

MATHEUS DE PAULA FERREIRA

**GEOESTATÍSTICA E AEROFOTOGRAMETRIA APLICADAS À SELEÇÃO DE
FAMÍLIAS DE CANA-DE-AÇÚCAR**

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

Orientador: Luiz Alexandre Peternelli

Coorientador: Gérson Rodrigues dos Santos

**VIÇOSA - MINAS GERAIS
2020**

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

T

F383g
2020
Ferreira, Matheus de Paula, 1988-
Geoestatística e aerofotogrametria aplicadas à seleção de
famílias de cana-de-açúcar / Matheus de Paula Ferreira. –
Viçosa, MG, 2020.
72 f. : il. (algumas color.) ; 29 cm.

Inclui anexos.

Orientador: Luiz Alexandre Peternelli.

Tese (doutorado) - Universidade Federal de Viçosa.

Referências bibliográficas: f. 60-68.

1. Análise de covariância. 2. Cana-de-açúcar - Seleção -
Métodos estatísticos. 3. Cana-de-açúcar - Melhoramento
genético. 4. Geologia - Métodos estatísticos. 5. Fertilidade do
solo. 6. Fotogrametria aérea. I. Universidade Federal de Viçosa.
Departamento de Estatística. Programa de Pós-Graduação em
Estatística Aplicada e Biometria. II. Título.

CDD 22. ed. 519.538

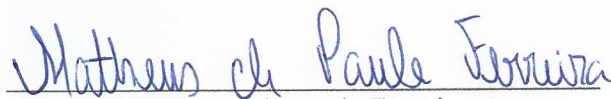
MATHEUS DE PAULA FERREIRA

**GEOESTATÍSTICA E AEROFOTOGRAMETRIA APLICADAS À SELEÇÃO DE
FAMÍLIAS DE CANA-DE-AÇÚCAR**

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

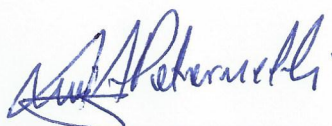
APROVADA: 19 de outubro de 2020.

Assentimento:



Matheus de Paula Ferreira

Autor



Luiz Alexandre Peternelli

Orientador

Dedico aos meus pais.

AGRADECIMENTOS

Agradeço, primeiramente, a Deus pelas graças e benções nessa longa jornada.

Aos meus pais, Mauro e Denísia, que nunca polparam esforços para que eu realizasse meus estudos.

Ao meu irmão Philippe, companheiro para toda a vida.

Ao meu sobrinho Lucas, que trouxe alegria em momentos difíceis.

Ao meu tio João, pelo exemplo e incentivo.

A todos meus familiares, pelas palavras de encorajamento.

Aos meus orientadores Luiz Alexandre Peternelli e Gérson Rodrigues dos Santos pela disponibilidade, amizade e conhecimentos transmitidos.

Aos professores e funcionários do Departamento de Estatística da Universidade Federal de Viçosa, sempre solícitos nas necessidades dos estudantes.

Aos membros da banca por aceitarem o convite e por estarem dispostos a dar suas contribuições para este trabalho.

Aos amigos que Viçosa me deu, dos tempos da graduação, passando pelo mestrado e finalizando com o doutorado.

À Universidade Federal de Viçosa e ao Programa de Pós-Graduação em Estatística Aplicada e Biometria pela oportunidade.

À Iowa State University, especialmente ao Doutor Mehari Tekeste, onde tive a oportunidade de fazer o doutorado sanduíche.

À CAPES, pela concessão da bolsa de estudos.

Por fim, a todos os que contribuíram direta ou indiretamente para a concretização deste trabalho.

“Tudo posso Naquele que me fortalece.”
(Filipenses 4:13)

RESUMO

FERREIRA, Matheus de Paula, D.Sc., Universidade Federal de Viçosa, outubro de 2020. **Geoestatística e aerofotogrametria aplicadas à seleção de famílias de cana-de-açúcar.** Orientador: Luiz Alexandre Peternelli. Coorientador: Gérson Rodrigues dos Santos.

Um dos desafios para os programas de melhoramento de cana-de-açúcar é a seleção inicial de genótipos. A fertilidade do solo pode influenciar no desenvolvimento das plantas e tem potencial de afetar essa seleção, porém a análise de parâmetros de fertilidade demanda mais tempo, mão de obra e gera custos adicionais. Uma possível alternativa a isto é a utilização de imagens de campo para explicar as informações do solo. Assim, o objetivo deste trabalho foi incorporar as informações dos atributos do solo, via Geoestatística e/ou Aerofotogrametria, e corrigir o ranqueamento das famílias de cana-de-açúcar com a incorporação de covariáveis no modelo de delineamento experimental, e verificar a viabilidade da utilização de imagens RGB como alternativa à análise do solo. Neste trabalho foram utilizados dados provenientes de um experimento com 60 famílias no delineamento em blocos casualizados, com quatro repetições, tendo como variável resposta a média de toneladas de colmos por hectare (TCH). Além disso, na área do experimento foram coletadas 36 amostras de solo e dois imageamentos realizados por VANT (veículo aéreo não tripulado). Os dados da análise de fertilidade foram modelados via Geoestatística e incorporados ao modelo de delineamento mediante análise de covariância. As variáveis incorporadas ao modelo foram o índice de saturação por alumínio, magnésio, fósforo, pH e a soma de bases. A inclusão dessas variáveis influenciou no ranqueamento das famílias demonstrando a sua importância na seleção durante as fases iniciais do melhoramento genético, além de possibilitar a criação de um mapa de correção do modelo para as parcelas do experimento. A partir das imagens RGB foram calculados índices e obtidos os mapas de valor médio, amplitude e desvio-padrão das unidades experimentais, porém quando comparados ao mapa de correção de médias do modelo proposto não foi observado grau de concordância satisfatório, indicando assim que imagens RGB não foram capazes de substituir os dados coletados do solo. Com a utilização dos mapas de krigagem dos atributos do solo foi possível verificar a não homogeneidade dos blocos do experimento. Dessa forma, é recomendado fazer uma investigação da qualidade do solo via análises de fertilidade e Geoestatística, para então corrigir o solo e definir os blocos experimentais corretamente, evitando assim o efeito das variáveis no ranqueamento das famílias. Caso não seja realizada esta análise anteriormente a implantação do experimento, após a colheita da cana-de-açúcar deve-se utilizar as análises de

fertilidade e Geoestatística para incluir o efeito do solo no modelo empregado para ranquear as famílias.

Palavras-chave: Imageamento RGB. Variabilidade Espacial. Fertilidade Do Solo. Ranqueamento De Famílias. *Saccharum* Spp.

ABSTRACT

FERREIRA, Matheus de Paula, D.Sc., Universidade Federal de Viçosa, October, 2020. **Geostatistics and images as alternatives for ranking sugarcane families.** Adviser: Luiz Alexandre Peternelli. Co-adviser: Gérson Rodrigues dos Santos.

One of the obstacles faced by the sugarcane breeding programs is the selection of the first genotypes. Soil fertility can influence the development of plants and has the potential to affect the selection of individuals, but incorporating this information requires more time, labor, and generates additional costs. A possible alternative to the analysis of fertility is the use of field images to explain the soil information. The objective of this work was to incorporate information on soil attributes, via Geostatistics and/or Aerophotogrammetry, and to correct the ranking of sugarcane families with the incorporation of covariables in the experimental design model, in addition to verifying the feasibility of using RGB images as an alternative to soil analysis. In this work, data from an experiment with 60 families were used in a randomized block design, with 4 repetitions, with variable response to the average of tons of stalks per hectare (TCH). Also, in the experiment area, 36 soil samples were collected and two UAV (unmanned aerial vehicle) imaging. The data from the fertility analysis were modeled using Geostatistics and incorporated into the design model through covariance analysis. The variables incorporated into the model were saturation index for aluminum, magnesium, phosphorus, pH and sum of bases. The inclusion of these variables influenced the ranking of families, demonstrating their importance in the selection of families in the early stages of genetic improvement, in addition to enabling the creation of a model correction map for the experimental plots. From the RGB images, indexes were calculated and maps of mean value, amplitude and standard deviation of the experimental units were obtained, however, when compared to the means correction map of the proposed model, a satisfactory degree of agreement was not observed, thus indicating that RGB images were unable to replace the data collected from the soil. Using the kriging maps of the soil attributes it was possible to verify the inhomogeneity of the experiment blocks. Thus, for future experiments, one must first investigate the soil quality via fertility and geostatistics analysis, then correct the soil and define the blocks of the experiments correctly, thus avoiding the effect of variables on the ranking of families. If such an analysis is not carried out before the experiment is carried out, after the sugarcane harvest, fertility and geostatistics analysis should be used to include the effect of the soil in the model used to rank families.

Keywords: RGB Imaging. Spatial Variability. Soil Fertility. Family Ranking. *Saccharum* Spp.

LISTA DE ILUSTRAÇÕES

Figura 1 – Exemplos de semivariogramas teórico e empírico.	19
Figura 2 – Representação do espaço tridimensional RGB no plano cartesiano	24
Figura 3 – Exemplo de imagens (a) Padrão RBG. (b) Falsa banda <i>Red</i> . (c) Falsa banda <i>Green</i> . (d) Falsa banda <i>Blue</i>	25
Figura 4 – Fluxograma dos processos envolvidos nas análises.....	32
Figura 5 – Município de Oratórios – Minas Gerais.....	33
Figura 6 – Blocos experimentais do experimento. (a) Bloco 1. (b) Bloco 2. (c) Bloco 3. (d) Bloco 4.....	34
Figura 7 – (a) Grade amostral de pontos para análise de fertilidade do solo. (b) Retira de amostra para análise de fertilidade do solo	35
Figura 8 – Ortomosaicos da área de estudo. (a) Voo realizado em 03/10/2018. (b) Voo realizado em 15/03/2019.....	36
Figura 9 – Recortes das parcelas 15, 42, 170, 217 e 238 do voo realizado em 15/03/2019 ...	36
Figura 10 – Mapas de krigagem das variáveis. (a) Alumínio. (b) Cálcio. (c) Acidez potencial. (d) Índice de saturação por alumínio.	44
Figura 11 – Mapas de krigagem das variáveis. (a) Magnésio. (b) Fósforo. (c) Potencial Hidrogeniônico. (d) Soma de bases. (e) Capacidade de troca catiônica. (f) Índice de saturação por base.....	45
Figura 12 – Mapas de valores a nível de parcela das variáveis. (a) Alumínio. (b) Cálcio. (c) Acidez potencial. (d) Índice de saturação por alumínio.....	46
Figura 13 – Mapas de valores a nível de parcela das variáveis. (a) Magnésio. (b) Fósforo. (c) Potencial Hidrogeniônico. (d) Soma de bases. (e) Capacidade de troca catiônica. (f) Índice de saturação por base.....	47
Figura 14 – Valores residuais do modelo ajustado em função das parcelas para os delineamentos em blocos casualizados (DBC).....	49
Figura 15 – Gráfico de dispersão do ranqueamento do modelo TRADICIONAL vs DBC ...	51
Figura 16 – Mapa de correção da variável TCH para o modelo DBC.....	53
Figura 17 – Índices RGB calculados para os dois voos. (a) VARI (<i>Visible Atmospherically Resistant</i>) - 03/10/2018. (b) GLI (<i>Green Leaf Index</i>) - 03/10/2018. (c) NGRDI (<i>Normalized Green Red Difference Index</i>) - 03/10/2018. (d) VARI (<i>Visible Atmospherically Resistant</i>) - 15/03/2019. (e) GLI (<i>Green Leaf Index</i>) - 15/03/2019. (f) NGRDI (<i>Normalized Green Red Difference Index</i>) - 15/03/2019.....	54
Figura 18 – Valor médio dos índices RGB para cada parcela do experimento. (a) VARI (<i>Visible Atmospherically Resistant</i>) - 03/10/2018. (b) GLI (<i>Green Leaf Index</i>) - 03/10/2018. (c) NGRDI (<i>Normalized Green Red Difference Index</i>) - 03/10/2018. (d) VARI (<i>Visible Atmospherically Resistant</i>) - 15/03/2019. (e) GLI (<i>Green Leaf Index</i>) - 15/03/2019. (f) NGRDI (<i>Normalized Green Red Difference Index</i>) - 15/03/2019.	55
Figura 19 – Valor da amplitude dos índices RGB para cada parcela do experimento. (a) VARI (<i>Visible Atmospherically Resistant</i>) - 03/10/2018. (b) GLI (<i>Green Leaf Index</i>) - 03/10/2018. (c) NGRDI (<i>Normalized Green Red Difference Index</i>) - 03/10/2018. (d) VARI (<i>Visible Atmospherically Resistant</i>) - 15/03/2019. (e) GLI (<i>Green Leaf Index</i>) - 15/03/2019. (f) NGRDI (<i>Normalized Green Red Difference Index</i>) - 15/03/2019.	56
Figura 20 – Valor do desvio-padrão dos índices RGB para cada parcela do experimento. (a) VARI (<i>Visible Atmospherically Resistant</i>) - 03/10/2018. (b) GLI (<i>Green Leaf Index</i>) - 03/10/2018. (c) NGRDI (<i>Normalized Green Red Difference Index</i>) - 03/10/2018. (d) VARI (<i>Visible Atmospherically Resistant</i>) - 15/03/2019. (e) GLI (<i>Green Leaf Index</i>) - 15/03/2019. (f) NGRDI (<i>Normalized Green Red Difference Index</i>) - 15/03/2019.	57

LISTA DE TABELAS

Tabela 1 – Comparação de ranqueamento das melhores famílias quanto ao TCH, em função do percentual de seleção de melhores famílias.....	30
Tabela 2 – Estatística descritiva dos valores de TCH.....	40
Tabela 3 – Estatística descritiva das variáveis da análise de fertilidade do solo	41
Tabela 4 – Método escolha do semivariograma	43
Tabela 5 – Parâmetros dos semivariogramas para as variáveis de análise do solo	43
Tabela 6 – Estatística descritiva da análise de fertilidade do solo	48
Tabela 7 – Resultado do quadro de ANOVA	48
Tabela 8 – Estatística descritiva dos resíduos	48
Tabela 9 – Teste estatísticos de normalidade e homogeneidade de variância para o DBC. ...	50
Tabela 10 – Tabela da ANOVA após execução do método Stepwise de seleção de variáveis	50
Tabela 11 – Ranqueamento das famílias quanto ao TCH nos modelos com correção no DBC e TRADICIONAL	52
Tabela 12 – Comparação de ranqueamento das melhores famílias quanto ao TCH, em função do percentual de seleção de melhores famílias.....	53
Tabela 13 – Valores do Índice Kappa para os mapas de estatísticas a nível de parcelas dos índices de imagens RGB com os mapas de correção de médias dos modelos DBC e TRADICIONAL.....	58

SUMÁRIO

INTRODUÇÃO GERAL.....	14
1.1. Motivação	14
1.2. Objetivos	17
1.3. Estrutura de tese.....	17
2. REVISÃO DE LITERATURA.....	18
2.1. Geoestatística	18
2.1.1. Semivariograma	19
2.1.2. Krigagem	21
2.2. Aerofotogrametria.....	23
2.3. Fertilidade do solo.....	25
2.4. Análise de covariância	27
2.5. Método Stepwise.....	29
2.6. Índice Kappa	30
3. MATERIAIS E MÉTODOS	31
3.1. Dados experimentais	32
3.2. Dados experimentais	37
3.3. Análises estatísticas.....	37
3.3.1. Modelagem espacial dos atributos da análise de fertilidade do solo.....	37
3.3.2. Obtenção dos valores médios a nível de parcela dos atributos do solo	37
3.3.3. Seleção das covariáveis para correção de TCH.....	38
3.3.4. Ranqueamento das famílias de cana-de-açúcar	38
3.3.5. Cálculo dos índices de imagens RGB.....	39
3.3.6. Comparação entre o mapa de correção e de índices de imagens.....	39
4. RESULTADOS	40
4.1. Análises Estatística.....	40
4.1.1. Análise dos dados experimentais de TCH	40
4.1.2. Análise de fertilidade do solo.....	40
4.1.3. Modelagem dos semivariogramas.....	42
4.1.4. Médias das variáveis de solo nível de parcela	46
4.1.5. Análise de Resíduos da Regressão	48
4.1.6. Método Stepwise.....	50
4.1.7. Ranqueamento da famílias.....	51
4.1.8. Índices de imagens RGB	54

4.1.9. Índice Kappa	58
5. CONCLUSÕES	59
REFERÊNCIAS	60
ANEXOS	69

INTRODUÇÃO GERAL

1.1. Motivação

O Brasil é o maior produtor de cana-de-açúcar do planeta com área de plantio de, aproximadamente, 8.400 mil hectares segundo estimativas da temporada 2020/21. Atualmente a cana-de-açúcar é ligada ao setor de biocombustíveis, com produtos como o etanol e derivados, e opera numa conjuntura positiva e sustentável (CONAB, 2020). Os principais produtos, açúcar e etanol, tem produção média estimada de 35,3 milhões de toneladas e 29,3 bilhões de litros respectivamente (CONAB, 2020). As últimas décadas apontam para uma tendência na utilização de fontes renováveis de energia devido a crescente preocupação com os impactos ambientais associados aos combustíveis fósseis. Consequentemente, a cana-de-açúcar foi apontada como uma alternativa viável na busca de energia limpa e sustentável (HOFSETZ e SILVA, 2012). Quanto ao desenvolvimento de biocombustíveis, o Brasil disputa o comando tecnológico com Alemanha, China e Estados Unidos (Milanez et al. 2015).

Com a adoção de tecnologias voltadas para o aumento da produtividade, o setor sucroenergético brasileiro é referência mundial, oferecendo diversas variedades ao mercado, sendo o trabalho dos programas de melhoramento genético o maior responsável por tal tarefa (MELO, 2014).

O desenvolvimento de novas variedade genéticas demanda anos de experimentação e estudo da variabilidade genética, pois diferentes genótipos apresentam particularidades para características de interesse no processo de seleção. São intrínsecos de cada genótipo, propriedades do colmo (modo de perfilhamento, número, diâmetro, altura) e das folhas (largura e comprimento). Além disso, fatores ambientais, climáticos e do próprio experimento podem exercer influência nas características observadas. (Santana, 2012).

Em conformidade com Rodrigues (1995), características que definem números de colmos, altura, diâmetro de colmos, comprimento e largura de folhas e a arquitetura da parte aérea, são inerentes a cada genótipo, sendo a sua expressão influenciada pelo clima e pelas práticas culturais

Um dos maiores obstáculos nos programas de melhoramento da cana-de-açúcar é a seleção dos primeiros genótipos, é de suma importância que a seleção seja eficaz (Moreira e Peternelli, 2015). Outro fator que pode limitar o processo de seleção são os aspectos computacionais envolvidos nas análises. Para a seleção de indivíduos nos programas de melhoramento de cana-de-açúcar a seleção massal tem sido amplamente utilizada, essa técnica

pode ser entendida com um método baseado na experiência em campo, na qual avalia-se nos clones aspectos de interesse da cana-de-açúcar acerca de produtividade e saúde. Uma alternativa a essa metodologia é a seleção de famílias promissoras para futuramente fazer a seleção individual dos clones, uma vez que a herdabilidade para as famílias é maior quanto ao rendimento da planta (Moreira, 2013).

No melhoramento o ganho genético pode ser entendido como o aumento no desempenho alcançado por unidade de tempo mediante da seleção artificial, uma das maiores limitações para que esse crescimento ocorra e se tenha um avanço genético é a identificação de genótipos superiores de forma eficiente baseada nas informações fenotípicas observadas no campo (ARAUS et al., 2018).

A caracterização morfoagronômica é efetuada manualmente devido a carência de tecnologias adequadas, tornando inviável a verificação em larga escala (TANGER et al., 2017).

Resumidamente, o período de obtenção de genótipos pode ser descrito como uma seleção a partir da hibridação de genitores previamente escolhidos, seguido de uma fase de obtenção de clones. O processo de seleção de clones perdura em média 10 anos até a obtenção de genótipos superiores (Ferreira, 2020, BARBOSA et al., 2005).

Na fase embrionária do processo de seleção, muitos genótipos são utilizados a fim de maximizar a probabilidade de se obter indivíduos superiores. A primeira fase do melhoramento é chamada T1, onde as plântulas oriundas dos cruzamentos são transplantadas individualmente para sulcos com espaçamento de 1,40 m, sendo cada plântula a 0,50 m da outra no sulco (Ferreira, 2020). Essa é a etapa mais importante dos programas de melhoramento da cana-de-açúcar, pois é quando os primeiros genótipos são selecionados e erros de seleção em T1 podem frustrar o sucesso de todo o programa (Barbosa e Silva, 2012).

Alta quantidade de área experimental, mão-de-obra e recursos financeiros são algumas das limitações no melhoramento genético, a avaliação dos possíveis candidatos é trabalhosa e onerosa. Afim de minimizar o espaço e o tempo necessário para a realização de experimentos nessa etapa do melhoramento genético, novos arranjos experimentais estão surgindo, como o sistema simplificado proposto por Melo (2014). Outra alternativa muito utilizada é a seleção de famílias ao invés de clones individuais, e posteriormente, a distinção dos genótipos superiores dentro das famílias com melhores características (Moreira e Peternelli, 2015).

Várias características podem auxiliar e possivelmente antecipar alguns resultados. A utilização de informações provenientes do solo pode contribuir na qualidade da seleção de genótipos, uma vez que os nutrientes presentes no solo exercem influência direta no desenvolvimento das plantas.

A regulação das atividades metabólicas de uma planta, demanda uma grande quantidade de nutrientes do solo, bem como uma absorção dos mesmos em proporções ideais (Ronquim, 2010). Na literatura o mapeamento da variabilidade dos atributos do solo é grande e engloba vários tipos de solo, clima e culturas. A análise do comportamento espacial dos atributos do solo no contexto agrícola permite um manejo otimizado, minimizando danos ambientais e maximizando a produtividade (Silva et al, 2013)

Bottega et al. (2013) e Alvez et al. (2018) estudaram a variabilidade espacial de atributos do solo em sistema de semeadura direta com rotação de culturas no cerrado brasileiro e cultivados com cana-de-açúcar no pontal do Paranapanema respectivamente. Machado et al. (2007) analisaram espacialmente o comportamento químico do solo em áreas sob sistema plantio convencional. No estudo químico do solo de cultivado com soja sob plantio direto, Dalchiavon et al. (2017) destaca que o conhecimento geoestatístico dos atributos do solo é fundamental para o desenvolvimento das culturas em geral, logo características do solo podem ser utilizados como fator de correção dos modelos experimentais, e para isso é necessário mapear a variabilidades desses atributos. O estudo da variabilidade do solo determina a quantidade de nutrientes em uma determinada região e expressa sua fertilidade, logo é extremamente importante os processos de amostragem e análise realizadas.

Os métodos tradicionais de obtenção de informações carecem de amostragem manual, trabalhosa e lenta, porém a utilização de veículos aéreos não tripulados e sensores embarcados tornaram essa tarefa rápida, eficiente e flexível (Yang, 2017), indicando um caminho promissor para a utilização dessas plataformas e seus sensores no âmbito agrário.

Atualmente os drones tem sido utilizados nas mais diversas atividades civis e científicas devido a facilidade de acesso e mobilidade que essas aeronaves proporcionam. Diversos estudos tem focado na utilização dessas aeronaves em áreas aplicadas do conhecimento, analisando a qualidade e viabilidade do seu uso (Ferreira et al., 2013; Santos, 2016; Fonseca Neto et al., 2017). A utilização dessas plataformas para fins civis é gigantesca, com atividades como monitoramento e localização em tempo real, conectividade, segurança e logística, porém ainda carece de maior entendimento e desenvolvimento, visto que ainda não há garantias de segurança operacional e financeira dos VANTS (SHAKHATREH et al, 2019).

