

MÉTODOS ESTATÍSTICOS FREQUENTISTAS E BAYESIANOS
NO MELHORAMENTO INTRAPOPULACIONAL DO
MARACUJAZEIRO-AZEDO

ANDRÉ OLIVEIRA SOUZA

UNIVERSIDADE ESTADUAL DO NORTE FLUMINENSE
DARCY RIBEIRO – UENF

CAMPOS DOS GOYTACAZES - RJ
NOVEMBRO - 2020

**MÉTODOS ESTATÍSTICOS FRENQUENTISTAS E BAYESIANOS
NO MELHORAMENTO INTRAPOPULACIONAL DO
MARACUJAZEIRO-AZEDO**

ANDRÉ OLIVEIRA SOUZA

Tese apresentada ao Centro de Ciências e
Tecnologias Agropecuárias da Universidade
Estadual do Norte Fluminense Darcy Ribeiro,
como parte das exigências para obtenção do
título de Doutor em Produção Vegetal.

Orientador: Prof. Alexandre Pio Viana

CAMPOS DOS GOYTACAZES - RJ
NOVEMBRO - 2020

FICHA CATALOGRÁFICA

UENF - Bibliotecas

Elaborada com os dados fornecidos pelo autor.

S729

Souza, André Oliveira.

MÉTODOS ESTATÍSTICOS FREQÜENTISTAS E BAYESIANOS NO MELHORAMENTO INTRAPOPULACIONAL DO MARACUJAZEIRO-AZEDO / André Oliveira Souza. - Campos dos Goytacazes, RJ, 2020.

84 f. : il.

Bibliografia: 60 - 67.

Tese (Doutorado em Produção Vegetal) - Universidade Estadual do Norte Fluminense Darcy Ribeiro, Centro de Ciências e Tecnologias Agropecuárias, 2020.
Orientador: Alexandre Pio Viana.

1. Bayesiana. 2. Componentes de variância. 3. REML. 4. post-hoc blocking Row-Col. 5. Precisão Experimental . I. Universidade Estadual do Norte Fluminense Darcy Ribeiro. II. Título.

CDD - 630

MÉTODOS ESTATÍSTICOS FRENQUENTISTAS E BAYESIANOS
NO MELHORAMENTO INTRAPOPULACIONAL DO
MARACUJAZEIRO-AZEDO

ANDRÉ OLIVEIRA SOUZA

Tese apresentada ao Centro de Ciências e Tecnologias Agropecuárias da Universidade Estadual do Norte Fluminense Darcy Ribeiro, como parte das exigências para obtenção do título de doutor em Produção Vegetal.

Aprovada em 30 de novembro de 2020.

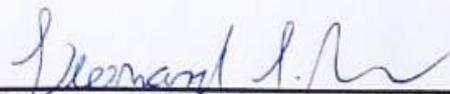
Comissão Examinadora:



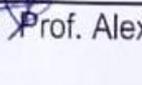
Prof. Antônio Teixeira do Amaral Júnior (D.Sc., Genética e Melhoramento) - UENF



Prof. Camila Ferreira Azevedo (D.Sc., Estatística Aplicada e Biometria) - UFV



Prof. Leonardo Siqueira Glória (D.Sc., Genética e Melhoramento) - UENF



Prof. Alexandre Pio Viana (D.Sc., Produção Vegetal) - UENF
(Orientador)

A Deus.
À minha filha Ana Beatriz.
Ao meu filho Henrique.
À minha esposa Andrea.
Aos meus pais.
A todos que de alguma forma torceram por mim.
DEDICO!

AGRADECIMENTOS

Primeiramente a Deus por ter me dado força, inspiração e sabedoria.

À minha esposa Andrea e aos meus filhos Henrique e Ana Beatriz pela paciência e compreensão.

À Universidade Estadual do Norte Fluminense Darcy Ribeiro (UENF) e ao Instituto Federal do Espírito Santo (Ifes) pela oportunidade de qualificação profissional.

Ao professor Dr. Alexandre Pio Viana pela oportunidade de aprimoramento profissional, sabedoria, paciência, sugestões e pela confiança depositada em mim para realização deste trabalho.

À coorientadora Dra. Camila Ferreira Azevedo pelos ensinamentos, paciência, sabedoria e contribuições.

Ao Natan Ramos Cavalcante pela colaboração, paciência e auxílio durante todo o desenvolvimento deste trabalho.

Ao professor Dr. Fabyano Fonseca e Silva pelas valiosas sugestões, paciência, contribuições e pela oportunidade de cursar a disciplina “Inferência Bayesiana Aplicada ao Melhoramento Genético”.

Aos professores Dr. Moyses Nascimento, Dr. Cosme Damião Cruz, Dra. Camila Ferreira Azevedo e Dra. Ana Carolina Campana Nascimento pela oportunidade de cursar a disciplina. Abordagens biométricas para reconhecimento de padrão, classificação e predição no melhoramento genético.

Ao professor Dr. Leonardo Bhering pela oportunidade de cursar a disciplina “Genética Quantitativa”.

Aos professores Dr. Antônio Teixeira do Amaral Júnior e Dra. Rosana Rodrigues pela oportunidade de cursar a disciplina “Melhoramento de Plantas”.

À professora Dra. Camila Ferreira Azevedo pelas valiosas sugestões, contribuições e oportunidade de cursar a disciplina de “Inferência Bayesiana”.

Aos membros da banca de qualificação, professor Dr. Fabyano Fonseca e Silva, professor Dr. Leonardo Siqueira Glória, Dr. Antônio Teixeira do Amaral Júnior e professor Dr. Alexandre Pio Viana, pela disponibilidade e sugestões para o aprimoramento deste trabalho.

Aos membros da banca de defesa, professores Dr. Leonardo Siqueira Glória, Dra. Camila Ferreira Azevedo, Dr. Antônio Teixeira do Amaral Júnior e Dr. Alexandre Pio Viana, pela disponibilidade e valiosas sugestões para a conclusão deste trabalho de pesquisa.

Aos professores do Programa de Pós-Graduação em Produção Vegetal da UENF pela contribuição em minha formação profissional.

A todos meus colegas pela amizade e apoio durante o curso, e em especial ao Adriano Posse Senhorelo que nunca mediu esforços para me ajudar quando precisei durante esta etapa de minha vida.

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

O presente trabalho de pesquisa também foi financiado pela Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro (Faperj).

SUMÁRIO

LISTAS DE FIGURAS	vii
LISTAS DE TABELAS	ix
RESUMO	xii
ABSTRACT	xiv
1. INTRODUÇÃO	1
1.1 Objetivo geral.....	3
1.2 Objetivos específicos	3
2. REVISÃO DE LITERATURA	5
2.1 Aspectos gerais do maracujazeiro-azedo	5
2.2 Melhoramento do maracujazeiro azedo na UENF via seleção recorrente intrapopulacional.....	7
2.3 Técnica Post-hoc blocking Row-Column (Row-Col)	8
2.4 Métodos de estimação de componentes de variância frequentistas e Bayesianos	9
2.4.1 Modelos Mistos	9
2.4.2 Modelos Bayesianos	12
2.4.2.1 Distribuição a priori	14
2.4.2.2 Extensão de Parâmetros	15
2.4.2.3 Verificação da Convergência de cadeias	16
2.4.2.4 Avaliadores e Critérios de seleção de modelos	17
2.4.2.5 Intervalo Highest Posterior Density (HPD)	18
2.4.3 Validação cruzada.....	18
2.5 Estimação de Valores Genéticos no melhoramento de Plantas.....	19

3. MATERIAL E MÉTODOS	22
3.1 População avaliada	22
3.1.1 Experimento I	22
3.1.2 Experimento II	22
3.2 Descrição e condução do experimento I	22
3.3 Descrição e condução do experimento II	23
3.4 Características agronômicas avaliadas no experimento I	23
3.5 Características agronômicas avaliadas no experimento II	24
3.6 Parâmetros genéticos	24
4. MODELOS ESTATÍSTICOS DA AVALIAÇÃO GENÉTICA	26
4.1 Inferência bayesiana do modelo completo	27
4.2 Inferência bayesiana do modelo reduzido	29
5. RESULTADOS E DISCUSSÕES	32
5.1 Experimento I	32
5.2 Experimento II	44
6. CONCLUSÕES	59
REFERÊNCIAS BIBLIOGRÁFICAS	60

LISTAS DE FIGURAS

Figura 1 - Frequência das 30 primeiras progênies para todas as características avaliadas: Modelo BLUP/REML (a) e Bayesiano (b).	43
Figura 2 - Frequência das 30 primeiras progênies para todas as características avaliadas: Modelo BLUP/REML (a) e Bayesiano (b).	55

LISTAS DE TABELAS

Tabela 1 - Distribuição a priori para os componentes de variância	29
Tabela 2 - Teste da razão de verossimilhança (TRV) para comparação entre os modelos blocos completos casualizados (DBC) e <i>post-hoc blocking Row-Col</i> BLUP/REML. Campos dos Goytacazes, RJ, 2020	34
Tabela 3 - Comparação dos modelos Bayesianos em blocos completos casualizados (DBC) e Row-Col pelo Deviance Information Criterion (DIC) - Bayes II. Campos dos Goytacazes, RJ, 2020	35
Tabela 4 - Comparação dos modelos pelas métricas raiz quadrada do erro quadrático médio (REQM) e correlação entre o predito e o observado (r) para os três cenários de modelos escolhidos. Campos dos Goytacazes, RJ, 2020	37
Tabela 5 - Estimativa dos parâmetros genéticos das 81 progênes de maracujazeiro-azedo de irmãos completos do terceiro ciclo de seleção recorrente via modelos mistos-REML/BLUP. Campos dos Goytacazes, RJ, 2020	38
Tabela 6 - Estimativa de parâmetros genéticos das 81 progênes de maracujazeiro-azedo do terceiro ciclo de seleção recorrente – Bayes II. Campos dos Goytacazes, RJ, 2020	39

Tabela 7 - Índice de coincidência dos indivíduos selecionados (%) entre as 81 progênes de maracujazeiro-azedo do terceiro ciclo de seleção recorrente via modelo Bayesiano e por Modelos mistos BLUP/REML. Campos dos Goytacazes, RJ, 2020.....	40
Tabela 8 - Tabela com ranks das 30 primeiras progênes de maracujazeiro-azedo de irmãos completos do terceiro ciclo de seleção recorrente obtidas via modelos mistos REML/BLUP. Campos dos Goytacazes, RJ, 2020	41
Tabela 9 - Tabela com ranks das 30 primeiras progênes de maracujazeiro-azedo de irmãos completos do terceiro ciclo de seleção recorrente obtidas via modelos Bayesiano. Campos dos Goytacazes, RJ, 2020	42
Tabela 10 - Teste da razão de verossimilhança (TRV) para comparação entre os dois modelos propostos, blocos completos casualizados (DBC) e post-hoc blocking Row-Col. Cambuci, RJ, 2020.....	45
Tabela 11 - Comparação dos modelos Bayesianos em blocos completos casualizados (DBC) e Row-Col pelo Deviance Information Criterion (DIC) – Bayes II. Cambuci RJ, 2020.....	46
Tabela 12 - Comparação dos modelos pelas métricas raiz quadrada do erro quadrático médio (REQM) e a correlação entre o observado e o predito (r) para os três cenários. Cambuci, RJ, 2020	48
Tabela 13 - Estimativa de parâmetros genéticos das 97 progênes de maracujazeiro-azedo do quarto ciclo de seleção recorrente da seleção recorrente via modelos mistos-REML/BLUP. Cambuci, RJ, 2020	50
Tabela 14 - Estimativa de parâmetros genéticos das 97 progênes de maracujazeiro-azedo do quarto ciclo de seleção recorrente-Bayes II. Cambuci, RJ, 2020	51
Tabela 15 - Índice de coincidência dos indivíduos selecionados (%) entre as 81 progênes de maracujazeiro-azedo do terceiro ciclo de seleção recorrente via modelo Bayesiano e por Modelos mistos REML/BLUP. Cambuci, RJ, 2020	52

Tabela 16 - Tabela com ranks das 30 primeiras progênies de maracujazeiro-azedo de irmãos completos do quarto ciclo de seleção recorrente via modelos mistos-REML/BLUP. Cambuci, RJ, 2020	53
Tabela 17 - Tabela com ranks e as médias estimadas das 30 primeiras progênies de maracujazeiro-azedo do terceiro ciclo de seleção recorrente da seleção recorrente via modelos Bayesiano. Cambuci, RJ, 2020	54
Tabela 18 - Estimativas das métricas obtidas pelos modelos na validação cruzada 5-folds	57

RESUMO

SOUZA, André Oliveira; D.Sc.; Universidade Estadual do Norte Fluminense Darcy Ribeiro; novembro de 2020; MÉTODOS ESTATÍSTICOS FREQUENTISTAS E BAYESIANOS NO MELHORAMENTO INTRAPOPULACIONAL DO MARACUJAZEIRO-AZEDO. Orientador: Prof. Alexandre Pio Viana; Coorientadora: Prof^a. Camila Ferreira Azevedo.

O maracujazeiro-azedo é originário da América Tropical, sendo mais de 150 espécies nativas do Brasil e, dentre estas, inclui-se a mais importante comercialmente que é o *Passiflora edulis*, representando cerca de 95% dos pomares no Brasil, sendo também a espécie mais plantada no mundo. A produção brasileira é direcionada para o consumo do fruto, sendo a espécie *Passiflora edulis* a mais cultivada, e, além disso, o maracujazeiro-azedo também é utilizado na indústria farmacêutica, cosmética e ornamental. Neste sentido, tornam-se necessários estudos direcionados ao desenvolvimento de cultivares resistentes a doenças e mais adaptadas às condições edafoclimáticas das regiões produtoras, melhorando assim, características relacionadas aos componentes de produção e aumento da renda dos produtores. A metodologia de modelos mistos tem sido empregada com sucesso na estimação de parâmetros genéticos e seleção de indivíduos, em diversas culturas. Assim, neste trabalho de pesquisa é proposto a modelagem de observações de dois experimentos com maracujazeiro-azedo sob seleção recorrente intrapopulacional: pelos métodos de estimação de modelos mistos e modelos bayesianos, concebidos com a técnica

post-hoc blocking Row-Column (Row-Col) com a finalidade de obter estimativas de herdabilidade fidedignas com a captação de efeito ambiental. O delineamento experimental utilizado foi em blocos ao acaso com 81 progênies e duas repetições com parcelas constituídas por cinco plantas no experimento I. No experimento II o delineamento experimental utilizado foi em blocos ao acaso com 97 progênies com duas repetições e cada parcela constituídas por três plantas. As variáveis modeladas foram: Número de frutos (NF), Produtividade (Prod.), Massa do Fruto (MF), Diâmetro Transversal do Fruto (DTF), Diâmetro Longitudinal do Fruto (DLF), Polpa de frutos (Polpa), Espessura da casca (EC), Sólidos Solúveis Totais (SST). O modelo bayesiano a *priori* com extensão de parâmetros (*priori* II) obteve a menor raiz quadrada do erro quadrático médio (REQM) se comparado com a Gama Inversa (*priori* I) e foi a que mais se aproximou das métricas no contexto modelos mistos. Sendo assim, o modelo com a *priori* II foi escolhido para *ranquear* as progênies via modelos bayesianos. Para esta estrutura de observações foi possível verificar que os dois modelos utilizados, no geral, selecionaram as mesmas progênies e tiveram índices de coincidência, na seleção, acima de 83,33% para todas as variáveis modeladas. A capacidade de predição (CP), e a eficiência seletiva (ES) foram obtidas pela técnica 5-folds e mostrou que os dois modelos testados têm a mesma capacidade de generalização para as variáveis modeladas.

Palavras-chave: Bayesiana, Componentes de variância, REML, *post-hoc blocking Row-Col*, Precisão Experimental.

ABSTRACT

SOUZA, André Oliveira; D.Sc.; Universidade Estadual do Norte Fluminense Darcy Ribeiro; 2020, november; FREQUENTIST AND BAYESIAN STATISTICAL METHODS IN THE INTRAPOPULATIONAL IMPROVEMENT OF PASSION FRUIT; Advisor: Prof. Alexandre Pio Viana; Co-advisor: Camila Ferreira Azevedo

Passion fruit is originally from Tropical America, with more than 150 native species from Brazil and among these, the most commercially important is *Passiflora edulis*, which represents about 95% of orchards in Brazil, and is also the most planted species in the world. Brazilian production is directed towards the consumption of the fruit, with the *Passiflora edulis* species being the most cultivated, and, in addition, passion fruit is also used in the pharmaceutical, cosmetic and ornamental industries. Therefore, studies directed to the development of cultivations resistant to diseases and more adapted to the edaphoclimatic conditions of the producing regions are necessary, improving characteristics related to the production components and increasing the producers' income. The mixed model methodology has been used successfully in the genetic parameters estimation and in the selection of individuals, in different cultures. Thus, in this research paper it is proposed the modeling of observations of two experiments with passion fruit under recurrent intrapopulation selection, by the estimation methods of mixed models and Bayesian models, where the models were designed with the post-hoc blocking Row-Column (Row-Col) technique in order to obtain reliable heritability estimates with the capture of environmental effect. The experimental design used was in

randomized blocks with 81 progenies and two replications with plots consisting of five plants in experiment I. In experiment II, the experimental design used was in randomized blocks with 97 progenies with two replications and each plot consisted of three plants. - The modeled variables were: Number of fruits (NF), Productivity (Prod.), Fruit Mass (MF), Transverse Fruit Diameter (DTF), Longitudinal Fruit Diameter (DLF), Fruit pulp (Pulp), Thickness of shell (EC), Total Soluble Solids (SST). The Bayesian model, in first place, with extension of parameters (*priori* II) obtained the smallest square root of the medium square error (RMSE) if compared to the Inverse Range (*priori* I) and was the one that came closest to the metrics in the context of mixed models, being thus, the model with the *priori* II was chosen to rank the progenies via Bayesian models. For this observations structure it was possible to observe that the two models used, in general, selected the same progenies and had coincidently rates, in the selection, above 83.33% for all modeled variables. The prediction capacity (CP), and the selective efficiency (ES) were obtained by the 5-fold technique and have shown that the two tested models have the same generalization capacity for all modelled variables.

Keywords: Bayesian, Variance Components, REML, post-hoc blocking Row-Col, Experimental Precision.

1. INTRODUÇÃO

O maracujazeiro-azedo é originário da América Tropical, sendo mais de 150 espécies nativas do Brasil (Bernacci et al., 2015). Dentre estas, inclui-se a mais importante comercialmente, o maracujazeiro-azedo (*Passiflora edulis*), que representa mais de 95% dos pomares no país, sendo também a espécie mais plantada no mundo (Meletti, 2011) com estimativa de que a espécie *Passiflora edulis* ocupe mais de 90% da área cultivada com maracujá no mundo (Faleiro et al., 2005).

A elevada proporção de cultivo desta espécie é devido à qualidade dos seus frutos, vigor, produtividade e rendimento em suco (Meletti e Brückner, 2001). O foco da produção brasileira é o consumo do fruto, sendo a espécie *Passiflora* a mais utilizada, e além disso, o maracujazeiro também é utilizado na indústria farmacêutica, cosmética e ornamental (Bernacci et al., 2015).

O Brasil é considerado o maior produtor mundial de maracujá-azedo, com produção de 602651 toneladas cultivada numa área aproximada de 43248 hectares, resultando assim, numa produtividade média de 13935 kg ha⁻¹. Essas estimativas mostram o potencial de produção e indicam a importância econômica do cultivo dessa fruteira para o país. Com cerca de sessenta e dois por cento da produção nacional a Região Nordeste lidera o *ranque*, seguida pelas regiões Sudeste, Sul, Norte e Centro-Oeste. No entanto, quando se trata de produtividade a liderança deste *ranque* é da Região Sul com cerca de vinte e cinco por cento da produção nacional, seguida pelas Regiões Centro-Oeste, Sudeste, Nordeste e Norte (Instituto Brasileiro de Geografia e Estatística, 2018).

A grande exploração comercial do *Passiflora edulis* no Brasil é devido ao seu maior rendimento industrial (Vianna-Silva et al., 2008) e desconhecimento das outras espécies comestíveis por parte dos consumidores. Além do maracujazeiro-azedo, várias outras espécies do gênero *passiflora* são cultivadas comercialmente em países como o Brasil, Colômbia, Peru, Equador, Austrália, África do Sul, Costa Rica, entre outros países (Faleiro et al., 2017).

Em 2018, a produção de maracujá no estado do Rio de Janeiro foi de 6959 toneladas, podendo ser considerada baixa se comparada com as 160902 toneladas do maior produtor que é o estado da Bahia. Quando se trata de produtividade média o estado do Rio de Janeiro produziu 13565 kg ha^{-1} , que é abaixo da média nacional e pode ser considerada baixa quando comparada com a maior produtividade brasileira que é do Distrito Federal com 27675 kg ha^{-1} (Instituto Brasileiro de Geografia e Estatística, 2018).

