Universidade de São Paulo
Escola Superior de Agricultura “Luiz de Queiroz”

Geotecnologias aplicadas em experimentação agronômica: ênfase em cana-de-açúcar

Danilo Eduardo Cursi

Tese apresentada para obtenção do título de Doutor em
Ciências. Área de concentração: Fitotecnia

Piracicaba
2020

Danilo Eduardo Cursi
Engenheiro Agrônomo

Geotecnologias aplicadas em experimentação agrônômica: ênfase em cana-de-açúcar

Orientador:
Prof. Dr. **DURVAL DOURADO NETO**

Tese apresentada para obtenção do título de Doutor em
Ciências. Área de concentração: Fitotecnia

Piracicaba
2020

Dados Internacionais de Catalogação na Publicação
DIVISÃO DE BIBLIOTECA – DIBD/ESALQ/USP

Cursi, Danilo Eduardo

Geotecnologias aplicadas em experimentação agrônômica: ênfase em cana-de-açúcar / Danilo Eduardo Cursi. - - Piracicaba, 2020.

78 p.

Tese (Doutorado) - - USP / Escola Superior de Agricultura "Luiz de Queiroz".

1. Sensoriamento proximal 2. Variabilidade espacial 3. Genética quantitativa 4. Geoestatística 5. Envirotyping 6. Saccharum officinarum L. . I.
Título

DEDICATÓRIA

À toda minha **família**.
Em especial, aos meus pais,
João C. Cursi e **Rosi M. M. Cursi**,
por todo suporte, estímulo e amor;
Aos meus avós,
Luiz Cursi (*in memoriam*) e **Lourdes P. Cursi**,
pelos gestos gratificantes e singelos de incentivo;
À minha esposa **Patrícia N. Barbosa**,
que nunca mediu esforços para me apoiar,
Dedico.

AGRADECIMENTOS

A Deus, por me cercar de pessoas inspiradoras.

À Escola Superior de Agricultura “Luiz de Queiroz” (Esalq), Universidade de São Paulo (USP), através do Programa de Pós-Graduação em “Fitotecnia”.

Ao Prof. Dr. Durval Dourado Neto, pela orientação e confiança depositada desde o início da Pós-Graduação.

Ao Prof. Dr. Lucas Rios do Amaral, pela co-orientação e oportunidade de desenvolver parte das atividades junto ao Grupo Interdisciplinar de Tecnologia em Agricultura de Precisão (GiTAP), da Faculdade de Engenharia Agrícola da Universidade Estadual de Campinas (Feagri/Unicamp).

Aos Professores dos diferentes Departamentos da Esalq que tive a oportunidade de cursar disciplinas.

Ao Programa de Melhoramento Genético da Cana-de-açúcar da Universidade Federal de São Carlos (UFSCar/RIDESa), pelo apoio financeiro.

Ao Professor Dr. Hermann Paulo Hoffmann, que desde o início de minha carreira acadêmica e profissional não mediu esforços para me guiar, aconselhar e proporcionar infinitas oportunidades. Minha eterna gratidão e admiração.

Ao Professor Dr. Rodrigo Gazaffi, pelas inúmeras conversas, ensinamentos e fundamental contribuição para a realização deste trabalho.

À Professora Dra. Monalisa Sampaio Carneiro, pela amizade, conselhos e direcionamentos.

A todos os alunos do Grupo Interdisciplinar de Tecnologia em Agricultura de Precisão (GiTAP), da Faculdade de Engenharia Agrícola da Universidade Estadual de Campinas (FEAGRI/UNICAMP), que me auxiliaram em diversas análises do trabalho. Em especial, ao colega Thiago Luis Brasco.