As informações de imagens aéreas para o mapeamento no contexto agrícola têm sido amplamente discutidas nos últimos anos. Ribeiro et al. (2017) estudou a variabilidade espaço-temporal das condições da vegetação por meio de imagens de satélite, Alface et al. (2019) utilizou índice calculados via de imagem de satélite para monitorar cana-de-açúcar e analisar a sensibilidade do índice no vigor vegetativo da cultura, concluindo a existência de semelhança

entre as variáveis estudadas e consequentemente a possibilidade de obtenção de um coeficiente de cultivo. Peruzzo et al. (2019) no monitoramento ambiental da bacia hidrográfica observou alterações uso do solo provenientes de variações climáticas e consequências hídricas, vegetais e no solo. Paludo et al. (2020) utilizaram sensoriamento remoto e análise em nuvem para mapear áreas de culturas de milho e soja via imagens de satélites landsat-8 e Sentinel-2. FREIRE-SILVA et al. (2019) verificou que nas revisões de literatura acerca de índices de vegetação nenhuma utilizou apenas as bandas do visível RGB, especialmente no Brasil.

Tal contexto legitima a busca por novas metodologias que viabilizem o processo de seleção de indivíduos no melhoramento genético. Na seleção de indivíduos considera-se apenas efeitos de repetição, genotípicos individuais e famílias. Uma maneira simples de incorporar informações do solo é a utilização das características da fertilidade do solo como covariáveis no modelo de delineamento experimental. Dessa forma, baseando-se apenas no delineamento, pode-se eliminar os efeitos do solo no ranqueamento das famílias e obter uma a melhor classificação.

1.2. Objetivos

O objetivo geral do presente trabalho foi avaliar e comparar metodologias de ranqueamento de famílias de cana-de-açúcar nas etapas iniciais do programa de melhoramento genético via Geoestatística e Aerofotogrametria. Mais especificamente, primeiramente comprovar que os efeitos de solo devem ser considerados ao ranquear as famílias de cana-de-açúcar de acordo com a quantidade de toneladas de colmos por hectare (TCH), e posteriormente, verificar a possibilidade de utilização de índices provenientes de imagens no padrão RGB como alternativa à análise de fertilidade do solo tradicional.

1.3. Estrutura de tese

O tópico 2 desta tese apresenta uma revisão da literatura em Geoestatística, fertilidade do solo, imagens RGB e análise de covariância. No tópico 3 é apresentado um método alternativo para ranqueamento de famílias de cana-de-açúcar incorporando a influência de variáveis de solo. Além disso, é proposta uma instigação para indicação do uso de imagens RGB como alternativa as variáveis do solo. O tópico 4 apresenta os principais resultados obtidos e discussão. No tópico 5 apresenta-se as conclusões desse estudo, bem como propostas para estudos futuros.

2. REVISÃO DE LITERATURA

2.1. Geoestatística

O conhecimento espacial de características de dados georreferenciados é hoje uma das grandes questões que envolvem diversas áreas do conhecimento, desde a agrícola até a área de humanas (Druck et al., 2004). A principal ferramenta utilizada para obter esse conhecimento é a Geoestatística, que é o ramo da estatística que associa o conceito de variáveis aleatórias com o conceito de variáveis regionalizadas, gerando o conceito de funções aleatórias. Diferente da estatística clássica, a Geoestatística tem como pressuposto a dependência entre vizinhos, ou seja, observações próximas exercem influências entre si.

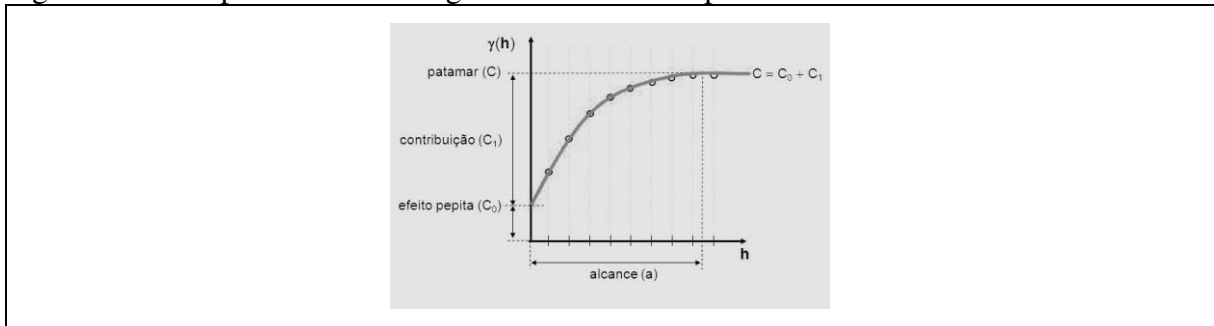
A Geoestatística pode ser entendida como a seção da estatística destinada a relatar o comportamento espacial dos dados, refletindo sobre as variáveis em uma perspectiva local. Essa metodologia busca discriminar a relação espacial de uma variável, provendo ferramentas estatísticas que possibilitem uma estruturação da aleatoriedade da variável por intermédio de uma função espacialmente correlacionada (Yamamoto e Landim, 2013). Idealizada por Daniel Gerhardus Krige (Krige, 1951) ao estudar dados de mineração de ouro e observar uma estrutura dependência dos dados com a localização dos pontos amostrais, essa metodologia foi posteriormente formalizada matematicamente pelo francês Matheron (1963), concedendo as noções de Krige uma generalização e extensão a teoria de variáveis regionalizadas.

Atualmente os procedimentos geoestatísticos são amplamente aplicados em diversas áreas do conhecimento, Bertolin et al. (2017) define agricultura de precisão como o conjunto de ferramentas e tecnologias direcionadas para a gestão agrícola no contexto espacial e temporal da variabilidade, visando caminhos alternativos de manejo de acordo com as especificidades da área em estudo. Dentre as vantagens do uso da agricultura de precisão, Silva et al. (2013) destaca o uso da ferramenta como método de seleção de cultivares considerando a adaptabilidade às condições ambientais, o manejo de nutrientes de forma lógica e monitoramento ervas daninhas e pragas. Mais especificamente nos estudos de solos, Trangmat et al. (1985) retrata de forma detalhada a Geoestatística e aplicações visando análise de variabilidade e classificação, Carvalho et al. (2003), Souza et al. (2004), Siqueira et al. (2008), Montanari et al. (2010), Salvador et al. (2012), Rodrigues et al. (2017), Dalchiavon et al. (2017), Noetzold et al. (2019) apresentam estudos referentes a variabilidade do solo na agricultura em diferentes culturas. O passo inicial para estudar e entender o comportamento espacial de variáveis é a modelagem da dependência espacial, feita através dos semivariogramas.

2.1.1. Semivariograma

O semivariograma ou variograma é uma ferramenta geoestatística que permite visualização o comportamento da dependência espacial em uma determinada área (Santos et al, 2011). O semivariograma empírico ou experimental é um gráfico bidimensional, onde o eixo x expressa a distância de separação entre pontos amostrais e o eixo y a estimativa para semivariância. A partir desses pontos é então ajustada uma curva de modelo teórico (semivariograma teórico) e definidos os valores dos parâmetros do semivariograma (efeito pepita, alcance, patamar e contribuição), conforme Figura 1 (Ferreira, 2015).

Figura 1 – Exemplos de semivariogramas teórico e empírico.



Fonte: Ferreira (2015).

Segundo Ferreira (2015), o efeito pepita (C_0) representa a descontinuidade na modelagem para pequenas distâncias, que teoricamente não deveria existir, e é proveniente de fatores não controláveis no processo de coletas de dados e/ou análises. O alcance (a) é o limiar entre dependência e aleatoriedade no contexto espacial, isto é, o valor limite para o qual a dependência de atributos em função das distâncias de separação das amostras pode ser modelada utilizando uma função matemática. O Patamar (C) é o valor correspondente da semivariância para uma distância de separação igual ao alcance. Já a contribuição (C_1) é definida pela diferença entre patamar e o efeito Pepita ($C_1 = C - C_0$).

As estimativas de semivariância para o semivariograma empírico são geralmente computadas pela fórmula proposta por (Matheron, 1963):

$$\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{i|h_{ij}=h} (Z_i - Z_{i+h})^2 \quad (1)$$

Onde:

$\hat{\gamma}(h)$ é o estimador de semivariância para uma distância h

$N(h)$ é o número de pares de pontos que estão separados por uma distância h

Z_i é o valor da variável no ponto amostrado i

Z_{i+h} é o valor da variável no ponto amostrado $i + h$

A escolha do melhor modelo teórico está intimamente ligada as menores distâncias de separação entre amostras. A literatura apresenta diversos modelos teóricos, sendo os modelos esférico, exponencial e gaussiano mais comumente utilizados. Caso os fenômenos em estudo apresente maior continuidade para pequenas distâncias de separação, o modelo gaussiano é mais indicado, caso contrário deve-se utilizar o modelo exponencial. Situações intermediárias são satisfatoriamente modeladas pelo modelo esférico.

As equações abaixo apresentam os modelos exponencial, esféricos e gaussiano, respectivamente:

$$\gamma(h) = \begin{cases} 0, & \text{se } h = 0 \\ c_0 + c_1 \left[1 - \exp \left(-\frac{3h}{a} \right) \right], & \text{se } h \neq 0 \end{cases} \quad (2)$$

$$\gamma(h) = \begin{cases} 0, & \text{se } h = 0 \\ c_0 + c_1 \left[1,5 \frac{h}{a} - 0,5 \left(\frac{h}{a} \right)^3 \right], & \text{se } 0 < h < a \\ c_0 + c_1, & \text{se } h \geq a \end{cases} \quad (3)$$

$$\gamma(h) = \begin{cases} 0, & \text{se } h = 0 \\ c_0 + c_1 \left[1 - \exp \left(-\frac{3h^2}{a^2} \right) \right], & \text{se } h \neq 0 \end{cases} \quad (4)$$

Os métodos utilizados para o ajuste de semivariogramas podem ser de dois tipos: feito via semivariograma experimental: métodos de quadrados mínimos ordinários (Ordinary Least Squares - OLS) e ponderados (*Weight Least Squares* - WLS); realizado diretamente a partir dos dados: métodos da máxima verossimilhança (*Maximum Likelihood* - ML) e máxima verossimilhança restrita (*Restricted Maximum Likelihood* - REML). Nos métodos OLS e WLS o ajuste das curvas de semivariograma é realizado ao estimar os parâmetros de covariância e minimizando a expressão de soma de quadrados da diferença entre os valores observados e os estimados. As metodologias ML e REML são técnicas de estimação onde a partir de uma amostra é obtido o estimador mais verossímil dos parâmetros de um certo modelo probabilístico.

Após a determinação de todos parâmetros do modelo, juntamente com os pontos amostrados, pode-se então obter estimativas para qualquer ponto na área de estudo de acordo com sua localização.

2.1.2. Krigagem

Para obtenção de mapas que representem o comportamento de uma determinada variável, após a modelagem estrutural de dependência espacial, deve-se realizar um processo de interpolação a partir dos valores amostrados. No contexto geoestatístico, os métodos de interpolação mais aplicados são: inverso da distância, triangulação e krigagem.

A krigagem é um método de interpolação que leva este nome em homenagem a Daniel G. Krige, cientista sul-africano que foi pioneiro em estudos georreferenciados (Ferreira, 2015). Segundo Isaaks & Srivastava (1989) a krigagem é o melhor interpolador linear geoestatístico, pois produz estimativas com a menor variância do erro de estimação. Uma grande vantagem da krigagem frente aos outros interpoladores é a possibilidade de obtenção da variância de krigagem ou interpolação, uma medida de erro que é calculada para cada valor estimado, consequentemente há uma medida de confiabilidade associada ao método (Soares, 2006).

Dentre as variantes dos estimadores de krigagem podemos citar: simples (pontual ou em bloco), ordinária, universal, cokrigagem (Oliver e Webster, 1990). A principal diferença entre a krigagem e outros métodos de interpolação está na maneira como os pesos são atribuídos a diferentes amostras, sendo eles determinados por uma análise espacial baseada no semivariograma experimental. As krigagens simples e ordinária são processos estacionários, sendo que na primeira considera a média local como um constante igual a média populacional. Já as krigagens universal e cokrigagem são processos não-estacionários. A krigagem simples é mais precisa que a universal (Santos et al, 2011).

A krigagem, de uma forma geral, pode ser entendida como parte da família de algoritmos de regressões de mínimos quadrados generalizados (Goovaerts, 1997) e a estimação é baseada no estimador definido por:

$$\mathbf{Z}^*(u) - \mathbf{m}(u) = \sum_{\alpha=1}^{n(u)} \lambda_{\alpha}(u) [\mathbf{Z}(u_{\alpha}) - \mathbf{m}(u_{\alpha})] \quad (5)$$

Onde $n(u)$ é o número de pares de pontos amostrais, $\lambda_{\alpha}(\mathbf{u})$ são os pesos definidos para os dados $z(\mathbf{u}_{\alpha})$ quando considerados como realizações da variável aleatória $Z(\mathbf{u}_{\alpha})$. Os valores esperados para as variáveis aleatórias $Z(\mathbf{u}_{\alpha})$ e $Z(\mathbf{u})$ são iguais a $m(\mathbf{u}_{\alpha})$ e $m(\mathbf{u})$ respectivamente. Com essa perspectiva é possível, então, obter erros de estimação para a diferença das variáveis $Z^*(\mathbf{u}) - Z(\mathbf{u})$.

O processo de estimação da krigagem parte da diferença $Z^*(\mathbf{u}) - Z(\mathbf{u})$, no qual minimiza-se a estimativa da variância do erro de $\mathbf{u}$ que é dado por:

$$\sigma_E^2(u) = \text{Var}\{Z^*(u) - Z(u)\} \quad (6)$$

Note que cada estimador de krigagem diferente (simples, ordinária, universal, cokrigagem) implica em um modelo diferente. Uma solução comum então é acrescentar um componente ao componente de tendência para representar $Z(\mathbf{u})$.

$$\mathbf{Z}(\mathbf{u}) = \mathbf{R}(\mathbf{u}) + \boldsymbol{\mu} \quad (7)$$

Esse componente $\mathbf{R}(\mathbf{u})$ é modelado por meio de uma função que apresenta estacionaridade, média nula e de covariância (ou semivariância) $C(\mathbf{h})$. Dessa maneira, temos:

$$\mathbf{E}[\mathbf{R}(\mathbf{u})] = \mathbf{0} \quad (8)$$

$$\mathbf{COV}(\mathbf{R}(\mathbf{u}), \mathbf{R}(\mathbf{u} + \mathbf{h})) = \mathbf{E}(\mathbf{R}(\mathbf{u}).\mathbf{R}(\mathbf{u} + \mathbf{h})) \quad (9)$$

Note que dessa forma a esperança de Z no local $\mathbf{u}$ corresponde ao valor médio de $\mathbf{u}$ analisado de forma local.

Conforme Landim (2002), para uma variável e um ponto de interesse, é possível obter uma estimativa dessa variável na sua respectiva localização:

$$\mathbf{Z}_{x_0}^* = \sum_{i=1}^n \boldsymbol{\mu} + \lambda_i \mathbf{Z}_{x_i} \quad (10)$$

Onde $\boldsymbol{\mu}$ é média de atributo, $\mathbf{Z}_{x_0}^*$ é o valor estimado do atributo na localização x_0 , λ_i é o peso da amostra na localização x_i , i o indexador da amostra x_i de localização conhecida.

O estimador linear na krigagem simples, denotado por $\mathbf{Z}_{SK}^*(\mathbf{u})$ é dado por:

$$\mathbf{Z}_{SK}^*(\mathbf{u}) = \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_{\alpha}^{SK}(\mathbf{u}) \mathbf{Z}(\mathbf{u}_{\alpha}) + \left[\mathbf{1} - \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_{\alpha}^{SK}(\mathbf{u}) \right] \mathbf{m} \quad (11)$$

Onde $\lambda_{\alpha}^{SK}(\mathbf{u})$ são determinados de tal maneira que a variância do erro seja mínima $\sigma_E^2(\mathbf{u}) = \text{Var}\{\mathbf{Z}_{SK}^*(\mathbf{u}) - \mathbf{Z}(\mathbf{u})\}$ mediante a condição de não tendenciosidade.

A variância de krigagem é dada pela equação:

$$\sigma_{KS}^2 = C(\mathbf{0}) - \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_{\alpha}^{SK}(\mathbf{u}) C(\mathbf{u} - \mathbf{u}_{\alpha}) \geq 0 \quad (12)$$

Maiores informações sobre os processos de krigagem podem ser encontradas em Santos (2010).

Após a modelagem da dependência espacial e interpolação é necessário validar o modelo geoestatístico (Vieira, 2010). A validação cruzada é uma ferramenta que possibilita tal análise. Essa técnica permite mensurar a incerteza da medição prévia dos dados (Almeida et al. 2007), tem por objetivo prever a acurácia do modelo para amostras futuras. Para isso os dados amostrados são retirados individualmente e estimados pelo modelo utilizando os demais pontos amostrais, conseqüentemente é possível comparar os valores amostrados com os estimados pelo modelo e obter uma estatística de erro de estimação (Isaaks e Srivastava, 1989).

Dentre as estatísticas de erro obtidas podemos citar: média do erro de estimação, média do erro de estimação padronizada, raiz quadrada do erro quadrático médio, média do desvio padrão e raiz quadrada da média do erro de estimação padronizado ao quadrado (Ferreira, 2015). De acordo com Vieira (2000) na autovalidação devemos ter: média dos erros e dos erros padronizados nula; variância dos erros finita; variância dos erros padronizados unitária.

2.2. Aerofotogrametria

Na literatura podemos encontrar diversas nomenclaturas para identificar as aeronaves não tripuladas controladas remotamente, como *drones* (zangão), Veículos Aéreos Não Tripulados (VANTS), *Unmanned Aerial Vehicle* (UAV), Veículo Aéreo Remotamente Pilotado (VARP) ou Sistema De Veículo Aéreo Não Tripulado (SISVANT). Segundo Eisenbeiss (2009) VANT pode ser entendido como um veículo motorizado conduzido por um piloto por controle remoto, podendo voar de forma autônoma, semiautônoma ou manual. Luz e Antunes (2015) definem um SISVANT como um sistema composto por VANT, equipamentos de localização espacial (receptores GNSS) e sensores fotográficos digitais, que tem por finalidade obter informações espaciais.

No contexto agrícola tem-se observado um grande entusiasmo pelo monitoramento e desenvolvimento das lavouras com o uso da robótica como instrumentos de maximização da produção, sendo os VANTS uma das ferramentas mais promissoras, auxiliando nas mais diversas atividades e tomadas de decisão. Essas aeronaves apresentam como principais vantagem as facilidades de logística e baixo custo operacional, sendo as com sensores *Red, Green and Blue* (RGB) os mais acessíveis segundo o quesito custo (Andrade et al., 2019). Com a utilização de imagens digitais no padrão RGB é possível a análise de plantas em um processo não destrutivo (Barbosa et al., 2016), fenotipagem de progênies (Rosas et al., 2019), estudo da degradação de biomas (Teixeira et al., 2018) e monitoramento dos estágios de desenvolvimento de culturas (Andrade et al., 2019). Araus et al. (2018) apresenta diversas ferramentas de sensoriamento remoto que tem potencial para avaliação fenotípica de culturas e destaca a exploração das imagens RGB devido ao seu baixo custo. Gebremedhin et al. (2019) diz que de posse das imagens é possível calcular índices e avaliar diversas características de maneira rápida, assertiva e com baixos custo e mão-de-obra. No que diz respeito às características do solo Gadú e Demattê (2011) destacam que a composição físico-química influi sobre o padrão espectral do solo. Freire-Silva et al. (2019) verificou que nas revisões de literatura acerca de

índices de vegetação nenhuma utilizou apenas as bandas do visível RGB, especialmente no Brasil, o que justifica a investigação de metodologias que envolvam tais comprimentos de onda.

Dentre os diversos índices calculados a partir imagens RGB encontrados na literatura podemos citar VARI (*Visible Atmospherically Resistant Index*) (Gitelson et al. 2002), GLI (*Green Leaf Index*) (Louhaichi et al., 2013), NGRDI (*Normalized Green Red Difference Index*) (Hunt r. et al., 2013). Outros índices podem ser encontrados em Júnior et al. (2008), Rosas et al. (2019) e Freire-Silva et al. (2019).

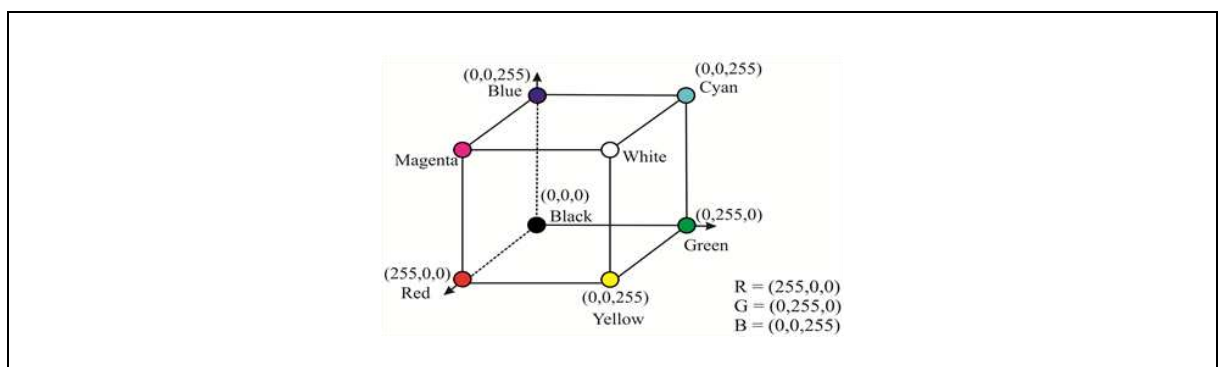
Matematicamente o modo mais fácil de representar as cores é valendo-se de modelos algébricos do espectro, nos quais cada cor é padronizada associando-se um valor numérico ou combinações de cores básicas. O universo de todas as cores expressas por um modelo é chamado de espaço ou *gamut* de cores (Park e Lee, 2011).

Existem diversos modelos matemáticos ou espaços de cores. O padrão RGB, que é bastante utilizado em eletrônicos como celulares, tablets, telas em geral, retroprojetores e câmeras fotográficas, é um espaço tridimensional onde qualquer cor é obtida pela soma de três cores primárias, a saber: vermelho (*red*), verde (*green*) e azul (*blue*). No padrão de cores RGB cada banda possui 256 cores (de 0 a 255).

Em uma imagem o valor atribuído a um pixel pode ser entendido como uma combinação linear de três coordenadas de um vetor, onde cada coordenada corresponde a intensidade de uma das bandas RGB, portanto a cor característica do pixel é o resultado da soma de três cores.

Podemos representar essa soma por meio de uma soma de funções no plano cartesiano (Figura 2), onde cada função individualmente retorna o valor correspondente a cada uma das bandas RGB.