A baixa produtividade, causada, entre outros fatores, pela falta de genótipos altamente produtivos, grande variabilidade existente em pomares comerciais, carência de cultivares melhoradas que sejam adaptadas às condições do solo, clima e ambiente da região de cultivo, além de problemas relacionados a aspectos fitossanitários, assim refletindo na necessidade do melhoramento genético para essa cultura (Gonçalves et al., 2007), e que tem ocasionado a diminuição da vida útil dos novos plantios (Lima e Cunha, 2004; Cavichioli et al., 2011).

Nesta perspectiva, e na busca para contornar os efeitos da baixa produtividade, a Universidade Estadual do Norte Fluminense Darcy Ribeiro (UENF) vem desenvolvendo há cerca de 20 anos o programa de melhoramento genético do maracujazeiro-azedo, via seleção recorrente intrapopulacional, com a obtenção de resultados promissores (Viana et al., 2004; Ferreira et al., 2016; Viana et al., 2016; Silva et al., 2017; Ribeiro et al., 2019).

Neste contexto, a metodologia de modelos mistos (REML/BLUP) vem sendo uma importante ferramenta que se tornou padrão para a estimação de parâmetros genéticos e na predição dos valores genotípicos em diversos programas de melhoramento, resultando assim, num processo de seleção com maior eficiência seletiva (Resende, 2002b; Alves e Resende, 2008; Viana e Resende, 2014).

A abordagem de modelos mistos surgiu no melhoramento animal (Henderson, 1959, 1984) e tem sido amplamente empregada no melhoramento vegetal (Resende et al., 2014; Silva et al., 2016; Machado et al., 2020).

Outra metodologia de estimação que atualmente é uma ferramenta de grande importância na estimação de parâmetros genéticos é a estatística Bayesiana, pois, segundo Thompson et al. (2005), uma razão para o uso desta abordagem é que ela permite o uso natural de distribuições *a priori* dos parâmetros genéticos em estudo. A análise de experimentos por esta abordagem tem se intensificado com o passar dos anos (Mila e Ngugi, 2011; Gouy et al., 2013; Silva, et al., 2013, 2020). A disponibilidade computacional atual torna a abordagem Bayesiana exequível e capaz de lidar com modelos estatísticos mais complexos, porém, devido à complexidade computacional envolvida na implementação deste tipo de metodologia, esta precisa ser implementada usando ferramentas específicas (Omer et al., 2015).

1.1 Objetivo geral

Ranquear as progênes de maracujazeiro-azedo pelo método frequentista de estimação de modelos mistos (REML/BLUP) e por modelos bayesianos.

1.2 Objetivos específicos

- a) Estimar os componentes de variância dos fatores (linha, coluna, bloco, progênie) e da variância residual pelos métodos de estimação modelos mistos (REML/BLUP) e modelos bayesianos.
- b) Avaliar o potencial agrônômico de famílias de irmãos completos de maracujazeiro-azedo pertencentes ao terceiro e quarto ciclos de seleção recorrente intrapopulacional, e, selecionar progênes promissoras para compor os genitores do próximo ciclo de seleção recorrente.
- c) Estimar parâmetros genéticos (herdabilidade, ganho de seleção), para cada variável aferida nos experimentos e *ranquear* as 30 primeiras progênes em cada ciclo (ambiente). Obter os intervalos de confiança (IC) aproximados pelo método delta e obter o *Highest Posterior*

Density (HPD) para herdabilidade de todas as características avaliadas nos experimentos.

- d) Obter estimativas da capacidade de predição (CP), raiz quadrada do erro quadrático médio (REQM) para os modelos mistos e modelos bayesianos para todas as características avaliadas nos experimentos via metodologia 5-folds.
- e) Avaliar a significância dos fatores linhas e colunas para as variáveis associadas a características produtivas quando a técnica *post-hoc blocking Row-Col* é inserida nos modelos.
- f) Obter estimativas dos índices de coincidência das progênes selecionadas por cada método de estimação.
- g) Selecionar as progênes superiores para todas as características avaliadas em cada ciclo (ambiente).

2.REVISÃO DE LITERATURA

2.1 Aspectos gerais do maracujazeiro-azedo

Passiflora edulis é uma espécie nativa do Brasil e conhecida popularmente como maracujá-amarelo, maracujá-azedo e maracujá-roxo. É uma espécie de maior importância econômica por causa de bons preços, principalmente, no mercado de frutíferas frescas. É cultivado em praticamente todo território nacional, sendo que o Brasil se destaca como maior produtor e consumidor mundial (Junghans, 2015).

Segundo a pesquisa Produção Agrícola Municipal, realizada pelo Instituto Brasileiro de Geografia e Estatística (Instituto Brasileiro de Geografia e Estatística, 2018) a produção chegou a 602651 toneladas cobrindo uma área cultivada de 43248 ha. A cultura do maracujazeiro que é amplamente disseminada no país é cultivada em todos os estados da federação, com destaque para o nordeste, que em 2018 correspondeu com a proporção de 62% de toda produção nacional.

Os estados da Bahia, Ceará, Santa Catarina, São Paulo e Minas Gerais são, pela ordem, os que mais produziram a fruta em 2018, e, entre esses, se destacam a Bahia (160902 t) e o Ceará (147458 t) que juntos, no mesmo ano, produziram aproximadamente 51,5% da produção nacional. Segundo o Instituto Brasileiro de Geografia e Estatística (2018) a área de plantio no país diminuiu em 31% entre os anos de 2010 (62401 ha) e 2018 (43248 ha), havendo um decréscimo, na produção, de 35% neste mesmo período. Considerando o volume

e a área total cultivada com maracujá, o *Passiflora edulis*, corresponde a quase que a totalidade da área cultivada e também ao volume de produção.

Estima-se que mais de 60% da produção brasileira seja destinada ao consumo *in natura*, comercializado através de sacolões, feiras e supermercados. O restante da produção é destinado às indústrias de processamento, onde o suco é o principal produto (Rossi et al., 2001). Também têm uso medicinal, culinário, ornamental, alimentar, doces, geleias, néctares, farinha, refrescos, sorvetes e iogurtes.

Em 2018, a produção de maracujá no estado do Rio de Janeiro foi de 6959 toneladas, podendo ser considerada baixa se comparada com as 160902 toneladas do maior produtor que é o estado da Bahia. Quando se trata de produtividade média, o estado do Rio de Janeiro produziu 13565 kg ha^{-1} , que é uma produtividade abaixo da média nacional, e pode ser considerada baixa quando comparada com a maior produtividade média brasileira que é do Distrito Federal com 27675 kg ha^{-1} (Instituto Brasileiro de Geografia e Estatística, 2018). O baixo rendimento médio se deve a alguns fatores, entre eles, o reduzido número de cultivares disponível para produtores no mercado, além de doenças virais e fúngicas (Santos et al., 2015).

A baixa produtividade, causada, entre outros fatores, pela falta de genótipos altamente produtivos, pela grande variabilidade existente em pomares comerciais, pela carência de cultivares melhoradas que sejam adaptadas às condições do solo, clima e ambiente das regiões de cultivo, tem refletido na necessidade do melhoramento genético (Gonçalves et al., 2007), e ocasionado a diminuição da vida útil dos novos plantios (Lima e Cunha, 2004; Cavichioli et al., 2011).

No Brasil, os inúmeros desafios e oportunidades para esta cultura se estendem às ações de pesquisas, transferência de tecnologia e inovação. Torna-se necessário, portanto, o desenvolvimento de novas cultivares resistentes a pragas e doenças e de tecnologias de produção e processamento que contribuam para maior sustentabilidade econômica da cultura (Faleiro et al., 2017).

Neste contexto, torna-se necessário o desenvolvimento de novas tecnologias com apoio de novas metodologias de estimação de parâmetros genéticos para que seja possível identificar progênies promissoras para compor os genitores dos próximos ciclos de seleção recorrente.

2.2 Melhoramento do maracujazeiro azedo na UENF via seleção recorrente intrapopulacional

Os esquemas de seleção recorrente são classificados em duas categorias: seleção recorrente intrapopulacional e interpopulacional.

A seleção recorrente é um método bastante eficaz e pode ser utilizado para melhoramento da cultura do maracujazeiro-azedo. Em seus estudos, Borém (2017) ressalta que os métodos intrapopulacionais, em geral, são mais comumente utilizados do que os interpopulacionais, pois são de mais fácil execução e aplicáveis à maioria das características agrônômicas. No caso do programa em desenvolvimento pela UENF, como já estimado, os efeitos de aditividade são controladores para a maioria das características sob seleção, sendo assim, é recomendado um método intrapopulacional (Ribeiro et al., 2019).

Segundo Souza Júnior (2001), cada ciclo da seleção recorrente é constituído de quatro etapas: obtenção de progênies, avaliação destas em experimento com repetições, seleção das progênies superiores e recombinação destas.

Nestas diversas etapas é de fundamental importância as estimativas dos componentes de variância, estimativas de herdabilidade, assim como, a predição dos ganhos genéticos a serem obtidos. Todos esses parâmetros genéticos são de fundamental importância para a condução do programa de melhoramento, pois permitem o monitoramento da variabilidade genética disponível para o melhorista na população sob seleção, visto que como seu sustentáculo, a seleção recorrente preconiza o aumento das medias populacionais mantendo a variância genética.

Devido à baixa produtividade de maracujá no estado do Rio de Janeiro e a busca para contornar os efeitos da mesma, na UENF vem se desenvolvendo, há cerca de 20 anos o Programa de Melhoramento Genético do maracujazeiro-azedo, via seleção recorrente intrapopulacional, com a obtenção de resultados promissores (Viana et al., 2004; Ferreira et al., 2016; Viana et al., 2016; Silva et al., 2017; Ribeiro et al., 2019).

O programa de melhoramento começou com uma amostragem inicial de vários genótipos em três regiões produtoras e distintas do Estado do Rio de Janeiro (Viana et al., 2004). Com base nestes estudos preliminares, procurou-se em uma segunda etapa realizar uma amostragem mais ampla em áreas

comerciais da Região Norte Fluminense, onde por meio de delineamentos genéticos apropriados obteve-se um maior número de progênies.

Assim, com base nos resultados destas análises iniciais, optou-se pela elaboração de um programa de seleção recorrente. O primeiro ciclo de seleção recorrente foi conduzido por Silva (2009), com a denominação de Maracujá-Amarelo. O segundo ciclo de seleção recorrente foi conduzido por Ferreira et al. (2016) onde foram avaliadas 28 progênies de meios-irmãos resultantes da recombinação de 30% das melhores progênies das 140 progênies de irmãos completos, selecionadas via índice de seleção no primeiro ciclo de seleção recorrente (Silva, 2009).

Pesquisadores como Silva et al. (2016) conduziram o terceiro ciclo de seleção recorrente do maracujazeiro-azedo, avaliando 81 progênies de irmãos completos. Após o terceiro ciclo de seleção recorrente do maracujazeiro-azedo, em agosto de 2015 foi disponibilizado pelo Ministério da Agricultura, o registro de um novo cultivar e a autorização de produção de sementes e o cultivo da cultivar UENF Rio Dourado, visando à recomendação de uma variedade produtiva e adaptada para as regiões Norte e Noroeste Fluminense.

Após essa etapa, os trabalhos de seleção recorrente foram continuados, com a estruturação de novo ciclo de seleção. Experimentos foram estruturados com foco em resultados obtidos nas etapas anteriores. Um dos pontos levantados foi a utilização de um maior controle experimental via uso de técnicas experimentais *a posteriori*. De como esse controle afetaria as estimativas dos componentes de variância e a consequente seleção dos novos genótipos superiores, para a continuidade dos ciclos de seleção, de acordo com trabalho realizado por Silva et al. (2016) e Machado et al. (2020).

2.3 Técnica Post-hoc blocking Row-Column (Row-Col)

O delineamento em blocos casualizados, em geral, é eficiente quando a variabilidade dentro das repetições é relativamente pequena, o que é raro quando um grande número de genótipos é testado.

A grande parte dos experimentos de maracujazeiro-azedo, incluindo os ensaios conduzidos na UENF, utilizam o delineamento em blocos casualizados (DBC). Uma alternativa de modelagem neste trabalho é a utilização da técnica

post-hoc blocking Row-Col, para modelar as variáveis via modelos mistos e bayesianos.

A técnica *post-hoc blocking Row-Col* é uma técnica *posteriori* e foi proposta inicialmente por Patterson e Hunter (1983). Consiste na sobreposição a *posteriori* de um *gride* de linhas e colunas sobre o croqui do delineamento em blocos ao acaso (Gezan et al., 2006). No entanto, em um experimento avaliado em *Row-Col*, além dos efeitos de blocos e tratamento, estão presentes os efeitos de linhas e colunas (Silva et al., 2016).

Alguns autores tem relatado a eficiência da técnica *post-hoc blocking Row-Col* (Silva et al., 2016; Machado et al., 2020) e concluíram que a mesma pode ser usada no melhoramento genético de maracujazeiro-azedo.

Nessa direção, Machado et al. (2020) concluíram que a técnica *post-hoc blocking Row-Col* apresentou melhor ajuste às observações quando comparados ao desenho original em blocos aleatórios, para características de massa, número de frutos sólidos solúveis totais e as demais variáveis não apresentaram diferença significativa entre os modelos. Ainda segundo estes autores é uma técnica de baixo custo que fornece melhorias na precisão experimental e, conseqüentemente melhorias na seleção de genótipos superiores.

Segundo Silva et al. (2016) numa comparação entre blocos casualizados e a técnica *post-hoc blocking Row-Col* as estimativas de herdabilidade eram melhores quando se trabalha com a técnica *post-hoc blocking Row-Col*. Segundo estes autores a técnica *post-hoc blocking Row-Col* foi eficaz fornecendo estimativas mais precisas em relação a análise original em blocos aleatórios. E, recomendaram que a técnica seja aplicada aos ciclos de seleção recorrente subsequentes do programa de melhoramento de maracujá da UENF.

2.4 Métodos de estimação de componentes de variância frequentistas e Bayesianos

2.4.1 Modelos Mistos

O melhoramento genético é uma ciência que tem como objetivo principal aumentar a proporção de alelos favoráveis para característica de interesse, e, assim, aumentar a eficiência de produção de um indivíduo ou de uma população

(Silva et al., 2014). Para isto, a utilização de métodos tradicionais de seleção, aliados a princípios biométricos e estatísticos, como modelos mistos, entre outros, foi uma das maiores contribuições para a genética quantitativa, pois, permite o pesquisador analisar observações experimentais e gerar informações úteis, de modo a potencializar o sucesso do esquema seletivo adotado (Cruz, 2012).

O uso de técnicas de melhoramento genético tem sido um grande aliado dos pesquisadores, pois, permitem a manipulação de caracteres quantitativos de interesse de modo a selecionar e recombinar as formas genéticas mais adaptadas, de melhor qualidade e mais eficientes (Cruz, 2012). Para tanto, torna-se necessário que estudos sobre os parâmetros populacionais envolvidos sejam realizados.

Neste contexto, a metodologia de modelos mistos REML/BLUP é uma alternativa para a estimação de parâmetros genéticos e de componentes de variância, pois, tal metodologia de estimação é pelo método de máxima verossimilhança residual (REML) e a predição dos valores genotípicos é a melhor predição linear não viciada (BLUP). Tal metodologia é denominada genericamente de modelos mistos resultando assim, em um processo de seleção mais acurado (Viana e Resende, 2014). Assim, o REML mantém as demais propriedades de máxima verossimilhança (ML) de suficiência, consistência, eficiência e invariância a translação e permite também a imposição de restrições de não negatividade (Resende et al., 2014).

De maneira geral, argumentos teóricos e evidências indicam que inferências pontuais realizadas a partir de funções de verossimilhança não são afetadas por algumas formas de seleção (Gialola et al., 1989). Essa propriedade fez com que o método REML se tornasse padrão para a estimação de componentes de variância em programas de melhoramento genético (Resende et al., 2014).

A metodologia de estimação REML/BLUP tem sido utilizada por diversos autores no melhoramento vegetal (Santos et al., 2015; Entringer et al., 2016; Ferreira et al., 2016; Silva et al., 2016, 2017; Machado et al., 2020).

Autores como Silva et al. (2017) comparou quatro índices de seleção com REML/BLUP e concluiu que o REML/BLUP foi mais eficiente e oferece maior ganho genético em relação aos índices testados. O índice de seleção de Mulamba e Mock (1978) proporcionou os melhores resultados para a seleção das

progênes. O método REML/BLUP mostrou-se mais eficiente, tendo selecionado progênes com ganhos genéticos preditos melhores que os dos índices de seleção testados.

Segundo Santos et al. (2015) o uso de modelos mais sofisticados, como REML/BLUP, possibilita a obtenção do melhores estimativas para os parâmetros genéticos. Tal método de estimação leva em consideração os valores de estimativa de componentes de variância que são estimados por REML/BLUP. Segundo Santos et al. (2015) a estratégia utilizada (REML/BLUP) foi eficaz na identificação genótipos superiores de maracujá, principalmente, para número de frutos por planta, massa média de polpa que são as principais características avaliadas no início do programa, e que permitirá ganhos mais rentáveis devido à sua seleção.

Segundo Entringer et al. (2016) o estudo teve como objetivo comparar as estimativas de ganhos genéticos utilizando quatro índices de seleção, com base no método dos mínimos quadrados e o índice aditivo através do REML/BLUP. Foram testados vários índices de seleção e dentre os quatro índices de seleção baseados na ANOVA (análise de variância), o Mulamba e Mock (1978) foi o mais adequado para a seleção de famílias de meios-irmãos em milho superdoce. O índice REML/BLUP apresentou ganhos genéticos preditos ainda melhores do que Mulamba e Mock (1978), e foi eficiente para selecionar famílias de meios-irmãos em milho superdoce. O alto nível de coincidência entre os índices REML/BLUP e Mulamba e Mock (1978) indica eficiência semelhante para fins de seleção. Porém, o método REML/BLUP apresentou melhores ganhos genéticos, podendo ser recomendado o uso para futuras atividades de seleção.

Análises de Silva et al. (2016) concluiu que o método REML/BLUP mostrou-se mais eficiente, tendo selecionado progênes com ganhos genéticos preditos melhores que os outros índices de seleção testados. Os autores concluíram que a técnica *blocking Row-Col* ajustam as observações melhor ou igual ao delineamento bloco completos casualizados fornecendo estimativas mais precisas. Os autores indicaram implementação da metodologia *blocking Row-Col* nos ciclos subsequentes do programa de melhoramento de maracujá da UENF, em vez de usá-lo como uma ferramenta de correção.

Segundo Ferreira et al. (2016), a metodologia REML/BLUP mostrou-se adequada para a avaliação, possibilitando obter estimativas dos parâmetros

genéticos e fenotípicos que revelaram a possibilidade de sucesso com a seleção de progênes superiores. Nas estimativas dos parâmetros genéticos via REML, as duas características ligadas diretamente à produção e, portanto, consideradas as mais importantes, número de frutos por parcela (NF) e massa total de frutos por parcela (MTF) apresentaram estimativas de herdabilidade média de progênes de 0,395 e 0,439, respectivamente.

2.4.2 Modelos Bayesianos

Segundo Congdon (2007), a inferência bayesiana tem várias vantagens na modelagem estatística e análise de observações sobre a inferência clássica. Tal método de estimação fornece uma maneira de formalizar o processo de aprender com os dados para atualizar as crenças de acordo com as noções recentes de síntese do conhecimento. A estatística bayesiana proporciona a utilização de dados como o aprendizado de uma geração anterior, a entrada desta geração e o que não é possível na estatística frequentista. No entanto, no passado, a análise estatística baseada no teorema de Bayes era assustadora e inexecutável por causa das integrações e aproximações numéricas necessárias. Contudo, recentemente os métodos bayesianos podem melhorar os estimadores clássicos em termos da precisão de estimativas.

Além de proporcionar inferências distintas a estimação Bayesiana difere da estimação por máxima verossimilhança devido ao fato de maximizar a distribuição a *posteriori* do parâmetro ao invés da função de verossimilhança (Resende et al., 2014). Segundo Paulino et al. (2018), a abordagem de estimação por máxima verossimilhança foi empregada de forma quase unânime durante a primeira metade do século XX.

Segundo Resende et al. (2014) se a informação a *priori* encontrar-se disponível, a estimação bayesiana deve ser preferível a estimação por máxima verossimilhança, e, além disto, se a distribuição a *priori* é não informativa e/ou a quantidade de observações é muito grande a estimação bayesiana converge para a estimação máxima verossimilhança, ou seja, ambas são equivalentes.

Devido às excelentes propriedades teóricas e práticas da análise bayesiana, ela tem se tornado rotineira em melhoramento de espécies perenes,

transformando-se em procedimento padrão de estimação de componentes de variância e de predição de valores genéticos (Resende et al., 2014).