Aos colegas de equipe do Programa de Melhoramento Genético da Cana-de-açúcar (RIDESA/UFSCar), tanto da Estação Experimental de Araras: Adriana Uccella, Carlos Loureiro, Flávio Costa, Igor Nunes, José Ciofi, José Adalberto da Cruz, Luiz Fernando Dias, Murillo Perticarari, Nathan Pinto, Plínio Zavaglia, Roberto Chapola, Sandro Ferrarez, Stela Scalabrin, Thiago Balsalobre e Valdir Guedes, também aos de Valparaíso: Antonio Ribeiro, Bruno Molina, Ismael Mantelato, João Civeiro, Lilian Mendonça, Lungas Menezes, Rafael Simões, Samuel Ferreira e Valdir Oliveira, pelo valioso desenvolvimento profissional ao longo desses anos.

Ao colega de disciplinas Matheus Gabriel Acorsi, pelos ensinamentos na utilização de diferentes ferramentas computacionais para geoprocessamento.

À minha esposa Patrícia Neves Barbosa, pela paciência, compreensão e suporte ao longo desses quase quatro anos de dedicação ao curso de Doutorado.

À minha família, Patrícia N. Barbosa (esposa), João C. Cursi (pai), Rosi M. M. Cursi (mãe), Diego E. Cursi (irmão), Franciane Ariele Cursi (cunhada), Enzo R. Cursi (sobrinho), Liliane C. Cursi (irmã), Alex de Carvalho (cunhado), Domingos Sávio C. Barbosa (sogro), Márcia R. N. Barbosa (sogra), Camila N. Barbosa (cunhada), Antonio N. Martinez (cunhado), Laís B. Martinez (afilhada) e para a nova integrante da família que está por vir, Sofia C. de Carvalho. Vocês são meus maiores presentes de Deus.

À minha sogra, Márcia R. N. Barbosa, pela correção ortográfica e gramatical do texto.

Aos colaboradores de campo do Programa de Melhoramento Genético da Cana-de-açúcar (RIDESA/UFSCar) das Estações Experimentais de Araras e Valparaíso.

ΕΠÍΓΡΑΦΕ

“Continuous effort - not strength or intelligence -
is the key to unlocking our potential.”
Winston S. Churchill (30/11/1874 – 24/01/1965)

“If I had asked people what they wanted, they would have said faster horses.”
Henry Ford (30/07/1863 – 07/04/1947)

SUMÁRIO

RESUMO	1
ABSTRACT	2
LISTA DE FIGURAS	3
1. INTRODUÇÃO	7
2. REVISÃO BIBLIOGRÁFICA	9
2.1. MELHORAMENTO GENÉTICO DA CANA-DE-AÇÚCAR	9
2.2. GANHO GENÉTICO NO MELHORAMENTO.....	10
2.3. AGRICULTURA DE PRECISÃO	11
2.3.2. <i>Sensor de condutividade elétrica aparente do solo (CEa)</i>	14
2.3.3. <i>Sensores aéreos multiespectrais (RPA, drones)</i>	15
2.4. GEOESTATÍSTICA	15
2.4.1. <i>Semivariograma</i>	17
2.4.2. <i>Métodos de interpolação</i>	18
2.4.2.1. <i>Krigagem</i>	19
2.4.2.2. <i>Ponderação do Inverso da Distância (IDW)</i>	19
3. OBJETIVO	20
3.1. GERAL.....	20
3.2. ESPECÍFICOS	20
4. MATERIAL E MÉTODOS	21
4.1. CARACTERIZAÇÃO DA ÁREA EXPERIMENTAL	21
4.1.1. <i>Área experimental 1</i>	21
4.1.2. <i>Área experimental 2</i>	21
4.1.3. <i>Delineamento experimental</i>	22
4.2. ESQUEMA DE AMOSTRAGEM DE SOLO	22
4.3. LEITURA DE CONDUTIVIDADE ELÉTRICA APARENTE DO SOLO (CEA).....	23
4.4. TRATAMENTO E ANÁLISE DOS DADOS	24
4.5. ÍNDICE DE VEGETAÇÃO	27
4.6. AVALIAÇÃO EXPERIMENTAL E AJUSTE DOS MODELOS GENÉTICO-ESTATÍSTICOS.....	28
5. RESULTADOS	31
5.1. ANÁLISE EXPLORATÓRIA DOS DADOS DA ÁREA EXPERIMENTAL 1	31
5.1.1. <i>Análise geoestatística do conjunto de dados</i>	34
5.1.2. <i>Condutividade elétrica aparente do solo (CEa)</i>	37
5.1.3. <i>Índice de vegetação</i>	39
5.1.4. <i>Ajuste dos modelos genético-estatísticos</i>	40
5.2. ANÁLISE EXPLORATÓRIA DOS DADOS DA ÁREA EXPERIMENTAL 2	44
5.2.1. <i>Análise geoestatística do conjunto de dados</i>	48
5.2.2. <i>Condutividade elétrica aparente do solo (CEa)</i>	51
5.2.3. <i>Avaliação experimental e ajuste dos modelos genético-estatísticos</i>	54
6. DISCUSSÃO	58
7. CONCLUSÕES.....	63
REFERÊNCIAS	64
APÊNDICES.....	70