Figura 2 – Representação do espaço tridimensional RGB no plano cartesiano

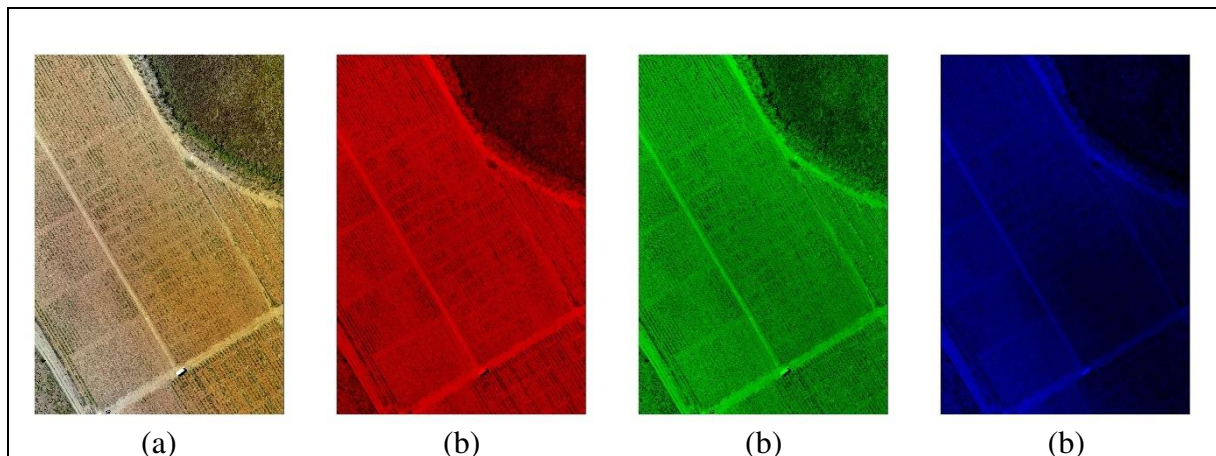


Fonte: Kumah et al. (2019)

Dessa forma, a cor característica de uma imagem é definida pelas bandas que compõem cada um de seus pixels individualmente (Figura 3), assim, em cada pixel de coordenada (x, y) o valor característico da imagem ($RGB(x, y)$) é função das bandas *red* ($R(x, y)$), *green* ($G(x, y)$) e *blue* ($B(x, y)$), expresso pela formula abaixo:

$$RGB(x, y) = R(x, y) + G(x, y) + B(x, y) \quad (1)$$

Figura 3 – Exemplo de imagens (a) Padrão RGB. (b) Falsa banda *Red*. (c) Falsa banda *Green*. (d) Falsa banda *Blue*.



Fonte: Preparado pelo autor

2.3. Fertilidade do solo

A fertilidade do solo pode ser compreendida como a eficácia do solo em fornecer nutrientes básicos à planta, acumulando e fornecendo os mesmos para a planta no período de seu desenvolvimento, mesmo que a planta não utilize ou absorva esses nutrientes (Raij, 1981). A variabilidade dos atributos presentes na composição do solo decorre da interação entre processo na formação, práticas de manejo e de cultura, sendo sua distribuição espacial de extrema importância. Ao se fazer afirmações acerca das propriedades do solo deve-se levar em consideração a variação do atributo em questão, uma vez que quanto maior a heterogeneidade do solo menos confiável se tornam as afirmativas, além disso, a escala com a qual se apresentam os dados e a metodologia para sua obtenção também pode ser um fator de influência.

No território brasileiro o solo geralmente apresenta menor fertilidade, sendo necessária à sua correção com o intuito de diminuir a acidez, ajustar o pH e regular as concentrações de nutrientes ligados ao desenvolvimento das plantas (Oliveira et al., 2005). Dentre esses nutrientes presentes no solo, alguns são altamente consumidos pelas plantas (macronutrientes),

enquanto que outros são menos exigidos (micronutrientes). Macro e micronutrientes são todos atributos importantes à planta em algum estágio do desenvolvimento.

Variáveis como fósforo (P), potássio (K), cálcio (Ca), enxofre (S), nitrogênio (N) e magnésio (Mg) estão no grupo dos macronutrientes. Cobre (Cu), zinco (Zn), ferro (Fe), boro (B) manganês (Mn), cloro (Cl) e molibdênio (Mo) são micronutrientes. Outros elementos presente no solo como o alumínio (Al) são prejudiciais ao desenvolvimento. Brito (2014) destaca que mesmo sendo tóxico e causando problemas para a produção agrícola, ainda há muito o que se discutir sobre o real efeito do alumínio tóxico do solo na cana-de-açúcar. Na forma que é encontrado no solo, sexidróxido de alumínio (um átomo de Al cercado por seis hidroxilas), é altamente radioativo e ao se hidrolisar no processo de calagem do solo libera íons de hidrogênio e faz com que o pH do solo se torne mais ácido. Dessa forma, o potencial que o alumínio tem de acidificar o solo no processo de hidrólise é chamado de acidez potencial e é definido pela soma $H + Al$.

O pH exerce grande influência no desenvolvimento das plantas, pois está associado à concentração, solubilidade, disponibilidade e capacidade de retenção de nutrientes essenciais, além de indicar possíveis problemas com a fertilidade do solo ou mesmo com toxicidade devido a relação do pH com o desenvolvimento de culturas (Lima et al. 2016). Quimicamente o potencial hidrogeniônico (pH) está relacionado a atividade iônica e intimamente ligado a todas as reações que ocorrem no solo. A soma de bases trocáveis (SB) exprime a soma de cátions permutáveis, ao passo que capacidade de troca catiônica (t) traduz a habilidade do solo em reter cátions. Ronquim (2010) destaca que a capacidade de troca iônica exprime a mensuração de liberação dos nutrientes presentes no solo, mantendo-o fértil e regulando a toxicidade proveniente do uso de fertilizantes. Quando capacidade de troca catiônica está relacionada a cátions de cálcio, magnésio e potássio o solo favorece a planta nutricionalmente, no entanto cátions tóxicos de hidrogênio e alumínio tornam o solo pobre. O índice de saturação por base (v) aponta a riqueza do solo e o índice de saturação de alumínio (m) indica a proporção de alumínio solúvel em relação aos teores de alumínio e de bases trocáveis.

Segundo Vitti et al. (2005) a quantidade disponível de nutrientes pode ser visualizada facilmente nas plantas de cana-de-açúcar. Sintomas de carência de fósforo, cálcio, potássio e magnésio podem ser observados nas folhas, com colorações e manchas específicas, podendo progredir até o estado de necrose. A carência de potássio diminui o teor de açúcar na cana, pouco fósforo implica em baixo crescimento e número de perfilhos. O fósforo tem grande importância no armazenamento e na transferência de energia, atua como componente dos ácidos nucleicos, fitina e fosfolipídios, sendo de extrema importância no metabolismo da planta

(Epstein e Bloom, 2006). Além disso, amplia a eficiência de uso da água pela planta, atua na utilização e assimilação de outros nutrientes, tornando a planta mais resistente a doenças e deficiência hídrica (Penatti, 2013).

Cambardella et al. (1994) resalta a importância do entendimento da variabilidade espacial das características do solo a fim de obter uma maior qualidade nas práticas de manejo e avaliação dos efeitos da agricultura sobre a qualidade do meio. Mangabeira et al (2005) mostram que solos com deficiência de fertilidade nutricional estão associados a áreas com baixa produtividade.

2.4. Análise de covariância

Em uma análise estatística muitas vezes as unidades experimentais podem conter informações de várias variáveis, ainda que somente uma delas é o foco de estudo, no caso, a variável dependente. Ainda mais, muitas vezes essas variáveis podem exercer efeito na variável resposta e, dessa forma, sua inclusão no modelo analisado pode trazer benefícios como aumento de precisão, redução da variabilidade do erro ou diminuição do efeito de fatores não controlados. A inclusão dessas covariáveis provocam um efeito que pode ser entendido como uma junção da ANOVA tradicional e de regressão. O papel dessas covariáveis na análise neste caso é corrigir os valores da variável dependente posteriormente a comparação de médias (Recnher e Schaalje, 2008).

De maneira geral a análise de covariância pode ser escrita como:

$$y = Z\alpha + X\beta + \varepsilon \quad (1)$$

Onde Z é uma matriz de 0 e 1, α contém a média μ e os parâmetros de efeito de tratamento, X contém os valores das covariáveis e β contém os coeficientes das covariáveis

Para um modelo balanceado, com um fator e k covariáveis, o modelo é dado por:

$$y_{ij} = \mu + \alpha_i + \beta_1 x_{ij1} + \beta_2 x_{ij2} + \dots + \beta_k x_{ijk} + \varepsilon_{ij}, 1 \leq i \leq k \text{ e } 1 \leq j \leq n \quad (2)$$

Onde α_i é o efeito de tratamento, x_{ij} é covariável observada na unidade experimental de y_{ij} , β é o coeficiente relacionando x_{ij} com y_{ij} .

Matricialmente, para as kn observações $y = Z\alpha + X\beta + \varepsilon$, é escrito como:

$$Z = \begin{pmatrix} 1 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 0 & \dots & 0 \\ 1 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & \dots & 1 \end{pmatrix}, \alpha = \begin{pmatrix} \mu \\ \alpha_1 \\ \vdots \\ \alpha_k \end{pmatrix}, X = \begin{pmatrix} x_{111} & x_{112} & \dots & x_{11p} \\ x_{121} & x_{122} & \dots & x_{12p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{ij1} & x_{ij2} & \dots & x_{ijp} \end{pmatrix}, \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{pmatrix} \quad (3)$$

A estimação de α e β no caso geral $y = Z\alpha + X\beta + \varepsilon$, assumindo $E(\varepsilon) = \mathbf{0}$ e $cov(\varepsilon) = \sigma^2 I$ pode ser obtido por:

$$\begin{aligned} y &= Z\alpha + X\beta + \varepsilon \\ &= (Z, X) \begin{pmatrix} \alpha \\ \beta \end{pmatrix} + \varepsilon \\ &= U\theta + \varepsilon \end{aligned} \quad (4)$$

Onde $U = (Z, X)$ e $\theta = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$.

A equação normal para (4) é dada por:

$$U'U\hat{\theta} = U'y \quad (5)$$

Que pode ser reescrita como:

$$\begin{aligned} \begin{pmatrix} Z' \\ X' \end{pmatrix} (Z, X) \begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix} &= \begin{pmatrix} Z' \\ X' \end{pmatrix} y \\ \begin{pmatrix} Z'Z & Z'X \\ X'Z & X'X \end{pmatrix} \begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix} &= \begin{pmatrix} Z'y \\ X'y \end{pmatrix} \end{aligned} \quad (6)$$

Da expressão (6) podemos escrever duas equações em função de $\hat{\alpha}$ e $\hat{\beta}$

$$Z'Z\hat{\alpha} + Z'X\hat{\beta} = Z'y \quad (7)$$

$$X'Z\hat{\alpha} + X'X\hat{\beta} = X'y \quad (8)$$

Utilizando a inversa generalizada de $Z'Z$, podemos obter o valor de $\hat{\alpha}$ em (7):

$$\begin{aligned} \hat{\alpha} &= (Z'Z)^{-}Z'y - (Z'Z)^{-}Z'X\hat{\beta} \\ &= \hat{\alpha}_0 - (Z'Z)^{-}Z'X\hat{\beta} \end{aligned} \quad (9)$$

Onde $\hat{\alpha}_0 = (Z'Z)^{-}Z'y$ é a solução da equação normal do modelo $y = Z\alpha + \varepsilon$ sem as covariáveis.

Para obter $\hat{\beta}$ basta substituir (9) em (8). Daí

$$\begin{aligned} X'Z[(Z'Z)^{-}Z'y - (Z'Z)^{-}Z'X\hat{\beta}] + X'X\hat{\beta} &= X'y \\ X'Z(Z'Z)^{-}Z'y + X'[I - Z(Z'Z)^{-}Z']X\hat{\beta} &= X'y \end{aligned} \quad (10)$$

Definindo

$$P = Z(Z'Z)^{-}Z' \quad (11)$$

Substituindo em (10) temos

$$X'(I - P)X\hat{\beta} = X'y - X'Py = X'(I - P)y$$

Assumindo as colunas de X como linearmente independente das colunas de Z , então

$X'(I - P)X$ é não singular e a solução de $\hat{\beta}$ é dada por:

$$\hat{\beta} = [X'(I - P)X]^{-1}X'(I - P)y \quad (12)$$

A equação 12 pode ser reescrita como:

$$\hat{\beta} = E_{xx}^{-1}e_{xy} \quad (13)$$

Onde

$$E_{xx} = X'(I - P)X \text{ e } e_{xy} = X'(I - P)y \quad (14)$$

Para o modelo de análise de covariância dado em (1) ou (4), denotando SSE como $SSE_{y.x}$, podemos escrever

$$\begin{aligned}
 SSE_{y.x} &= y'y - \hat{\theta}'Uy = y'y - (\hat{\alpha}', \hat{\beta}') \begin{pmatrix} Z'y \\ X'y \end{pmatrix} \\
 &= y'y - \hat{\alpha}'Z'y - \hat{\beta}'X'y \\
 &= y'y - [\hat{\alpha}'_0 - \hat{\beta}'X'Z(Z'Z)^{-1}]Z'y - \hat{\beta}'X'y \text{ (por (9))} \\
 &= y'y - \hat{\alpha}'_0Z'y - \hat{\beta}'X'[I - Z(Z'Z)^{-1}Z']y \\
 &= SSE_y - \hat{\beta}'X'(I - P)y
 \end{aligned} \tag{15}$$

Na equação acima o termo SSE_y é obtido na ANOVA do modelo sem covariáveis e é igual a $y'y - \hat{\alpha}'_0Z'y$, utilizando a equação 14 ainda pode-se reescrever a equação 15 como $SSE_{y.x} = e_{yy} - e'_{yy}E_{xx}^{-1}e_{xy}$, onde $e_{yy} = SSE_y = y'(I - P)y$.

2.5. Método Stepwise

Quando se trabalha com um conjunto de dados que envolve um grande número de variáveis, um questionamento imediato é quais dessas variáveis realmente influenciam na variável resposta? Em resposta a esse questionamento temos os métodos de seleção de variáveis, que tem como finalidade obter quais variáveis são mais importantes e consequentemente, as demais podem ser descartadas. Nos problemas que envolvem seleção de variáveis no contexto de regressões lineares múltiplas, o método *Stepwise* é o mais amplamente difundido (MONTGOMERY & RUNGER, 2009). O método pode ser entendido como uma metodologia linear de seleção de variáveis, é amplamente utilizado nas ciências da computação e demais áreas que trabalham com modelos que possuem grande número de variáveis ((SCHUSTER e CRUZ, 2013). Segundo James et al. (2013) o método é utilizado tanto na solução de problemas computacionais com várias variáveis, bem como na redução do modelo de regressão, uma vez que seleciona as variáveis mais influentes no conjunto de saída. Esse método é bastante utilizado quando se deseja obter informações sobre a relação entre variáveis e não necessariamente se tem como objetivo explicar essa relação.

No método *stepwise* pode ocorrer a inclusão ou exclusão de variáveis, passos *forward*, *backward* respectivamente. A inclusão ou remoção de variáveis no modelo deve ser baseada em um critério estatístico (TABACHNICK; FIDEL, 2007). Dentre os critérios estatísticos mais utilizados podemos citar o teste F, coeficiente correlação linear, erro quadrático, critério de informação de Akaike (HOCKING, 1976). O critério de informação de Akaike (AIC) (Akaike, 1974) é definido por:

$$AIC = -2 \ln(\hat{\theta}) + 2p \quad (1)$$

O termo $\ln(\hat{\theta})$ é a log-verossimilhança maximizada do modelo, calculada com base nos estimadores de máxima verossimilhança dos parâmetros, e p é a quantidade de parâmetros do modelo. O melhor ajuste é aquele que apresenta menores valores de AIC.

2.6. Índice Kappa

O índice Kappa é a principal técnica de validação para mapas categóricos, proposto por Cohen (1960), pode ser definido como uma medida de avaliação do grau de concordância de classificação. O índice é calculado a partir de uma matriz de confusão que resume a concordância na categorização entre um mapa de referência e o mapa temático analisado (MARINHO et al., 2012). Landis e Koch (1977) propuseram uma classificação da acurácia para os valores do índice Kappa.

Tabela 1 – Comparação de ranqueamento das melhores famílias quanto ao TCH, em função do percentual de seleção de melhores famílias.

Coeficiente Kappa	Classificação
< 0,00	Péssima
0,00-0,20	Ruim
0,21-0,40	Razoável
0,41-0,60	Boa
0,61-0,80	Muito Boa
0,81-1,00	Excelente

3. MATERIAIS E MÉTODOS

Visando comparar o método usual de seleção de famílias de cana-de-açúcar, doravante chamado de modelo TRADICIONAL e presente em Ferreira (2020), com a metodologia proposta e descrita abaixo, foram utilizadas informações acerca de famílias de irmãos completos de cana-de-açúcar, análise de fertilidade do solo e imagens obtidas via VANT da área de experimento.

Primeiramente foi feito um estudo espacial das variáveis de fertilidade do solo (potencial hidrogeniônico, fósforo, potássio, cálcio, magnésio, alumínio, acidez potencial, soma de bases trocáveis, capacidade de troca catiônica efetiva, capacidade de troca catiônica a pH 7,0, índice de saturação de bases e índice de saturação de alumínio). Para as variáveis que apresentaram padrão de dependência espacial foram obtidos mapas de interpolação por krigagem simples e posteriormente calculados valores médios dos atributos para as unidades experimentais ($n = 240$).

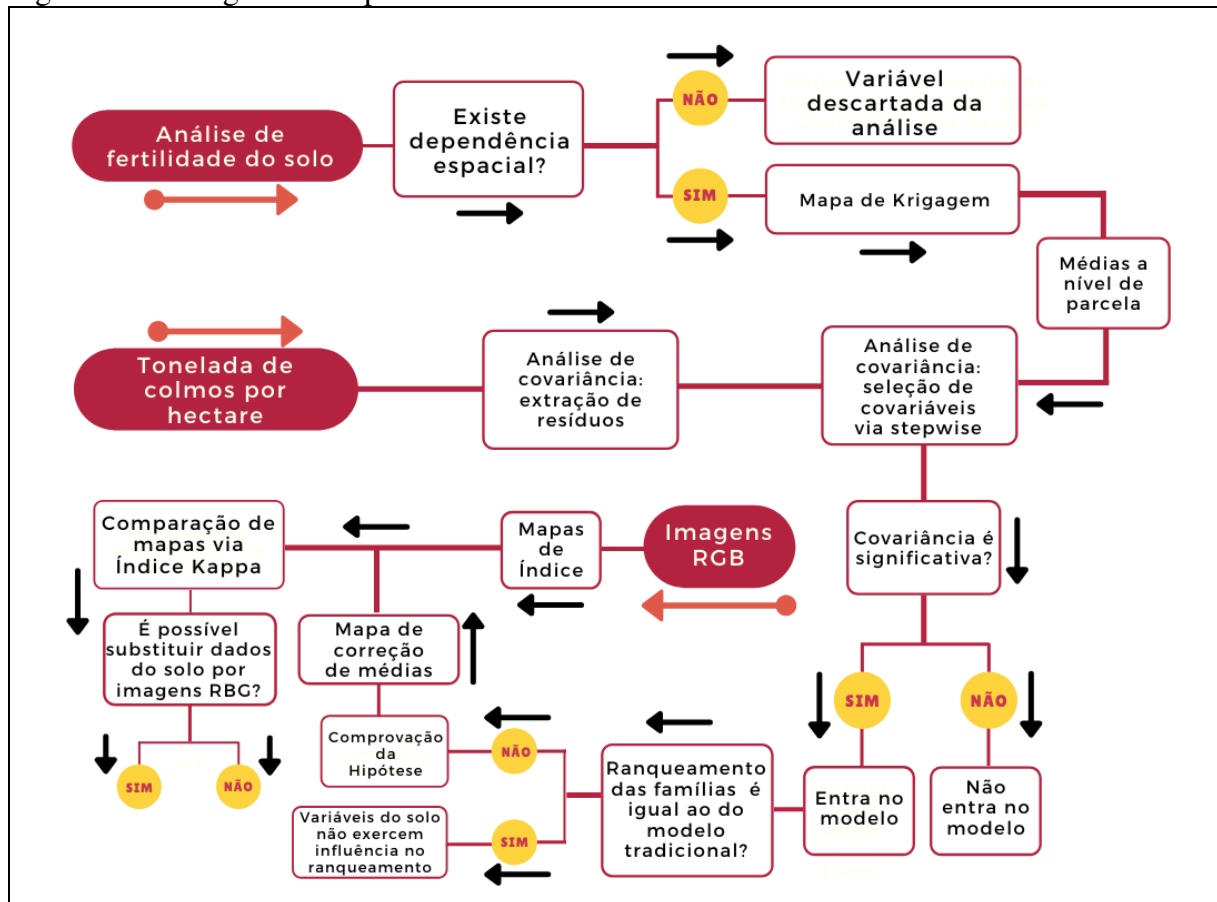
Utilizando a variável toneladas de colmos por hectare (TCH) como variável dependente foram extraídos os resíduos do modelo de delineamento em blocos casualizados (DBC).

Esses resíduos foram então utilizados como variável resposta em uma análise de covariância (ANCOVA), tomando as variáveis de solo como variáveis independentes. Com o modelo obtido na ANCOVA, as famílias de cana-de-açúcar foram ranqueadas, e dessa forma, comparou-se o ranqueamento do modelo tradicional com o modelo incorporando as covariáveis. Além disso, foi calculado o valor da correção de cada unidade experimental segundo o modelo de delineamento experimental supracitado, permitindo assim a confecção de um mapa de correção de TCH.

Posteriormente, a partir das imagens aéreas, foram calculados os índices VARI, GLI, NGRDI e obteve-se os respectivos mapas de valor médio, amplitude e desvio-padrão a nível de parcela para cada um dos índices. O índice Kappa foi utilizado para analisar concordância entre os mapas de correção da variável e os de estatísticas dos índices verificando, assim, a possibilidade de utilização dos índices como método alternativo à análise de fertilidade do solo.

O fluxograma apresentado a seguir (Figura 4) ilustra todo o processo descrito acima para ranqueamento das famílias e análise da indicação do uso de imagens RGB como alternativa as informações de amostra do solo.

Figura 4 – Fluxograma dos processos envolvidos nas análises

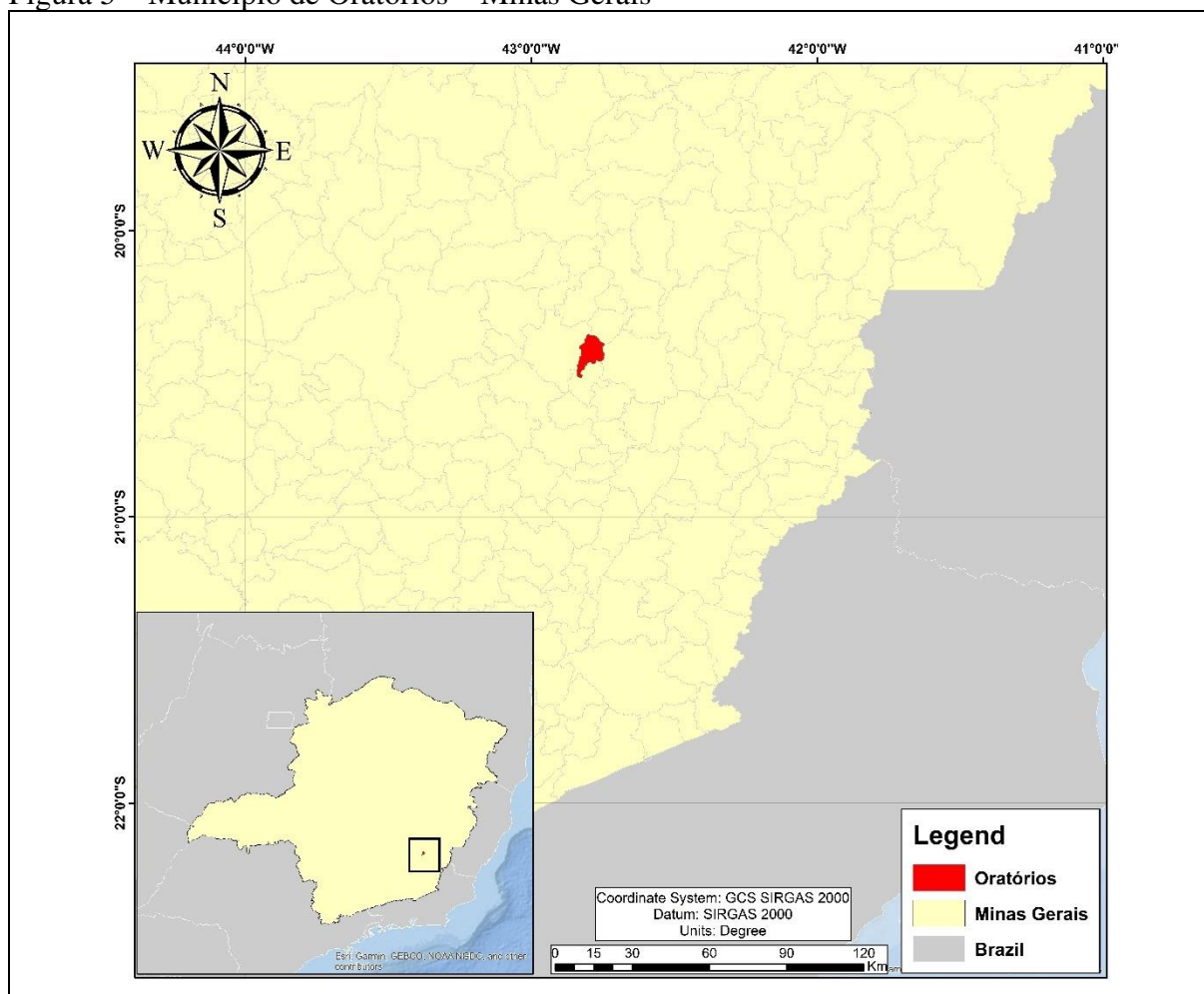


Fonte: Preparado pelo autor

3.1. Dados experimentais

O experimento foi conduzido no Centro de Pesquisa e Melhoramento da Cana-de-açúcar (CECA) da Universidade Federal de Viçosa (UFV), localizado no município de Oratórios, MG (latitude 20°25' S, longitude 42°48' W e 494 m de altitude).

