhante, RANK_YLD_BLUE_mod_4 também apresentou ganho elevado ($\Delta G = 2,39$), reforçando o valor preditivo dessas variáveis (COBB *et al.*, 2013; SHAKOOR *et al.*, 2017).

Nos modelos BLUP, o efeito das variáveis secundárias também é destacado. Os modelos RANK_YLD_BLUP_mod_5 e RANK_YLD_BLUP_mod_7 apresentaram ganhos um pouco superiores ao modelo base RANK_YLD_BLUP_mod_3, sugerindo que as covariáveis contribuem para maior acurácia, embora o efeito seja amortecido pelo *shrinkage* (PIEPHO *et al.*, 2008; RESENDE, 2016). Já o modelo com duas variáveis secundárias, RANK_YLD_BLUP_mod_8, apresentou o menor ganho ($\Delta G = 2,18$), possivelmente indicando redundância de informação ou presença de distorções no modelo, fenômeno também relatado em abordagens multivariadas com variáveis correlacionadas (BERNARDO, 2020).

É possível destacar que a inclusão de duas variáveis secundárias no modelo BLUE (RANK_YLD_BLUE_mod_9) não superou os modelos com apenas uma variável, sugerindo que há um ponto ótimo de complexidade. A adição excessiva de covariáveis pode não resultar em ganhos adicionais e pode até reduzir a eficiência do modelo, especialmente quando há colinearidade ou redundância de informação (MONTGOMERY *et al.*, 2021; BERNARDO, 2020).

Considerando o contexto mais amplo da fenotipagem de alto rendimento (HTP), os resultados corroboram a premissa apresentada por Cobb *et al.* (2013) e Shakoob *et al.* (2017), de que o aumento na precisão fenotípica, aliado a métodos estatísticos robustos, potencializa o ganho genético. A elevada concordância entre modelos sugere que os dados utilizados apresentam boa qualidade experimental, permitindo que tanto BLUE quanto BLUP capturem de forma consistente o desempenho relativo das linhagens.

Contudo, em cenários mais complexos, como múltiplos ambientes, interação genótipo $\times$ ambiente ou dados desbalanceados, espera-se que os modelos mistos apresentem vantagens mais pronunciadas sobre médias simples. Conforme destacado por Bernardo (2020), a acurácia seletiva depende diretamente do aproveitamento máximo das informações disponíveis, o que é mais eficientemente realizado via modelos mistos com estimativas por máxima verossimilhança restrita. Assim, mesmo que neste conjunto específico as diferenças entre abordagens tenham sido pequenas, a fundamentação teórica e empírica sustenta a adoção preferencial de BLUP para fins de seleção genética.

5 CONCLUSÕES

Os resultados demonstram concordância no ranqueamento dos genótipos entre os modelos, indicando boa qualidade dos dados e adequado controle da variabilidade ambiental. A inclusão de variáveis secundárias provenientes da fenotipagem digital aumentou a discriminação entre genótipos e os ganhos genéticos estimado. No entanto, o aumento da complexidade nem sempre resultou em ganhos adicionais, sugerindo a existência de um ponto ótimo na inclusão de covariáveis.

Embora os modelos BLUE tenham apresentado ganhos significativos e melhor ajuste estatístico, os modelos BLUP se destacaram pela maior estabilidade e robustez

biológica, sendo mais adequados para seleção genética, sobretudo em cenários complexos.

No entanto, a aplicação prática desses modelos depende da viabilidade operacional da obtenção das variáveis secundárias. Custos, limitações logísticas e dependência de condições ambientais podem restringir seu uso. Assim, considerando a similaridade no ranqueamento dos genótipos, modelos mais simples e operacionalmente viáveis podem representar a alternativa mais eficiente, assegurando escalabilidade e sustentabilidade no processo de seleção.

REFERÊNCIAS

AKAIKE, H. A new look at the statistical model identification. **IEEE Transactions on Automatic Control**, Notre Dame, v. 19, n. 6, p. 717-723, 1974.

ALLARD, R. W. **Plant Breeding**. Disponível em: <https://www.britannica.com/science/plant-breeding>. Acesso em: 22 maio 2023.