RESUMO

Geotecnologias aplicadas em experimentação agrônômica: ênfase em cana-de-açúcar

Tecnologias que permitem remover fontes de variação sistemática do resíduo apresentam grande potencial de aplicação em experimentação agrônômica. Assim, este trabalho teve como objetivo avaliar de forma comparativa diferentes métodos e geotecnologias para ajuste da dependência espacial em experimentos de fase inicial do melhoramento genético da cana-de-açúcar. Para tanto, considerou-se dois experimentos localizados em regiões com diferentes classificações de solos. Buscando investigar uma possível estrutura de variabilidade espacial, ambas as áreas foram previamente georreferenciadas e as amostragens de solo realizadas em alta densidade, majoritariamente, em grade regular para análise dos elementos físicos e químicos. Uma cultura de rotação (*Crotalaria juncea* L.), antecedendo à implantação do experimento de fase inicial do melhoramento, foi considerada para o mapeamento, a priori, da variabilidade espacial da Área experimental 1 (experimento de uniformidade). Para isso, calculou-se o *NDVI* considerando os valores obtidos do comportamento espectral da cultura de rotação através da aquisição de imagens *RGB* e *NIR* a partir de uma aeronave remotamente pilotada - RPA. Para o mapeamento, a posteriori, realizaram-se leituras de condutividade elétrica aparente do solo (CE_a), em ambas as áreas experimentais, utilizando o sensor EM38-MK2®, que tem como princípio, a Indução Eletromagnética (IEM). Ademais, optou-se pela redução da dimensionalidade dos dados físicos e químicos do solo por meio da metodologia de análise de componentes principais (PCA). Com base nos coeficientes extraídos dos semivariogramas, obtidos pelo pacote estatístico *geoR* versão 1.8-1, determinou-se o índice de dependência espacial para cada PCA. Devido à alta densidade de pontos amostrados, utilizou-se o interpolador estatístico Inverso do Quadrado da Distância IDW. Obtidos os mapas temáticos provenientes das informações do PCA e dos diferentes sensores proximais, determinou-se a posição e a localização exata de cada parcela experimental. Para cada experimento, dados de TCH foram obtidos e analisados por meio de diferentes modelos lineares mistos, i.e., modelo independente (ID) e autorregressivo separável de primeira ordem (AR1 x AR1). Para ambos os modelos, utilizou-se a metodologia de seleção “*forward*” para incorporação das informações do PCA e dos diferentes sensores proximais. O uso do PCA, a partir de dados de amostragem de solo em alta densidade, permitiu capitalizar parte da variabilidade total dos dados para a Área experimental 1 e mostrou-se como um eficiente índice a ser incorporado como covariável no modelo genético-estatístico, possibilitando redução do erro experimental (CV_e). O uso do comportamento espectral da cultura de rotação permitiu a identificação, a priori, de diferentes classes de variabilidade na Área experimental 1. Porém, não foi capaz de remover as variações ambientais não controladas pelo delineamento experimental. Informações de CE_a mostraram-se capazes de incrementar a acurácia seletiva (AC_s) dos clones em experimentação quando incorporadas nos diferentes modelos genéticos-estatísticos, com destaque para o modelo AR1 x AR1. Portanto, este tipo de geotecnologia pode ser amplamente empregado em experimentação agrônômica e nas diversas áreas de estudos dentro do melhoramento genético de plantas, e.g., fases de experimentação, seleção genômica e estudos de interação genótipo x ambiente.