Figura 5 – Município de Oratórios – Minas Gerais



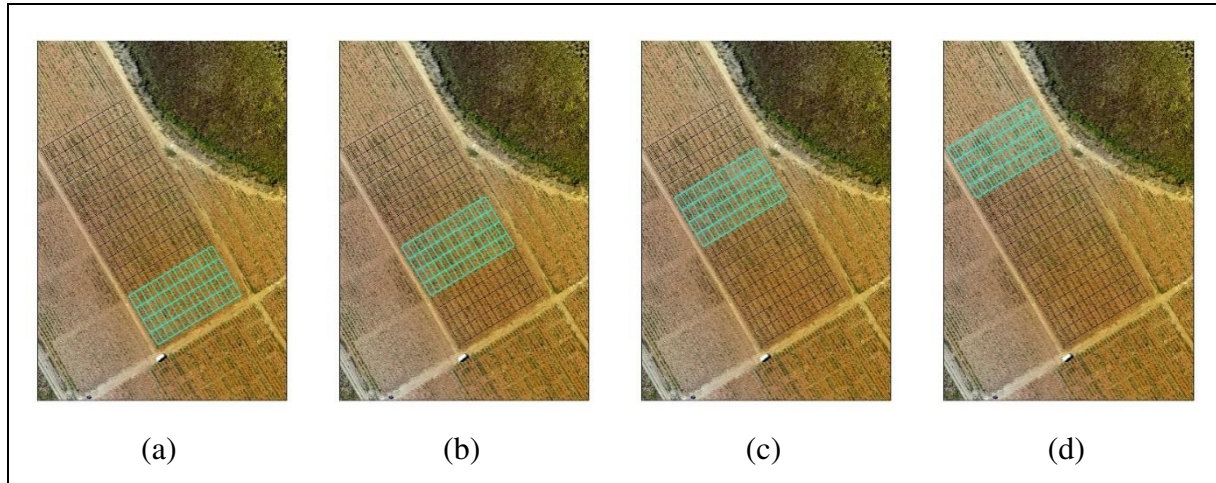
O delineamento experimental adotado foi o de blocos casualizados, contando com 4 blocos. As unidades experimentais foram compostas por dois sulcos com 5,0 m de comprimento, espaçados a 1,4 m, cada sulco contendo dez plântulas equidistantes a 0,5 m. O número de famílias de irmãos completos de cana-de-açúcar foi igual a 60, totalizando assim 240 parcelas. Dessa forma, as informações referentes a cada parcela foram obtidas da média das 20 plântulas. Para cada parcela após a colheita foi então extraído o valor médio da quantidade de toneladas de colmos por hectare (TCH), calculada por meio da equação:

$$TCH = (NCM \times MMC \times 10) / TP \quad (1)$$

onde NCM corresponde ao número de colmos por metro, MMC a massa média dos colmos, TP ao tamanho da parcela em m² e 10 é uma constante para conversão de massa em toneladas por hectare.

Os anexos A e B apresentam as tabelas com os valores de TCH para cada família nos blocos 1 e 2, e 3 e 4, respectivamente. Na Figura 6 temos a representação dos blocos com as respectivas parcelas.

Figura 6 – Blocos experimentais do experimento. (a) Bloco 1. (b) Bloco 2. (c) Bloco 3. (d) Bloco 4.



Para a análise de fertilidade do solo da área experimental foram coletadas 36 amostras de solo na profundidade 0-20 cm no dia 21/08/2019, sendo essas amostras provenientes de uma grade amostral 4x9 (Figura 7). As variáveis analisadas foram: potencial hidrogeniônico (pH), fósforo (P), potássio (K), cálcio (Ca), magnésio (Mg), alumínio (Al), acidez potencial (H+Al), soma de bases trocáveis (SB), capacidade de troca catiônica efetiva (t), capacidade de troca catiônica a pH 7,0 (T), índice de saturação de bases (V) e índice de saturação de alumínio (m). A análise de solo foi realizada em um laboratório particular no município de Viçosa-MG. Para extração de P e K utilizou-se o extrator Mehlich 1, os atributos Ca, Mg e Al com o extrator KCL – 1 mol/L e H+Al foi extraído usando acetato de cálcio 0,5 mol/L a pH 7 conforme indica a literatura (Embrapa, 2011). Segundo Ronquim (2010) a capacidade de troca catiônica total ou capacidade de troca catiônica a pH 7,0 (T) é obtida quando se considera todos os cátions permutáveis do solo ($Ca^{2+}, Mg^{2+}, K^+, H^+, + Al^{3+}$). Quando o hidrogênio é retirado temos a capacidade de troca catiônica efetiva (t). A soma de bases trocáveis (SB) é definida como a soma dos cátions de cálcio (Ca), magnésio (Mg) e potássio (P). Os índices de saturação de bases (V) e índice de saturação de alumínio (m) são obtidos em função de SB, T, t e Al. As fórmulas para obtenção dessas variáveis são apresentadas abaixo:

$$SB = Ca^{2+} + Mg^{2+} + K^+ \quad (2)$$

$$t = Ca^{2+} + Mg^{2+} + K^+ + Al^{3+} \quad (3)$$

$$T = Ca^{2+} + Mg^{2+} + K^+ + H^+ + Al^{3+} \quad (4)$$

$$V(\%) = \frac{SB}{T} \times 100 \quad (5)$$

$$m(\%) = \frac{Al}{t} \times 100 \quad (6)$$

Figura 7 – (a) Grade amostral de pontos para análise de fertilidade do solo. (b) Retira de amostra para análise de fertilidade do solo



Imagens no padrão RGB ou ortomosaicos do experimento foram obtidas por um VANT do tipo RPA (*Remotely Piloted Aircraft*) de 4 rotores da marca DJI, modelo Phantom 4, equipado com uma câmera RGB modelo FC330 de resolução igual a 12 megapixel, distância focal 3.61 mm, pixel de $1.56 \times 1.56 \mu\text{m}$ e sensor CMOS. A aeronave possui capacidade de voo igual 27 minutos e de carga 1.368kg, apresenta bússola, rádio, alimentação via bateria, o receptor de rádio para comunicação com a base possui alcance de 7km. Durante a realização do voo o sistema de comunicação via rádio envia as informações dos componentes do SISVANT para uma estação de recepção no solo, na qual o operador pode controlar e monitorar o voo fotogramétrico, sendo as rotas executadas de forma manual ou automática.

Os voos para obtenção de imagens foram realizados nos dias 03/10/2018 e 15/03/2019 no período de 11:00 às 13:00, na primeira data ainda havia predominância de solo exposto na área experimental devido ao pouco desenvolvimento da cultura, já na data posterior a cultura estava totalmente desenvolvida e havia predominância de área verde. A câmera para coleta de imagens foi programada para obter imagens a cada dois segundos para esse estudo. O dispositivo fotográfico ficou fixo em uma plataforma estabilizadora de 3 eixos que permite rotação e estabilidade, uma vez o suporte permite a movimentação da mesma de forma independente ao VANT. O ângulo de movimento da câmera foi de 0° a -90° para foto panorâmica ou nadir (Fonseca Neto et al., 2017). Antes de realizar o aerolevantamento foram materializados alvos e feito o levantamento geodésico dos mesmos, dessa forma foi possível, simultaneamente ao processo de reconstrução do modelo 3D, obter o georreferenciamento da área estudada.

A Figura 8 apresenta os ortomosaicos obtidos a partir dos voos com VANT. No primeiro levantamento, a altura de voo foi de 110m, foram coletadas 228 imagens combinadas culminaram em uma resolução de pixel igual a 4,75 cm ao cobrir uma área de 0,078 km². O segundo levantamento, teve altura de voo igual a 60m, totalizando 425 imagens de uma área igual a 0,022 km² e após processamento obteve-se resolução de 1,58 cm por pixel. A Figura 9 apresenta recortes do ortomosaico nas parcelas 15, 42, 170, 217 e 238 em estágio avançado de crescimento.

Figura 8 – Ortomosaicos da área de estudo. (a) Voo realizado em 03/10/2018. (b) Voo realizado em 15/03/2019.

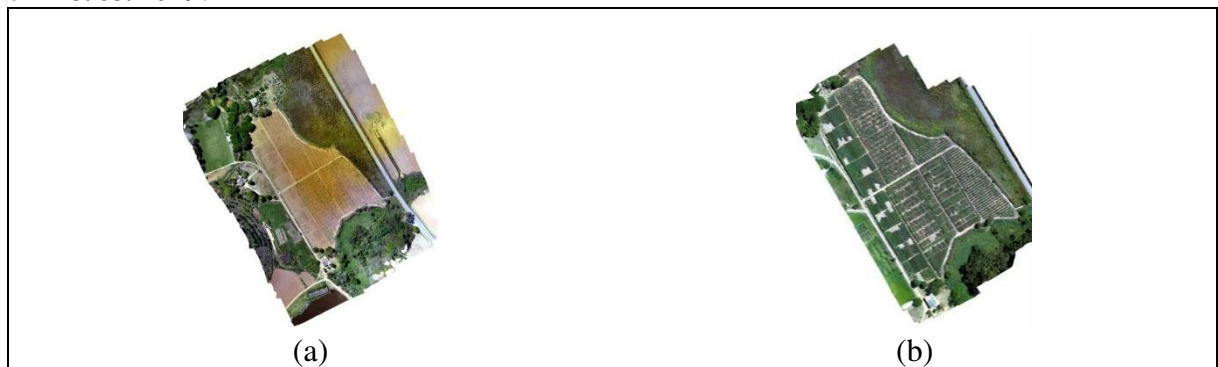
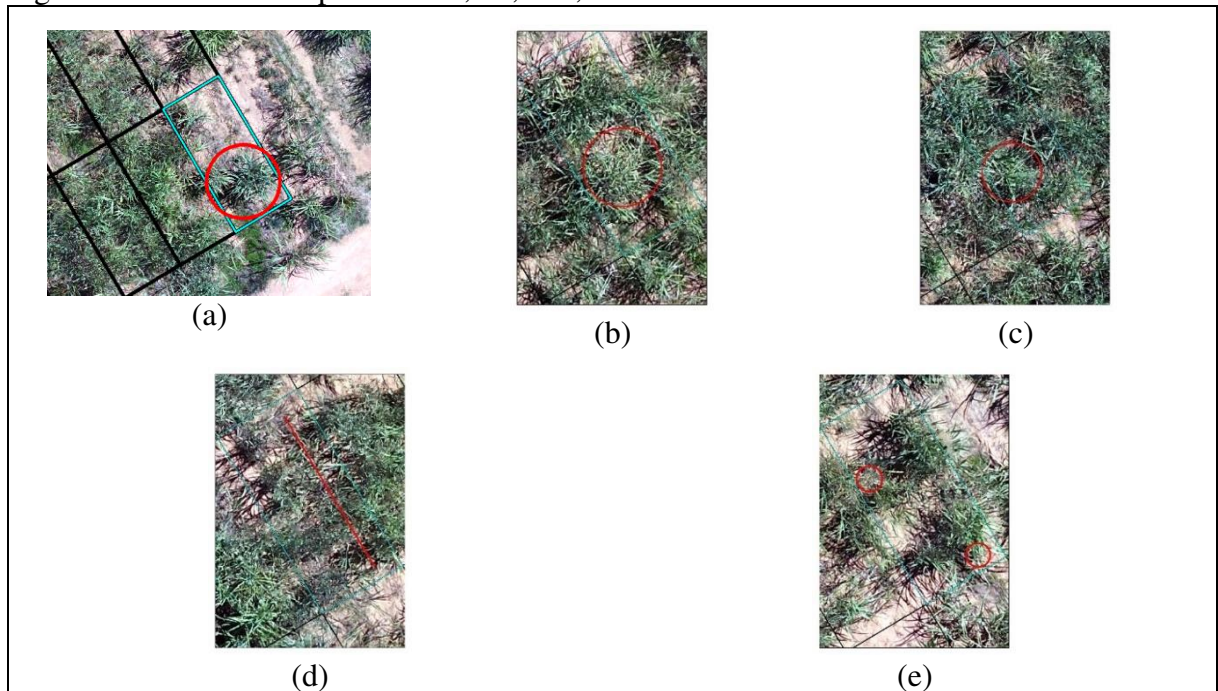


Figura 9 – Recortes das parcelas 15, 42, 170, 217 e 238 do voo realizado em 15/03/2019



3.2. Dados experimentais

O experimento foi conduzido no Centro de Pesquisa e Melhoramento da Cana-de-açúcar (CECA) da Universidade Federal de Viçosa (UFV), localizado no município de Oratórios, MG (latitude 20°25' S, longitude 42°48' W e 494 m de altitude).

3.3. Análises estatísticas

Todas as análises foram realizada nos softwares R (R Development Core Team, 2019), versão 3.6.0, e Arcgis (ESRI, 2004), versão 10.2.2.

3.3.1. Modelagem espacial dos atributos da análise de fertilidade do solo

O ajuste dos modelos teóricos de semivariograma e elaboração dos mapas de krigagem para as variáveis de fertilidade do solo foram realizados com o pacote geoR do R (Ribeiro Junior & Diggle, 2001). Foram utilizados 4 tipos de ajustes de semivariograma: por mínimos quadrados ordinários e ponderados; por verossimilhança e verossimilhança restrita. Além disso, analisaram-se 3 tipos de modelos teóricos: Exponencial, Gaussiano e Esférico. A escolha do melhor modelo seu deu por análise visual do ajuste do modelo teórico aos pontos amostrais do semivariograma experimental seguida de auto validação. De acordo com Santos et al. (2017) ao utilizar a auto validação deve-se atentar para algumas medidas: média dos erros e erros padronizados nulos; variância dos erros finita; variância dos erros padronizados unitária.

3.3.2. Obtenção dos valores médios a nível de parcela dos atributos do solo

De posse dos mapas de krigagem, para cada variável do solo, foi extraído o valor médio correspondente a cada uma das 240 parcelas com a ferramenta *Zonal Statistics*. Essa ferramenta faz parte do *Spatial Analyst Tools* presente no *ArcToolbox* do *software* Arcgis, e tem como funcionalidade o cálculo de estatísticas dos valores de um *raster* a partir das zonas determinadas por outro conjunto de dados, neste caso o *raster* dos mapas de krigagem e as zonas definidas pelas parcelas do experimento.

3.3.3. Seleção das covariáveis para correção de TCH

Para determinar quais variáveis seriam utilizadas como covariáveis no modelo explicativo dos valores de TCH, realizou-se uma análise em duas etapas (STICH, 2008).

A primeira etapa consistiu em obter os resíduos de uma ANOVA para modelo DBC. A segunda etapa da análise consistiu na seleção de variáveis de fertilidade do solo que apresentaram efeito significativo no modelo de explicação do TCH, para essa seleção utilizou-se o método *Stepwise*. Os resíduos obtidos na primeira etapa de seleção foram utilizados como variável dependente em função das variáveis de fertilidade do solo que apresentaram dependência espacial. As variáveis selecionadas foram aquelas que apresentam significância na modelagem dos valores de TCH em função dessas covariáveis.

A modelagem envolvida na seleção em duas etapas pode ser entendida como segue abaixo:

O modelo inicial é dado por:

$$y_{ij} = \mu + F_i + B_j + \varepsilon_{ij}, \quad 1 \leq i \leq 60, \quad 1 \leq j \leq 4 \quad (1)$$

Onde i denota a família, j denota o bloco. y_{ij} é valor do atributo de interesse na família i no bloco j , μ é a média geral, F_i é o efeito da família, B_j é o efeito de bloco, ε_{ij} é o resíduo na família i no bloco j .

Para inserir o efeito das covariáveis basta fazer a decomposição dos resíduos ε_{ij} como:

$$\varepsilon_{ij} = \beta_k X_{ijk} + \varepsilon_{ij}^*, \quad 1 \leq k \leq p \quad (2)$$

Onde β_k é o coeficiente da covariável k , X_{ijk} é o valor da covariável k na família i no bloco j , ε_{ij}^* é o novo resíduo desconsiderando o efeito de covariáveis e p é o número de covariáveis utilizadas.

Dessa forma podemos obter o valor corrigido para a média das famílias em cada uma das parcelas pela equação abaixo:

$$y_{ij}^* = y_{ij} - (\beta_k X_{ijk}) = \mu + F_i + B_j + \varepsilon_{ij}^* \quad (3)$$

De posse do termo $\beta_k X_{ijk}$ é possível obter o mapa de correção do modelo para as médias de TCH das parcelas do experimento, os valores de correção classificados em 5 classes com o intuito de compará-los com mapas baseado em imagens RGB apresentados posteriormente.

3.3.4. Ranqueamento das famílias de cana-de-açúcar

Para o ranqueamento das famílias utilizou o pacote *emmeans* (Lenth, 2020) e a função *rank*. O pacote *emmeans* permite obter, de forma rápido, os valores das médias corrigidas para cada família em função dos blocos. Além disso fornece intervalos de confiança para as médias com probabilidade de 95% (Lenth, 2020).

3.3.5. Cálculo dos índices de imagens RGB

Para cada uma das imagens obtidas por VANT foram calculados três índices RGB: VARI (Gitelson et al. 2002); GLI (Louhaichi et al., 2013) e NGRDI (Hunt et al. 2013). Os respectivos índices podem ser obtidos conforme as fórmulas abaixo:

$$VARI = \frac{G - R}{G + R + B} \quad (1)$$

$$GLI = \frac{2G - R - B}{2G + R + B} \quad (2)$$

$$NGRDI = \frac{G - R}{G + R} \quad (3)$$

Visando a possibilidade substituição as informações da análise de fertilidade do solo pelos índices de imagem RGB com a ferramenta *Zonal Statistics* foram extraídos a média, a amplitude e o desvio-padrão dos valores de cada índice para cada parcela do experimento. Os valores obtidos foram então classifica-los em 5 classes com o intuito de compará-los com os mapas de correção de médias obtidos a partir do modelo do DBC.

3.3.6. Comparação entre o mapa de correção e de índices de imagens

O Índice Kappa foi utilizado para comparar as classes dos mapas de correção com as classes dos mapas de média, amplitude e desvio-padrão.

4. RESULTADOS

4.1. Análises Estatística

4.1.1. Análise dos dados experimentais de TCH

A Tabela 2 apresenta algumas estatísticas descritivas da variável tonelada de colmos por hectare (TCH) referente as médias obtidas para as 240 parcelas do experimento. O valor médio de TCH foi de 119,087 t/ha. As medidas de dispersão apresentaram valores altos, fato já esperado uma vez que os dados são provenientes de uma etapa inicial do processo de seleção, onde diversos cruzamentos são analisados na busca de indivíduos superiores, logo há uma grande heterogeneidade entre os indivíduos. Observou-se um desvio-padrão de 30,23 e um coeficiente de variação de 25,38. Os valores mínimo e máximo foram iguais a 46,83 e 249,29 respectivamente. Na literatura os valores médios de TCH para estudos de genótipos de cana-de-açúcar são bem variados, com valores de 56 a 210 t/ha (Dutra Filho et al. (2020); Dellabiglia et al. (2019)).

Tabela 2 – Estatística descritiva dos valores de TCH.

Variável	Média	Mínimo	Máximo	Desvio	Variância	CV	Curtose	Assimetria
TCH	119,087	46,83	249,29	30,23	913,793	25,38	1,33	0,58

4.1.2. Análise de fertilidade do solo

O Anexo C apresenta a tabela com os dados amostrais da análise de fertilidade do solo para os 36 pontos amostrados. Foram analisadas as variáveis potencial hidrogeniônico (pH), fósforo (P), potássio (K), cálcio (Ca), magnésio (Mg), alumínio (Al), acidez potencial (H+Al), soma de bases trocáveis (SB), capacidade de troca catiônica (t), capacidade de troca catiônica a pH 7,0 (T), índice de saturação por base (V) e índice de saturação de alumínio (m). As variáveis Al, Ca, H+Al, K, Mg, P, SB, t, T presentes nesse estudo são expressas em cmolc/dm^3 , M e V em porcentagem e pH é adimensional. A Tabela 3 apresenta a estatística descritiva dessas variáveis, esses resultados são similares aos encontrados em Brito (2014) para um solo apresentando estresse com saturação de alumínio.

Tabela 3 – Estatística descritiva das variáveis da análise de fertilidade do solo

Variável	Média	Mínimo	Máximo	Desvio	Variância	CV	Curtose	Assimetria
Al	1,04	0,20	1,90	0,38	0,14	37,14	2,90	-0,05
Ca	0,85	0,22	5,00	0,82	0,66	95,78	20,21	3,98
H+Al	6,01	3,63	8,25	0,81	0,66	13,54	4,43	-0,21
K	43,61	23,00	92,00	15,99	255,62	36,66	3,89	1,08
m	49,43	5,60	79,50	18,03	325,05	36,47	2,75	-0,44
Mg	0,27	0,07	0,81	0,13	0,018	49,55	8,68	1,85
P	11,15	6,54	19,97	3,40	11,54	30,47	2,83	0,75
pH	4,48	4,10	5,60	0,28	0,077	6,19	8,47	1,91
SB	1,11	0,36	3,40	0,54	0,29	48,94	10,15	2,16
t	2,14	1,76	3,60	0,34	0,12	15,90	11,10	2,59
T	7,12	5,44	8,97	0,65	0,42	9,16	3,80	0,04
V	15,63	5,40	48,4	7,67	58,85	49,07	10,33	2,20

pH = potencial hidrogeniônico; P = fósforo; K = potássio; Ca = cálcio; Mg = magnésio; Al = alumínio; H+Al = acidez potencial; SB = soma de bases trocáveis; t = capacidade de troca catiônica; T = capacidade de troca catiônica a pH 7,0; V índice de saturação por base; m = índice de saturação de alumínio

De acordo com Warrick e Nielsen (1980), quanto ao coeficiente de variação pH e T são classificados como baixos ($CV < 12\%$), Al, H+Al, K, m, Mg, P, SB, t e V apresentaram valores considerados médios ($12\% < CV < 60\%$) e somente o Ca obteve valor alto ($CV > 60\%$). Os valores mínimo, médio e máximo do pH foram igual a 4,1, 4,48 e 5,6 respectivamente, classificados como de acidez média a elevada (PREZOTTI & GUARÇONI, 2013). Segundo (Ageitec, 2020) o valor recomendado de pH para a cultura de cana-de-açúcar é de 6,5.