BERNARDO, R. **Breeding for quantitative traits in plants**. 3. ed. Woodybury, MN: Stemma Press, 2020. 422 p.

BURNHAM, K. P.; ANDERSON, D. R. **Model selection and multimodel inference: a practical information-theoretic approach**. 2. ed. New York: Springer, 2002.

COBB, J. N. *et al.* Next-generation phenotyping: requirements and strategies for enhancing our understanding of genotype–phenotype relationships and its relevance to crop improvement. **Theoretical and Applied Genetics**, v. 126, p. 867–887, 2013. DOI: <https://doi.org/10.1007/s00122-013-2066-0>.

CRUZ, C. D.; REGAZZI, A. J.; CARNEIRO, P. C. S. Modelos biométricos aplicados ao melhoramento genético. v. 2. 3. ed. Viçosa: **Editora Viçosa**, 2014. 668p.

DE CARVALHO SILVEIRA, J. M. *et al.* Uso de imagens multiespectrais e termográficas para monitoramento das condições hídricas da cana-de-açúcar. **Irriga**, v. 25, n. 4, p. 689-696, 2020. DOI: 10.15809/irriga 2020 v25 n4 p689-696

DUBCOVSKY, J.; DVORAK, J. Genome plasticity: a key factor in the success of polyploid wheat under domestication. **Science**, v. 316, n. 5833, p. 1862-1866, 2007.

EL BAIDOURI, M. *et al.* Reconciling the evolutionary origin of bread wheat (*Triticum aestivum*). **New Phytologist**, v. 213, n. 3, p. 1477-1486, 2017.

EMILIANO, P. C. **Critérios de informação: como eles se comportam em diferentes modelos**. 2013. 193 p. Tese (Doutorado em Estatística e Experimentação Agropecuária) – Universidade Federal de Lavras, Lavras, 2013.

FELDMAN, M. Wheats. In: SMARTT, J.; SIMMONDS, N. W. (Ed.). Evolution of crop plants. Harlow, **UK: Longman Scientific and Technical**, 1995. p. 185-192.

HAWKESFORD, M. J. *et al.* Prospects of doubling global wheat yields. **Food and Energy Security**, v. 2, n. 1, p. 34-48, 2013.

HENDERSON, C. R. Best linear unbiased estimation and prediction under a selection model. **Biometrics**, v. 31, n. 2, p. 423, 1975.

ISIK, F.; HOLLAND, J.; MALTECCA, C. Genetic data analysis for plant and animal breeding. Cham: **Springer International Publishing**, 2017.

PEREIRA, L. A. **Estratégias de condução de populações segregantes utilizando o método BULK/F₂ ou S₀**. 2018. 118 f. Tese (Doutorado) - Curso de Pós-graduação em Genética e Melhoramento de Plantas, Universidade Federal de Lavras, Lavras, MG, 2018.

PETERSEN, G. *et al.* Phylogenetic relationships of Triticum and Aegilops and evidence for the origin of the A, B, and D genomes of common wheat (*Triticum aestivum*). **Molecular Phylogenetics and Evolution**, v. 39, n. 1, p. 70-82, 2006.

PIANA, C. F. B.; CARVALHO, F. I. F. Trigo: a cultura que deu suporte à civilização. In: BARBIERI, R. L.; STUMPF, E. R. T. Origem e evolução de plantas cultivadas. Brasília, DF: **Embrapa Clima Temperado**, 2008. p. 909.

PIEPHO, H. P. *et al.* BLUP for phenotypic selection in plant breeding and variety testing. **Euphytica**, 2008.

PONT, C.; SALSE, J. Wheat paleohistory created asymmetrical genomic evolution. **Current Opinion in Plant Biology**, v. 36, p. 29-37, 2017.

MARCUSSEN, T. *et al.* Ancient hybridizations among the ancestral genomes of bread wheat. **Science**, v. 345, n. 6194, p. 1250092-1-4, 2014.

MATIAS, F. I. *et al.* Bison-Fly: an open-source UAV pipeline for plant breeding data collection. **The Plant Phenome Journal**, v. 5, n. 1, p. 1-9, 2022.