Palavras-chave: Sensoriamento proximal, Variabilidade espacial, Genética quantitativa, Geoestatística, *Envirotyping*, *Saccharum officinarum* L.

ABSTRACT

Geotechnology applied in agronomic field experiments: emphasis on sugarcane

Technologies that allow the removal of systematic source of variation of the residue have great potential for application in agronomic experimentation. Thus, this study aimed to evaluate comparatively different methods and geotechnologies for adjusting spatial dependence in early sugarcane breeding stages. Two experiments located in regions with different soil classifications were considered. In order to investigate a possible structure of spatial variability, both areas were previously georeferenced and soil samples were carried out in high density, mostly in a regular grid for analysis of physical and chemical properties. A rotation crop (*Crotalaria juncea* L.), before to the implementation of the early breeding stage experiment, was considered for the prior mapping of the spatial variability of Experimental Area 1 (uniformity trial). For this, the NDVI was calculated considering the values obtained from the spectral behavior of the rotation crop through the acquisition of RGB and NIR images from a remotely piloted aircraft - RPA. For the posterior mapping, measurements of apparent soil electrical conductivity (EC_a) were performed in both experimental areas, using the EM38-MK2® sensor, which uses the Electromagnetic Induction (EMI) technique. In addition, a reduction on the dimensionality of the physical and chemical soil dataset through the principal component analysis (PCA) methodology was performed. Based on the extracted semivariogram coefficients, obtained by the statistical package geoR version 1.8-1, the spatial dependency index for each PCA was determined. Due to the high density of sampled points, the Inverse Distance Weighting (IDW) was considered as the interpolation method. After obtained the PCA and the different proximal sensors thematic maps, the exact position and location of each experimental plot was determined. For each experiment, TCH data were obtained and analyzed using different mixed linear models, i.e., independent model (ID) and separable first-order autoregressive (AR1 x AR1). For both models, the “forward” selection methodology was used to incorporate information from the PCA and the different proximal sensors. The use of PCA based on high density soil sampling data allowed to capitalize part of the total variability of the data for Experimental Area 1 and proved to be an efficient index to be incorporated as a covariate in the statistical model, allowing reduction of experimental error (CV_e). The use of the spectral behavior of the rotation crop allowed the prior identification of different classes of variability in Experimental Area 1. However, it was not able to remove environmental variations not controlled by the experimental design. EC_a information was able to increase the selective accuracy (r) of the experimental clones when incorporated into the different statistical models, with emphasis on the AR1 x AR1 model. Therefore, this type of geotechnology can be widely used in agronomic experimentation and in the various study areas within plant breeding, e.g., experimentation phases, genomic selection and genotype x environment interaction studies.

Keywords: Proximal sensing, Spatial variability, Quantitative genetics, Geostatistics, Envirotyping, *Saccharum officinarum* L.