Solos são caracterizados como distróficos ou inférteis quando $V < 50$, caso contrário são solos férteis. A avaliação da toxicidade do solo pode ser feita valendo-se da saturação em alumínio (m), que é a porcentagem de Al^{3+} em relação a capacidade de troca efetiva (t). Concentrações de m superiores a 50 indicam teor elevado de alumínio e baixas quantidades de cátions de base, consequentemente os solos são considerados muito pobres ou álicos (Ronquim, 2010). Prezotti & Guarçoni (2013) classificam o teor de Al^{3+} como baixo (menor 0,3), médio (entre 0,3 e 1) e alto (maior que 1). Para esse experimento as médias de V, m, Al foram respectivamente iguais a 15,63, 49,43 e 1,04, logo o solo não estava na condição fértil.

Prezotti & Guarçoni (2013) destacam que solos com pH inferior a 5,5 tem como característica elevados teores de alumínio, baixos teores de cálcio e magnésio e baixa saturação por bases. Além disso, valores de P entre 10 e 20 satisfazem a demanda de grande parte das culturas perenes, mas ainda é um fator limitante, K disponível inferior a 60 é considerado baixo. Os valores médios encontrados para P e K foram 11,15 e 43,61 respectivamente.

Esses resultados indicam que o solo não apresentava condições favoráveis ao desenvolvimento da cultura. Segundo Marin (2020) a cana-de-açúcar é capaz de se desenvolver em solos com condições físicas desfavoráveis ou baixa fertilidade, porém somente atinge altas produtividades em solos férteis e fisicamente adequados.

4.1.3. Modelagem dos semivariogramas

A análise geoestatística das amostras do solo via modelagem de semivariogramas identificou dependência espacial para as variáveis Al, Ca, H+Al, m, Mg, P, pH, SB, t e V. Já as variáveis K e T não apresentaram dependência espacial, consequentemente são variáveis de comportamento espacial aleatório e não foram utilizadas para obtenção de mapas de krigagem.

A Tabela 4 apresenta o método para estimação da curva de semivariograma e estatísticas da validação cruzada relacionadas a escolha dos modelos dos respectivos semivariogramas apresentados no Anexo D. Quanto aos métodos de ajuste do semivariograma as variáveis AL, Ca, Mg, pH, SB e V obtiveram melhores resultados com o WLS. Já os atributos m, P e t foram definidos via de ML e H+Al pela REML. Conforme Santos et al. (2017), na validação cruzada a média dos erros e erros padronizados deve ser nula, variância dos erros finita e variância dos erros padronizados unitária. Para todas as modelagens de semivariograma os erros foram próximos de zero, variância dos erros padronizados próxima de um, somente a média dos erros para as variáveis m, P e V não apresentaram valores próximos a zero (16,37, 33,1 3 7,25 respectivamente). Essas variáveis apresentam maiores valores de semivariância em magnitude, na casa das dezenas e centenas, consequentemente pontos distantes da curva de semivariograma podem implicar nesse tipo de ocorrência. Ao utilizar os erros padronizados esse efeito de escala não foi observado. Pode-se observar pelos semivariogramas relativos a essas variáveis que pontos com essa característica ocorreram para valores de distâncias de separação próximas ou maiores ao valor do alcance. O alcance é a medida que representa a maior distância de separação entre pares de pontos para a qual se deve utilizar a Geoestatística (Ferreira, 2015). Portanto esses pontos foram observados em distâncias de separação onde a grau de dependência espacial é menor ou nulo, implicando em um comportamento aleatório dos valores da semivariância.

Tabela 4 – Método escolha do semivariograma

Variável	Método	Média dos erros	Média dos erros padronizados	Variância dos erros	Variância dos erros padronizados
Al	WLS	-0,0009	-0,0012	0,37	1,02
Ca	WLS	0,0011	0,0008	0,81	1,12
H+Al	REML	-0,0077	-0,0050	0,86	1,13
m	ML	-0,0850	-0,0026	16,37	1,05
Mg	WLS	0,0021	0,0110	0,12	1,28
P	ML	0,0300	0,0049	3,11	1,02
pH	WLS	0,0030	0,0063	0,27	1,13
SB	WLS	0,0075	0,0089	0,51	1,21
t	ML	0,0015	0,0022	0,33	1,03
V	WLS	0,1100	0,0090	7,25	1,21

Al = alumínio; Ca = cálcio; H+Al = acidez potencial; m = índice de saturação de alumínio; Mg = magnésio; P = fósforo; pH = potencial hidrogeniônico; SB = soma de bases trocáveis; t = capacidade de troca catiônica; V índice de saturação por base.

A Tabela 5 apresenta os respectivos parâmetros dos semivariogramas ajustados e índice de dependência espacial (IDE) e Figura 10 e Figura 11 apresentam os mapas de krigagem para as 10 variáveis de fertilidade do solo que apresentaram dependência espacial.

Tabela 5 – Parâmetros dos semivariogramas para as variáveis de análise do solo

Variável	C_0	C_1	Alcance(m)	Modelo	IDE (%)	Classe
Al	0,06	0,086	23,51	Esférico	0,61	Moderada
Ca	0,35	0,180	24,76	Esférico	0,34	Moderada
H+Al	0,30	0,390	32,01	Esférico	0,57	Moderada
m	65,32	247,410	21,00	Exponencial	0,79	Alta
Mg	0,006	0,018	40,36	Gaussiano	0,75	Alta
P	2,00	9,250	24,71	Esférico	0,82	Alta
pH	0,042	0,037	32,24	Gaussiano	0,47	Moderada
SB	0,12	0,220	36,53	Gaussiano	0,65	Moderada
t	0,05	0,069	13,22	Exponencial	0,58	Moderada
V	25,00	44,460	36,36	Gaussiano	0,64	Moderada

Al = alumínio; Ca = cálcio; H+Al = acidez potencial; m = índice de saturação de alumínio; Mg = magnésio; P = fósforo; pH = potencial hidrogeniônico; SB = soma de bases trocáveis; t = capacidade de troca catiônica; V índice de saturação por base.

Das 10 variáveis que apresentaram dependência, 4 foram melhor explicadas pelo modelo esférico (Al, Ca, H+Al e P), 4 pelo gaussiano (Mg, pH, SB e V) e 2 pelo exponencial (m e t), demonstrando assim um menor grau de continuidade para variáveis m e t na área de estudo quando comparadas às demais. Quanto ao IDE, conforme a classificação de Zimback

(2001), houve predominância de dependência espacial moderada ($25 < IDE < 75 \%$), com 7 variáveis incluídas nesse intervalo, enquanto as demais foram classificadas como Alta ($IDE > 75\%$). Dessa forma, conforme Guimarães (2013) pode-se concluir que a dependência espacial foi bem capturada pela malha de amostragem utilizada e que malha utilizada foi adequada.

Figura 10 – Mapas de krigagem das variáveis. (a) Alumínio. (b) Cálcio. (c) Acidez potencial. (d) Índice de saturação por alumínio.

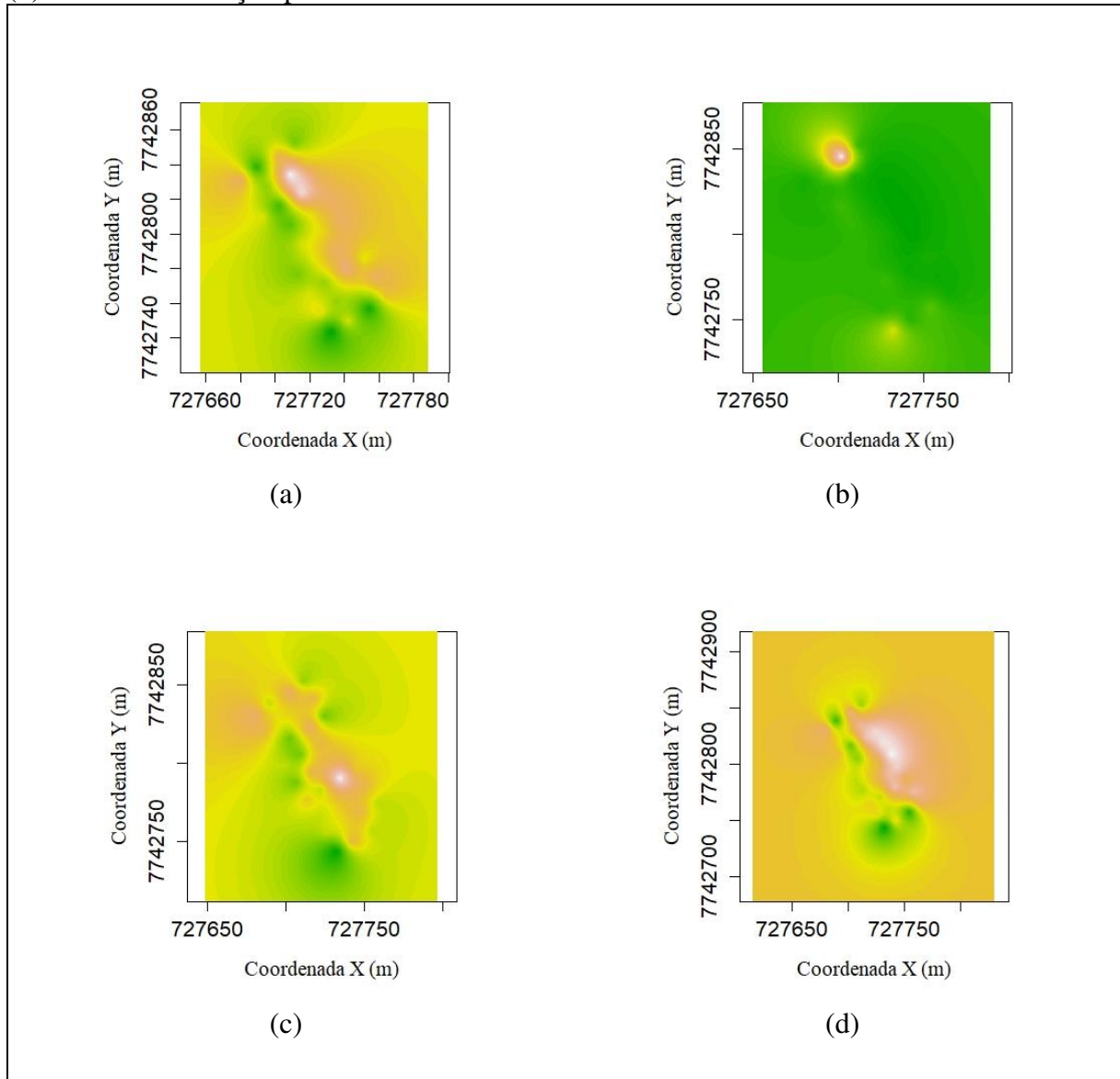
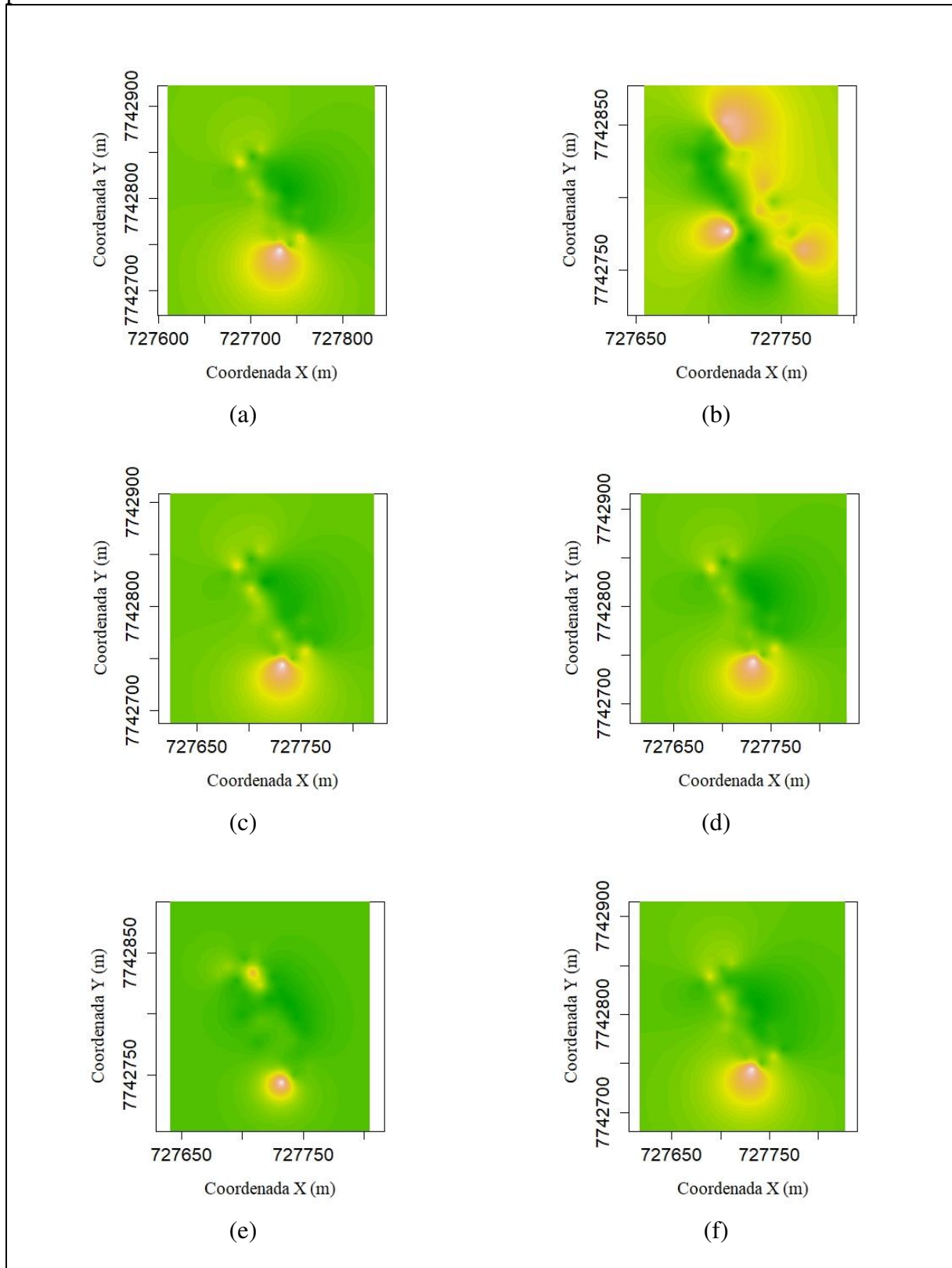


Figura 11 – Mapas de krigagem das variáveis. (a) Magnésio. (b) Fósforo. (c) Potencial Hidrogeniônico. (d) Soma de bases. (e) Capacidade de troca catiônica. (f) Índice de saturação por base.



4.1.4. Médias das variáveis de solo nível de parcela

Os mapas com os valores médios a nível de parcela das variáveis que apresentaram dependência espacial são apresentados em Figura 12 e Figura 13. De acordo com a distribuição espacial dos atributos de fertilidade do solo, devido a heterogeneidade dentro dos blocos, conclui-se que os blocos do experimento não foram definidos corretamente, uma vez que não determinam regiões homogêneas para esses atributos. A Tabela 6 apresenta a estatística descritiva dos valores médios das variáveis de fertilidade a nível de parcela. Quando comparados com os valores da área total do experimento (Tabela 3) pode-se observar que não houveram alterações nos valores médios, porém ocorreu diminuição dos coeficientes de variação, ou seja, as unidades experimentais apresentaram maior homogeneidade, mas mantiveram o valor médio das variáveis no campo.

Figura 12 – Mapas de valores a nível de parcela das variáveis. (a) Alumínio. (b) Cálcio. (c) Acidez potencial. (d) Índice de saturação por alumínio.

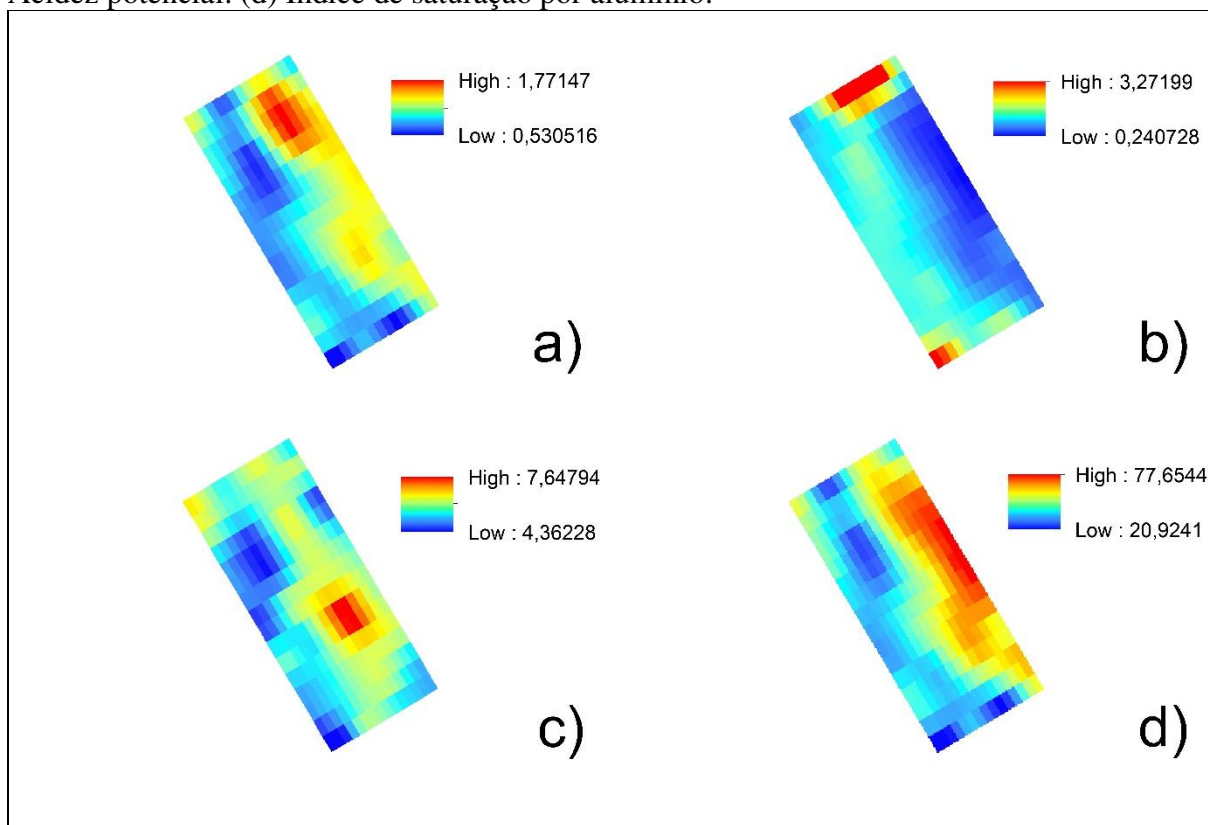
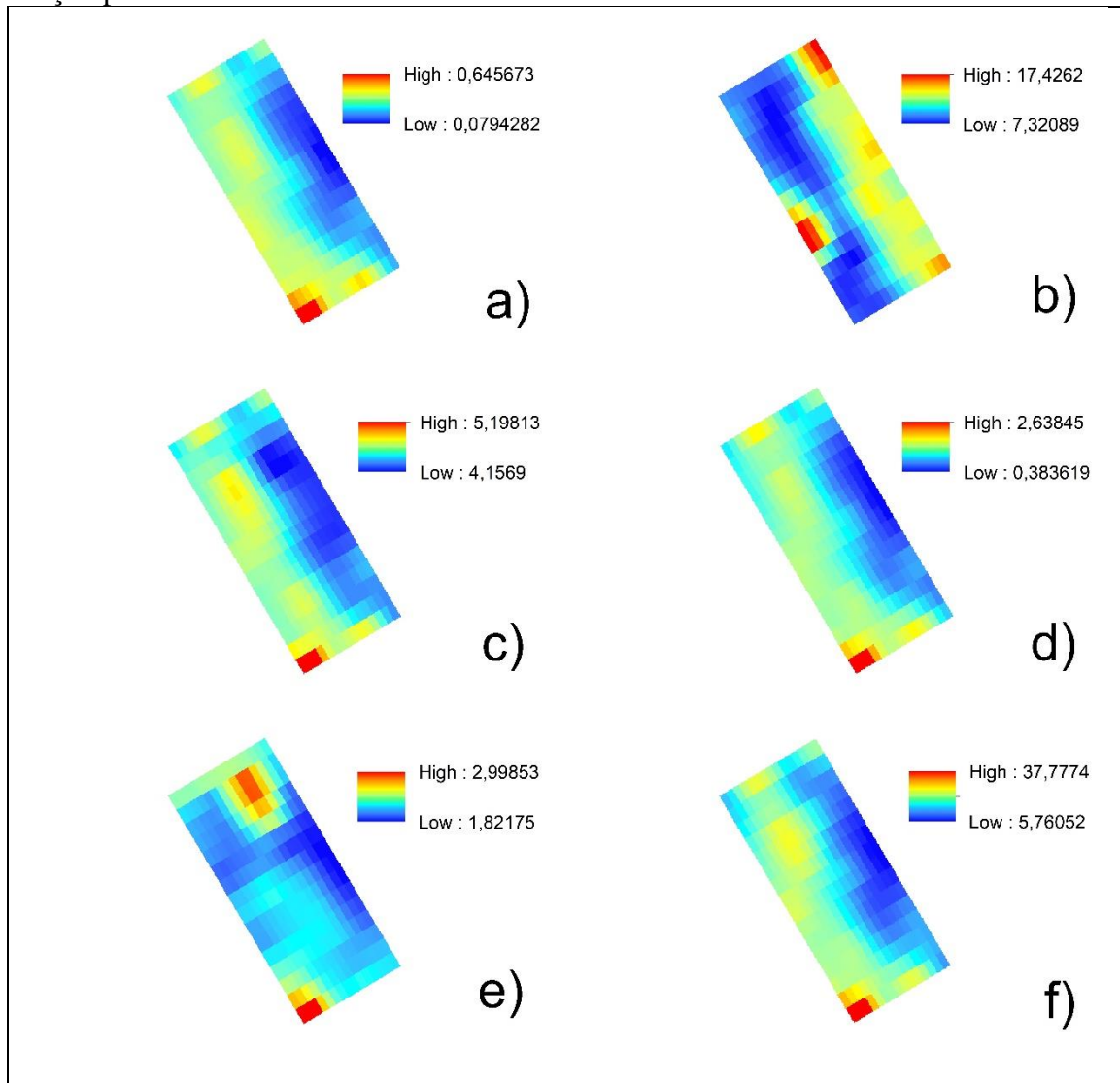


Figura 13 – Mapas de valores a nível de parcela das variáveis. (a) Magnésio. (b) Fósforo. (c) Potencial Hidrogeniônico. (d) Soma de bases. (e) Capacidade de troca catiônica. (f) Índice de saturação por base.



A Tabela 6 apresenta a estatística descritiva das variáveis da análise de fertilidade a nível de parcela após processo de interpolação.