A inferência Bayesiana baseia-se fundamentalmente no Teorema de Bayes e, utiliza probabilidades condicionais (Paulino et al., 2018). Considerando a amostra de uma população Y_1, Y_2, \dots, Y_n indexada pelo parâmetro θ (ou um vetor de parâmetros). Sob o ponto de vista Bayesiano o parâmetro θ é uma variável aleatória ou um conjunto de variáveis aleatórias com dada distribuição de probabilidade, enquanto que em inferência frequentista o mesmo parâmetro θ é uma quantidade fixa e desconhecida. Os modelos Bayesianos estão comumente preocupados com as inferências sob um conjunto de parâmetros $\theta = (\theta_1, \theta_2, \theta_3, \dots, \theta_n)$ de dimensão n , em que os parâmetros são quantidades aleatórias (Congdon, 2007). Pelo teorema de Bayes, o conhecimento prévio sobre os parâmetros é identificado como, informação *a priori* ou distribuição *a priori*, representado por $h(\theta)$. A informação *a priori* ao ser atualizada com a verossimilhança dos dados $f(Y|\theta)$ resulta na distribuição *a posteriori* $h(\theta|Y)$. Assim pelo teorema de Bayes, tem-se:

$$h(\theta|Y) = \frac{f(Y|\theta) \times h(\theta)}{\int_{\Theta} f(Y|\theta) \times h(\theta) d\theta}; \theta \in \Theta$$

em que, o denominador representa a verossimilhança marginal e, pode ser considerada como uma constante normalizadora, para garantir que $h(\theta|Y)$ seja uma densidade adequada. Assim, o teorema de Bayes pode ser apresentado como:

$$h(\theta|Y) \propto f(Y|\theta) \times h(\theta); \theta \in \Theta$$

em \propto que representa proporcionalidade. A partir de então, os resultados são avaliados e validados através da distribuição *a posteriori*, em que, de acordo com alguns autores como Congdon (2007), Paulino et al. (2018) a influência relativa gerada pela distribuição *a posteriori* depende do peso que é dado à *a priori* e do tamanho da amostra utilizada.

A distribuição *a priori* é qualquer informação que se tem sobre o problema antes da realização do experimento e segundo Congdon (2007) como escolher a densidade ou informação *a priori* é uma questão importante na inferência bayesiana. Muitos autores tem usado a inferência Bayesiana na estimação de

parâmetros no melhoramento de diversas culturas (Silva et al., 2013, 2018). Silva et al. (2013) discutem que tanto a escola Bayesiana e a frequentista estão bem estabelecidas, mas que há vantagens da escola Bayesiana em alguns aspectos, por exemplo o uso de *prioris* para os parâmetros em estudo. Mas que em geral nenhuma abordagem pode ser considerada superior em todas as situações.

Em suas investigações Silva et al. (2018) compararam a abordagem bayesiana e os métodos frequentistas para estimar as médias e os parâmetros genéticos em experimentos multiambientes de soja. A soja avaliada pela abordagem bayesiana apresentou estimativas de parâmetros genéticos próximos de modelos mistos (REML/BLUP) para as seguintes métricas, tais como: herdabilidade média, acurácia da seleção dos genótipos, coeficiente de variação genético e coeficiente de variação ambiental. Os autores concluíram, em experimentos multiambientes, que a metodologia de modelos mistos e a abordagem bayesiana produzem resultados similares de estimativas dos parâmetros genéticos.

2.4.2.1 Distribuição a priori

Segundo Congdon (2007) o conhecimento da distribuição *a priori* sob o parâmetro de interesse é uma importante etapa do processo de inferência, e pode apresentar influência significativa na obtenção da distribuição *a posteriori* dependendo de como for estabelecida. A distribuição, *a priori*, trata-se de uma distribuição de probabilidade do parâmetro (ou conjunto de parâmetros) de interesse, θ , antes do processo de experimentação, em que θ é uma variável aleatória (Paulino et al., 2018), e, segundo estes mesmos autores as *prioris* podem ser classificadas basicamente em informativas e não informativas. A escolha e decisão por *prioris* informativas é bem vinda quando existe alguma informação relevante sobre o tema. No entanto, a elicitación destas informações deve ser realizada com cuidado, pois informações equivocadas podem levar a resultados também equivocados e que apresentam maior impacto quando o número de observações for pequeno, pois, neste caso a função de verossimilhança terá um peso menor que a distribuição *a priori*.

Uma característica interessante que pode contribuir no cálculo da *posteriori* é a utilização de famílias conjugadas, que ocorre quando a distribuição

a *posteriori* tem a mesma distribuição da *priori* escolhida, porém, com parâmetros atualizados (Gelman et al., 2003; Paulino et al., 2018). Segundo Gelman et al. (2003) a utilização de *prioris* conjugadas são matematicamente convenientes, uma vez que a distribuição a *posteriori* tem uma forma paramétrica conhecida e os cálculos computacionais tornam-se mais simples, assim, determinando assim qual algoritmo (Metropolis-Hastings e Amostrador de Gibbs) a ser usado para aproximar integrais múltiplas, e, solucionar problemas complexos computacionalmente do espaço multidimensional para obter a aproximação das distribuições marginal a *posteriori* pelas distribuições condicionais completas a *posteriori* (DCCP).

A etapa de escolha e de discussão de informações prévias parece não ser mais relevante, com a diferença entre as duas abordagens (frequentistas e bayesianas) bem compreendidas e as virtudes de cada uma delas utilizadas quando mais convenientes (Gamerman, 1996). Segundo Thompson et al. (2005), a evidência a *priori* nem sempre é bem explicada. Por exemplo, no Congresso Mundial de Genética Aplicada à Produção Pecuária de 2002, mais de trinta artigos apresentaram análises bayesianas de parâmetros quantitativos e genéticos. Mais da metade dos trabalhos não quantificou o conhecimento prévio, e um terceiro usou conhecimento prévio vago ou plano para os parâmetros genéticos.

2.4.2.2 Extensão de Parâmetros

A expansão de parâmetros é um método que foi originalmente desenvolvido por Liu et al. (1998) para acelerar o algoritmo *Expectation Maximization* (EM). Foi considerado particularmente para modelos de efeito aleatório por van Dyk e Meng (2001), Browne (2004) e Gelman et al. (2008). E, que segundo Gelman et al. (2004) desde então foi adaptado com algoritmos de amostrador de Gibbs e Metropolis (Liu e Wu 1999; van Dyk e Meng 2001; Gelman et al., 2003). A extensão de parâmetros não é necessária em todas as análises Bayesiana, mas neste presente trabalho foi necessária, pois, existem componentes de variância próximos de zero e as cadeias não convergiam sem a utilização da mesma.

Neste trabalho a *priori* (Bayes II) foi usado a extensão de parâmetros como está implementado na biblioteca MCMCgIimm através dos parâmetros $\alpha.\mu=0$ e $\alpha.V=25^2$ como sugerido por (Hadfield, 2010).

2.4.2.3 Verificação da Convergência de cadeias

A análise da convergência das cadeias é importante na inferência bayesiana, pois, assim, garante que as distribuições condicionais completas a *posteriori* (DCCPs) tem uma boa aproximação da distribuição marginal a *posteriori* (DMP) dos parâmetros. Os indicativos de estacionariedade de uma cadeia podem ser avaliados utilizando alguns métodos difundidos na literatura. Autores como Nogueira et al. (2004) e outros sugeriram os seguintes testes, para análise da convergência:

- a) Raftery e Lewis (1992) em uma amostra piloto e determinação do tamanho ideal da série;
- b) Monitorar a convergência das sequências nas proximidades do tamanho ideal, indicado pelo critério de Raftery e Lewis (1992), por meio do critério de Geweke (1992);
- c) Obter o tamanho do burn-in, ou seja, do número de iterações descartadas para se retirar o efeito dos valores iniciais do parâmetro, pelo critério de Heidelberger e Welch (1983).

O método proposto por Raftery e Lewis (1992) tem como objetivo, indicar o menor número de iterações que indica a estabilização da cadeia, utilizando-se do valor de descarte estimado para a amostra piloto (*burn-in*) e o número de iterações. A dupla Raftery e Lewis (1992) basearam-se em um valor estimado (Fator de Dependência), em que se este valor for muito maior que 1, há indicativo de não estacionariedade e estes autores obtiveram ainda como valor máximo de referência o número 5, sendo que, para valores superiores a 5, sugere-se que não há estacionariedade da cadeia.

Heidelberger e Welch (1983) parte da hipótese de que, há estacionariedade na cadeia. Se a hipótese for rejeitada o teste indica o descarte inicial de 10% das iterações, realizando o teste novamente e, se ainda for

rejeitada a hipótese de estacionariedade, faz-se o descarte progressivo de mais 10%, chegando até no máximo à 50% de descarte dos valores iniciais. Assim após esta metodologia, se ocorrer a rejeição da hipótese nulidade, existe indicativo de falha na convergência da cadeia, sugerindo um número maior de iterações.

Segundo Geweke (1992), as amostras geradas pelo algoritmo de Gibbs e outros algoritmos, em geral, não são independentes e identicamente distribuídas, o que dificulta a análise da convergência das cadeias e gera aproximações numéricas com baixa precisão para o valor esperado das distribuições a *posteriori*. O teste de Geweke (1992), basicamente, estabelece a comparação de duas partes da cadeia, dividindo-a entre a parte 1 composta por 10% da sequência inicial da amostra gerada e, a parte 2, com os 50% finais da mesma sequência. As médias entre as partes são comparadas via teste de hipótese baseando-se na estimativa do valor-p (Pereira et al., 2009). Se não existir evidências para rejeitar a hipótese de que as médias são iguais, há indícios da estacionariedade da cadeia.

2.4.2.4 Avaliadores e Critérios de seleção de modelos

O critério de informação Deviance (DIC) Spiegelhalter et al. (2002), é uma estatística muito utilizada em inferência Bayesiana. Este critério serve como medida de comparação quanto à qualidade do ajuste de modelos baseados no máximo da função de verossimilhança, sendo dependentes do número de observações e dos parâmetros do modelo em estudo, ou ainda na dificuldade de ajuste do modelo (Sousa et al., 2014). Os critérios partem de princípios semelhantes, baseados no máximo da função de verossimilhança. Eles levam em consideração a quantidade de parâmetros dos modelos e têm como característica evitar modelos superparametrizados.

O critério de informação Deviance (DIC) foi proposto por Spiegelhalter et al. (2002) como uma generalização do critério de informação de Akaike - AIC (Akaike, 1973) para ser utilizado sob o enfoque Bayesiano. O critério de informação de Akaike é um critério para ser usado sob enfoque frequentista.

O DIC é dado por: $DIC = D(\bar{\theta}) + 2p_D$, em que: $p_D = D(\bar{\theta}) - \bar{D}(\theta)$, $D(\theta) = -2 \cdot \ln L(\hat{\theta}) + V$, $L(\hat{\theta})$ corresponde ao máximo da função de

verossimilhança ao utilizar as médias a *posteriori*, $D(\bar{\theta})$ consiste na média dos valores da *deviance* aplicada em cada iteração, $\bar{D}(\theta)$ consiste no valor da *deviance* aplicada no valor médio de θ (estimador bayesiano usual) considerando as iterações após o *burn-in*, e V é uma constante. O DIC é utilizado para a comparação de modelos, de forma que quanto menor o valor de DIC melhor o modelo, ou seja, mais indicado é o modelo para descrever o fenômeno estudado. Segundo Spiegelhalter et al. (2002) dois modelos com a diferença entre DICs menores ou igual a 2 (dois) podem ser considerados igualmente parcimoniosos.

2.4.2.5 Intervalo Highest Posterior Density (HPD)

Depois de obtida a informação a *posteriori* do parâmetro θ faz-se necessário obter um intervalo (uma região) de estimativas prováveis com certo grau de confiabilidade, e, assim, eliminando o peso de somente uma estimativa pontual como a média, moda e mediana. Esta região de credibilidade é similar ao intervalo de confiança (*IC*) da inferência frequentista, porém, com interpretação diferente. O intervalo de credibilidade (*ICr*) de máxima densidade a *posteriori* é denominado *Highest posterior density* (HPD).

2.4.3 Validação cruzada

As medidas estatísticas utilizadas para avaliar a qualidade de ajuste dos modelos sob o ponto de vista frequentistas (BIC, AIC, coeficiente de determinação ajustado, dentre outras) não permitem informações a respeito da previsibilidade de um modelo.

A previsibilidade dos modelos e o poder de generalização podem ser obtidos pela metodologia k-folds. Segundo James et al. (2013) sugerem que $\frac{2}{3}$ das informações estejam presentes no conjunto de treinamento e conseqüentemente, $\frac{1}{3}$ no conjunto de validação. As únicas restrições do método k-folds é que estes procedimentos assumem é que os dados sejam identicamente distribuídos e que o treinamento e validação sejam independentes, fato que pode ser contornado (Arlot e Celisse, 2010).

Segundo Borra e Ciaccio (2010), o viés do método k-folds diminui quanto maior for o valor de k, porém, um valor de k muito elevado acaba elevando o

custo computacional da técnica, além de implicar em uma amostra de teste pequena o que aumenta a variância. Na literatura, se discute qual valor de k seria o ideal, sendo as opções mais usuais os valores de $k = 2, 5$ ou 10 , (Kohavi, 1995) faz esta discussão e afirma que em seus estudos $k=10$ tem o melhor desempenho. Para avaliar a acurácia, confiabilidade e eficiência dos modelos os trabalhos podem contemplar estudos de algumas estatísticas (métricas) que informam sobre as acurácias seletivas e preditivas dos modelos.

2.5 Estimação de Valores Genéticos no melhoramento de Plantas

Um motivo para fazer uso do modelo misto é a possibilidade de fazer predição de efeitos aleatórios, através do BLUP e obter estimativa BLUE para efeitos fixos. O procedimento ótimo de estimação de componentes de variância é o REML (máxima verossimilhança residual). Na maioria das vezes, a estimação dos componentes de variância e a predição de valores genéticos são realizados, simultaneamente, pelo procedimento REML/BLUP, e, neste caso, o BLUP é denominado de BLUP empírico (Resende et al., 2014).

O uso da metodologia de modelos mistos é uma tendência no melhoramento genético de plantas por fornecer parâmetros que auxiliam na identificação de genótipos superiores. A metodologia REML/BLUP permite a predição acurada e não viesada dos valores genéticos mesmo sob desbalanceamento, propiciando estimativas mais precisas dos componentes de variância e dos valores genéticos individuais. O melhoramento genético depende do sucesso na escolha dos melhores indivíduos para serem os genitores das próximas gerações (Cruz e Carneiro, 2006).

No melhoramento de plantas, as técnicas de avaliação genética desempenham papel fundamental, pois permitem a predição dos valores genéticos dos candidatos à seleção, propiciando uma seleção mais acurada (Resende, 2000).

O procedimento de predição de valores genéticos é basicamente o BLUP individual para os casos de dados balanceados ou desbalanceados. A metodologia dos modelos mistos foi proposta por Henderson (1949) para ser utilizada na avaliação genética de bovinos, e foi apresentada pela primeira vez em 1973 (Henderson, 1973), passando a ser utilizada na prática a partir da década de

80, devido aos avanços tecnológicos computacionais que permitiram seu uso, conforme descrito por Resende (2002a).

O conhecimento prévio dos componentes de variância e covariância é necessário na predição dos valores genéticos, quando se faz o uso de métodos de predição como BLUP (melhor predição linear não-viesada). Entretanto, estes componentes geralmente não são conhecidos e podem ser estimado por vários métodos, entre eles o da máxima verossimilhança residual (REML), que é o recomendado por modelos lineares mistos e observações desbalanceadas (Meyer, 1986).

Uma das dificuldades para estimação de ganhos genéticos mais expressivos no melhoramento de plantas é a baixa acurácia seletiva. Para resolver isto, o procedimento mais preciso de predição de valores genéticos é o BLUP, com uso de componentes de variância estimados via REML - máxima verossimilhança residual - (Resende, 2004).

A inferência bayesiana é uma alternativa ao método clássico para estimação dos componentes de variância e dos efeitos fixos, bem como para predição dos efeitos aleatórios do modelo (Lindley e Smith, 1972; Box e Tião, 1973, Resende et al., 2014). Dentro do contexto bayesiano, qualquer inferência em relação às quantidades desconhecidas do modelo é obtida, a partir da distribuição marginal a *posteriori* (Sorensen e Gianola, 2002).

Modelos lineares mistos e inferência bayesiana podem prever o valor genético com alta acurácia, permitindo aumentar a intensidade de seleção, sem receio de deixar de selecionar os melhores indivíduos (Ludke et al., 2017).

A metodologia REML/BLUP vem sendo empregada na estimação de parâmetros genéticos no melhoramento de diversas culturas (Sousa et al., 2017; Santos et al., 2018; Stida et al., 2018).

Segundo Sousa et al. (2017) a seleção via REML/BLUP para a produção de vagens e grãos possibilitou a predição e obtenção de ganhos genéticos significativos para o melhoramento do feijão-vagem das próximas gerações. Permitiu ainda a obtenção de progênies superiores, a partir de cruzamentos, ambos para produção de vagens e grãos, e, foi possível mesmo em condições de baixa herdabilidade, o que torna REML/BLUP uma metodologia eficaz para obter ganhos genéticos no melhoramento do feijão-vagem.

Análises de Stida et al. (2018) citam o uso de modelos mistos do tipo REML/BLUP que tem sido aplicado com sucesso no melhoramento de diferentes culturas. Neste mesmo trabalho foi obtido dez genótipos superiores para a característica rendimento de capim-elefante quando estimou parâmetros genéticos via metodologia REML/BLUP.

Por sua vez, Santos et al. (2018) relata que a estimação de parâmetros genéticos via REML/BLUP foram adequados para a seleção de genótipos de videira com resistência genética a *P. Brachyurus*.

3. MATERIAL E MÉTODOS

3.1 População avaliada

3.1.1 Experimento I

Foram avaliadas 81 progênies de irmãos completos oriundas do terceiro ciclo do programa de melhoramento genético da UENF via seleção recorrente. Essas progênies foram obtidas a partir de 27 parentais (progênies de meios-irmãos) selecionadas por Silva (2009) via índice de seleção Mulamba e Mock (1978).

3.1.2 Experimento II

Foram avaliadas 97 progênies de irmãos completos oriundas do quarto ciclo do programa de melhoramento genético da UENF via seleção recorrente. Essas progênies foram obtidas a partir do cruzamento de 23 progênies de meio irmãos selecionadas por Cavalcante et al. (2019).

3.2 Descrição e condução do experimento I

O estudo foi conduzido entre 2011-2013, na Unidade Experimental da UENF em Itaocara – Ilha Barra do Pomba (21°40' S, 42°04' W e altitude de 76 m),

Região Noroeste Fluminense. Durante o período experimental a temperatura anual média foi 22,5°C e a precipitação média foi 1041mm.

O delineamento experimental utilizado foi em blocos ao acaso com duas repetições e parcelas constituídas por cinco plantas (810 plantas). Com objetivo de trabalhar com a técnica *post-hoc blocking Row-Col*, foi implantado um *grid* com 13 linhas e 18 colunas sobre o croqui experimental.

3.3 Descrição e condução do experimento II

O experimento foi conduzido em Cambuci - RJ. Em Cambuci, na fazenda Santo Antônio, que pertence também ao Instituto Federal de Educação, Ciência e Tecnologia Fluminense (IFF), situada nas coordenadas 21° 34' 31"S de latitude e 41° 54' 40"W de longitude com altitude de 35 metros ao nível do mar, clima predominante quente e úmido no verão e seco no inverno com precipitação anual média de 1200 mm e temperatura média anual de 23 °C. O delineamento experimental utilizado foi em blocos ao acaso com duas repetições e cada parcelas constituídas por três plantas (582 plantas). Com objetivo de trabalhar com a técnica *post-hoc blocking Row-Col*, foi implantado um *grid* com 19 linhas e 15 colunas sobre o croqui experimental.

3.4 Características agronômicas avaliadas no experimento I

As progênies foram fenotipadas para as seguintes características: Número de frutos (NF): obtido pela contagem do número de frutos totais em cada parcela; produtividade total (Prod.) em kg por parcela, realizando-se colheitas semanais, com posterior pesagem dos frutos; massa do fruto (MF), em gramas, obtida por meio da média aritmética de massa de 15 frutos amostrados por parcela, com o auxílio de uma balança digital; diâmetro longitudinal do fruto (DLF/Comp.), em mm, obtido por meio de medidas das dimensões longitudinais dos frutos com paquímetro digital, amostrando 15 frutos por parcela; diâmetro transversal do fruto (DTF/Diam.), em mm, obtido por meio de medida da dimensão transversal dos frutos com paquímetro digital, amostrando 15 frutos por parcela; porcentagem de polpa do fruto (PP), obtida pela razão entre a massa de polpa e a massa total do fruto, amostrando 15 frutos por parcela; espessura da

casca (EC), em mm, determinada na porção mediana dos frutos (cortados transversalmente, no sentido de maior diâmetro), com utilização de paquímetro digital, utilizando amostras de 15 frutos por parcela; teor de sólidos solúveis totais (SST), em °Brix, obtido por refratometria utilizando-se refratômetro digital portátil ATAGO N1, com leitura na faixa de 0 a 32° graus de Brix. As leituras foram feitas em alíquotas de suco da polpa em amostras de 15 frutos.