MOHAMMED, E. A. *et al.* Emerging business intelligence framework for a clinical laboratory through big data analytics. **Emerging Trends in Computational Biology, Bioinformatics, and Systems Biology**, p. 577-602, 2015.

MOLIN, J. P.; DO AMARAL, L. R.; COLAÇO, A. Sensoriamento e sensores. In: MOLIN, J. P.; DO AMARAL, L. R.; COLAÇO, A. (Eds.). **Agricultura de precisão**. São Paulo: Oficina de Textos, 2015. p. 119-131.

MONTGOMERY, D. C. *et al.* **Introduction to Linear Regression Analysis**. 6. ed. Hoboken: Wiley, 2021.

NOGUEIRA, A. P. O. *et al.* Análise de trilha e correlações entre caracteres em soja cultivada em duas épocas de semeadura. **Bioscience Journal**, Uberlândia, v. 28, n. 6, 2012.

NETO, F. A. *et al.* Grain yield, adaptability and stability of soybean genotypes in different Cerrado environments of Piauí, Brazil. **Comunicata Scientiae**, v. 9, n. 2, p. 226-234, 2018. DOI: <https://doi.org/10.14295/cs.v9i2.2674>.

NUNES, J. A. R. *et al.* Inclusion of genetic relationship information in the pedigree selection method using mixed models. **Genetics and Molecular Biology**, v. 31, n. 1, p. 73-78, 2008.

R CORE TEAM. **R: A language and environment for statistical computing**. 2019. Disponível em: <https://www.r-project.org/>.

RAMALHO, M. A. P. *et al.* Aplicações da genética quantitativa no melhoramento de plantas autógamas. Lavras: **Editora UFLA**, 2012. 522 p.

RESENDE, M. D. V. **Genética biométrica e estatística no melhoramento de plantas perenes**. Brasília: Embrapa, 2016.

ROCHA, G. N. **Seleção de progênies estáveis e produtivas de *Eucalyptus urophylla* em multiambientes**. 2024. Dissertação (Mestrado) - Universidade de São Paulo, Piracicaba, 2024. Disponível em: <https://www.teses.usp.br/teses/disponiveis/11/11150/tde-04112024-154103/>. Acesso em: 27 jan. 2025.

SCHWARZ, G. Estimating the dimensional of a model. *Annals of Statistics*, **Hayward**, v. 6, n. 2, p. 461-464, 1978. DOI: 10.1214/aos/1176344136

SHAKOOR, N.; LEE, S.; MOCKLER, T. C. High throughput phenotyping to accelerate crop breeding and monitoring of diseases in the field. **Current Opinion in Plant Biology**, v. 38, p. 184-192, 2017. DOI: 10.1016/j.pbi.2017.05.006.

SHEWRY, P. R. Wheat. **Journal of Experimental Botany**, v. 60, n. 6, p. 1537-1553, 2009.

SUGIURA, N. Further analysis of the data by Akaike's information criterion and the finite corrections: Further analysis of the data by Akaike's. **Communications in Statistics - Theory and Methods**, v. 7, n. 1, p. 13-26, 1978. DOI: <https://doi.org/10.1080/03610927808827599>.

THOMPSON, R. Estimation of quantitative genetic parameters. *Proceedings of the Royal Society B: Biological Sciences*, v. 275, n. 1635, p. 679-686, 2008. DOI: 10.1098/rspb.2007.1417.

THOMPSON, R. Relationship between the cumulative difference and best linear unbiased predictor methods of evaluating bulls. **Animal Production**, v. 23, n. 1, p. 15-24, 1976.

VAIDA, F.; BLANCHARD, S. Conditional Akaike information for mixed-effects models. **Biometrika**, v. 92, n. 2, p. 351-370, 2005.

WHITE, T. L.; HODGE, G. R. **Predicting breeding values with applications in forest tree improvement**. Dordrecht: Kluwer Academic Publishers, 1989. 363 p.

YANG, W. *et al.* Crop phenomics and high-throughput phenotyping: Past decades, current challenges, and future perspectives. **Molecular Plant**, v. 13, n. 2, p. 187–214, 2020. DOI: 10.1016/j.molp.2020.01.008.