Tabela 6 – Estatística descritiva da análise de fertilidade do solo

Variável	Média	Mínimo	Máximo	Desvio	Variância	CV	Curtose	Assimetria
Al	1,07	0,53	1,77	0,27	0,07	24,81	-0,54	0,38
Ca	0,76	0,24	3,26	0,42	0,18	55,46	13,12	3,02
H+Al	6,11	4,36	7,64	0,53	0,28	8,59	0,70	0,08
m	50,89	20,17	77,76	12,58	158,19	24,72	-0,87	0,25
Mg	0,25	0,08	0,65	0,08	0,01	32,78	2,35	0,59
P	10,84	7,32	17,43	2,39	5,70	22,05	-0,68	0,42
pH	4,45	4,16	5,20	0,16	0,03	3,70	1,80	0,58
SB	1,05	0,38	2,67	0,32	0,10	30,59	3,62	0,75
t	2,12	1,82	3,00	0,18	0,03	8,41	4,89	1,76
V	14,79	5,76	37,78	4,58	20,94	30,95	3,27	0,82

Al = alumínio; Ca = cálcio; H+Al = acidez potencial; m = índice de saturação de alumínio; Mg = magnésio; P = fósforo; pH = potencial hidrogeniônico; SB = soma de bases trocáveis; t = capacidade de troca catiônica; V índice de saturação por base.

4.1.5. Análise de Resíduos da Regressão

A Tabela 7 apresenta o quadro da ANOVA para o DBC, onde pode-se observar que o efeito de família foi significativo para o modelo (p -valor = 0,000). Conforme a Tabela 8 o modelo apresentou resíduos com média zero ($1,19^{-15}$), valor mínimo de -50,1, máximo de 101,09 e desvio-padrão de 20,19. Os valores dos resíduos do modelo em função das parcelas são apresentados na Figura 14.

Tabela 7 – Resultado do quadro de ANOVA

Modelo	Fonte de variação	gl	SQ	QMR	Valor F	p-valor
$y = \mu + F + B + \varepsilon$	Família	59	119598	2027,1	3,689	0,000 ***
	Bloco	3	1404	468,0	0,851	0,468
	Resíduo	177	97295	550,3	—	—

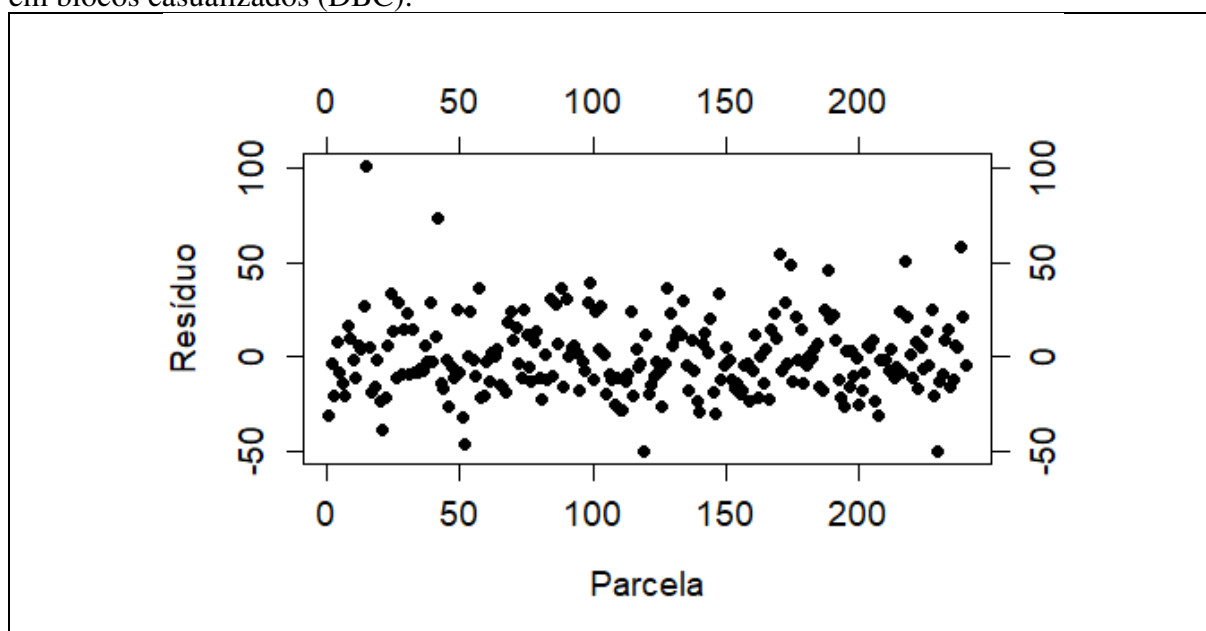
μ = média geral; F = efeito de família; B = efeito de bloco; ε = resíduo; *** significativo a 0,001%

Tabela 8 – Estatística descritiva dos resíduos

Modelo	Média	Mínimo	Máximo	Desvio
$y = \mu + F + B + \varepsilon$	$1,19^{-15}$	-50,1	101,09	20,19

μ = média geral; F = efeito de família; B = efeito de bloco; ε = resíduo.

Figura 14 – Valores residuais do modelo ajustado em função das parcelas para os delineamentos em blocos casualizados (DBC).



Os testes de normalidade e homogeneidade de variâncias dos resíduos não apresentaram resultados satisfatórios aos pressupostos da ANOVA ($3,88 \times 10^{-7}$ no teste de Shapiro-Wilk; 0,0038 no teste de Kolmogorov–Smirnov; 0,0026 no teste Bartlett) conforme a Tabela 9. Esse resultado negativo apontado pelos testes está associada aos resíduos das parcelas 15, 42, 170, 217 e 238. Ao excluir os resíduos referentes a essas parcelas, verificou-se a presença de normalidade e igualdade de variância dos resíduos conforme a Tabela 9 (0,04 e 0,08 no teste de Shapiro; 0,081 e 0,13 no teste de Kolmogorov–Smirnov; 0,96 e 0,97 no teste Bartlett). Porém como o objetivo desse trabalho é somente a comparação do ranqueamento das famílias e não a realização de testes de comparação de médias, optou-se pela não exclusão desses valores na análise. Além disso, a presença de outliers para este estudo pode ser um indicativo importante, uma vez que pode estar associado a alguma anomalia nas parcelas do experimento, como morte de indivíduos ou condições ambientais distintas devido a variação da fertilidade do solo. Dentro do contexto das pressuposições da ANOVA, Pino (2014) destaca que é comum a suposição de que as amostras populacionais sigam uma distribuição normal ou aproximadamente normal, porém a análise de variância nem sempre carece de tal suposição, sendo necessária somente quando se necessita dos resultados de testes de significância para comparação de médias. Dessa forma é possível admitir um pequeno afastamento da normalidade implicando em efeito prático pouco significativo na análise de variância convencional sem que haja erros sérios de interpretação de médias simples.

Tabela 9 – Teste estatísticos de normalidade e homogeneidade de variância para o DBC.

Modelo	sem exclusão de parcelas			com exclusão de parcelas		
	SW	KS	Bartlett	SW	KS	Bartlett
$y = \mu + F + B + \varepsilon$	$3,88 \times 10^{-7}$	0,0038	0,0026	0,04	0,081	0,96

μ = média geral; F = efeito de família; B = efeito de bloco; ε = resíduo; *** significativo a 0,001%; SW = teste de Shapiro-Wilk; KS = teste de Kolmogorov-Smirnov.

4.1.6. Método Stepwise

A seleção das covariáveis da ANCOVA foi feita via método Stepwise e os resultados são apresentados na Tabela 10. Para o DBC cinco variáveis apresentaram significância, m, Mg, P, pH e SB. Os coeficientes betas foram respectivamente -0,94, -350,32, 2,16, -99,97 e 96,97. Dessa forma as variáveis m, Mg e pH penalizaram os valores das médias, ao passo que P e SB bonificaram os valores médios. Conforme já dito anteriormente o valor de m está associado a avaliação da toxicidade do solo (Ronquim, 2010), o pH exerce grande influência no desenvolvimento de culturas (Lima et al. 2016) e P é um micronutriente importante no metabolismo das plantas (EPSTEIN; BLOOM, 2006). Lima et al. (2016) estudou a relação entre pH e TCH para diferentes camadas do solo, porém a relação só observou resultados significativos para camada 0,8 – 1,0 m, Parente et al. (2016) concluiu que existe correlação linear positiva entre dose de fosfato aplicado e produtividade, altura de planta e distância entre nós da cana-de-açúcar. Embora não seja diretamente representado pelo teor de alumínio, o índice de saturação por alumínio está intimamente ligado ao seu valor. Brito (2014) verificou que o estresse de Al pode causar diversos efeitos negativos na cana-de-açúcar, como redução da massa da planta seca, da parte aérea seca e as raízes axiais, além de aumentar a formação de raízes laterais.

Tabela 10 – Tabela da ANOVA após execução do método Stepwise de seleção de variáveis

Modelo	Fonte de variação	gl	SQ	QM	Valor F	p-valor	Beta
$y = \mu + F + B + \varepsilon$	m	1	2415	2415	6,46	0,01*	-0,94
	Mg	1	1167	1167	3,11	0,08.	-350,32
	P	1	1876	1876	5,01	0,03*	2,16
	pH	1	2044	2044	5,46	0,02*	-99,67
	SB	1	2229	2229	5,95	0,02*	96,97
	Resíduo	234	87664	374,63	—	—	—

μ = média geral; F = efeito de família; B = efeito de bloco; ε = resíduo; gl = graus de liberdade; SQ = soma de quadrados; QM = quadrado médio; ** significativo a 1%, * significativo a 5%, . significativo a 10%

4.1.7. Ranqueamento da famílias

Um dos objetivos desse trabalho é a comparação do ranqueamento das famílias de cana-de-açúcar nos modelos propostos com o ranqueamento tradicional (Ferreira, 2020). A Tabela 11 apresenta os ranqueamentos dos modelos (BDC e TRADICIONAL), a Figura 15 o gráfico de dispersão comparando os ranqueamentos e a Tabela 12 apresenta as taxas de concordâncias dos modelos. Pelo gráfico de dispersão dos modelos (Figura 15) observa-se que 49 famílias foram ranqueadas em posições diferentes e que a nuvem de pontos se afasta da reta identidade, indicando dissimilaridade entre os ranqueamentos e que o modelo proposto teve efeito no ranqueamento das famílias. A porcentagem de concordância dos modelos é apresentada na Tabela 12, para as 20% melhores famílias conformidade foi de 83,33%, passando a 90% de concordância para 50% das melhores famílias, seguindo a 95% para as 70% melhores famílias e atingindo 98,15% nas 90% melhores famílias. Isso mostra que a inclusão dos efeitos de fertilidade de solo no ranqueamento das famílias pode levar a correções e melhoria no ranqueamento das famílias. Ferreira (2020) dentre as 60 famílias de cana-de-açúcar selecionou as 24 melhores, isto é, 40%, o que corresponde a um erro de 87% quando comparado ao modelo proposto. Nos programa de melhoramento genético a seleção em T1 é de extrema importância para o prosseguimento do programa, logo uma boa quantidade de material genético pode estar sendo desperdiçada, implicando em perdas valiosas no processo de melhoramento.

Figura 15 – Gráfico de dispersão do ranqueamento do modelo TRADICIONAL vs DBC

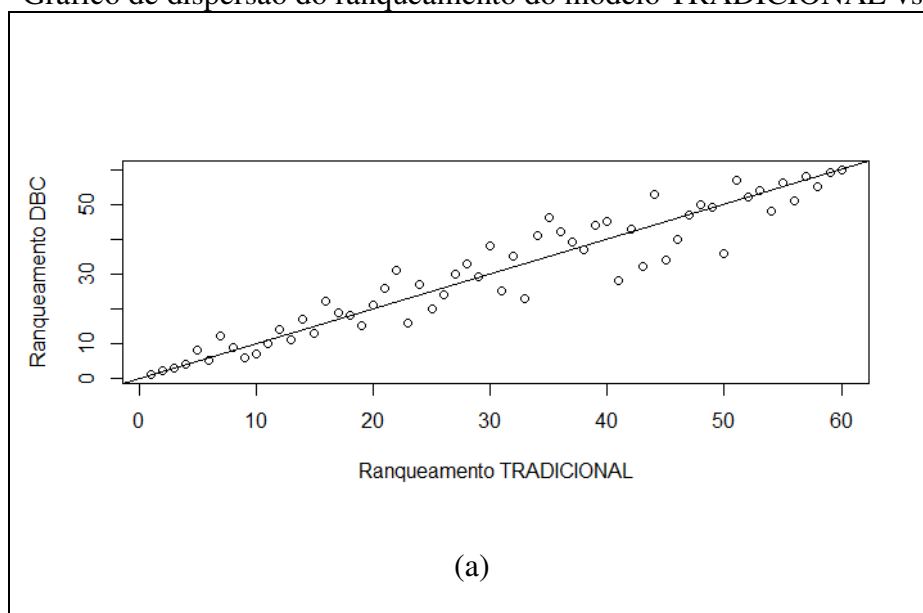


Tabela 11 – Ranqueamento das famílias quanto ao TCH nos modelos com correção no DBC e TRADICIONAL

Ranque	DBC	TRADICIONAL	Ranque	DBC	TRADICIONAL
1	289	289	31	314	305
2	499	499	32	413	413
3	473	285	33	530	527
4	285	470	34	305	444
5	424	497	35	476	489
6	497	473	36	269	445
7	439	219	37	444	476
8	219	466	38	329	394
9	514	514	39	558	472
10	470	424	40	489	223
11	348	539	41	223	269
12	384	384	42	533	447
13	539	439	43	309	309
14	343	343	44	447	535
15	512	234	45	472	329
16	463	293	46	426	533
17	466	475	47	535	323
18	280	443	48	457	426
19	443	474	49	323	324
20	527	468	50	368	457
21	293	512	51	244	368
22	475	559	52	564	342
23	232	314	53	342	279
24	559	232	54	324	244
25	474	348	55	515	564
26	234	363	56	279	515
27	468	463	57	317	317
28	394	530	58	224	224
29	363	280	59	555	555
30	445	558	60	250	250

DBC = Delineamento em blocos casualizados; TRADICIONAL = modelo tradicional de seleção famílias de cana-de-açúcar.

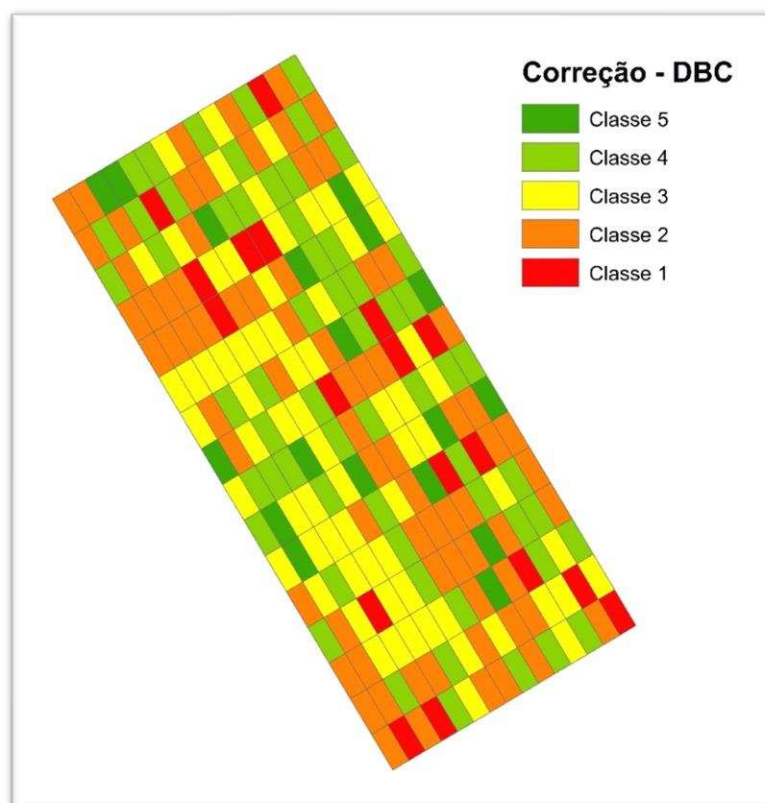
Tabela 12 – Comparação de ranqueamento das melhores famílias quanto ao TCH, em função do percentual de seleção de melhores famílias.

% de melhores famílias	DBC vs TRADICIONAL
10%	83,33%
20%	83,33%
30%	83,33%
40%	87,50%
50%	90,00%
60%	94,44%
70%	95,24%
80%	97,92%
90%	98,15%
100%	100.00%

DBC = Delineamento em blocos casualizados; TRADICIONAL = modelo tradicional de seleção famílias de cana-de-açúcar.

A Figura 16 apresenta o mapa de classes de correção da variável TCH para as 240 parcelas do experimento realizado em DBC.

Figura 16 – Mapa de correção da variável TCH para o modelo DBC.



4.1.8. Índices de imagens RGB

Utilizando as imagens da Figura 8 foram calculados os índices VARI (*Visible Atmospherically Resistant*), GLI (*Green Leaf Index*) e NGRDI (*Normalized Green Red Difference Index*) apresentados na Figura 17. A partir desses índices foram então calculados a nível de parcela o valor médio, a amplitude e o desvio-padrão de cada um dos índices (Figura 18, Figura 19 e Figura 20). Dos 18 mapas obtidos a partir dos índices de imagens RGB 5 não apresentaram variação com a predominância de uma única classe.

Figura 17 – Índices RGB calculados para os dois voos. (a) VARI (*Visible Atmospherically Resistant*) - 03/10/2018. (b) GLI (*Green Leaf Index*) - 03/10/2018. (c) NGRDI (*Normalized Green Red Difference Index*) - 03/10/2018. (d) VARI (*Visible Atmospherically Resistant*) - 15/03/2019. (e) GLI (*Green Leaf Index*) - 15/03/2019. (f) NGRDI (*Normalized Green Red Difference Index*) - 15/03/2019.

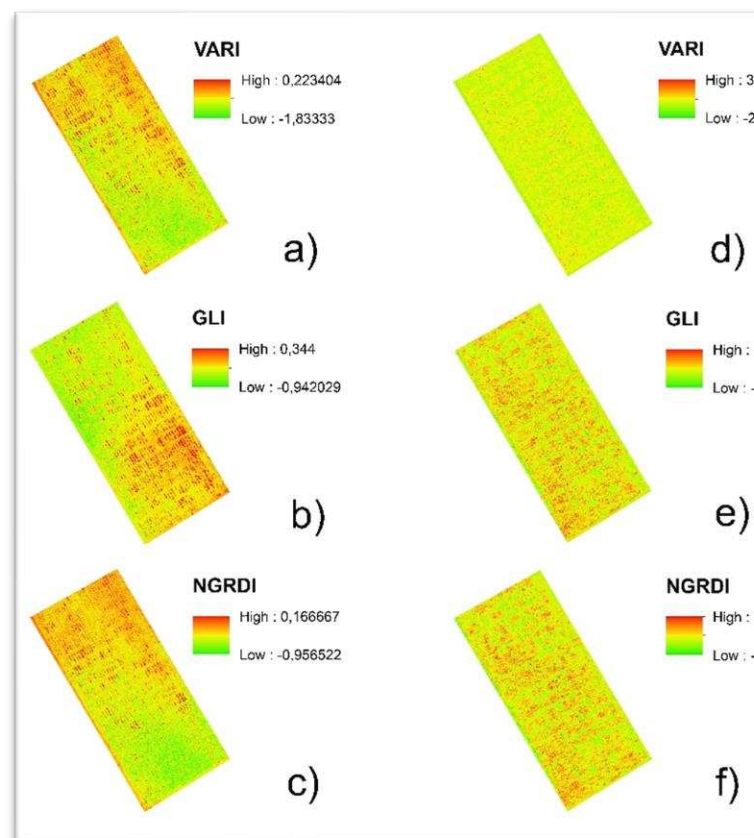


Figura 18 – Valor médio dos índices RGB para cada parcela do experimento. (a) VARI (*Visible Atmospherically Resistant*) - 03/10/2018. (b) GLI (*Green Leaf Index*) - 03/10/2018. (c) NGRDI (*Normalized Green Red Difference Index*) - 03/10/2018. (d) VARI (*Visible Atmospherically Resistant*) - 15/03/2019. (e) GLI (*Green Leaf Index*) - 15/03/2019. (f) NGRDI (*Normalized Green Red Difference Index*) - 15/03/2019.

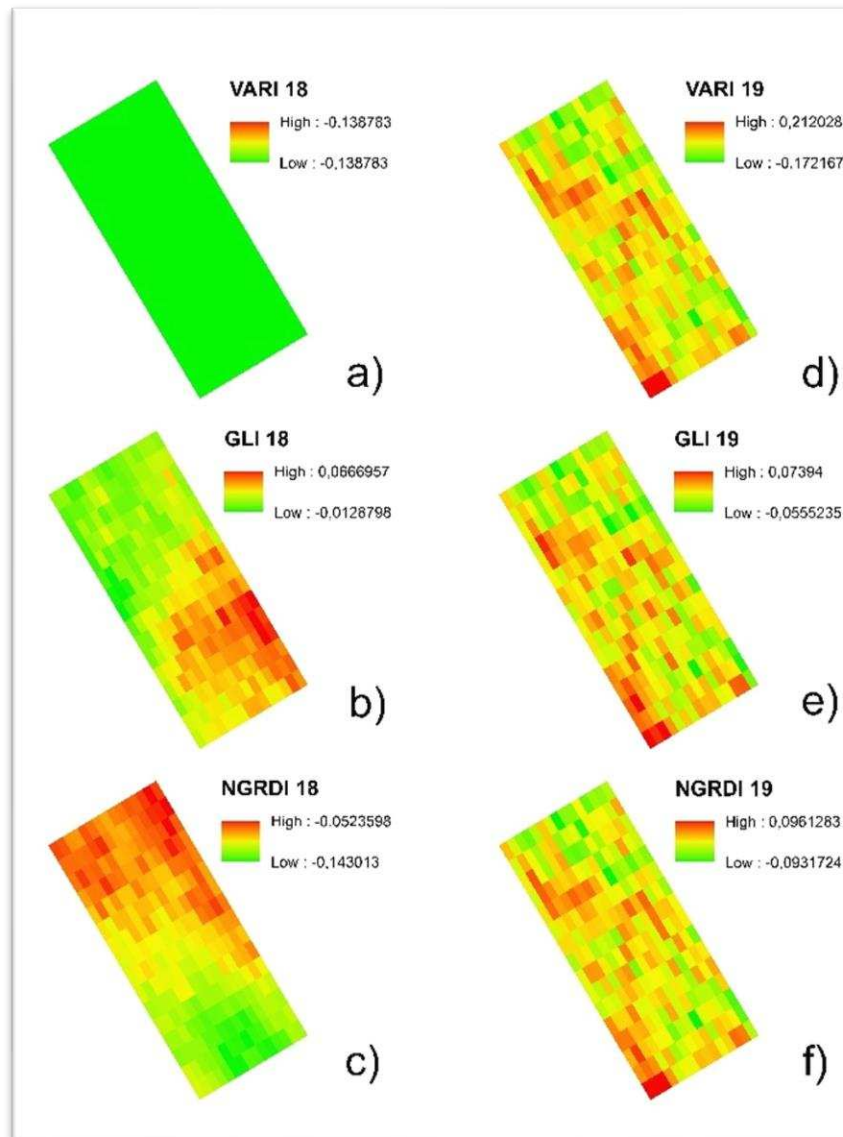


Figura 19 – Valor da amplitude dos índices RGB para cada parcela do experimento. (a) VARI (*Visible Atmospherically Resistant*) - 03/10/2018. (b) GLI (*Green Leaf Index*) - 03/10/2018. (c) NGRDI (*Normalized Green Red Difference Index*) - 03/10/2018. (d) VARI (*Visible Atmospherically Resistant*) - 15/03/2019. (e) GLI (*Green Leaf Index*) - 15/03/2019. (f) NGRDI (*Normalized Green Red Difference Index*) - 15/03/2019.