3.5 Características agronômicas avaliadas no experimento II

As progênes foram fenotipadas para as seguintes características: Número de frutos (NF): obtido pela contagem do número de frutos totais em cada parcela; produtividade total em um ano (Prod.) em kg por parcela, realizando-se colheitas semanais, com posterior pesagem dos frutos; massa do fruto (MF), em gramas, obtida por meio da média aritmética de massa de 15 frutos amostrados por parcela, com o auxílio de uma balança digital; diâmetro longitudinal do fruto (DLF/Comp.), em mm, obtido por meio de medidas das dimensões longitudinais dos frutos com paquímetro digital, amostrando 15 frutos por parcela; diâmetro transversal do fruto (DTF/Diam.), em mm, obtido por meio de medida da dimensão transversal dos frutos com paquímetro digital, amostrando 15 frutos por parcela; porcentagem de polpa do fruto (PP), obtida pela razão entre a massa de polpa e massa total do fruto, amostrando 15 frutos por parcela; espessura da casca (EC), em mm, determinada na porção mediana dos frutos (cortados transversalmente, no sentido de maior diâmetro), com utilização de paquímetro digital, utilizando amostras de 15 frutos por parcela; teor de sólidos solúveis totais (SST), em °Brix, obtido por refratometria utilizando-se refratômetro digital portátil ATAGO N1, com leitura na faixa de 0 a 32° graus de Brix. As leituras foram feitas em alíquotas de suco da polpa em amostras de 15 frutos.

3.6 Parâmetros genéticos

Dentre os parâmetros genéticos de maior importância, destacam-se as variâncias genéticas, as correlações e as herdabilidades (Cruz e Carneiro, 2006). A herdabilidade (h^2) consiste em um parâmetro que expressa a proporção da variabilidade observada devido aos efeitos aditivos dos genes, no caso restrito, ou

a variância genética total, no caso amplo, ou seja, representa a proporção herdável da variabilidade genética (Borém, 2017). Esta proporção da variância é quantificada pela equação abaixo.

$$h^2 = \frac{\sigma_G^2}{\sigma_F^2}$$

O valor genético de cada progênie foi obtido pela soma de cada efeito genotípico (g) à média geral do experimento (u). O ganho de seleção foi estimado considerando a diferença entre a média das progênies selecionadas em relação à média de todas as progênies do experimento, ou seja, $G_s = 100 \cdot \left(\frac{\bar{X}_s - \bar{X}_G}{\bar{X}_G} \right)$ em que G_s é o ganho de seleção; (\bar{X}_G) é a média geral e (\bar{X}_s) é a média das primeiras progênies selecionadas pelo modelo.

4. MODELOS ESTATÍSTICOS DA AVALIAÇÃO GENÉTICA

Para modelar as variáveis de cada experimento foi proposto dois modelos. O modelo linear misto completo: $y = \mathbf{1}\mu + Z_1u_1 + Z_2u_2 + Z_3u_3 + Z_4u_4 + \varepsilon$ em que: $y_{n \times 1}$ é o vetor de observações fenotípicas; $\mathbf{1}_{n \times 1}$ é uma matriz coluna de uns; u_1 é o vetor de efeitos aleatórios para blocos associado a matriz de incidência Z_1 com distribuição de probabilidade $u_1 \sim N(\mathbf{0}, I\sigma_{bloco}^2)$; u_2 é o vetor de efeitos aleatórios para progênie associado a matriz de incidência Z_2 e com distribuição de probabilidade $u_2 \sim N(\mathbf{0}, I\sigma_{progênie}^2)$; u_3 é o vetor de efeitos aleatórios para linha associado a matriz de incidência Z_3 e com distribuição de probabilidade $u_3 \sim N(\mathbf{0}, I\sigma_{linha}^2)$; u_4 é o vetor de efeitos aleatórios para coluna associado a matriz de incidência Z_4 e com distribuição de probabilidade $u_4 \sim N(\mathbf{0}, I\sigma_{coluna}^2)$; ε é o vetor de erros aleatórios e com distribuição de probabilidade $\varepsilon \sim N(\mathbf{0}, I\sigma_{erro}^2)$.

O modelo linear misto reduzido: $y = \mathbf{1}\mu + Z_1u_1 + Z_2u_2 + \varepsilon$ em que: $y_{n \times 1}$ é o vetor de observações fenotípicas; $\mathbf{1}_{n \times 1}$ é uma matriz coluna de uns; u_1 é o vetor de efeitos aleatórios para blocos associado a matriz de incidência Z_1 e com distribuição de probabilidade $u_1 \sim N(\mathbf{0}, I\sigma_{bloco}^2)$; u_2 é o vetor de efeitos aleatórios para progênie associado a matriz de incidência Z_2 e com distribuição de probabilidade $u_2 \sim N(\mathbf{0}, I\sigma_{progênie}^2)$; ε é o vetor de erros aleatórios e com distribuição de probabilidade $\varepsilon \sim N(\mathbf{0}, I\sigma_{erro}^2)$; em que $\varepsilon | \sigma_{erro}^2 \sim N(\mathbf{0}, I\sigma_{erro}^2)$ e a distribuição conjunta dos dados observados é $E(y) \sim N(\mathbf{1}\mu + Z_1u_1 + Z_2u_2, I\sigma_e^2)$ e $E(y) \sim N(\mathbf{1}\mu + Z_1u_1 + Z_2u_2 + Z_3u_3 + Z_4u_4, I\sigma_e^2)$ para o modelo reduzido e completo respectivamente.

O modelo Bayesiano completo proposto foi: $y = \mathbf{1}\mu + Z_1u_1 + Z_2u_2 + Z_3u_3 + Z_4u_4 + \varepsilon$ em que: $y_{n \times 1}$ é o vetor de observações fenotípicas onde n é o número de progênies; $\mathbf{1}_{n \times 1}$ é uma matriz coluna de uns; u_1 é o vetor de efeitos aleatórios para blocos associado a matriz de incidência Z_1 com distribuição de probabilidade $u_1 \sim N(\mathbf{0}, I\sigma_{bloco}^2)$; u_2 é o vetor de efeitos aleatórios para progênie associado a matriz de incidência Z_2 e com distribuição de probabilidade $u_2 \sim N(\mathbf{0}, I\sigma_{progênie}^2)$; u_3 é o vetor de efeitos aleatórios para linha associado a matriz de incidência Z_3 e com distribuição de probabilidade $u_3 \sim N(\mathbf{0}, I\sigma_{linha}^2)$; u_4 é o vetor de efeitos aleatórios para coluna associado a matriz de incidência Z_4 e com distribuição de probabilidade $u_4 \sim N(\mathbf{0}, I\sigma_{coluna}^2)$; ε é o vetor de erros aleatórios e com distribuição de probabilidade $\varepsilon \sim N(\mathbf{0}, I\sigma_{erro}^2)$; em que $\varepsilon | \sigma_{erro}^2 \sim N(\mathbf{0}, I\sigma_{erro}^2)$.

E a distribuição conjunta dos dados observados é obtida a partir da distribuição: $y | \mu, u_1, u_2, u_3, u_4, \sigma_{progênie}^2, \sigma_{bloco}^2, \sigma_{linha}^2, \sigma_{coluna}^2, \sigma_{erro}^2 \sim N(\mathbf{1}\mu + Z_1u_1 + Z_2u_2 + Z_3u_3 + Z_4u_4, I\sigma_e^2)$ em que: $\mu \sim N(\mu_0, \sigma_0^2)$; $u_1 | \sigma_{progênie}^2 \sim N(\mathbf{0}, I\sigma_{progênie}^2)$; $\sigma_{progênie}^2 \sim GI(v = v_1, V = V_1)$; $u_2 | \sigma_{bloco}^2 \sim N(\mathbf{0}, I\sigma_{bloco}^2)$; $\sigma_{bloco}^2 \sim GI(v = v_2, V = V_2)$; $u_3 | \sigma_{coluna}^2 \sim N(\mathbf{0}, I\sigma_{coluna}^2)$; $\sigma_{coluna}^2 \sim GI(v = v_3, V = V_3)$; $u_4 | \sigma_{linha}^2 \sim N(\mathbf{0}, I\sigma_{linha}^2)$; $\sigma_{linha}^2 \sim GI(v = v_4, V = V_4)$ e ; $\sigma_{erro}^2 \sim GI(v = v_0, V = V_0)$.

4.1 Inferência bayesiana do modelo completo

Para o modelo completo a formulação do Teorema de Bayes é apresentada a seguir:

$$\begin{aligned}
 & P(\mu, u_1, u_2, u_3, u_4, \sigma_{progênie}^2, \sigma_{bloco}^2, \sigma_{coluna}^2, \sigma_{erro}^2 | y) \propto \\
 & P(y | \mu, u_1, u_2, u_3, u_4, \sigma_{progênie}^2, \sigma_{bloco}^2, \sigma_{linha}^2, \sigma_{coluna}^2, \sigma_{erro}^2) \\
 & \quad \times P(\mu | \mu_0, \sigma_0^2) \times P(u_1 | \sigma_{progênie}^2) \times P(u_2 | \sigma_{bloco}^2) \\
 & \quad \times P(u_3 | \sigma_{linha}^2) \times P(u_4 | \sigma_{coluna}^2) \times P(\sigma_{progênie}^2 | v_1, V_1) \times P(\sigma_{bloco}^2 | v_2, V_2) \times P(\sigma_{linha}^2 | v_3, V_3) \cdot \\
 & \quad \times P(\sigma_{coluna}^2 | v_4, V_4) \times P(\sigma_{erro}^2 | v_0, V_0)
 \end{aligned}$$

Após avaliar a estacionariedade das cadeias as distribuições marginal a *posteriori* (DMP) para cada parâmetro foi aproximada pelas distribuições condicionais completas a *posteriori* (DCCP) para cada parâmetro.

$$\begin{aligned}
& P(\mu|y, u_1, u_2, u_3, u_4, \sigma_{progenie}^2, \sigma_{bloco}^2, \sigma_{linha}^2, \sigma_{coluna}^2, \sigma_{erro}^2) \propto \\
& P(y|\mu, u_1, u_2, u_3, u_4, \sigma_{progenie}^2, \sigma_{bloco}^2; \sigma_{linha}^2, \sigma_{coluna}^2, \sigma_{erro}^2) \times P(\mu|\mu_0, \sigma_0^2); \\
& P(\sigma_{progenie}^2|y, \mu, u_1, u_2, u_3, u_4, \sigma_{bloco}^2, \sigma_{linha}^2, \sigma_{coluna}^2, \sigma_{erro}^2) \propto \\
& P(y|\mu, u_1, u_2, u_3, u_4, \sigma_{progenie}^2, \sigma_{bloco}^2; \sigma_{linha}^2, \sigma_{coluna}^2, \sigma_{erro}^2) \\
& \quad \times P(u_1|\sigma_{progenie}^2) \times P(\sigma_{progenie}^2|v_1, V_1); \\
& \sigma_{bloco}^2|y, \mu, u_1, u_2, u_3, u_4, \sigma_{progenie}^2, \sigma_{linha}^2, \sigma_{coluna}^2, \sigma_{erro}^2) \propto \\
& P(y|\mu, u_1, u_2, u_3, u_4, \sigma_{progenie}^2, \sigma_{bloco}^2; \sigma_{linha}^2, \sigma_{coluna}^2, \sigma_{erro}^2) \times P(\\
& \quad P(u_2|\sigma_{bloco}^2) \times P(\sigma_{bloco}^2|v_2, V_2); \\
& \sigma_{linha}^2|y, \mu, u_1, u_2, u_3, u_4, \sigma_{progenie}^2, \sigma_{bloco}^2, \sigma_{coluna}^2, \sigma_{erro}^2) \propto \\
& P(y|\mu, u_1, u_2, u_3, u_4, \sigma_{progenie}^2, \sigma_{bloco}^2; \sigma_{linha}^2, \sigma_{coluna}^2, \sigma_{erro}^2) \times P(\\
& \quad \times P(u_3|\sigma_{linha}^2) \times P(\sigma_{linha}^2|v_3, V_3); \\
& P(\sigma_{coluna}^2|y, \mu, u_1, u_2, u_3, u_4, \sigma_{progenie}^2, \sigma_{bloco}^2, \sigma_{linha}^2, \sigma_{erro}^2) \propto \\
& P(y|\mu, u_1, u_2, u_3, u_4, \sigma_{progenie}^2, \sigma_{bloco}^2; \sigma_{linha}^2, \sigma_{coluna}^2, \sigma_{erro}^2) \times P(u_3|\sigma_{coluna}^2) \times P(\sigma_{coluna}^2|v_4, V_4); \\
& P(\sigma_{erro}^2|y, \mu, u_1, u_2, u_3, u_4, \sigma_{progenie}^2, \sigma_{bloco}^2, \sigma_{linha}^2, \sigma_{coluna}^2) \propto \\
& P(y|\mu, u_1, u_2, u_3, u_4, \sigma_{progenie}^2, \sigma_{bloco}^2; \sigma_{linha}^2, \sigma_{coluna}^2, \sigma_{erro}^2) \times P(\sigma_{erro}^2|v_0, V_0).
\end{aligned}$$

O modelo Bayesiano reduzido proposto foi: $y = \mathbf{1}\mu + Z_1u_1 + Z_2u_2 + \varepsilon$ em que: $y_{n \times 1}$ é o vetor de observações fenotípicas; $\mathbf{1}_{n \times 1}$ é uma matriz coluna de uns; u_1 é o vetor de efeitos aleatórios para blocos associado a matriz de incidência Z_1 e com distribuição de probabilidade $u_1 \sim N(\mathbf{0}, I\sigma_{bloco}^2)$; u_2 é o vetor de efeitos aleatórios para progênie associado a matriz de incidência Z_2 e com distribuição de probabilidade $u_2 \sim N(\mathbf{0}, I\sigma_{progênie}^2)$; ε é o vetor de erros aleatórios e com distribuição de probabilidade $\varepsilon \sim N(\mathbf{0}, I\sigma_{erro}^2)$; em que $\varepsilon|\sigma_{erro}^2 \sim N(0, I\sigma_{erro}^2)$ a distribuição conjunta dos dados observados é obtida a partir da distribuição: $y|\mu, u_1, u_2, \sigma_{progenie}^2, \sigma_{bloco}^2, \sigma_{erro}^2 \sim N(\mu + Z_1u_1 + Z_2u_2, I\sigma_e^2)$ em que:

$$\begin{aligned}
& \mu \sim N(\mu_0, \sigma_0^2); u_1|\sigma_{progenie}^2 \sim N(\mathbf{0}, I\sigma_{progenie}^2); \sigma_{progenie}^2 \sim GI(v = v_1, V = V_1); \\
& u_2|\sigma_{bloco}^2 \sim N(\mathbf{0}, I\sigma_{bloco}^2); \sigma_{bloco}^2 \sim GI(v = v_2, V = V_2); \sigma_{erro}^2 \sim GI(v = v_0, V = V_0).
\end{aligned}$$

4.2 Inferência bayesiana do modelo reduzido

Para o modelo reduzido a formulação do Teorema de Bayes é apresentada a seguir:

$$\begin{aligned}
 & P(\mu, u_1, u_2, \sigma_{progenie}^2, \sigma_{bloco}^2, \sigma_{erro}^2 | y) \propto \\
 & P(y | \mu, u_1, u_2, \sigma_{progenie}^2, \sigma_{bloco}^2, \sigma_{erro}^2) \times P(\mu | \mu_0, \sigma_0^2) \times P(u_1 | \sigma_{progenie}^2) \times P(u_2 | \sigma_{bloco}^2) \\
 & P(\sigma_{progenie}^2 | \nu_1, V_1) \times P(\sigma_{bloco}^2 | \nu_2, V_2) \times P(\sigma_{erro}^2 | \nu_0, V_0)
 \end{aligned}$$

Após avaliar a estacionariedade das cadeias as distribuições marginal a *posteriori* (DMP) para cada parâmetro foram aproximadas pela distribuição condicional completas a *posteriori* (DCCP) para cada parâmetro.

$$\begin{aligned}
 & P(\mu | y, u_1, u_2, \sigma_{progenie}^2, \sigma_{bloco}^2, \sigma_{erro}^2) \propto \\
 & P(y | \mu, u_1, u_2, \sigma_{progenie}^2, \sigma_{bloco}^2, \sigma_{erro}^2) \times P(\mu | \mu_0, \sigma_0^2); \\
 & P(\sigma_{progenie}^2 | y, \mu, u_1, u_2, \sigma_{bloco}^2, \sigma_{erro}^2) \propto \\
 & P(y | \mu, u_1, u_2, \sigma_{progenie}^2, \sigma_{bloco}^2, \sigma_{erro}^2) \times P(u_1 | \sigma_{progenie}^2) \times P(\sigma_{progenie}^2 | \nu_1, V_1); \\
 & P(\sigma_{bloco}^2 | y, \mu, u_1, u_2, \sigma_{progenie}^2, \sigma_{erro}^2) \propto \\
 & P(y | \mu, u_1, u_2, \sigma_{progenie}^2, \sigma_{bloco}^2, \sigma_{erro}^2) \times P(u_2 | \sigma_{bloco}^2) \times P(\sigma_{bloco}^2 | \nu_2, V_2). \\
 & P(\sigma_{erro}^2 | y, \mu, u_1, u_2, \sigma_{progenie}^2, \sigma_{bloco}^2) \propto \\
 & P(y | \mu, u_1, u_2, \sigma_{progenie}^2, \sigma_{bloco}^2, \sigma_{erro}^2) \times P(\sigma_{erro}^2 | \nu_0, V_0).
 \end{aligned}$$

As *prioris* Bayes I e Bayes II para os parâmetros foram as indicadas na tabela abaixo e ambas são *prioris* informativas.

Tabela 1 - Distribuição a priori para os componentes de variância

<i>Priori</i>	ν	V	alpha.mu	Alpha.V	Classificação	μ
Bayes I	1	0,002	0	0	Gama Inversa (GI)	$N(0, 10^8)$
Bayes II	1	1	0	25^2	Half-Cauchy (HC)	$N(0, 10^8)$

Fonte: Elaborada pelo autor.

A escolha da *priori* foi fundamentada na menor raiz quadrada do erro quadrático médio e na correlação entre o observado e predito pelo modelo, assim a *priori* II foi escolhida por ter apresentado tais métricas mais acuradas. Vale ressaltar que uma *priori* não informativa os modelos não convergiam, pois, alguns componentes de variância são baixos.

O trabalho contemplou com a validação cruzada 5-folds. Através do qual se obteve a métrica raiz quadrada do erro quadrático médio da validação

$$\text{cruzada} \left(\text{eficiência preditiva } EP = \sqrt{\frac{\sum_{i=1}^N (\hat{Y}_i - Y_i)^2}{N}} \right), \text{ coeficiente de correlação}$$

capacidade preditiva $CP = 100 \cdot \left(\frac{\text{cov}(Y_{\text{predito}}; Y_{\text{observado}})}{S_{Y_{\text{predito}}} \cdot S_{Y_{\text{observado}}}} \right)$ como descrito por Silva et al. (2018).

As observações foram divididas de forma independente em k partes onde cada parte foi usada uma única vez para validação, e as outras $k - 1$ partes foram utilizadas para treinamento até que todas as k partes fizessem parte da validação. Todas as métricas foram obtidas para o conjunto de observações de validação e para o conjunto de observações de treinamento através da média dos k valores em cada etapa da validação e treinamento.

Com o objetivo de verificar os efeitos dos fatores linha e colunas cada modelo foi ajustado com e sem a técnica (*post-hoc blocking row-column*). A comparação entre os modelos foi por meio do teste da razão de verossimilhança (TRV) usando a estatística $TRV = 2[\ln(\text{Modelo de interesse}) - \ln(\text{Modelo sob } H_0)]$, onde TRV sob H_0 tem distribuição aproximada de Qui-Quadrado com 2 (dois) graus de liberdade. Já para modelos bayesianos a escolha e comparação de modelos foi feita pelo Deviance Information Criterion (DIC) com número de iterações igual a 7000000, *thin* = 20, *burnin* = 200000.

Para avaliar o percentual (índice) de coincidências das progênes selecionadas foi contabilizado o percentual em comum de progênes selecionadas pelas duas metodologias de estimação para cada um dos três níveis de intensidade de seleção 37%, 25% e 15%.

Todas as análises estatística foram realizadas através do software R (R Core Team, 2019) usando as bibliotecas *Sommer* (Covarrubias, 2016, 2018), *MCMCglmm* (Hadfield, 2010; Hadfield e Nakagawa, 2010) e *Imer4* (Bates et al., 2015). A convergência das cadeias MCMC foram checadas pelo critério de

Geweke (1992) usando o pacote Coda (*convergence diagnosis and output analysis for MCMC*; Plummer et al., 2006) e as DCCPs foram obtidas de acordo com (Hadfield, 2010). Todo o trabalho foi realizado utilizando uma máquina com processador Intel Core i3 4005u 1.7GHz e 8GB de RAM com *Windows 10* onde o processamento de cada modelo bayesiano demandou aproximadamente 42 minutos, enquanto cada modelo misto demandou em torno de 4 segundos.