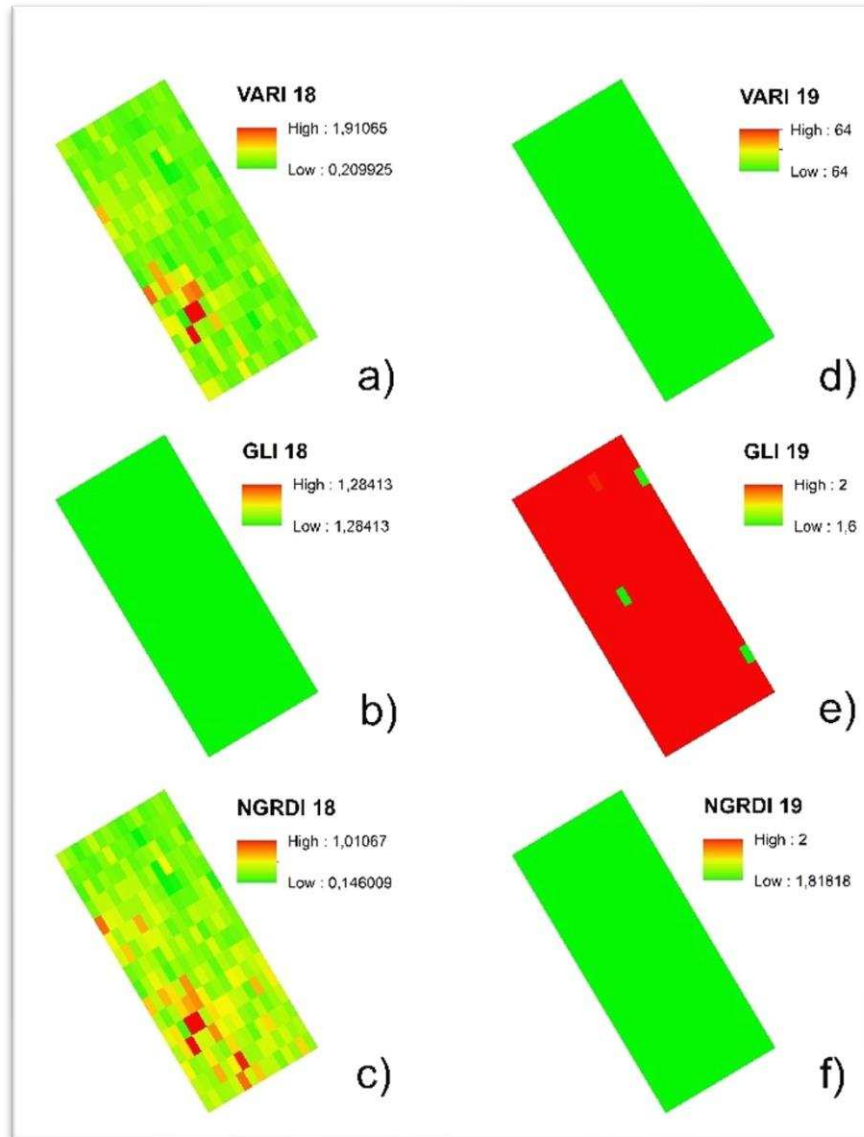
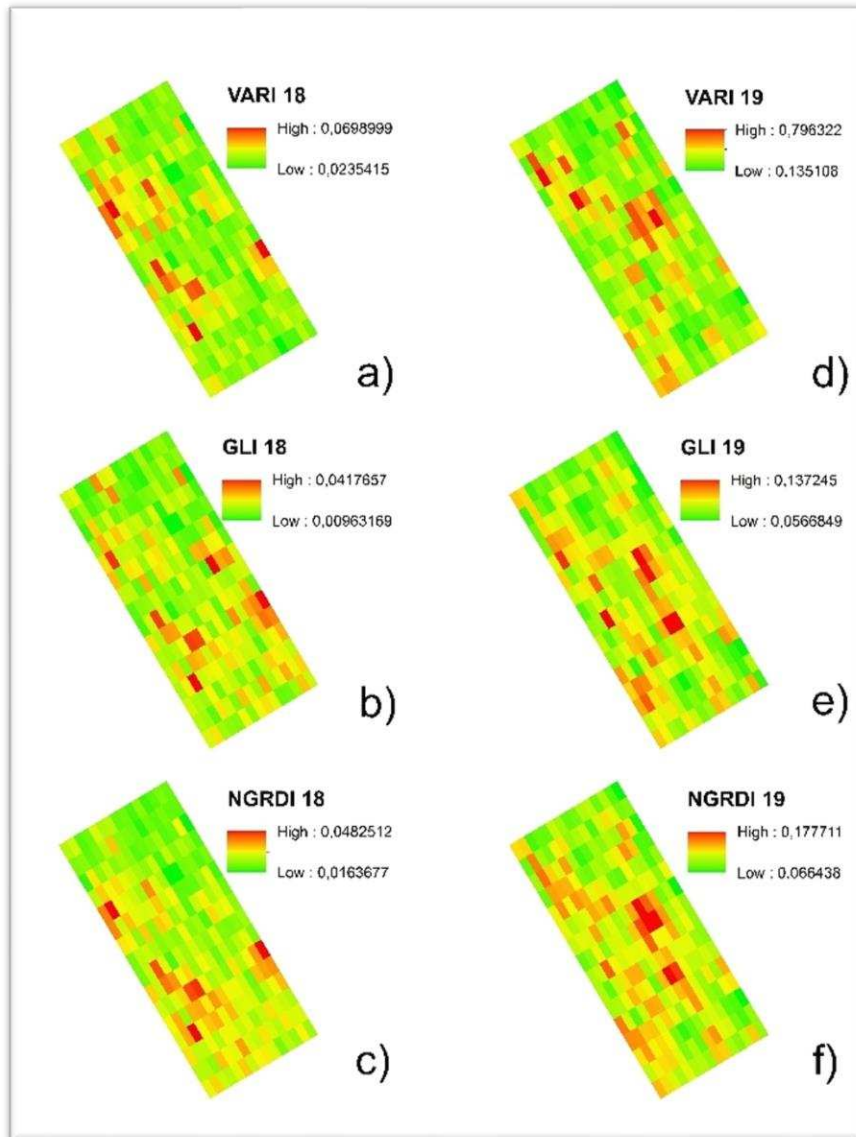


Figura 20 – Valor do desvio-padrão dos índices RGB para cada parcela do experimento. (a) VARI (*Visible Atmospherically Resistant*) - 03/10/2018. (b) GLI (*Green Leaf Index*) - 03/10/2018. (c) NGRDI (*Normalized Green Red Difference Index*) - 03/10/2018. (d) VARI (*Visible Atmospherically Resistant*) - 15/03/2019. (e) GLI (*Green Leaf Index*) - 15/03/2019. (f) NGRDI (*Normalized Green Red Difference Index*) - 15/03/2019.



4.1.9. Índice Kappa

Utilizando como referência o mapa de correção de parcelas (Figura 16) os mapas de valor médio, amplitude e desvio-padrão de cada um dos índices foram comparados quanto às classes, exceto os 5 com a predominância de uma única classe. Segundo a classificação proposta por Landis e Koch (1977) o índice Kappa apresentou valores de concordância mínima para todos os casos, impossibilitando assim a indicação de utilização dos índices RGB como alternativa ao mapa de correção de médias das parcelas, consequentemente inviabilizando a substituição da análise de fertilidade do solo pelos índices de imagens RGB. Golçalvez (2019) observou resultados similares ao obtido nesse estudo, concluindo que os valores das bandas espectrais RGB apresentaram baixa correlação com dados de fertilidade do solo. Por outro lado, Gracia-Romero et al. (2017) concluiu que índices de vegetação derivados de RGB podem monitorar os efeitos das aplicações de fertilizantes de fósforo e que existe correlações entre diferenças na biomassa e na produtividade de milho causadas pela fertilização com fósforo. Além disso, índices de vegetação RGB são capazes prever traços fenotípicos nos estágios iniciais de crescimento da cultura de milho e destaca que os índices RGB tiveram um desempenho melhor do que os índices de vegetação multiespectral.

Tabela 13 – Valores do Índice Kappa para os mapas de estatísticas a nível de parcelas dos índices de imagens RGB com os mapas de correção de médias dos modelos DBC e TRADICIONAL.

Mapa	Correção BDC		
	Média	Amplitude	SD
VARI 18	-	-0,03	0,06
VARI 19	0,16	-	0,08
GLI 18	0,06	-	0,05
GLI 19	0,03	-	0,06
NGRDI 18	0,00	-0,03	0,06
NGRDI 19	0,16	-	0,09

DBC = Delineamento em blocos casualizados; SD = desvio-padrão; VARI 18 = *Visible Atmospherically Resistant* em 03/10/2018; VARI 19 = *Visible Atmospherically Resistant* em 15/03/2019; GLI 18 = *Green Leaf Index* em 03/10/2018; GLI 19 = *Green Leaf Index* em 15/03/2019; NGRDI 18 = *Normalized Green Red Difference Index* em 03/10/2018; NGRDI 19 = *Normalized Green Red Difference Index* em 15/03/2019.

5. CONCLUSÕES

Conforme os resultados obtidos observou-se que a metodologia proposta foi capaz de apresentar correções no modelo tradicionalmente utilizado, comprovou-se a relevância dos efeitos do solo ao ranquear as famílias de cana-de-açúcar quanto a produtividade em toneladas de colmos por hectare (TCH), sendo as variáveis índice de saturação por alumínio, magnésio, fósforo, pH e soma de bases as que apresentaram efeito significativo neste estudo.

Adicionalmente, a partir da análise geoestatística dos atributos do solo foi possível verificar não homogeneidade das parcelas dos blocos utilizados neste estudo. Com efeito, para futuros experimentos indica-se a realização de correções mais exatas do solo, via análises de fertilidade e Geoestatística, para posteriormente iniciar os experimentos, permitindo assim a adequação da estrutura experimental e evitar o efeito do solo no ranqueamento das famílias. Caso não seja realizada tal correção do solo, indica-se após a colheita da cana-de-açúcar utilizar as análises de fertilidade do solo e Geoestatística para incluir o efeito do solo no modelo experimental, e assim fazer o ranqueamento das famílias de forma mais assertiva.

A utilização de imagens no padrão RGB não foi capaz de substituir as informações adquiridas via análise de fertilidade do solo, uma vez que os mapas de índices VARI, GLI e NGRDI não apresentaram grau de concordância significativo com os mapas temáticos das variáveis do solo. Em estudos futuros sensores NIR podem ser uma opção para substituição dos sensores RGB, e assim relacionar imagens às informações das propriedades do solo através de outros índices.

REFERÊNCIAS

- AGEITEC, Agência Embrapa de Informação Tecnológica. *Árvore do conhecimento – cana-de-açúcar*. Embrapa, disponível em: < https://www.agencia.cnptia.embrapa.br/gestor/cana-de-acucar/arvore/CONTAG01_18_3112006152934.html> acesso em 30/09/2020.
- AKAIKE, H. A new look at the statistical model identification. **IEEE Transactions on Automatic Control**, v.19, p.716-723, 1974.
- ALFACE, Angelo B. et al. Sugarcane spatial-temporal monitoring and crop coefficient estimation through NDVI. **Rev. bras. eng. agríc. ambient.**, Campina Grande, v. 23, n.5, p.330-335, May 2019. Available from <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1415-43662019000500330&lng=en&nrm=iso>. access on 26 June 2020. Epub May 06, 2019. <https://doi.org/10.1590/1807-1929/agriambi.v23n5p330-335>.
- ALVES, M. R.; DE MIRANDA, L. P. M.; ROCHA, A. C.; SENA, K. N.; TROLEIS, M. J. B.; MONTANARI, R. (2018, July). Espacialidade de atributos de solos cultivados com cana-de-açúcar no Pontal do Paranapanema, SP. In Colloquium Agrariae. ISSN: 1809-8215 (Vol. 14, No. 2, pp. 81-97).
- ANDRADE, R. G.; HOTT, M. C.; MAGALHAES JUNIOR, W. C. P. de; OLIVEIRA, P. S. d’; OLIVEIRA, J. S. e. **Uso de veículo aéreo não tripulado (VANT) como plataforma para monitoramento da produção agropecuária: estudo de caso para o milho forrageiro**. Juiz de Fora: EMBRAPA Gado de Leite, 2019. 20p. (Documentos, 233). Disponível em: <<http://ainfo.cnptia.embrapa.br/digital/bitstream/item/195875/1/DOC-233-Uso-de-VANT-como-plataforma-p-monitoram-da-prodagropecuaria.pdf>>. Acesso em: 17 mai. 2019.
- ANDRADE, R. G.; HOTT, M. C.; MAGALHÃES JUNIOR, W. C. P. Uso de veículo aéreo não tripulado (VANT) no monitoramento dos estágios de desenvolvimento da cultura do milho. **Embrapa Goda de leite – Capítulo em livro técnico (INFOTECA-E)**, 2019.
- ARAUS, J. L.; KEFAUVER, S. C.; ZAMAN-ALLAH, M.; OLSEN, M. S.; CAIRNS, J. E. Translating highthroughput phenotyping into genetic gain. **Trends in plant science**, v. 23, n. 5, p. 451-466, 2018.
- BARBOSA, M. H. P.; RESENDE, M. D. V.; BRESSIANI, J. A.; SILVEIRA, L. C. I.; PETERNELLI, L. A. Selection of sugarcane families and parents by Reml/Blup. **Crop Breeding and Applied Biotechnology**, v. 5, n. 4, p. 443-450, 2005.
- BARBOSA, M. H. P.; SILVEIRA, L. C. I. Melhoramento genético e recomendação de cultivares. In Santos F, Borém A and Caldas C (eds) Cana-de-açúcar, bioenergia, açúcar e etanol - tecnologias e perspectivas. **Folha de Viçosa**, Viçosa, p. 313-331, 2012.
- BARBOSA, Julierme Z. et al. Uso de imagens digitais obtidas com câmeras para analisar plantas. *Rev. de Ciências Agrárias*, Lisboa, v. 39, n. 1, p. 15-24, mar. 2016. Disponível em <http://www.scielo.mec.pt/scielo.php?script=sci_arttext&pid=S0871-018X2016000100003&lng=pt&nrm=iso>. acessos em 14 fev. 2020. <http://dx.doi.org/10.19084/RCA15006>.

BERTOLIN, N.O.; FILGUEIRAS, R.; VENANCIO, L.P.; MANTOVANI, E.C. Predição da produtividade de milho irrigado com auxílio de imagens de satélite. **Revista Brasileira de Agricultura Irrigada**, Fortaleza, v.11, n.4, p.1627-1638, 2017.

BRITO, C. M. **Variabilidade genética em cana-de-açúcar na tolerância ao alumínio**. 2014. 24p. Tese (Doutorado em Genética e melhoramento) – Universidade Federal de Viçosa, Viçosa, MG.

BNDES/CGEE (Org.). **Bioetanol de cana-de-açúcar: energia para o desenvolvimento sustentável**. Rio de Janeiro: BNDES, 2008. 316p.

BOTTEGA, E. L.; QUEIROZ, D. M.; PINTO, F. A. C.; SOUZA, C. M. A. Variabilidade espacial de atributos do solo em sistema de semeadura direta com rotação de culturas no cerrado brasileiro. **Revista Ciência Agronômica**, v. 44, n.1, p.1-9, 2013.

CAMBARDELLA, C.A, et al. (1994) Field-scale variability of soil properties in central Iowa soils. **Soil science society of America journal** 58.5:1501-1511.

CARVALHO, M. P.; TAKEDA, E. Y.; FREDDI, O. S. Variabilidade espacial de atributos de um solo sob videira em Vitória Brasil (SP). **Rev. Bras. Ciênc. Solo**, Viçosa, v. 27, n. 4, p. 695-703, Aug. 2003.

CONAB – Acomp. safra bras. cana, v. 7 - Safra 2019/20, n. 1 - Primeiro levantamento, Brasília, p. 1-62 maio de 2020.

COHEN, J. A coefficient of agreement of nominal scales. **Educational and Psychological Measurement**, v. 20, n. 1, p. 37–46, 1960.

DALCHIAVON, F.C.; RODRIGUES, A.R.; LIMA, E.S.; LOVERA, L.H; MONTANARI, R. Variabilidade espacial de atributos químicos do solo cultivado com soja sob plantio direto. **Revista de Ciências Agroveterinárias**, Lages, v.16, n.2, p.144-154, 2017.

DELLABIGLIA, W. J.; GAVA, G. J. C.; ARLANCH, A. B.; VILLAS BOAS, R. L.; CANTARELLA, H.; ROSSETTO, R. Produtividade de cana-de-açúcar fertirrigada com doses de N e inoculadas com bactérias diazotróficas. **Irriga**, v.1, p.29-34, 2018. <https://doi.org/10.15809/irriga.2018v1n1p28-41>

DRUCK, S.; CARVALHO, M.; CÂMARA, G. **Análise Espacial de Dados Geoestatísticos**, EMBRAPA, Brasília, 2004.

DUTRA FILHO, J. de A.; CALSA JÚNIOR, T.; SIMÕES NETO, D. E.; SOUTO, L. S.; DE LUNA, R. G.; SOUZA, A. dos S.; OLIVEIRA, O. H. de; SILVA, F. A. C. Sugarcane genotypes selection in three harvest times. **Research, Society and Development**, [S. l.], v. 9, n. 9, p. e461997187, 2020. DOI: 10.33448/rsd-v9i9.7187. Disponível em: <https://www.rsdjournal.org/index.php/rsd/article/view/7187>. Acesso em: 30 ago. 2020.

EISENBEISS, H. **UAV Photogrammetry**. Doctor of sciences, University of Technology Dresden, Germany, 2009, 237p.

EMBRAPA - Empresa Brasileira de Pesquisa Agropecuária. **Manual de métodos de análise de solo**. 2. ed. Rio de Janeiro: Embrapa Solos. 2011, 230 p.

EPSTEIN, E.; BLOOM, A. J. **Nutrição Mineral. Princípios e perspectivas**. 2.ed., Londrina: Editora Planta, 2006.

ESRI. **USING ArcGIS Geostatistical Analyst**. Redlands, CA – USA: ESRI Press, 2004a. 300p.

FERREIRA, Í. O.; SANTOS, G. R. DOS; RODRIGUES, D. D. Estudo sobre a utilização adequada da krigagem na representação computacional de superfícies batimétricas. **Revista Brasileira de Cartografia**, v. 5, n. 65/5, 2013.

FONSECA NETO, F.D.; Oliveira, J.C.; Botelho, M.F.; Santos, A.P.; Nascimento, L.A.; Fonseca, A.L.B. Avaliação da qualidade posicional de ortofotos geradas por SISVANT com diferentes alturas de voo para aplicações cadastrais. **Revista Brasileira de Cartografia**, n.69/7, p.1247-1261, 2017.

FREIRE-SILVA, Jadson et al. Índices de vegetação do Sensoriamento Remoto para processamento de imagens na faixa do visível (RGB). **Journal of Hyperspectral Remote Sensing**, [S.l.], v. 9, n. 4, p. 228-239, dec. 2019. ISSN 2237-2202. Available at: <<https://periodicos.ufpe.br/revistas/jhrs/article/view/242924>>. Date accessed: 06 jun. 2020. doi:<https://doi.org/10.29150/jhrs.v9.4.p228-239>.

GRACIA-ROMERO, A., Kefauver, S.C., VERGARA-DÍAZ, O., ZAMAN-ALLAH, M.A., PRASANNA, B.M., CAIRNS, J.E., ARAUS, J.L., 2017. Comparative performance of ground vs. aerially assessed RGB and multispectral indices for early-growth evaluation of maize performance under phosphorus fertilization. **Front. Plant Sci.** 8, 1–13. <https://doi.org/10.3389/fpls.2017.02004>.

GEBREMEDHIN, A.; BADENHORST, P. E.; WANG, J.; et al. Prospects for Measurement of Dry Matter Yield in Forage Breeding Programs Using Sensor Technologies. **Agronomy**. v. 9, p. 1-19, 2019.

GENÚ, A. M.; DEMATTÊ, J. A. M. Espectrorradiometria de solos e comparação com sensores orbitais. **Bragantia**, v. 71, n. 1, p. 82–89, 2012.

GITELSON, A. A.; STARK, R.; GRITS, U.; RUNDQUIST, D.; KAUFMAN, Y.; DERRY, D. Vegetation and soil lines in visible spectral space: a concept and technique for remote estimation of vegetation fraction. **International Journal of Remote Sensing**, v. 23, n. 13, p. 2537-2562, 2002.

GONÇALVES, A. K. **Sensoriamento remoto para estimativa dos parâmetros físicos e químicos para fertilidade do solo da Fazenda Experimental Lageado, Botucatu/SP**. 2019. Tese (Doutorado em Agronomia) – Universidade Estadual Paulista Júlio de Mesquita Filho, Botucatu, SP, 2019.

GUIMARÃES, W.D. **Geoestatística para o mapeamento da variabilidade espacial de atributos físicos do solo**. 2013. 60p. Tese (Doutorado em Engenharia Civil) – Universidade Federal de Viçosa, Viçosa, MG.

HUNT JR., E. R.; DORAISWAMY, P. C.; MCMURTREY, J. E.; DAUGHTRY, C. S. T.; PERRY, E. M. A visible band index for remote sensing leaf chlorophyll content at the canopy scale. **International Journal of Applied Earth Observation and Geoinformation**, v. 21, p. 103-112, 2013.

HOCKING, R. R. **The analysis and selection of variables in linear regression biometrics**. Washington, v. 32, n. 1, p. 1-49, 1976.

HOFSETZ, K.; SILVA M.A. Brazilian sugarcane bagasse: Energy and non-energy consumption. **Biomass Bioenerg**, v.46, p.564-573, 2012.

ISAAKS, E. H.; SRIVASTAVA, R. M. Applied geostatistics. Nova York: Oxford Univ. Press, p. 561, 1989.

JAMES, G.; WITTEN, D.; HASTIE, T.; TIBSHIRANI, R. **An Introduction to Statistical Learning with Applications in R**. 2013.

JÚNIOR, Sena et al. Discriminação entre estágios nutricionais na cultura do trigo com técnicas de visão artificial e medidor portátil de clorofila. **Engenharia Agrícola**, v.28, n. 1, p. 187-195, 2008.

KUMAH, C., ZHANG, N., RAJI, R. and PAN, R. (2019) Color Measurement of Segmented Printed Fabric Patterns in Lab Color Space from RGB Digital Images. *Journal of Textile Science and Technology*, 5, 1-18. doi: 10.4236/jtst.2019.51001.

KRIGE, D.G. A statistical approach to some basic mine evaluation problems on the Witwatersrand. **Journal of South African Institution of Mining and Metallurgy, Johannesburg**. v. 52, p. 119-139, 1951.

LANDIS, J.R. e KOCH, G.G. The measurement of observer agreement for categorical data. **Biometrics**, v.33, n.1, p. 159-174, 1977.

LENTH, R. **EMMs: Estimated Marginal Means: aka Least-Squares Means: R package version 1.3.4**. Disponível em: <<https://cran.r-project.org/web/packages/emmeans/emmeans.pdf>>. Acesso em: 16 maio. 2020.

LIMA, C. G. R.; CARVALHO, M. DE P. E; SOUZA, A.; COSTA, N. R.; MONTANARI, R. Correlação entre componentes da produtividade da cana-de-açúcar com ph de um argissolo vermelho distrófico do noroeste paulista. **Revista Engenharia na Agricultura - REVENG**, v. 24, n. 2, p. 120-130, 30 jun. 2016.

LUZ C. C.; ANTUNES A. F. B. Validação da tecnologia VANT na atualização de bases de dados cartográficos geológicos – estudo de caso: sistema cárstico do rio João Rodrigues. **Revista Brasileira de Cartografia**, Rio de Janeiro, n. 67, v. 7, p. 1439-1452, 2015.