5. RESULTADOS E DISCUSSÕES

5.1 Experimento I

Atualmente o modelo misto é a ferramenta padrão para a estimação de parâmetros genéticos e predição do valor genotípico em diversos programas de melhoramento, resultando assim, num processo de seleção com maior eficiência seletiva. Segundo Resende et al. (2014) em algumas situações existe uma superioridade do modelo linear misto quando abordado do ponto de vista da Inferência Estatística Bayesiana. Alguns autores tem relatado a eficiência da técnica *post-hoc blocking Row-Col*, notadamente como uma ferramenta auxiliar no controle das variações ambientais, como forma de aumentar as estimativas de herdabilidade em experimentos com plantas perenes (Silva et al., 2016; Machado et al., 2020).

Para avaliação dos parâmetros estimados no método Bayesiano pode ser feito o cálculo do *Highest posterior density* (HPD). Neste trabalho foi feito apenas para a herdabilidade média de progênie. No entanto, pode ser obtida para qualquer estimativa de parâmetro genético de interesse. No método REML é possível obter o intervalo de confiança para herdabilidade utilizando um método denominado de delta (Oehlert, 1992).

A significância dos fatores (*DBC + Row-Col*) foi testada pelo teste da razão de verossimilhança (TRV) - Tabela 2 - pela estatística de Qui-Quadrado. Apontando que o modelo completo foi significativo somente para as variáveis relacionadas aos componentes da produção (Número de Frutos e Produtividade),

indicando que o modelo completo teve melhor ajuste para apenas estas duas variáveis ($p < 0,05$). Por outro lado, a *Deviance Information Criterion* (DIC) na (Tabela 3) mostra que o modelo completo teve melhor ajuste para as variáveis número de frutos, produtividade, polpa, diâmetro longitudinal do fruto e sólidos solúveis totais, ou seja, o modelo completo capturou a variação espacial para estas variáveis quando em comparação ao de blocos completos casualizados.

Sobre o enfoque de modelos mistos Cavalcante et al. (2019) encontraram resultados divergentes deste trabalho onde não houve significância do modelo completo para nenhuma variável quando trabalhou com a técnica *post-hoc blocking Row-Col*. Observa-se neste caso que os autores trabalharam com um número menor de progênies (fase de obtenção de progênies de IC, para posterior avaliação), o que acarretou em uma composição de campo experimental com maior uniformidade. Entretanto, Machado et al. (2020) verificaram superioridade da técnica *post-hoc blocking Row-Col* em comparação ao de blocos casualizados em ensaios com maracujá-azedo, quando avaliaram um maior quantitativo de progênies, sendo essa a fase de avaliação das progênies recombinadas/selecionadas do programa de seleção recorrente, o que inexoravelmente é feito em condições experimentais com maior desuniformidade do campo experimental, reportando resultados similares com este trabalho quanto à significância dos fatores (*DBC + Row-Col*) onde o modelo completo foi significativo para a variável número de frutos e não significativo para as demais variáveis.

Tabela 2 - Teste da razão de verossimilhança (TRV) para comparação entre os modelos blocos completos casualizados (DBC) e *post-hoc blocking Row-Col* BLUP/REML. Campos dos Goytacazes, RJ, 2020

Variável	Modelo	Deviance	GL	Estatística (D)	p-value
NF	<i>DBC + Row-Col</i>	1935,20	2	16,70*	<0,001
	DBC	1951,90			
Prod.	<i>DBC + Row-Col</i>	1296,50	2	17,10*	<0,001
	DBC	1313,40			
MF	<i>DBC + Row-Col</i>	1553,6	2	0,70 ^{ns}	0,708
	DBC	1554,3			
DTF	<i>DBC + Row-Col</i>	1602,00	2	1,00 ^{ns}	0,61
	DBC	1603,00			
DLF	<i>DBC + Row-Col</i>	970,10	2	0,65 ^{ns}	0,95
	DBC	969,45			
Polpa	<i>DBC + Row-Col</i>	1026,5	2	0,83 ^{ns}	0,66
	DBC	1027,3			
EC	<i>DBC + Row-Col</i>	496,87	2	0,00 ^{ns}	1,00
	DBC	496,87			
SST	<i>DBC + Row-Col</i>	523,24	2	4,36 ^{ns}	0,11
	DBC	527,60			

Fonte: Elaborada pelo autor.

Número de frutos (**NF**), Produtividade (**Prod.**), Massa do Fruto (**MF**), Diâmetro Transversal do Fruto (**DTF**), Diâmetro Longitudinal do Fruto (**DLF**), Polpa de frutos (**Polpa**), Espessura da casca (**EC**), Sólidos Solúveis Totais (**SST**). (*) significativo ao nível de 5% de probabilidade pelo teste da razão de verossimilhança (TRV) via estatística de Qui-Quadrado com dois graus de liberdade; (^{ns}) não significativo ao nível de 5% de probabilidade pelo teste da razão de verossimilhança (TRV) via teste de Qui-Quadrado com dois graus de liberdade.

Tabela 3 - Comparação dos modelos Bayesianos em blocos completos casualizados (DBC) e Row-Col pelo Deviance Information Criterion (DIC) - Bayes II. Campos dos Goytacazes, RJ, 2020

Variável	Modelo	DIC	W	Δ	Convergência	
					Geweke (1992)	Heidenberg e Welch (1983)
NFrutos	<i>DBC+ Row-Col</i>	1898,68	1,00	-44,90 ^d	Convergiu	Convergiu
	DBC	1943,58	0,00		Convergiu	Convergiu
Prod.	<i>DBC+ Row-Col</i>	1266,68	0,99	-42,35 ^d	Convergiu	Convergiu
	DBC	1309,03	0,01		Convergiu	Convergiu
MF	<i>DBC+ Row-Col</i>	1550,31	0,40	0,87 ⁱ	Convergiu	Convergiu
	DBC	1549,44	0,60		Convergiu	Convergiu
DLF	<i>DBC+ Row-Col</i>	1036,75	0,70	1,68 ⁱ	Convergiu	Convergiu
	DBC	1038,43	0,30		Convergiu	Convergiu
DTF	<i>DBC+ Row-Col</i>	942,73	0,34	1,38 ⁱ	Convergiu	Convergiu
	DBC	941,35	0,66		Convergiu	Convergiu
Polpa	<i>DBC+ Row-Col</i>	1019,72	0,75	-2,16 ^d	Convergiu	Convergiu
	DBC	1021,88	0,25		Convergiu	Convergiu
EC	<i>DBC+ Row-Col</i>	480,03	0,21	2,70 ^d	Convergiu	Convergiu
	DBC	477,33	0,79		Convergiu	Convergiu
SST	<i>DBC+ Row-Col</i>	496,38	0,98	-8,25 ^d	Convergiu	Convergiu
	DBC	504,63	0,02		Convergiu	Convergiu

Fonte: Elaborada pelo autor.

Número de frutos (**NF**), Produtividade (**Prod.**), Massa do Fruto (**MF**), Diâmetro Transversal do Fruto (**DTF**), Diâmetro Longitudinal do Fruto (**DLF**), Polpa de frutos (**Polpa**), Espessura da casca (**EC**), Sólidos Solúveis Totais (**SST**). $\Delta = DIC_C - DIC_R$ em que DIC_C é *Deviance Information Criterion* do modelo completo, DIC_R *Deviance Information Criterion* do modelo reduzido. *Priori* como sugerido por (Hadfield, 2019) com inserção dos parâmetros de extensão alpha.mu, alpha.V: (i) Modelos são iguais para $|\Delta| < 2$ (Spiegelhalter et al., 2002). (d) Modelos são diferentes para $|\Delta| \geq 2$ (Spiegelhalter et al., 2002). A escolha entre dois modelos foi fundamentada no menor valor de Deviance Information Criterion (DIC) entre os modelos. W é a probabilidade a *posteriori* do modelo obtido como descrito por (Wilberg e Bence, 2008).

A escolha de *priori* é tema central na estatística bayesiana, neste presente trabalho o modelo com a *priori* com extensão de parâmetros (Tabela 4) mostrou-se eficaz ao obter menor raiz quadrada do erro quadrático médio (REQM), maior correlação entre o predito e o observado em relação a *priori* Gama Inversa (GI), e, além disto, para *priori bayes II* no monitoramento das cadeias observou-se que no geral para um $lag = 20$ a correlação entre as amostras da cadeia foi $r < 0,02$, assim, indicando que as amostras podem ser consideradas independentes.

Ao comparar os valores das métricas correlação entre observado e predito (r) e a raiz quadrada do erro quadrático médio (REQM) da Tabela 4 observa-se que quando ajustou os modelos com a *priori II* tais métricas se aproximam das métricas dos modelos ajustados via modelos mistos-REML. Assim, a partir disto todos os modelos bayesianos foram ajustados com a *priori bayes II*, visto que as métricas foram mais acuradas em relação aos modelos ajustados com a *priori I* (Gama Inversa). Por outro lado, Silva et al. (2018) encontrou a distribuição a *priori* mais adequada para os estimação dos valores genéticos, utilizando o método *Markov chain Monte Carlo (MCMC)*, ela foi a half-normal positiva e que apresentou menor critério de informação da deviance (DIC) em relação a distribuição uniforme.

Tabela 4 - Comparação dos modelos pelas métricas raiz quadrada do erro quadrático médio (REQM) e correlação entre o predito e o observado (r) para os três cenários de modelos escolhidos. Campos dos Goytacazes, RJ, 2020

Variável	Métrica	Bayes I	Bayes II	REML
Número de Frutos – Nfrutos	REQM	55,64	54,87	53,33
	<i>r</i>	89,39	89,81	89,59
Produtividade em kg – Prod.	REQM	8,95	7,98	8,05
	<i>r</i>	85,53	88,05	87,45
Massa do Fruto (g) – MF	REQM	26,10	23,91	23,11
	<i>r</i>	64,14	72,79	75,23
Comprimento de frutos (mm) – DLF	REQM	4,34	4,27	4,24
	<i>r</i>	83,83	83,93	84,21
Diâmetro de frutos (mm) – DTF	REQM	9,64	3,08	3,06
	<i>r</i>	85,76	85,56	85,74
Quantidade de Polpa (%) – PP	REQM	4,99	4,29	4,51
	<i>r</i>	71,24	80,77	76,62
Espessura da Casca (mm) – EC	REQM	0,77	0,76	0,75
	<i>r</i>	84,36	84,24	84,36
Sólidos Solúveis Totais(°Brix) – SST	REQM	0,83	0,74	0,81
	<i>r</i>	84,60	88,08	84,61

Fonte: Elaborada pelo autor.

$$REQM = \sqrt{\frac{\sum_{i=1}^N (\hat{Y}_i - Y_i)^2}{N}}$$
;
$$r = 100 \cdot \left(\frac{cov(Y_{predito} : Y_{observado})}{S_{predito} \cdot S_{observado}} \right)$$

Y_i é o valor observado; \hat{Y}_i é o valor estimado pelo modelo Bayesiano com menor DIC para cada *priori* ou pelo modelo misto escolhido pelo teste da razão de verossimilhança via estatística Qui-Quadrado; r é a correlação entre o observado e o predito pelo modelo.

As estimativas de parâmetros genéticos herdabilidade e ganho de seleção (Tabela 5 e Tabela 6) mostram que os maiores ganhos de seleção foram para as variáveis produtividade e números de frutos revelando a existência de variabilidade genética e potencial seletivo entre as progênies de maracujá-azedo em estudo, e, nos mostram também que os intervalos de confiança frequentista (IC) para herdabilidade têm menores amplitudes que o *Highest Posterior Density* (HPD).

As estimativas de herdabilidade sobre enfoque de modelos mistos variaram de 0,21 até 0,43 sendo que número de frutos teve a maior herdabilidade (0,43). Estimativas de herdabilidade com mesma magnitude foram encontradas por Cavalcante et al. (2019) para esta mesma cultura e mesmo delineamento experimental com 23 famílias de meio-irmãos. Por outro lado, o ganho de seleção variou de -7,31% a 34,12%, sendo as variáveis número de frutos e produtividade que as que tiveram maiores ganhos de seleção 28,17% e 34,12%, respectivamente. Tais estimativas de herbabilidade sobre o enfoque bayesiano

variaram no intervalo de 0,11 a 0,35 e o ganho de seleção de -7,16% a 31,98%. De forma geral, as estimativas de herdabilidade não coincidiram entre as duas metodologias de estimação, porém, as estimativas de ganho de seleção se aproximaram entre as duas metodologias de estimação. Pode-se observar que as estimativas de herdabilidade aqui reportadas são inferiores às reportadas por Viana et al. (2004) fato que é explicado, pois, neste trabalho são dados do terceiro ciclo de seleção recorrente e as plantas já estão mais homogêneas, via aumento da frequência de alelos favoráveis, para as características desejáveis, acrescido a isso os autores trabalharam com plantas clonadas, o que eventualmente convergem para maiores estimativas de herdabilidade.

Tabela 5 - Estimativa dos parâmetros genéticos das 81 progênes de maracujazeiro-azedo de irmãos completos do terceiro ciclo de seleção recorrente via modelos mistos-REML/BLUP. Campos dos Goytacazes, RJ, 2020

Variável	$\hat{\sigma}_{progenie}^2$	$\hat{\sigma}_{bloco}^2$	$\hat{\sigma}_{linha}^2$	$\hat{\sigma}_{coluna}^2$	$\hat{\sigma}_{erro}^2$	\hat{h}^2	$\hat{G}_s(\%)$	IC- $\hat{h}^2(95\%)$
NFrutos	4410,00	0,00	2156,00	0,00	4651,00	0,39	34,12	0,22; 0,56
Produtividade	66,28	1,44	44,74	0,00	99,78	0,31	28,17	0,14; 0,48
MF	200,96	87,37	-	-	665,67	0,21	5,38	0,01; 0,41
DLF	18,59	7,99	-	-	25,81	0,35	4,09	0,13; 0,58
DTF	12,00	1,88	-	-	13,90	0,43	3,57	0,25; 0,62
Polpa	8,16	2,52	-	-	25,54	0,23	4,47	0,03; 0,42
EC	0,58	0,01	-	-	0,81	0,42	-7,31	0,24; 0,60
SST	0,71	0,09	-	-	0,95	0,41	4,50	0,22; 0,59

Fonte: Elaborada pelo autor.

Número de frutos (**NF**), Produtividade (**Prod.**), Massa do Fruto (**MF**), Diâmetro Transversal do Fruto (**DTF**), Diâmetro Longitudinal do Fruto (**DLF**), Polpa de frutos (**Polpa**), Espessura da casca (**EC**), Sólidos Solúveis Totais (**SST**); (\hat{h}^2) é a estimativa pontual da herdabilidade no sentido amplo ($\hat{h}^2 = \frac{\hat{\sigma}_{progenie}^2}{\hat{\sigma}_{progenie}^2 + \hat{\sigma}_{bloco}^2 + \hat{\sigma}_{coluna}^2 + \hat{\sigma}_{linha}^2 + \hat{\sigma}_{erro}^2}$); ($\hat{\sigma}_{erro}^2$) é a estimativa da variância residual; ($\hat{\sigma}_{bloco}^2$) é a estimativa da variância devido ao fator bloco; ($\hat{\sigma}_{coluna}^2$) é a estimativa da variância devido ao fator coluna; ($\hat{\sigma}_{linha}^2$) é a estimativa da variância devido ao fator linha; $\hat{\sigma}_{progenie}^2$ é a estimativa da variância genética; IC- \hat{h}^2 é o intervalo de confiança frequentista para herdabilidade obtida pelo método delta e $G_s = 100 \cdot \left(\frac{\bar{X}_{30} - \bar{X}_G}{\bar{X}_G} \right)$ é o ganho de seleção em relação a média geral (\bar{X}_G) e a média das 30 (trinta) primeiras progênes selecionadas (\bar{X}_{30}).

As estimativas das métricas herdabilidade e intervalo de *Highest Posterior Density* diferiram quando obtidas pelo método bayesiano do método frequentista (Tabela 5 e Tabela 6). Este resultado é divergente dos resultados encontrados por Silva et al. (2018), onde foi estimada a herdabilidade para a cultura da soja de experimentos em blocos casualizados em 6 ambientes com 3 repetições para a

característica produtividade de grãos da soja kg ha⁻¹, e obteve resultados bem próximos para a métrica herdabilidade para cada uma das metodologias de estimação (0,51 e 0,53). Por outro lado, a métrica ganho de seleção teve magnitude similar para os dois métodos de estimação (Tabela 5 e Tabela 6).

Tabela 6 - Estimativa de parâmetros genéticos das 81 progênies de maracujazeiro-azedo do terceiro ciclo de seleção recorrente – Bayes II. Campos dos Goytacazes, RJ, 2020

Variável	$\hat{\sigma}_{progenie}^2$	$\hat{\sigma}_{bloco}^2$	$\hat{\sigma}_{linha}^2$	$\hat{\sigma}_{coluna}^2$	$\hat{\sigma}_{erro}^2$	\hat{h}^2	$\hat{G}_s(\%)$	HPD-95%
NFrutos	4178,00	2873,00	2075,00	147,10	4975,00	0,35	31,98	0,15; 0,53
Produtividade	67,27	1559,00	53,07	5,61	102,20	0,21	27,00	0,00; 0,39
MF	180,00	3658,00	-	-	702,20	0,14	4,61	0,00; 0,31
DLF	19,14	1847,00	-	-	26,45	0,18	4,02	0,00; 0,43
DTF	12,41	2183,00	-	-	14,21	0,22	5,53	0,00; 0,50
Polpa	8,20	1339,00	2,00	1,39	24,89	0,11	4,21	0,00; 0,31
EC	0,60	1067,00	-	-	0,84	0,19	-7,16	0,00; 0,48
SST	0,67	868,60	0,06	0,18	0,88	0,14	4,01	0,00; 0,41

Fonte: Elaborada pelo autor.

Número de frutos (**NF**), Produtividade (**Prod.**), Massa do Fruto (**MF**), Diâmetro Transversal do Fruto (**DTF**), Diâmetro Longitudinal do Fruto (**DLF**), Polpa de frutos (**Polpa**), Espessura da casca (**EC**), Sólidos Solúveis Totais (**SST**); (\hat{h}^2) é a estimativa da herdabilidade obtida pela média da distribuição a posteriori; ($\hat{\sigma}_{erro}^2$) estimativa da variância residual; ($\hat{\sigma}_{bloco}^2$) estimativa da variância devido ao fator bloco, ($\hat{\sigma}_{coluna}^2$) estimativa da variância devido ao fator coluna; ($\hat{\sigma}_{linha}^2$) estimativa da variância devido ao fator linha; $\hat{\sigma}_{progenie}^2$ é a estimativa da variância genética; *HPD Highest posterior density* e G_s é o ganho de seleção em relação a média geral estimada. *Priori* como sugerido por (Hadfield, 2019) com inserção de extensão de parâmetros alpha.mu, alpha.V.

Com objetivo de comparar os *ranks* de progênies formados pelas duas metodologias de estimação definiu-se três graus de intensidade de seleção (37%, 25%, 15%) como mostrado na Tabela 7. Pode-se observar o índice de coincidência entre as progênies selecionadas pelas metodologias REML/BLUP, *Bayes I*, *Bayes II* varia de 83,33% a 100,00%. Pode-se concluir que independente dos métodos de estimação adotados as progênies com maiores valores genéticos que foram selecionadas são praticamente as mesmas para todos os cenários.

Cada metodologia tem virtudes e pontos fracos, entretanto, cabe ao pesquisador usá-las de acordo com os objetivos e necessidades. Se o tempo de execução da análise é prioridade, se sua amostra é pequena ou se deseja obter uma distribuição para o parâmetro em estudo, cabendo assim ao pesquisador

decidir qual metodologia adotar de acordo com a dificuldade imposta pelo problema.

Tabela 7 - Índice de coincidência dos indivíduos selecionados (%) entre as 81 progênes de maracujazeiro-azedo do terceiro ciclo de seleção recorrente via modelo Bayesiano e por Modelos mistos BLUP/REML. Campos dos Goytacazes, RJ, 2020

Variável	Intensidade de Seleção (%)	BayesI/REML	BayesII/REML	Bayes I e II
NFrutos	37	100,00	96,67	96,67
	25	95,00	95,00	100,00
	15	100,00	100,00	100,00
Produtividade	37	100,00	96,67	96,67
	25	100,00	85,00	85,00
	15	100,00	100,00	100,00
MF	37	100,00	100,00	100,00
	25	100,00	100,00	100,00
	15	100,00	100,00	100,00
DLF	37	96,67	100,00	96,67
	25	100,00	100,00	100,00
	15	100,00	100,00	100,00
DTF	37	100,00	100,00	100,00
	25	95,00	100,00	95,00
	15	100,00	100,00	96,67
Polpa	37	100,00	96,67	96,67
	25	100,00	95,00	95,00
	15	100,00	91,67	91,67
EC	37	100,00	100,00	100,00
	25	100,00	100,00	100,00
	15	100,00	100,00	100,00
SST	37	100,00	93,33	93,33
	25	95,00	90,00	95,00
	15	91,67	83,33	91,67

Fonte: Elaborada pelo autor.

Número de frutos (**NF**), Produtividade (**Prod.**), Massa do Fruto (**MF**), Diâmetro Transversal do Fruto (**DTF**), Diâmetro Longitudinal do Fruto (**DLF**), Polpa de frutos (**Polpa**), Espessura da casca (**EC**), Sólidos Solúveis Totais (**SST**).

O rank dos indivíduos (Tabela 8 e Tabela 9) revela que não existe divergência na seleção das progênes pelos dois métodos de estimação, como pode ser observado na Tabela 7 o índice de coincidência entre as progênes selecionadas variou de 88,33% a 100,00%.

Tabela 8 - Tabela com ranks das 30 primeiras progênies de maracujazeiro-azedo de irmãos completos do terceiro ciclo de seleção recorrente obtidas via modelos mistos REML/BLUP. Campos dos Goytacazes, RJ, 2020

Rank	NFrutos	Prod.	MF	DLF	DTF	Polpa	EC	SST
1	56	26	69	68	3	81	7	6
2	26	56	5	8	8	19	54	75
3	1	47	3	44	27	54	26	38
4	11	11	8	27	42	68	27	53
5	24	1	68	5	48	6	76	69
6	47	24	42	41	41	17	37	68
7	68	68	47	69	69	18	18	44
8	38	70	72	40	5	41	50	35
9	16	16	75	37	78	28	29	63
10	70	38	51	7	60	36	28	47
11	9	31	78	23	68	50	14	52
12	23	9	2	52	2	30	23	51
13	34	22	41	10	72	45	6	56
14	10	69	6	73	52	65	36	66
15	5	48	33	13	53	7	81	79
16	31	23	7	42	75	16	70	11
17	22	10	53	53	50	35	12	5
18	60	60	73	48	24	34	34	20
19	48	5	1	32	9	40	19	25
20	15	15	24	29	59	67	41	48
21	19	34	52	35	11	23	43	8
22	12	19	44	75	47	57	64	50
23	69	27	29	78	71	9	8	43
24	79	78	59	4	23	25	16	21
25	27	50	48	54	33	29	68	33
26	80	79	27	72	61	43	55	39
27	67	36	31	24	7	64	65	2
28	28	80	35	59	22	69	69	60
29	54	75	60	80	13	20	62	65
30	45	41	9	50	51	79	59	41
\bar{X}_G	160,97	22,18	167,47	82,35	74,75	42,25	7,93	14,20
\bar{X}_{30}	215,89	28,43	176,48	85,71	77,42	44,14	7,35	14,84

Fonte: Elaborada pelo autor.

Número de frutos (**NF**), Produtividade (**Prod.**), Massa do Fruto (**MF**), Diâmetro Transversal do Fruto (**DTF**), Diâmetro Longitudinal do Fruto (**DLF**), Polpa de frutos (**Polpa**), Espessura da casca (**EC**), Sólidos Solúveis Totais (**SST**); *ranks* das 30 (trinta) progênies formadas a partir do valor médio somado ao valor genotípico predito ($u + g$) via modelos mistos REML/BLUP; \bar{X}_G é a média geral; \bar{X}_{30} média das 30 (trinta) primeiras progênies selecionadas.

Tabela 9 - Tabela com ranks das 30 primeiras progênies de maracujazeiro-azedo de irmãos completos do terceiro ciclo de seleção recorrente obtidas via modelos Bayesiano. Campos dos Goytacazes, RJ, 2020

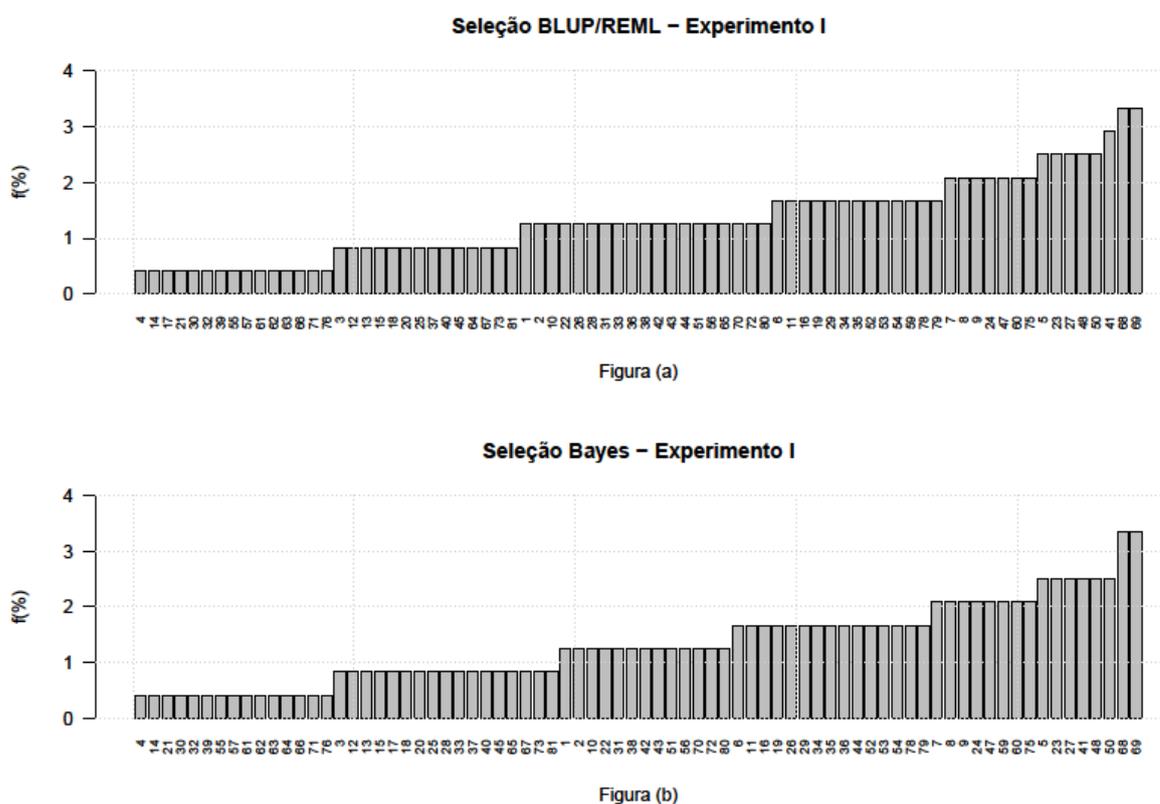
Rank	NFrutos	Prod.	MF	DLF	DTF	Polpa	EC	SST
1	56	26	69	68	3	19	7	6
2	26	56	5	8	8	68	54	44
3	1	47	3	44	27	81	26	53
4	11	11	8	27	42	54	27	35
5	47	1	68	5	48	6	76	75
6	24	24	42	41	41	17	37	38
7	68	70	47	69	69	41	18	68
8	38	68	72	40	5	18	50	63
9	70	16	75	37	78	28	29	69
10	16	38	51	7	60	30	28	47
11	9	31	78	23	68	36	14	66
12	23	9	2	52	2	65	23	5
13	34	22	41	10	72	45	6	52
14	10	69	6	73	52	50	36	51
15	22	48	33	13	53	16	81	8
16	31	23	7	42	75	35	70	56
17	5	10	53	53	50	40	12	20
18	60	19	73	48	24	7	34	50
19	48	34	1	32	9	67	19	79
20	19	27	24	29	59	57	41	25
21	12	60	52	35	11	34	43	11
22	15	5	44	75	47	23	64	48
23	69	15	29	78	71	79	8	21
24	79	79	59	4	23	43	16	39
25	27	36	48	72	33	69	68	43
26	80	50	27	54	61	9	55	2
27	67	78	31	59	7	29	65	41
28	45	80	35	24	22	25	69	59
29	36	75	60	80	13	26	62	60
30	54	44	9	50	51	20	59	17
\bar{X}_G	160,60	22,03	167,30	82,33	74,76	42,30	7,96	14,14
\bar{X}_{30}	211,96	27,98	175,02	85,65	77,70	44,08	7,39	14,72

Fonte: Elaborada pelo autor.

Número de frutos (**NF**), Produtividade (**Prod.**), Massa do Fruto (**MF**), Diâmetro Transversal do Fruto (**DTF**), Diâmetro Longitudinal do Fruto (**DLF**), Polpa de frutos (**Polpa**), Espessura da casca (**EC**), Sólidos Solúveis Totais (**SST**); *ranks* das 30 (trinta) progênies formadas a partir do valor médio somado ao valor genotípico predito $u + g$ via modelos bayesianos; \bar{X}_G é a média geral; \bar{X}_{30} média das 30 (trinta) primeiras progênies selecionadas.

Para uma intensidade de seleção de aproximadamente de 37%, o *rank* das 30 melhores progênies para cada característica está descrito na Tabela 8 e Tabela 9. A Figura 1 (a) nos mostra as frequências entre as 30 progênies selecionadas pelo metodologia de modelos mistos para cada característica (Tabela 8) e observa-se que as progênies 69, 68, 41, 50, 48, 27, 23, 5, 75, 60, 47,

24, 9, 8 e 7 formam um bloco de progênies com alta probabilidade de serem as melhores progenies para todas as características avaliadas. Por outro lado a Figura 1 (b) mostra as frequências das 30 progênies selecionadas pela metodologia de metodologia de modelos bayesianos (Tabela 9), e, pode-se observar que as progênies que formam o bloco com alta probabilidade de serem as melhores progenies para todas as características é praticamente o mesmo bloco selecionado via metodologia de modelos mistos-REML/BLUP. Fato que está relacionado à estrutura balanceada das observações do experimento.



modeladas usando o método de estimação bayesiano. Uma das possibilidades frente a esses resultados diferentes se dá porque cabe ressaltar que no caso da goiabeira, trata-se de cultura extremamente perene, fato que não é observado para o maracujazeiro-azedo que hoje é classificado como cultura semi-perene. Os efeitos de ambiente permanente no caso da goiabeira são mais importantes de serem avaliados ao longo de várias safras, onde pode-se eventualmente capitalizar uma melhor informação experimental via uso de modelos bayesiano (Silva et al., 2020).

5.2 Experimento II

Neste ambiente por ser o quarto ciclo de seleção recorrente onde algumas variáveis já estão estabilizadas e homogêneas para as características desejáveis, visto que a técnica preconiza a concentração de alelos favoráveis, mantendo a variância genética. Neste experimento, a magnitude de herdabilidade e ganho de seleção foram muito menores que as estimativas obtidas no experimento I.

A significância dos fatores (*DBC + Row-Col*) foi testada pelo teste da razão de verossimilhança (TRV) via estatística de Qui-Quadrado (Tabela 10). Apontando que o modelo completo foi significativo somente para as variáveis de produção (Número de Frutos e Produtividade), indicando que o modelo completo teve melhor ajuste para estas duas variáveis ($p < 0,05$). Por outro lado, a *Deviance Information Criterion* (DIC) (Tabela 11) mostra que o modelo completo teve melhor ajuste para todas as variáveis exceto (Espessura da casca e Sólido Solúveis Totais).

Sobre o enfoque de modelos mistos Silva et al. (2016) encontraram os mesmos resultados quando inserida a técnica post-hoc blocking Row-Col no modelo, ou seja, apenas para as variáveis produtividade e número de frutos foram significativa ($p < 0,05$). Os argumentos para esses resultados já foram apontados na discussão dos resultados obtidos no experimento I. Os resultados favoráveis obtidos com a técnica *post-hoc blocking Row-Col* para a avaliação dos dados estão consistentes com estudos anteriores de outros autores. Por exemplo, Gezan et al. (2006) e Machado et al. (2020), comparando a eficiência do *post-hoc blocking Row-Col* com o delineamento de blocos casualizados obtiveram

estimativas mais precisas de herdabilidade utilizando o *post-hoc blocking Row-Col*. Entretanto, o delineamento em blocos casualizados, em geral, é eficiente quando a variabilidade dentro das repetições é relativamente pequena, o que é raro quando um grande número de genótipos é avaliado, notadamente em experimentos de grandes dimensões, como é o caso dos trabalhos desenvolvidos nesta pesquisa.

Neste estudo, observou-se que a técnica *post-hoc blocking Row-Col* foi mais eficiente para estimar parâmetros genéticos, pois, quando as observações foram modeladas via modelos bayesianos somente para as variáveis diâmetro, espessura da casca e sólido solúveis totais não houve significância para o fator *DBC + Row-Col*.

Tabela 10 - Teste da razão de verossimilhança (TRV) para comparação entre os dois modelos propostos, blocos completos casualizados (DBC) e *post-hoc blocking Row-Col*. Cambuci, RJ, 2020

Variável	Modelo	Deviance	GL	Estatística (D)	p-value
Prod.	DBC+ <i>Row-Col</i>	1289,00	2	33,50*	<0,001
	DBC	1322,50			
NF	DBC+ <i>Row-Col</i>	2051,00	2	29,00*	<0,001
	DBC	2022,00			
MF	DBC+ <i>Row-Col</i>	1789,10	2	2,20 ^{ns}	0,3343
	DBC	1791,30			
DLF	DBC+ <i>Row-Col</i>	1212,10	2	5,10 ^{ns}	0,0772
	DBC	1217,20			
DTF	DBC+ <i>Row-Col</i>	1075,80	2	1,79 ^{ns}	0,6353
	DBC	1074,01			
EC	DBC+ <i>Row-Col</i>	618,07	2	0,00 ^{ns}	1,000
	DBC	618,07			
Polpa do fruto	DBC+ <i>Row-Col</i>	1156,60	2	2,30 ^{ns}	0,3257
	DBC	1158,90			
SST	DBC+ <i>Row-Col</i>	622,16	2	0,06 ^{ns}	0,9722
	DBC	622,22			

Fonte: Elaborada pelo autor.

(*) significativo ao nível de 5% de probabilidade pelo teste da razão de verossimilhança (TRV); (^{ns}) não significativo nível de 5% de probabilidade pelo teste da razão de verossimilhança (TRV) via estatística de Qui-Quadrado com dois graus de liberdade.

Número de frutos (**NF**), Produtividade (**Prod.**), Massa do Fruto (**MF**), Diâmetro Transversal do Fruto (**DTF**), Diâmetro Longitudinal do Fruto (**DLF**), Polpa de frutos (**Polpa**), Espessura da casca (**EC**), Sólidos Solúveis Totais (**SST**).

Tabela 11 - Comparação dos modelos Bayesianos em blocos completos casualizados (DBC) e Row-Col pelo Deviance Information Criterion (DIC) – Bayes II. Cambuci RJ, 2020

Variável	Modelo	DIC	W	Δ	Convergência	
					Geweke (1992)	Heidenberg e Welch (1983)
NF	<i>DBC + Row-Col</i>	2005,57	1,00	-48,30 ^d	Convergiu	Convergiu
	DBC	2053,87	0,00		Convergiu	Convergiu
Prod.	<i>DBC + Row-Col</i>	1271,72	1,00	-54,99 ^d	Convergiu	Convergiu
	DBC	1326,71	0,00		Convergiu	Convergiu
MF	<i>DBC + Row-Col</i>	1789,65	0,85	-3,43 ^d	Convergiu	Convergiu
	DBC	1793,08	0,15		Convergiu	Convergiu
DLF	<i>DBC + Row-Col</i>	1206,08	0,98	-7,59 ^d	Convergiu	Convergiu
	DBC	1213,67	0,02		Convergiu	Convergiu
DTF	<i>DBC + Row-Col</i>	1073,92	0,60	0,76 ⁱ	Convergiu	Convergiu
	DBC	1074,68	0,40		Convergiu	Convergiu
Polpa	<i>DBC + Row-Col</i>	1158,35	0,84	-3,37 ^d	Convergiu	Convergiu
	DBC	1161,72	0,16		Convergiu	Convergiu
EC	<i>DBC + Row-Col</i>	609,87	0,17	3,22 ^d	Convergiu	Convergiu
	DBC	606,65	0,83		Convergiu	Convergiu
SST	<i>DBC + Row-Col</i>	671,18	0,25	2,15 ^d	Convergiu	Convergiu
	DBC	669,03	0,75		Convergiu	Convergiu

Fonte: Elaborada pelo autor.

Número de frutos (**NF**), Produtividade (**Prod.**), Massa do Fruto (**MF**), Diâmetro Transversal do Fruto (**DTF**), Diâmetro Longitudinal do Fruto (**DLF**), Polpa de frutos (**Polpa**), Espessura da casca (**EC**), Sólidos Solúveis Totais (**SST**). $\Delta = DIC_C - DIC_R$ em que DIC_C é Deviance Information Criterion do modelo completo, DIC_R Deviance Information Criterion do modelo reduzido. *Priori* como sugerido por (Hadfield, 2021) com inserção dos parâmetros de extensão alpha.mu, alpha.V: (i) Modelos são iguais melhor para $|\Delta| < 2$ (Spiegelhalter et al., 2002). (d) Modelos são diferentes para $|\Delta| \geq 2$ (Spiegelhalter et al., 2002). A escolha entre dois modelos foi fundamentada no menor valor de Deviance Information Criterion (DIC). W é a Probabilidade a *posteriori* do modelo obtido como descrito por (Wilberg e Bence, 2008).

A escolha de *priori* é ponto central da estatística bayesiana. A Tabela 12 nos mostra que o modelo com a *priori* com extensão de parâmetros tem menor REQM e maior correlação entre o predito e o observado em relação a *priori* I (Gama Inversa), além disto, no monitoramento das cadeias observa-se que para um

$lag = 20$ a correlação entre as amostras da cadeia é $r < 0,02$. Indicando assim que as amostras da cadeia podem ser consideradas independentes.

Ao comparar os valores das métricas correlação entre observado e predito (r), raiz quadrada do erro quadrático médio (REQM) da Tabela 12 observa-se que quando ajustou os modelos com a *priori* II tais métricas se aproximam das métricas dos modelos ajustados via modelos mistos-REML/BLUP. A partir de disto, todos os modelos bayesianos ajustados foram com a *priori* bayes II, visto que as métricas (Tabela 12) foram mais acuradas em relação aos modelos ajustados com *priori* I (Gama Inversa). Por outro lado, Silva et al. (2018) encontrou a distribuição a *priori* mais adequada para os estimação dos valores genéticos, utilizando o método MCMC, foi a *half-normal* positiva que apresentou menor critério de informação da *deviance* (DIC) e relação a distribuição uniforme.

Tabela 12 - Comparação dos modelos pelas métricas raiz quadrada do erro quadrático médio (REQM) e a correlação entre o observado e o predito (r) para os três cenários. Cambuci, RJ, 2020

Variável	Métrica	Bayes I	Bayes II	REML
Número de Frutos – Nfrutos	REQM	39,76	35,57	35,12
	r	57,21	70,29	71,32
Produtividade em kg – Prod.	REQM	6,01	5,29	5,35
	r	58,06	71,73	70,92
Massa do Fruto (g) – MF	REQM	23,84	20,92	23,00
	r	32,39	66,78	46,56
Comprimento de fruto (mm) – DTF	REQM	5,01	4,20	4,55
	r	67,90	79,45	74,56
Diâmetro de frutos (mm) –DLF	REQM	3,68	3,51	5,50
	r	38,95	52,34	53,71
Quantidade de Polpa (%) – PP	REQM	4,52	3,96	4,14
	r	75,48	76,09	75,68
Espessura da Casca (mm) – EC	REQM	0,94	0,91	0,89
	r	77,10	77,46	78,65
Sólidos Solúveis Totais(°Brix) – SST	REQM	1,25	1,19	1,16
	r	75,17	75,00	75,75

Fonte: Elaborada pelo autor.

$$REQM = \sqrt{\frac{\sum_{i=1}^N (\hat{Y}_i - Y_i)^2}{N}}; Y_i \text{ é o valor observado; } \hat{Y}_i \text{ é o valor estimado pelo modelo com menor Bayesiano com menor DIC ou pelo modelo escolhido pelo teste da razão de verossimilhança (REML) e } r = 100 \cdot \left(\frac{\text{cov}(Y_{\text{predito}}; Y_{\text{observado}})}{S_{\text{predito}} \cdot S_{\text{observado}}} \right).$$

O experimento II por ser o quarto ciclo de seleção recorrente as estimativas de parâmetros genéticos (Tabela 13 e Tabela 14) mostra que os maiores ganhos de seleção foram para as variáveis (produtividade e números de frutos), porém, muito inferiores que as mesmas métricas no experimento I. Pelo método frequentista revelam ainda a não existência de variabilidade genética e pouco potencial seletivo entre as progênies de maracujá-azedo em estudo, e, o que é comprovado pelos intervalos de confiança frequentista para herdabilidade contendo o zero. Este resultado, também pode ser devido o fator perturbador de alta variação ambiental. Essas variações de ambiente são de diversas naturezas, notadamente no manejo do experimento e na tomada das observações, no caso deste segundo experimento o mesmo foi conduzido em associação com o IFF, no seu *campus* de Cambuci-RJ. A equipe envolvida na tomada das observações não tinha o mesmo treinamento das equipes já treinadas e com experiência que conduzem os trabalhos junto as unidades da UENF. Isto, possivelmente pode ser

computado como um dos efeitos perturbadores, para a obtenção desses resultados.

Por outro lado, as estimativas pelo método bayesiano apresentaram HPD com menores amplitudes e sem conter o zero, fato que é explicado devido o espaço paramétrico assumido pela *priori II*, mas as estimativas de ganho de seleção são bem similares ao método frequentista.