LOUHAICHI, M.; YOUNG, W. C.; JOHNSON, D. E. Reliability of Yield Mapping System for Estimating Perennial Ryegrass Seed Yield. **Australian Journal of Basic and Applied Sciences**, v. 7, p. 133-138, 2013.

MACHADO, L. O.; LANA, A. M. Q.; LANA, R. M. Q.; GUIMARÃES, E. C.; FERREIRA, C. V. Variabilidade espacial de atributos químicos do solo em áreas sob sistema plantio convencional. **Revista Brasileira de Ciência do Solo**, Viçosa, v. 31, n. 3, p. 591-599, 2007.

MANGABEIRA, J. A. de C.; VALLADARES, G. S.; BATISTELLA, M.; GOMES, E. G. Relação espacial entre produtividade agrícola e fertilidade do solo com uso de imagens de satélite e técnicas de geoprocessamento. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 12., 2005, Goiânia. **Anais...** Goiânia: SBSR, 2005. p. 193-195.

MARIN, F. R., **Árvore do conhecimento cana-de-açúcar: características**. Agência Embrapa de Informação Tecnológica (Ageitec), [s.d.] Disponível em: <http://www.agencia.cnptia.embrapa.br/gestor/cana-deacucar/arvore/CONTAG01_20_3112006152934.html>, Data de acesso: 30 de abril de 2020.

MARINHO, E.; FASBENDER, D.; KOK, R. D. **Spatial assessment of categorical maps a proposed framework**. In: Proceedings of the 4thGEOBIA. Rio de Janeiro, 2012.

MATHERON, G. Principles of geostatistics. **Economic Geology**. N. 5, p. 1246-1266, 1963.

MELO, L. J. O. T. **Sistema simplificado de seleção para a fase inicial do melhoramento genético da cana-de-açúcar**. 2014. Tese (Doutorado em Produção Vegetal) – Universidade Federal do Paraná, Curitiba, PR, 2014.

MILANEZ, A.Y.; NYKO, D.; VALENTE, M.S.; SOUSA, L.C.; BONOMI, A.; JESUS, C.D.F.; WATANABE, M.D.B.; CHAGAS, M.F.; REZENDE, M.C.A.F.; COVALETT, O.; JUNQUEIRA, T.L.; GOUVÊIA, V.L.R. De promessa à realidade: como o etanol celulósico pode revolucionar a indústria de cana-de-açúcar – uma avaliação do potencial competitivo e sugestões de política pública. **Biocombustíveis BNDES Setorial**, 41, 237-294, 2015.

MONTANARI, Rafael et al. Aspectos da produtividade do feijão correlacionados com atributos físicos do solo sob elevado nível tecnológico de manejo. **Revista Brasileira de Ciência do Solo**, v. 34, n. 6, p. 1811-1822, 2010.

MONTGOMERY, D. C.; RUNGER, G. C. Estatística aplicada e probabilidade para engenheiros. 4. ed. Rio de Janeiro: LTC – Livros Técnicos e Científicos Editora S.A., 2009. 493 p.

MOREIRA, E.F.A. **Redes Neurais Artificiais E Análise Discriminante Linear Como Alternativas Para Seleção Entre Famílias De Cana-De-Açúcar**. 2013. 60p. Tese (Doutorado em Engenharia Civil) – Universidade Federal de Viçosa, Viçosa, MG.

MOREIRA, E. F. A.; e PETERNELLI, L. A. Sugarcane families selection in early stages based on classification by linear discriminant analysis. **Revista Brasileira de Biometria** 33: 484-493, 2005.

NOETZOLD, R., ALVES, M. C., GOUSSAIN JÚNIOR, M. M., GOUSSAIN, R. C. S. (2019). Variabilidade espacial da eficiência do uso de potássio e fósforo na cultura da soja 1. *Revista Engenharia na Agricultura*, 27(6), 529-541.

OLIVER M.; A. WEBSTER, R. Statistical methods in soil and land resource survey. Oxford Univ. Press, New York. 1990.

OLIVEIRA, I. P.; COSTA K. P.; RODRIGUES, C.; MACEDO, R. F.; MOREIRA, F. P. e SANTO, K. J. G. Manutenção e correção da fertilidade do solo para inserção do cerrado no processo produtivo. **Revista Eletrônica Montes Velos**, Goiás, ISSN 1808-8597, 2005, Vol 1(1), p. 50-64, 2005.

PALUDO. A.; BECKER W. R; RICHETTI J.; SILVA L. C A; JOHANN J. A. (2020) Mapping summer soybean and corn with remote sensing on Google Earth Engine cloud computing in Parana state – Brazil. **International Journal of Digital Earth**, DOI: 10.1080/17538947.2020.1772893

PARENTE, T. L. et al. Residual de fósforo em cana soca para produção de forragem no norte de Mato Grosso. **Revista de Ciências Agroambientais**, v. 14, n. 1, 2016.

PARK, C.-B.; LEE, S.-W. Real-time 3D pointing gesture recognition for mobile robots with cascade HMM and particle filter. **Image and Vision Computing**, 29(1):51–63, 2011.

PENATTI, P. C. **Adubação da cana-de-açúcar: 30 anos de experiência**. São Paulo: Ottoni Editora, 2013.

PERUZZO, J. S., PEREIRA, M. C. S., SILVA, L. D. R., OLIVEIRA, B. S., SILVINO, G. S. (2019). Sensoriamento remoto aplicado ao monitoramento ambiental da bacia do Alto Piranhas, Semiárido Nordeste (Brasil). **Revista Brasileira de Meio Ambiente**, v.7, n.3 (Edição Especial – XV SEAGROCCA), p.28-37.

PINO, F. A. A questão da normalidade: uma revisão. *Revista de economia agrícola*, v. 61, n. 2, p. 17-33, 2014. Disponível em: <<http://www.iea.sp.gov.br/ftpica/publicar/rea2014-2/rea2-2014.pdf>>. Acesso em: 07 out. 2019.

PREZOTTI, L.C.; GUARÇONI, A. **Guia de interpretação de análise de solo e foliar**. Vitória: Incaper, 2013. 104 p.

R Development Core Team. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2019. ISBN 3-900051-07-0. Disponível em: <<http://www.R-project.org>>

RENCER, A. C., SCHAALJE, G. B. 2008. *Linear models in statistics*, 2nd edn. John Wiley & Sons, Hoboken, NJ.

RIBEIRO JUNIOR, P. J.; DIGGLE, P. J. *geoR: a package for geostatistical analysis*. **R-NEWS**, London, v. 1, n. 2, p. 15-18, June 2001.

RIBEIRO, R. B.; FILGUEIRAS, R.; RAMOS, M. C. A.; ALMEIDA, L. T. de; GENEROSO, T. N.; MONTEIRO, L. I. B. Variabilidade espaço-temporal da condição da vegetação na agricultura irrigada por meio de imagens sentinel. **Revista Brasileira de Agricultura Irrigada**, v.11, p.1884-1893, 2017. <https://doi.org/10.7127/rbai.v11n600648>

RODRIGUES, R. A. S.; MONTENEGRO, A. A. A.; SANTOS, D. P. dos; MAGALHÃES, A. G.; FONTES JÚNIOR, R. P. Variabilidade espacial da umidade e das frações granulométricas do solo em um plantio de bananeiras irrigado no semiárido pernambucano. **Conexões - Ciência e Tecnologia**, v.11, p.134-143, 2017. <https://doi.org/10.21439/conexoes.v11i3.884>

RONQUIM, C.C. 2010. **Conceitos de Fertilidade do Solo e Manejo Adequado para as Regiões Tropicais**. Campinas: Embrapa Monitoramento por Satélite, 26p.

ROSAS, J. T. F.; JUNIOR, E. M.; SERAFINI, G. A. D.; LORENZONI, R. M.; SANTOS, C. E. M. Potencial do uso imagens digitais na caracterização fenotípica de progênies de maracujazeiro azedo. **Anais da 30ª Semana Acadêmica do Curso de Agronomia do CCAE/UFES-SEAGRO**, 2019.

SANTANA, P. B. **Desempenho agrônomo e tecnológico de variedades de cana-de-açúcar submetidas a diferentes lâminas de irrigação no Norte de Minas Gerais**. (2012). Tese de Doutorado. Dissertação (Mestrado) Janaúba, UNIMONTES.

SANTOS, A. P., MEDEIROS, N. G., SANTOS, G. R., RODRIGUES, D. D. 2016. Avaliação da Acurácia Posicional Planimétrica em Modelos Digitais de Superfície com o Uso de Feições Lineares. **Boletim de Ciências Geodésicas**, 22(1), pp.157-174. doi: 10.1590/S1982-21702016000100009

SANTOS, D. P.; MONTENEGRO, A. A. de A.; RODRIGUES, R. A. S.; ARAÚJO, D. C. dos S.; SANTOS, C. S.; CRUZ NETO, J. F. da C. Variabilidade especial de atributos físicos do solo em vale aluvial na região de Pernambuco. **Revista Brasileira de Agricultura Irrigada**, v.12, p.2271-2282, 2018. <https://doi.org/10.7127/rbai.v12n100647>

SANTOS, G. R. **Hierarquização geométrica dos preditores geoestatísticos**. Tese (Doutorado em Estatística e Experimentação Agropecuária) – Setor de Ciências Exatas, Universidade Federal de Lavras, Lavras. 2010. 200p.

SANTOS, G. R., et al. Krigagem simples versus krigagem universal: qual o preditor mais preciso? **Energia na Agricultura**, 26, 2011.

SALVADOR, M. M. S.; LIBARDI, P. L.; BRITO, A. S.; MOREIRA, N. B. Estabilidade temporal e variabilidade espacial da distribuição da armazenagem de água no solo numa sucessão feijão/aveia-preta. **Revista Brasileira de Ciências do Solo**, v.36, p.1801-1810, 2012. <https://doi.org/10.1590/S0100-06832012000500007>

SCHUSTER, I.; CRUZ, C.D. Estatística genômica aplicada a populações derivadas de cruzamentos controlados. Viçosa: UFV, 568p. 2013.

SHAKHATREH, Hazim et al. Unmanned aerial vehicles (UAVs): A survey on civil applications and key research challenges. **Ieee Access**, v. 7, p. 48572-48634, 2019.

SILVA, D. J.; ARAÚJO, C. A. S. Agricultura irrigada: a importância da adubação. Embrapa Semiárido. In: **Anais do Congresso Brasileiro de Ciência do Sol**, Recife, 2005.

SILVA, J.S.; MONTENEGRO, A.A.A.; SILVA, E.F.F.; ANDRADE, C.W.L.; SILVA, J.R.L. Distribuição espacial da condutividade elétrica e matéria orgânica em Neossolo Flúvico. **Revista Brasileira de Geografia Física**. v.6, p.764-776, 2013.

SIQUEIRA, G. M.; VIEIRA, S. R.; CEDDIA, M. B. Variabilidade de atributos físicos do solo determinados por métodos diversos. **Bragantia**, Campinas, v. 67, n. 1, p. 203-211, 2008.

SOARES A. Geoestatística para as ciências da terra e do ambiente. 2a ed. Lisboa: IST Press; 2006.

SOUZA, Z. M.; MARQUES JÚNIOR, J.; PEREIRA, G. T. Variabilidade espacial de atributos físicos do solo em diferentes formas do relevo sob cultivo de cana-de-açúcar. **Revista Brasileira de Ciência do Solo**, p. 937-944, 2004.

STICH, B; MORHRING, J; PIEPHO, H. P., HECKENBERGER, M.; Buckler, E. S.; Melchinger, A. E. Comparison of mixed-model approaches for association mapping. **Genetics**, v. 178, n. 3, p. 1745-1754, 2008.

TABACHNICK, B.G.; FIDELL, L. S. **Using Multivariate Statistics**. 5. ed. Boston: Pearson 2007.

TANGER, P. et al. Field-based high throughput phenotyping rapidly identifies genomic regions controlling yield components in rice. **Scientific Reports**, v. 7, p. 42839-42847, 2017.

TRANGMAR, B.B.; YOST, R.S. & UEHARA, G. Application of geostatistics to spatial studies of soil properties. *Adv. Agron.*, 38:45-93, 1985.

TEIXEIRA, M. B.; CAMARGO, P. L. T.; MARTINS JUNIOR, P. P. Uso das Imagens de Sensores Remotos para Análise Crítica da Degradação do Cerrado no Alto Médio São Francisco. Minas Gerais. **Anuário do Instituto de Geociências (UFRJ Impresso)**, Rio de Janeiro, v. 41, p. 245-254, 2018.

VIEIRA, S. R. Geoestatística em estudos de variabilidade espacial do solo. In: NOVAIS, R. F. de; ALVAREZ, V. H.; SCHAEFER, C. E. G. R. (Ed.). **Tópicos em ciência do solo**. Viçosa, MG: Sociedade Brasileira de Ciência do Solo, 2000. v. 1, p. 1-54.

VIEIRA, S.R.; CARVALHO, J.R.P. de; PAZ GONZÁLEZ, A. Jack knifing for semivariogram validation. **Bragantia**, v.69, p.97-105, 2010.

VITTI G. C; QUEIROZ F. E. C; OTTO R.; QUINTINO T. A; **Nutrição e adubação da cana-de-açúcar**; Departamento de Solos e Nutrição de Plantas – ESALQ/USP – Piracicaba – SP-2005.

WARRICK, A. W.; NIELSEN, D. R. Spatial variability of soil physical properties in the field. In: HILLED, D. (ed). **Applications of soil physics**. New York: Academic, 1980.

YAMAMOTO, J.K.; LANDIM, P. M. B. Geoestatística: conceitos e aplicações. 1ª ed. São Paulo: Oficina de Textos. 215p., 2013.

YANG, G. J. et al. Unmanned aerial vehicle remote sensing for field-based crop phenotyping: current status and perspectives. **Frontiers in plant science**, v. 8, p. 1111, 2017.

ZIMBACK, C.R.L. Análise espacial de atributos químicos de solos para fins de mapeamento da fertilidade. Botucatu, Universidade Estadual Paulista, 2001. 114p. Tese (livre-Docência).

ANEXOS

ANEXO A – Tabela com valor médio de TCH a nível de parcela nos blocos 1 e 2

BLOCO 1				BLOCO 2			
Código da Famílias	TCH	Código da Famílias	TCH	Código da Famílias	TCH	Código da Famílias	TCH
384	111,3697	280	113,6286	424	127,0924	457	98,1571
289	166,7460	527	133,4662	444	115,1053	250	55,8762
473	133,8655	445	112,8571	368	88,9286	489	113,7698
514	148,0357	426	97,4286	527	115,6786	447	110,1339
564	82,9911	472	99,3304	280	101,8797	269	89,1176
463	112,0982	309	106,1786	413	95,8647	443	122,2180
470	129,0714	457	111,1310	476	95,1587	555	70,8791
475	150,9341	342	91,9231	463	137,6786	558	143,5714
515	100,2778	223	143,3810	445	138,5065	305	155,2747
497	150,5952	250	55,2857	473	156,4796	293	120,5238
535	99,2208	363	139,1667	514	149,4388	314	142,2449
468	137,0879	293	212,6374	426	95,5357	224	84,2262
279	95,1020	269	100,2143	234	117,7551	232	149,2308
343	167,0879	314	107,7143	515	108,4286	343	134,9490
348	249,2857	224	85,0595	470	154,2262	244	68,7755
444	124,9412	329	89,2460	363	116,3304	539	124,9444
512	113,6639	439	135,9624	324	79,4603	329	97,2353
368	79,2857	394	111,3025	535	111,7347	394	90,6767
413	116,4286	499	195,4286	223	121,7063	533	93,1092
424	123,5119	533	102,9832	559	108,6161	323	66,3095
476	81,5179	539	109,3233	499	141,1429	439	107,0159
474	109,8214	558	75,8403	472	100,1339	309	94,5238
559	132,6891	305	122,6190	474	112,4176	219	136,5816
317	122,2857	489	138,4694	475	157,9121	342	112,3980
234	149,2857	443	130,2101	317	72,2959	285	133,9048
219	141,6071	555	75,0893	497	173,9732	466	139,0336
324	127,6531	323	137,5893	564	91,9048	530	110,9694
232	119,7959	530	102,2078	384	172,2527	468	121,3839
285	175,9524	447	93,2381	289	147,9121	348	91,3839
244	117,9464	466	139,5536	512	155,8163	279	96,4286

TCH = toneladas de colmos por hectare

ANEXO B – Tabela com valor médio de toneladas de colmos por hectare a nível de parcela nos blocos 3 e 4

BLOCO 3				BLOCO 4			
Código da Famílias	TCH	Código da Famílias	TCH	Código da Famílias	TCH	Código da Famílias	TCH
279	18,6360	457	99,7744	463	124,2857	232	137,1875
497	19,4169	329	99,6429	527	115,7857	279	121,3736
232	19,9665	559	106,0924	439	142,7143	497	138,6364
317	20,7740	489	97,2269	447	118,9098	368	111,4286
445	20,9247	363	104,8214	499	152,4176	219	141,8571
348	21,6867	527	96,8908	305	103,0252	309	133,0000
244	20,1921	463	118,5714	426	128,0357	329	136,2185
289	21,5093	219	146,0000	539	184,1837	559	65,5495
530	19,8636	475	107,5325	285	179,6875	280	53,6429
384	18,1360	234	124,4156	558	141,2857	564	100,1681
470	20,0889	555	92,7381	535	116,6518	443	109,4388
499	20,2733	305	97,0408	384	128,3766	514	116,8182
447	20,2738	466	138,4184	473	130,1190	234	140,9740
439	18,7007	285	143,5238	342	66,1161	444	96,8304
539	20,5786	342	95,4762	413	118,5238	474	116,5714
512	20,0805	424	120,1278	515	72,1597	343	135,7589
468	21,4788	394	133,0102	530	124,4365	269	89,0476
224	19,0369	280	142,5630	314	112,5446	317	202,5974
343	21,7447	533	117,2619	244	92,1849	472	141,2500
223	18,8933	472	156,2245	348	120,5357	223	98,0952
269	19,7565	514	129,8352	289	149,8095	445	137,1875
564	19,3600	309	138,6735	250	46,8304	533	121,3736
314	20,8085	443	125,1020	555	88,0102	363	138,6364
474	20,1050	476	165,1020	323	104,2857	489	111,4286
515	19,4217	323	84,7619	476	126,9328	468	141,8571
293	20,2574	368	113,2143	475	108,2967	512	133,0000
473	19,5321	444	114,6875	293	105,7619	224	136,2185
426	19,3162	413	129,1327	470	145,6044	424	65,5495
535	20,1062	324	81,0204	466	138,5294	394	53,6429
250	18,8833	558	113,3766	324	94,5536	457	100,1681

TCH = toneladas de colmos por hectare

ANEXO C – Tabela com valores amostrais das variáveis de fertilidade do solo

Amostra	pH	P	K	Ca	Mg	Al	H+Al	SB	t	T	V	m
1	5,60	8,68	55	2,45	0,81	0,2	3,63	3,40	3,60	7,03	48,4	5,6
2	4,50	7,58	39	0,85	0,35	1,1	5,61	1,30	2,40	6,91	18,8	45,8
3	4,50	8,68	30	0,78	0,31	1,0	5,61	1,17	2,17	6,78	17,3	46,1
4	4,50	19,97	41	0,83	0,35	0,7	6,27	1,28	1,98	7,55	17,0	35,4
5	4,60	13,22	55	0,84	0,33	0,8	4,95	1,31	2,11	6,26	20,9	37,9
6	4,50	9,56	27	0,75	0,27	0,8	5,61	1,09	1,89	6,7	16,3	42,3
7	4,50	8,68	24	0,69	0,28	1,0	5,61	1,03	2,03	6,64	15,5	49,3
8	4,50	9,86	29	0,7	0,27	1,0	6,60	1,04	2,04	7,64	13,6	49,0
9	4,40	8,97	56	0,55	0,24	1,4	6,94	0,93	2,33	7,86	11,8	60,1
10	4,50	7,58	52	0,76	0,27	1,0	6,60	1,16	2,16	7,76	14,9	46,3
11	4,60	9,26	61	0,81	0,34	0,8	5,94	1,31	2,11	7,25	18,1	37,9
12	4,70	6,54	30	0,96	0,32	0,8	5,94	1,36	2,16	7,3	18,6	37,0
13	4,50	8,4	31	0,74	0,26	1,0	5,61	1,08	2,08	6,69	16,1	48,1
14	4,60	7,84	41	0,82	0,28	1,0	6,60	1,20	2,20	7,8	15,4	45,5
15	4,70	7,31	35	0,91	0,36	0,6	4,95	1,36	1,96	6,31	21,6	30,6
16	4,80	7,05	27	1,06	0,35	0,5	4,95	1,48	1,96	6,43	23,0	25,3
17	4,40	7,05	50	0,75	0,27	0,8	5,94	1,15	1,95	7,09	16,2	41,0
18	4,90	9,56	77	1,36	0,46	0,4	5,61	2,02	2,42	7,63	26,5	16,5
19	4,90	11,46	45	1,31	0,48	0,3	5,61	1,91	2,21	7,52	25,4	13,6
20	4,30	13,6	72	0,58	0,26	1,0	6,60	1,02	2,02	7,62	13,4	49,5
21	4,30	12,5	39	0,38	0,16	1,5	6,60	0,64	2,14	7,24	8,8	70,1
22	4,20	14,78	47	0,44	0,16	1,4	8,25	0,72	2,12	8,97	8,0	66,0
23	4,30	12,14	57	0,58	0,23	1,1	6,60	0,96	2,06	7,56	12,7	53,4
24	4,30	11,46	37	0,48	0,16	1,1	6,60	0,74	1,84	7,34	10,1	59,8
25	4,10	12,86	67	0,46	0,17	1,8	6,97	0,80	2,60	7,73	14,5	69,2
26	4,40	8,68	47	0,7	0,24	1,9	6,27	1,06	2,96	7,33	9,8	64,2
27	4,30	9,56	31	5	0,17	1,3	6,93	0,75	2,05	7,68	14,1	63,4
28	4,30	16,95	92	0,49	0,19	1,2	5,61	0,92	2,12	6,53	10,2	56,6
29	4,30	11,46	34	0,39	0,16	1,4	5,61	0,64	2,04	6,25	11,8	68,6
30	4,40	13,98	39	0,57	0,21	1,0	6,60	0,88	1,88	7,48	11,8	53,2
31	4,20	10,8	43	0,27	0,11	1,3	6,27	0,49	1,79	6,76	7,2	72,6
32	4,20	14,78	28	0,22	0,07	1,4	6,27	0,36	1,76	6,63	5,4	79,5
33	4,20	13,6	23	0,25	0,11	1,4	5,94	0,42	1,82	6,36	6,6	76,9
34	4,20	12,14	29	0,3	0,12	1,4	4,95	0,49	1,89	5,44	9,0	74,1
35	4,40	16,95	34	0,59	0,2	1,2	6,60	0,88	2,08	7,48	11,8	57,5
36	4,70	17,9	46	1,02	0,37	0,7	5,28	1,51	2,21	6,79	22,2	31,7

pH = potencial hidrogeniônico; P = fósforo; K = potássio; Ca = cálcio; Mg = magnésio; Al = alumínio; H+Al = acidez potencial; SB = soma de bases trocáveis; t = capacidade de troca catiônica; T = capacidade de troca catiônica a pH 7,0; V índice de saturação por base; m = índice de saturação de alumínio.

ANEXO D – Semivariogramas ajustados das variáveis. (a) Alumínio. (b) Cálcio. (c) Acidez potencial. (d) Índice de saturação por alumínio. (e) Magnésio. (f) Fósforo. (g) Potencial Hidrogeniônico. (h) Soma de bases. (i) Capacidade de troca catiônica. (j) Índice de saturação por base.