As estimativas de herdabilidade sobre enfoque de modelos mistos variaram de 0,05 até 0,26 sendo que espessura da casca teve a maior estimativa de herdabilidade (0,26). Estimativas divergentes foram obtidas por Silva et al. (2016) para a mesma cultura e delineamento experimental para famílias de irmãos completo variando no intervalo 0,14 a 0,53. Por outro lado, o ganho de seleção variou de -6,54% a 6,06%, sendo que as variáveis número de frutos e produtividade tiveram maiores ganhos de seleção 6,06% e 5,89, respectivamente. Tais estimativas de herdabilidade sobre o enfoque bayesiano variaram no intervalo de 0,04 a 0,11 e o ganho de seleção de -6,27% a 6,06%. De forma geral, a magnitude das estimativas de herdabilidade não coincidiram entre as duas metodologias de estimação, porém, as estimativas de ganho de seleção se aproximaram entre as duas metodologias de estimação. Pode-se observar que para todas as características as estimativas de herdabilidade encontradas estão muito abaixo do encontrado na literatura, este fato pode ser devido à alta influência ambiental/manejo do experimento ou pela natureza poligênica dessas variáveis. Pode observar que as estimativas de herdabilidade aqui encontradas são muito inferiores a reportada por Viana et al. (2004) fato que é explicado por ser o quarto ciclo de seleção, recorrente onde as variáveis estão homogêneas para as características desejáveis, além das argumentações já apresentadas na discussão do experimento I.

Tabela 13 - Estimativa de parâmetros genéticos das 97 progênies de maracujazeiro-azedo do quarto ciclo de seleção recorrente da seleção recorrente via modelos mistos-REML/BLUP. Cambuci, RJ, 2020

Variável	$\hat{\sigma}_{progenie}^2$	$\hat{\sigma}_{bloco}^2$	$\hat{\sigma}_{linha}^2$	$\hat{\sigma}_{coluna}^2$	$\hat{\sigma}_{erro}^2$	\hat{h}^2	$\hat{G}_s(\%)$	IC- $\hat{h}^2(95\%)$
Prod.	4,50	1,26	24,01	0,00	34,21	0,07	5,89	-0,06; 0,20
NF	197,31	89,17	947,14	16,58	1486,45	0,07	6,06	-0,06; 0,20
MF	30,98	31,89	-	-	561,98	0,05	1,30	-0,14; 0,24
DLF	6,28	1,65	-	-	25,07	0,19	2,22	0,01; 0,38
DTF	1,26	1,73	-	-	13,45	0,08	0,73	-0,10; 0,26
Polpa	3,50	0,00	-	-	19,87	0,15	2,24	-0,04; 0,35
EC	0,41	0,13	-	-	1,04	0,26	-6,54	0,08; 0,44
SST	0,28	0,00	-	-	1,57	0,15	2,41	-0,04; 0,35

Fonte: Elaborada pelo autor.

Número de frutos (**NF**), Produtividade (**Prod.**), Massa do Fruto (**MF**), Diâmetro Transversal do Fruto (**DTF**), Diâmetro Longitudinal do Fruto (**DLF**), Polpa de frutos (**Polpa**), Espessura da casca (**EC**), Sólidos Solúveis Totais (**SST**). (\hat{h}^2) é a estimativa pontual da herdabilidade no sentido amplo ($\hat{h}^2 = \frac{\hat{\sigma}_{progenie}^2}{\hat{\sigma}_{progenie}^2 + \hat{\sigma}_{bloco}^2 + \hat{\sigma}_{coluna}^2 + \hat{\sigma}_{linha}^2 + \hat{\sigma}_{erro}^2}$); ($\hat{\sigma}_{erro}^2$) é a estimativa da variância residual; ($\hat{\sigma}_{bloco}^2$) é a estimativa da variância devido ao fator bloco; ($\hat{\sigma}_{coluna}^2$) é a estimativa da variância devido ao fator coluna; ($\hat{\sigma}_{linha}^2$) é a estimativa da variância devido ao fator linha; $\hat{\sigma}_{progenie}^2$ é a estimativa da variância genética e IC- \hat{h}^2 é a aproximação do intervalo de confiança para a herdabilidade pelo método delta e $G_s = 100 \cdot \left(\frac{\bar{X}_{30} - \bar{X}_G}{\bar{X}_G} \right)$ é o ganho de seleção em relação a média geral (\bar{X}_G) e a média das 30 (trinta) primeiras progênies selecionadas (\bar{X}_{30}).

De forma geral, as estimativas das métricas herbabilidade e o ganho de seleção diferiram quando obtidas pelo método frequentista e bayesiano (Tabela 13 e Tabela 14). Este resultado é divergente dos resultados encontrados por Silva et al. (2016), onde foi estimada a herdabilidade para a mesmas variáveis e encontrado estimativas com magnitude superior para as mesmas variáveis desta cultura, pois os autores trabalharam no terceiro ciclo de seleção recorrente.

Tabela 14 - Estimativa de parâmetros genéticos das 97 progênies de maracujazeiro-azedo do quarto ciclo de seleção recorrente-Bayes II. Cambuci, RJ, 2020

Variável	$\hat{\sigma}_{progenie}^2$	$\hat{\sigma}_{bloco}^2$	$\hat{\sigma}_{linha}^2$	$\hat{\sigma}_{coluna}^2$	$\hat{\sigma}_{erro}^2$	\hat{h}^2	$\hat{G}_s(\%)$	HPD-95%
NF	170,90	2593,00	1098,00	55,46	1533,00	0,05	6,06	0,00; 0,15
Prod.	4,75	1332,00	32,68	0,93	34,51	0,04	5,67	0,00; 0,13
MF	51,89	2847,00	44,91	26,10	515,50	0,06	1,99	0,00; 0,18
DTF	6,39	986,30	1,14	3,08	23,37	0,11	2,22	0,00; 0,28
DLF	1,35	13,42	-	-	13,61	0,04	0,72	0,00; 0,15
Polpa	3,26	11,50	1,82	0,72	19,27	0,08	1,96	0,00; 0,24
EC	0,41	682,00	-	-	1,08	0,09	-6,27	0,00; 0,30
SST	0,26	1034,00	-	-	1,63	0,07	2,04	0,00; 0,24

Fonte: Elaborada pelo autor.

Número de frutos (**NF**), Produtividade (**Prod.**), Massa do Fruto (**MF**), Diâmetro Transversal do Fruto (**DTF**), Diâmetro Longitudinal do Fruto (**DLF**), Polpa de frutos (**Polpa**), Espessura da casca (**EC**), Sólidos Solúveis Totais (**SST**). (\hat{h}^2) é a estimativa da herdabilidade obtiva pela média da distribuição a *posteriori*; ($\hat{\sigma}_{erro}^2$) estimativa da variância residual; ($\hat{\sigma}_{bloco}^2$) estimativa da variância devido ao fator bloco, ($\hat{\sigma}_{coluna}^2$) estimativa da variância devido ao fator coluna; ($\hat{\sigma}_{linha}^2$) estimativa da variância devido ao fator linha; $\hat{\sigma}_{progenie}^2$ é a estimativa da variância genética; HPD Highest posterior density e G_s é o ganho de seleção em relação a média geral estimada.

Tabela 15 - Índice de coincidência dos indivíduos selecionados (%) entre as 81 progênes de maracujazeiro-azedo do terceiro ciclo de seleção recorrente via modelo Bayesiano e por Modelos mistos REML/BLUP. Cambuci, RJ, 2020

Variável	Intensidade de Seleção (%)	BayesI/REML	BayesII/REML	Bayes I e II
NFrutos	37	96,67	93,33	93,33
	25	100,00	100,00	100,00
	15	100,00	91,67	91,67
Produtividade	37	100,00	100,00	100,00
	25	100,00	95,00	95,00
	15	100,00	91,67	91,67
MF	37	100,00	100,00	100,00
	25	100,00	90,00	90,00
	15	100,00	100,00	100,00
Comp.	37	100,00	93,33	93,33
	25	100,00	85,00	85,00
	15	100,00	83,33	83,33
Diam.	37	100,00	100,00	100,00
	25	95,00	95,00	100,00
	15	100,00	100,00	100,00
Polpa	37	100,00	93,33	93,33
	25	100,00	85,00	85,00
	15	100,00	83,33	83,33
EC	37	100,00	100,00	100,00
	25	100,00	100,00	100,00
	15	100,00	100,00	100,00
SST	37	100,00	100,00	100,00
	25	100,00	100,00	100,00
	15	100,00	100,00	100,00

Fonte: Elaborada pelo autor.

Número de frutos (**NF**), Produtividade (**Prod.**), Massa do Fruto (**MF**), Diâmetro Transversal do Fruto (**DTF**), Diâmetro Longitudinal do Fruto (**DLF**), Polpa de frutos (**Polpa**), Espessura da casca (**EC**), Sólidos Solúveis Totais (**SST**).

Com objetivo de comparar os *ranks* de indivíduos formados pelas duas metodologias de estimação foi definido três graus de intensidade de seleção (37%, 25%, 15%) conforme Tabela 15. Pode-se observar o índice de coincidência entre as progênes selecionadas por BLUP/REML, *Bayes I* e *Bayes II* varia de 83,33% a 100%. Pode-se também concluir que independente de um destes três métodos adotados as progênes selecionadas foram praticamente as mesmas em todos os cenários o que não interfere de forma significativa na seleção dos genitores da próxima geração.

Tabela 16 - Tabela com ranks das 30 primeiras progênies de maracujazeiro-azedo de irmãos completos do quarto ciclo de seleção recorrente via modelos mistos-REML/BLUP. Cambuci, RJ, 2020

Rank	NFrutos	Prod.	MF	Comp.	Diam.	Polpa	EC	SST
1	23	67	89	32	32	49	45	90
2	9	85	40	40	14	41	49	45
3	35	23	99	90	99	44	31	72
4	71	90	60	59	89	72	41	3
5	90	9	6	99	92	23	100	11
6	67	71	32	72	98	45	92	34
7	70	40	72	6	72	31	9	21
8	88	4	92	89	40	51	29	55
9	49	35	59	76	76	38	47	19
10	48	48	58	95	95	43	73	41
11	85	46	14	33	3	30	40	31
12	12	88	4	21	6	59	3	100
13	46	58	16	31	51	92	23	98
14	19	19	85	38	81	61	91	92
15	63	62	41	58	85	9	69	27
16	65	25	3	39	41	89	1	37
17	40	49	39	61	59	91	59	52
18	91	65	35	4	96	27	80	68
19	55	63	90	3	25	98	30	64
20	28	12	48	14	69	100	12	18
21	4	89	98	98	16	54	68	43
22	20	28	50	8	60	4	87	53
23	25	27	100	9	4	84	22	89
24	52	22	21	88	58	47	84	96
25	58	3	76	65	83	11	24	33
26	62	52	52	34	66	55	98	78
27	15	39	64	92	22	93	79	87
28	8	70	83	50	33	24	61	35
29	72	6	96	23	100	57	89	47
30	22	55	56	75	39	58	70	51
\bar{X}_G	116,44	18,35	158,57	80,97	72,23	46,84	7,34	13,40
\bar{X}_{30}	123,49	19,43	160,63	82,77	72,76	47,89	7,82	13,72

Fonte: Elaborada pelo autor.

Número de frutos (NF), Produtividade (Prod.), Massa do Fruto (MF), Diâmetro Transversal do Fruto (DTF), Diâmetro Longitudinal do Fruto (DLF), Polpa de frutos Polpa), Espessura da casca (EC), Sólidos Solúveis Totais (SST). *ranks* das progênies formado a partir do valor médio somado ao valor genotípico predito ($u + g$) via modelos mistos-REML/BLUP; \bar{X}_G é a média geral e \bar{X}_{30} média das 30 (trinta) primeiras progênies selecionadas.

Tabela 17 - Tabela com ranks e as médias estimadas das 30 primeiras progênes de maracujazeiro-azedo do terceiro ciclo de seleção recorrente da seleção recorrente via modelos Bayesiano. Cambuci, RJ, 2020

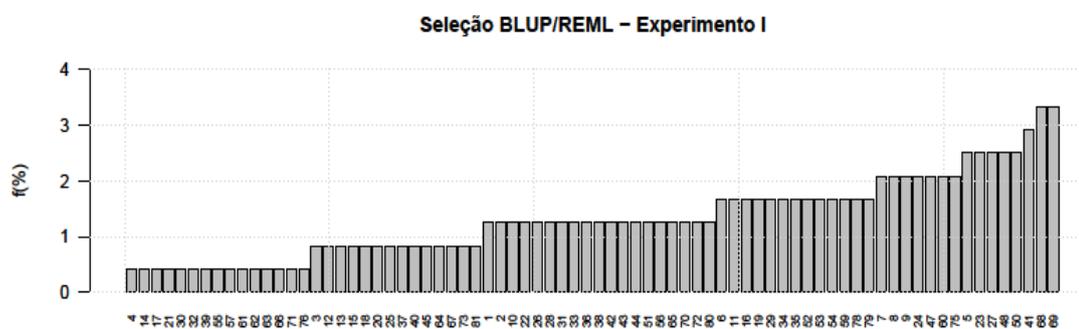
Rank	NFrutos	Prod.	MF	Comp.	Diam.	Polpa	EC	SST
1	23	67	89	32	32	49	45	90
2	9	85	99	59	14	41	49	45
3	35	23	60	90	99	45	31	72
4	71	90	6	40	89	72	41	3
5	90	9	40	89	92	23	100	11
6	67	71	32	6	98	44	92	34
7	70	4	72	72	72	89	9	21
8	88	40	14	38	40	51	29	55
9	49	35	92	95	76	30	47	19
10	48	46	59	99	95	31	73	41
11	85	48	4	21	3	38	40	31
12	46	19	58	14	6	27	3	100
13	19	58	39	76	51	92	23	98
14	63	88	3	39	81	91	91	92
15	12	62	41	4	85	61	69	27
16	65	25	85	98	41	84	1	37
17	91	63	16	8	59	59	59	52
18	40	65	48	58	96	43	80	68
19	55	49	21	9	25	54	30	64
20	28	28	50	3	16	52	12	18
21	4	89	100	65	69	9	68	43
22	52	27	98	31	60	98	87	53
23	25	52	76	33	4	11	22	89
24	58	12	90	61	58	4	84	96
25	20	22	35	66	83	100	24	33
26	62	70	52	23	22	93	79	78
27	15	3	64	88	66	24	98	87
28	8	39	83	85	33	57	61	35
29	93	6	56	50	100	20	89	47
30	68	55	96	92	39	47	70	51
\bar{X}_G	116,43	18,40	159,15	80,97	72,25	46,87	7,34	13,44
\bar{X}_{30}	123,49	19,45	162,32	82,77	72,80	47,79	6,89	13,72

Fonte: Elaborada pelo autor.

Número de frutos (**NF**), Produtividade (**Prod.**), Massa do Fruto (**MF**), Diâmetro Transversal do Fruto (**DTF**), Diâmetro Longitudinal do Fruto (**DLF**), Polpa de frutos (**Polpa**), Espessura da casca (**EC**), Sólidos Solúveis Totais (**SST**). *ranks* das progênes formado a partir do valor médio somado ao valor genotípico predito ($u + g$) via modelo bayesiano; \bar{X}_G é a média geral e \bar{X}_{30} média das 30 (trinta) primeiras progênes selecionadas.

Para uma intensidade de seleção de aproximadamente de 37%, o *rank* das 30 melhores progênes para cada característica está descrito nas Tabelas 16 e 17. Na Figura 2 (a) mostra as frequências entre as 30 progênes selecionadas para cada característica e observa-se que as progênes 89, 98, 92, 72, 58, 40, 4, 3, 100, 90, 59, 41, 23, 9 formam um bloco de progênes com alta probabilidade de serem as melhores progênes em todas as características avaliadas. Por outro

lado na Figura 2 (b) nos mostra as frequências das 30 progênes selecionadas pela metodologia de metodologia de modelos bayesianos (Tabela 17), e, pode-se observar que as progênes que formam o bloco com alta probabilidade de serem as melhores progênes em todas as características é praticamente o mesmo selecionado via metodologia de modelos mistos. Resultado divergente foi encontrado por Silva et al. (2020) na cultura da goiaba onde indivíduos que não haviam sido selecionados pelo método de estimação REML foram selecionados quando as observações foram modeladas pelo método de estimação bayesiano, algumas observações já foram alocadas nesta discussão quando reportado a avaliação do experimento I.



acurácias seletivas e preditivas dos modelos via metodologia 5-folds. No geral, as métricas de capacidade preditiva (CP) e a eficiência preditiva (REQM) - Tabela 18 - mostra que para cada variável tais métricas têm praticamente a mesma magnitude para os dois modelos. Indicando assim, que estes modelos testados tem a mesma capacidade de generalização.

Apenas as variáveis polpa, sólido solúvel total, massa do fruto do experimento II pode se dizer que são variáveis, onde os modelos tem baixa capacidade preditiva ($r < 0,14$).

Observando a Tabela 18 e fazendo a comparação entre as métricas sobre a ótica de metodologia de estimação, observa-se que, em geral, as métricas têm variação pequena, indicando assim, que os modelos têm a mesma capacidade de generalização, independente do método de estimação. Por exemplo, para a variável espessura da casca experimento I a $CP_{validação} = 0,38$ (Modelo Mistos) e $CP_{validação} = 0,39$ (Modelo Bayesiano), ou seja, uma variação pequena de 0,01. Como se pode observar é uma variação de baixa magnitude que ocorre para quase todas as variáveis. Assim, pode-se concluir que os dois modelos têm a mesma capacidade preditiva para todas as variáveis com exceção das variáveis massa de fruto, polpa de fruta e Sólidos Solúveis Totais do experimento II.

Nas etapas de validação e treinamento da validação cruzada (5-folds) os modelos têm a capacidade preditiva similar para todas as variáveis, com exceção das variáveis (MF, Polpa, SST) do experimento II. A capacidade preditiva para os modelos nas variáveis do experimento I varia de 21% a 54% e no experimento II de 6% a 40%. No experimento II para as variáveis (MF, Polpa, SST) observou-se uma baixa capacidade preditiva para os dois modelos.

Tabela 18 - Estimativas das métricas obtidas pelos modelos na validação cruzada 5-folds

Variável	Variável	Modelo	CP _{treino}	CP _{validação}	REQM _{treino}	REQM _{validação}
Experimento I	Número de Frutos	Misto	0,91	0,54	54,43	87,89
		Bayesiano	0,90	0,51	61,33	91,10
	Produtividade	Misto	0,89	0,50	7,98	12,68
		Bayesiano	0,89	0,49	8,19	12,64
	Massa de fruto	Misto	0,76	0,28	22,82	29,40
		Bayesiano	0,73	0,28	23,58	29,27
	DLF/Comprimento	Misto	0,86	0,46	4,22	6,14
		Bayesiano	0,85	0,46	4,29	6,14
	DTF/Diâmetro	Misto	0,88	0,46	2,97	4,73
		Bayesiano	0,88	0,47	3,00	4,72
Espessura da Casca	Misto	0,87	0,38	0,74	1,13	
	Bayesiano	0,87	0,39	0,75	1,12	
Polpa de fruta	Misto	0,77	0,26	4,43	6,04	
	Bayesiano	0,81	0,21	4,12	6,09	
Sólidos Solúveis Totais	Misto	0,87	0,47	0,77	1,16	
	Bayesiano	0,90	0,49	0,71	1,13	
Experimento II	Número de Frutos	Misto	0,74	0,39	33,44	44,43
		Bayesiano	0,75	0,39	33,74	44,23
	Produtividade	Misto	0,74	0,40	5,15	6,69
		Bayesiano	0,75	0,40	5,07	6,68
	Massa de fruto	Misto	0,70	0,07	20,43	24,85
		Bayesiano	0,71	0,13	20,21	24,52
	DLF/Comprimento	Misto	0,83	0,26	4,05	5,46
		Bayesiano	0,82	0,26	4,10	5,46
	DTF/Diâmetro	Misto	0,51	0,20	3,50	3,82
		Bayesiano	0,57	0,22	3,44	3,81

Tabela 18, Cont.

Variável	Variável	Modelo	CP _{treino}	CP _{validação}	REQM _{treino}	REQM _{validação}
Experimento II	Espessura da Casca	Misto	0,81	0,31	0,89	1,18
		Bayesiano	0,78	0,31	0,92	1,18
	Polpa de fruta	Misto	0,78	0,10	3,75	4,99
		Bayesiano	0,79	0,06	3,72	5,01
	Sólidos Solúveis Totais	Misto	0,79	0,14	1,19	1,35
		Bayesiano	0,78	0,12	1,19	1,34

Fonte: Elaborada pelo autor.

Capacidade preditiva do treinamento (**CP_{treino}**); Capacidade preditiva de validação (**CP_{validação}**); Raiz quadrada do erro quadrático médio do treinamento (**REQM_{treino}**); Raiz quadrada do erro quadrático médio do validação (**REQM_{validação}**)

6. CONCLUSÕES

A metodologia frequentista REML/BLUP e a metodologia bayesiana com *prioris* informativa selecionaram as mesmas progênies de maracujá-azedo.

REFERÊNCIAS BIBLIOGRÁFICAS

- Akaike, H. (1973) Information theory and an extension of the maximum likelihood principle. *In: Petrov, B.N., Csaki, F. (eds). Proceedings of the International Symposium on Information Theory, 2*, Budapest: Akademiai Kiado, p. 267-281.
- Alves, F.M., Resende, M.D.V. (2008) Avaliação genética de indivíduos e progênies de cupuaçuzeiro no estado do Pará e estimativas de parâmetros genéticos. *Ver Bras Frutic.* 30:696-701.
- Araújo, F.P., Junghans, T.G. (2017) Espécies de maracujazeiro no mercado internacional. *In: Junghans, T.G., Jesus, O.N. (eds.) Maracujá: do cultivo à comercialização.* Brasília, DF: Embrapa, p. 15-37.
- Arlot, S., Celisse, A. (2010) A survey of cross-validation procedures for model selection. *Statist. Aust J Crop Sci.* 10:411–417.
- Bates, B., Maechler, M., Bolker, B., Walker, S. (2015) Fitting linear mixed-effects models using lme4. *J Stat Softw.* 67:1-48.
- Bernacci, L.C., Cervi, A.C., Milward-de-Azevedo, M.A., Nunes, T.S., Imig, D.C., Mezzonato, bootstrap and covariance penalty methods. *Comput Statist Data Anal.* 54:2976-2989.
- Borém, A. (2017) *Melhoramento de plantas.* Viçosa: Ed. UFV, 543p.
- Borra, S., Ciaccio, A. (2010) Measuring the prediction error. A comparison of cross-validation, bootstrap and covariance penalty methods. *Comput Statist Data Anal.* 54:2976-2989.
- Browne, W.J. (2004) An illustration of the use of reparameterisation methods for improving MCMC efficiency in crossed random effects model. *Multilevel Modelling Newsletter*, 16:13-25.

- Box, G.E.P., Tião, G.C. (1973) *Bayesian inference in statistical analysis*. Reading: Addison-Wesley, 588p.
- Cavalcante, N.R. (2019) *Ganho genético em diferentes estratégias de obtenção de famílias de irmãos completos e estrutura populacional em maracujazeiro azedo*. Tese (Doutorado em Genética e Melhoramento de Plantas) - Campos dos Goytacazes-RJ, Universidade Estadual do Norte Fluminense - UENF, 87p.
- Cavalcante, N.R., Viana, A.P., Almeida Filho, J.E.A., Pereira, M.G., Ambrósio, M., Santos, E.A., Ribeiro, R.M., Rodrigues, D.L., Sousa, C.M.B. (2019) Novel selection strategy for half-sib families of sour passion fruit *Passiflora edulis* (Passifloraceae) under recurrent selection. *Genet Mol Res*.18:1-12.
- Cavichioli, J.C., Corrêa, L.S., Garcia, M.J.M., Fischer, I.H. (2011) Desenvolvimento, produtividade e sobrevivência de maracujazeiro-amarelo enxertado e cultivado em área com histórico de morte prematura de plantas. *Rev Bras Frutic*. 33:567-574.
- Congdon, P. (2007) *Bayesian statistical modelling*. 2nd ed. New York: Wiley Blackwell, 596p.
- Covarrubias, P.G. (2016). Genome assisted prediction of quantitative traits using the R Package Sommer. *PLoS One*, 11:1-15.
- Covarrubias, P.G. (2018) Software update: moving the R Package Sommer to multivariatmixed models for genome-assisted prediction. *bioRxiv*, 1-14.
- Crossa, J., Campos, G. de L., Perez, P., Gianola, D., Burgueno, J., Araus, J.L., Braun, H.J.
- Cruz, C.D. (2012) *Princípios de genética quantitativa*. Viçosa: Ed. UFV, 394p.
- Cruz, C.D., Carneiro, P.C.S. (2006) *Modelos biométricos aplicados ao melhoramento genético*. 2. ed. Viçosa: Ed. UFV, 585p.
- Cruz, C.D., Regazzi A.J., Carneiro, P.C.S. (2003) *Modelos biométricos aplicados ao melhoramento genético*. Viçosa: Ed. UFV, 585p.
- Entringer, G.C., Vettorazzi, J.C.F., Santos, E.A. (2016) Genetic gain estimates and selection of S1 progenies based on selection indices and REML/BLUP in super sweet corn. *Aust J Crop Sci*. 10:411–417.
- Faleiro, F.G., Junqueira, N.T.V., Braga, M.F. (2005) *Maracujá: germoplasma e melhoramento genético*. Planaltina: Embrapa Cerrados, 677p.
- Faleiro, F.G., Junqueira, N.T.V., Jesus, O.N., Costa, A.M., Machado, C.F., Junqueira, K.P., Araújo, F.P., Junghans, T.G. (2017) Espécies de maracujazeiro no mercado internacional. *In: Junghans, T.G., Jesus, O.N. (eds.) Maracujá: do cultivo à comercialização*. Brasília, DF: Embrapa, p. 15-37.

- Ferreira, R.T., Viana, A.P., Silva, F.H.L., Santos, E.A., Santos, J.O. (2016) Seleção recorrente intrapopulacional em maracujazeiro-azedo via modelos mistos. *Ver Bras Frutic.* 38:158-166.
- Gamerman, D. (1996) *Simulação estocástica via cadeia de Markov*. Caxambu: Associação Brasileira de Estatística, 196p.
- Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., Rubin, D.B. (2003) *Bayesian data analysis*. Boca Raton: Chapman and Hall/CRC, 661p.
- Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B. (2004) *Bayesian data analysis*. 2nd ed. Boca Raton: Chapman and Hall/CRC, 690p.
- Gelman, A., van Dyk, D.A., Huang, Z.Y., Boscardin, W.J. (2008) Using redundant parameterizations to fit hierarchical models. *J Comput Graph Stat.* 17:95-122.
- Geweke, J. (1992) Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. *In: Bernardo, J.M., Berger, J.O., Dawid, A.P., Smith, A.F.M. (eds.) Bayesian statistics 4*. New York: Oxford University Press, p. 625-631.
- Gezan, S.A., Huber, D.A., White, T.L. (2006) Post hoc blocking to improve heritability and precision of best linear unbiased genetic predictions. *Can J For Res.* 36:2141-2147.
- Gianola, D., Fernando, R.L., IM, S., Foulley, J.L. (1989) Likelihood estimation of quantitative genetic parameters when selection occurs: models and problems. *Genome*, 31:768-777.
- Gonçalves, G.M., Viana, A.P., Bezerra Neto, F.V., Pereira, M.G., Pereira, T.N.S. (2007) Seleção e herdabilidade na predição de ganhos genéticos em maracujá-amarelo. *Pesq. Agropecu. Bras.* 42:193-198.
- Gouy, M., Rousselle, Y., Bastianelli, D., Lecomte, P., Bonnal, L., Roques, D., Costet, L. (2013) Experimental assessment of the accuracy of genomic selection in sugarcane. *Theor Appl Genet.* 126:2575-2586.
- Hadfield, J. (2021) MCMCglmm Course Notes. [S.l.]: The Comprehensive R Archive Network, 144p.
- Hadfield, J.D. (2010) MCMC methods for multi-response generalized linear mixed models: the MCMCglmm R Package. *J Stat Softw.* 33:1-22.
- Hadfield, J.D., Nakagawa, S. (2010) General quantitative genetic methods for comparative biology: phylogenies, taxonomies and multi-trait models for continuous and categorical characters. *J Evol Biol.* 23:494-508.
- Heidelberger, P., Welch, P.D. (1983) Simulation run length control in the presence of an initial transient. *Oper Res.* 31:1109-1144.

- Henderson, C.R. (1949) Estimates of changes in herd environment. *J Dairy Sci.* 32:706.
- Henderson, C.R. (1973) Sire evaluation and genetic trends. *Proceedings of the Animal Breeding and Genetics Symposium in Honor of Jay L. Lush*, Champaign, Ill: American Society of Animal Science, p.10-41.
- Henderson, C.R. (1984) *Applications of linear models in animal breeding*. Ontario: University of Guelph, 462p.
- Henderson, C.R., Kempthorne, O., Searle, S.R., von Krosigk, C.M. (1959) The estimation of environmental and genetic trends from records subject to culling. *Biometrics*,13:192-218.
- models: the MCMCglimm R package. *J Stat Softw.* 33:1-22.
- Instituto Brasileiro de Geografia e Estatística (2018) Sistema IBGE de Recuperação Automática – SIDRA. *Prod Agric Munic.* 45:1-8.
- James, G., Witten, D., Hastie, T., Tibshirani, R. (2013) *An introduction to statistical learning: with applications in R*. New York: Springer Texts in Statistics, 441p.
- Junghans, T.G. (2015) *Guia de planta e propágulos de maracujazeiro*. Brasília, DF: Embrapa, 95p.
- Kohavi, R. (1995) A study of cross-validation and bootstrap for accuracy estimation and model selection. *Proceedings of the International Joint Conference on Artificial Intelligence*, 14, San Francisco: Morgan Kaufmann Publishers Inc., v. 2, p. 1137-1145.
- Lima, A.A., Cunha, M.A.P. (2004) Práticas culturais. *In: Lima, A.A., Cunha, M.A.P. Maracujá: produção e qualidade na passicultura*. Cruz das Almas: Embrapa Mandioca e Fruticultura, p. 169-178.
- Lindley, D.V., Smith, A.F.M. (1972) Bayes estimates for the linear model. *J R Stat Soc Series B Stat Methodol.* 34:1-41.
- Liu, C., Rubin, D.B., Wu, Y.N. (1998) Parameter expansion to accelerate EM: the PX-EM algorithm. *Biometrika*, 85:755-770.
- Liu, J.S., Wu, Y.N. (1999) Parameter expansion for data augmentation. *J Am Stat Assoc.* 94:1264-1274.
- Lousada Neto, F. (1991) Testes de sobrevivência acelerados: uma análise bayesiana do *Maracujá: produção e qualidade na passicultura*. Cruz das Almas: Embrapa Mandioca e *melhoramento genético*. Viçosa: Ed. UFV, 585p.
- Ludke, W. H., Andrade, A.C.B., Volpato, L., Oliveira, I.C.M. de, Paiva, J.T. de, Silva, M.J. da, del Conte, M.V., Silva, D.C., Almeida, V.C., Pinto, V.B. (2017). *Desafios biométricos no melhoramento genético*. Viçosa: Ed. UFV, 166p.

- Machado, R.F.C., Krause, W., Silva, C.A. Fachi, L.R., Silva, F.H.D.L., Viana, A.P. (2020). Implications of the post hoc blocking row-col technique on the intrapopulation improvement of the passion fruit. *Euphytica*, 216:1-11.
- Meletti, L.M.M. (2011) Avanços na cultura do maracujá no Brasil. *Rev Bras Frutic.* 33:83-91.
- Meletti, L.M.M., Brückner, C.H. (2001) Melhoramento genético. In: Brückner, C.H., Picanço, M.C. *Maracujá: tecnologia de produção, pós-colheita, agroindústria, mercado*. Porto Alegre: Cinco Continentes, p. 345-385.
- Meyer, K. (1986) Between algorithms: a "Short Cut" restricted maximum likelihood procedure to estimate variance components. *J Dairy Sci.* 69:1904-1916.
- Mila, A.L., Ngugi, H.K. (2011) A Bayesian approach to meta-analysis of plant pathology studies. *Phytopathology*, 101:42-51.
- Mulamba, N.N., Mock, J.J. (1978) Improvement of yield potential of the Eto Blanco maize (*Zea mays* L.) population by breeding for plant traits. *Egypt J Genet Cytol.* 7:40-51.
- Nogueira, D.A., Sáfadi, T., Ferreira, D.F. (2004) Avaliação de critérios de convergência para o método de Monte Carlo via Cadeias de Markov. *Rev Bras Estat.* 65:59-88.
- Oehlert, G.W. (1992) A note on the delta method. *Am Stat.* 46:27-29.
- Omer, S.O., Abdalla, A.W.H., Mohmmed, M.H., Singh, M. (2015) *Bayesian estimation of genotype-by-environment interaction in sorghum Variety trials. Commun Biometry Crop Sci.* 10:82-95.
- Patterson, H.D., Hunter, E.A. (1983) The efficiency of incomplete block designs in National List and Recommended List cereal variety trials. *J Agric Sci Camb.* 103:427-433.
- Paulino, C.D., Turkman, M.A.A., Murteira, B., Silva, G.L. (2018) *Estatística Bayesiana*. 2. ed. Lisboa: Fundação Calouste Gulbenkian, 616p.
- Pereira, J.M., Muniz, J.A., Sáfadi, T., Silva, C.A. (2009) Comparação entre modelos para predição do nitrogênio mineraliza do: uma abordagem Bayesiana. *Ciênc Agrotec.* 33:1792-1797.
- Piepho, H.P., Möhring, J., Melchinger, A.E., Büchse, A. (2008) BLUP for phenotypic selection in plant breeding and variety testing. *Euphytica*, 161:209-228.
- Plummer, M., Best, N., Cowles, K., Vines, K. (2006) CODA: Convergence Diagnosis and Output Analysis for MCMC. *R News*, 6:7-11.
- Raftery, A.E., Lewis, S. (1992) Comment: one long run with diagnostics implementation strategies for markov chain Monte Carlo. *Stat Sci.* 4:493-497.

- R Core Team (2019) *R: a language and environment for statistical computing*. Vienna: Foundation for Statistical Computing.
- Resende, M.D.V. de (2000) *Análise estatística de modelos mistos via REML/BLUP na experimentação em melhoramento de plantas perenes*. Colombo: Editora Embrapa Florestas, 101p. (Embrapa Florestas. Documentos, 47).
- Resende, M.D.V. de (2002a) *Genética biométrica e estatística no melhoramento de plantas perenes*. Brasília: Embrapa Informação Tecnológica, 975p.
- Resende, M.D.V. de (2002b) Procedimentos ótimos de seleção com dados balanceados e desbalanceados. *In*: Resende, M.D.V. de. *Genética biométrica e estatística no melhoramento de plantas perenes*. Brasília: Embrapa Informação Tecnológica, p. 209-347.
- Resende, M.D.V. de (2004) *Métodos estatísticos ótimos na análise de experimentos de campo*. Colombo: Embrapa Florestas, 57p. (Embrapa Florestas. Documentos, 100).
- Resende, M.D.V. de, Silva, F.F., Azevedo, C.F. (2014) *Estatística matemática, biométrica e computacional: modelos mistos, multivariados, categóricos e generalizados (REML/BLUP), inferência bayesiana, regressão aleatória, seleção genômica, QTL- GWAS, estatística espacial e temporal, competição, sobrevivência*. Viçosa, MG: Ed. UFV, 881p.
- Ribeiro R.M., Viana A.P., Santos E.A., Rodrigues, D.L., Preisigke, S. da C. (2019) Breeding passion fruit populations: review and perspectives. *Funct Plant Breed J.* 1:16-29.
- Rossi, A.D., Rossi, F.S., Silva, J.R. (2001) *Produção de sucos tropicais: maracujá*. Vera Cruz: AFRUVEC, 47p.
- Santos, E.A., Viana, A.P., Freitas, J.C.O., Rodrigues, D.L., Tavares, R.F., Paiva, C.L., Souza, M.M. (2015) Genotype selection by REML/BLUP methodology in a segregating population from an interspecific *Passiflora* spp. crossing. *Euphytica* 204:1-11.
- Santos, P.R., Viana, A.P., Gomes, V.M., Preisigke, S. da C. (2018) Clonal selection in interspecific *Vitis* spp. hybrids resistant to the root-lesion nematode *Pratylenchus brachyurus* by REML/BLUP. *Fruits*, 73:191–197.
- Silva, A. J. (2017) *Abordagem bayesiana e modelos mistos para experimentos multiambientes na cultura da soja*. Tese (Doutorado em Genética e Melhoramento de Plantas) - Faculdade de Ciências Agrárias e Veterinárias, Universidade Estadual Paulista - UEP, 33p.
- Silva, F.F., Viana, J.M., Faria, V.R, Resende M.D. de (2013) Bayesian inference of mixed models in quantitative genetics of crop species. *Theor Appl Genet.* 126:1749-1761.

- Silva, F.H. de L., Viana, A.P., Corrêa, C.C.G., Carvalho, B.M., Sousa, C.M.B. de, Amaral, B.D., Ambrósio, M., Glória, L.S. (2020) Impact of Bayesian inference on the selection of psidium guajava. *Sci Rep.* 10:1-9.
- Silva, F.H. de L., Viana, A.P., Freitas, J.C. de O., Santos, E.A., Rodrigues, D.L., Amaral Júnior, A.T. do (2017) Prediction of genetic gains by selection indexes and REML/BLUP methodology in a population of sour passion fruit under recurrent selection. *Acta Sci, Agron.* 39:183-190.
- Silva, F.H. de L., Muñoz, P.R., Vincent, C.I., Viana, A.P. (2016) Generating relevant information for breeding *Passiflora edulis*: genetic parameters and population structure. *Euphytica* 208:609–619.
- Silva, G.N., Tomaz, R.S., Sant'Anna, I.C., Nascimento, M., Bhering, L.L., Cruz, C.D. (2014) Neural networks for predicting breeding values and genetic gains. *Sci Agric.* 71:494-498.
- Silva, J.A., Sanches, A., Carla, A., Andrade, B. (2018) Bayesian approach, traditional method, and mixed models for multienvironment trials of soybean. *Pesq Agropec Bras.* 53:1093-1100.
- Silva, M.G.M. (2009) *Seleção recorrente intrapopulacional no maracujazeiro amarelo*. Tese (Doutorado em Produção Vegetal) - Universidade Estadual do Norte Fluminense – UENF, 157p.
- Silva, T.V., Resende, E.D., Viana, A.P., Pereira, S.M.F., Carlos, L.A., Vitorazi, L. (2008) Qualidade do suco de maracujá-amarelo em diferentes épocas de colheita. *Sci Technol.* 28:545-550.
- Sorensen, D., Gianola, D. (2002) *Likelihood, Bayesian and MCMC methods in quantitative genetics*. Berlin: Springer Science & Business Media, 760p.
- Sousa, C.M., Gravina, G.A., Viana, A.P., Daher, R.F., Souza, C.L. (2017) Selection of snap bean F2 progenies for production using the REML/BLUP methodology. *Hortic Bras.* 35:33-40.
- Sousa, I.F., Johan, I., Kunzle Neto, J.E., Muniz, J.A., Guimarães, R.M., Savian, T.V., Muniz, F.R. (2014). Fitting nonlinear autoregressive models to describe coffee seed germination Ajuste de modelos não lineares autorregressivos na descrição da germinação se sementes de café. *Ciênc Rural*, 44:2016-2021.
- Souza Júnior, C.D. (2001) *Melhoramento de espécies alógamas*. In: Nass, L.L., Valois, A.C.C., Melo, I.S. de, Valadares-Inglis, M.C. (orgs.) Recursos genéticos e melhoramento de plantas. Rondonópolis: Fundação MT, p. 159-199.
- Spiegelhalter, D.J., Best, N.G., Carlin, B.P., van der Linde, A. (2002) Bayesian measures of model complexity and fit (with discussion). *J R Statist Soc B.* 64, 583–639.

- Stida, W.F., Daher, R.F., Viana, A.P., Vidal, A.K.F., Freitas, R.S., Silva, V.B. da, Pereira, A.V., Cassaro, S., Menezes, B.R. de S., Furlani, E.P. (2018) Estimation of genetic parameters and selection of elephant-grass (*Pennisetum purpureum* Schumach.) for forage production using mixed models. *Chil J Agri. Res.* 78:198-204.
- Thompson, R., Brotherstone, S., White, I.M.S. (2005) Estimation of quantitative genetic parameters. *Philos Trans R Soc Lond, Ser B, Biol Sci.* 275:679-686.
- van Dyk, D.A., Meng, X.L. (2001) The art of data augmentation. *J Comput Graph Stat.* 10:1-50.
- Viana, A.P., Pereira, T.N.S., Pereira, M.G., Amaral Júnior, A.T. do, Souza, M.M. de., Maldonado, J.F.M. (2004) Parâmetros genéticos em populações de maracujazeiro-amarelo. *Rev Ceres*, 51:545-555.
- Viana, A.P., Resende, M.D.V. (2014) *Genética quantitativa no melhoramento de fruteiras*. Rio de Janeiro: Interciência, 296p.
- Viana, A.P., Silva, F.H.L.E., Gonçalves, G.M., Silva, M.G.M., Ferreira, R.T., Pereira, T.N.S., Pereira, M.G., Amaral Júnior, A.T. do, Carvalho, G.F. (2016) UENF Rio 87 Dourado: a new passion fruit cultivar with high yield potential. *Crop Breeding and Applied Biotechnology*, 16:250-253.
- Vianna-Silva, T., Resende, E.D. de, Vianna, A.P., Pereira, S.M.F., Rosa, R.C.C., Carlos, L. de A., Vitorazzi, L. (2008) Qualidade do suco de maracujá amarelo em diferentes épocas de colheita. *Ciênc Tecnol Aliment.* 28:545-550.
- Wilberg, M.J., Bence, J.R. (2008). Performance of deviance information criterion model selection in statistical catch-at-age analysis. *Fish Res.* 93:212-221.