LISTA DE FIGURAS

Figura 1. Bandas utilizadas em sensoriamento remoto (MUNDO GEO, 2004).....	13
Figura 2. Exemplos de assinaturas espectrais da água (<i>water</i>), vegetação (<i>vegetation</i>) e do solo (<i>soil</i>), nos diferentes comprimentos de onda. Fonte: Siegmund e Menz (2005).	14
Figura 3. Semivariograma (a) modelo de tendência global; (b) não relacionada; e (c) modelo com tendência ajustada.	17
Figura 4. Semivariograma experimental, modelo teórico, mostrando os seguintes componentes: efeito pepita (C0), semivariância estrutural (C1), patamar ((C) = C0 + C1) e alcance (a) (OLIVEIRA, 2003).	18
Figura 5. Localização das duas áreas experimentais utilizadas para o estudo. Área experimental 1, localizada no município de Araras/SP, nas dependências da Universidade Federal de São Carlos (UFSCar), no Centro de Ciências Agrárias (CCA); região administrativa de Campinas. Área experimental 2, localizada no município de Valparaíso, no Noroeste Paulista, também nas dependências da Estação Experimental da UFSCar; região administrativa de Araçatuba.	22
Figura 6. (A) Área experimental 1, com 6,5 ha, localizada no município de Araras/SP, nas dependências do Centro de Ciências Agrárias (CCA), da Universidade Federal de São Carlos (UFSCar); (B) Área experimental 2, com 9,7 ha, localizada no município de Valparaíso/SP, nas dependências da estação experimental, também da UFSCar; no total, foram amostrados e georreferenciados 56 (A) e 68 pontos (B), respectivamente.	23
Figura 7. Equipamento utilizado para mensuração da condutividade elétrica aparente do solo, na orientação do dipolo horizontal (0,75 m e 0,375 m), área experimental 1 (A) e 2 (B); coletor de dados para registro das informações coletadas em campo (C).	24
Figura 8. Esquema da posição e localização de cada parcela experimental para extração do valor médio de todos os pixels, Área experimental 1. A mesma rotina foi empregada para a Área experimental 2.	26
Figura 9. Esquema do plano de voo na área experimental 1 para obtenção das imagens RGB e NIR, considerando velocidade de voo definida em 5 m/s, altitude de voo de 50 m e sobreposição longitudinal (A) e lateral (B) das imagens em 75% e 65%, respectivamente; pixel com resolução igual a 1,5 cm.	27
Figura 10. Esquema para análise de correlação entre os valores médios dos componentes principais (PCA) e/ou <i>NDVI</i> com os atributos do solo nos diferentes pontos amostrais; valores obtidos a partir do valor médio dos pixels dentro do raio de 5 m contornando cada ponto amostral.	28
Figura 11. (A) Gráfico <i>scree</i> dos autovalores dos componentes principais e (B) Gráfico bi-dimensional (<i>biplot</i>) dos componentes principais 1 e 2 (PC ₁ e PC ₂) das 17 variáveis físicas e químicas do solo, amostradas em 55 pontos, na profundidade de 0 a 20 cm, na Área experimental 1.	33
Figura 12. (A) Correlações entre as diferentes variáveis físicas e químicas do solo: areia grossa (areia_g); areia fina (areia_f); areia total (areia_t); silte; fósforo (P); matéria orgânica (mo); pH; potássio (K); cálcio (Ca); magnésio (Mg); acidez potencial (Ha); alumínio (Al); saturação por base (SB); capacidade de troca de cátions (CTC); percentual de saturação por base (V); porcentagem de saturação por alumínio (m), com os diferentes sensores: condutividade elétrica aparente do solo a 0,75 m (Eca_1m); condutividade elétrica aparente do solo a 0,375 m (Eca_05m); com os componentes principais: componente principal 1 (PC ₁); componente principal 2 (PC ₂); componente principal 3 (PC ₃); e o Índice de Vegetação da Diferença Normalizada (<i>NDVI</i>), combinado com o teste de significância considerando um nível de 0,05% (P<0.05). (B) Valores de correlação entre as diferentes variáveis estudadas. Informações referente a Área experimental 1.	34
Figura 13. (A) Mapa temático gerado pela Ponderação do Inverso das Distâncias (<i>Inverse Distance Weighting</i> - <i>IDW</i>) para o componente principal 1 (PC ₁), Área experimental 1. (B) Parcelas	

experimentais georreferenciadas. (C) Sobreposição das parcelas experimentais sobre o mapa interpolado e obtenção dos valores médios dos pontos de interesse.	36
Figura 14. (A) Mapa temático gerado pela Ponderação do Inverso das Distâncias (<i>Inverse Distance Weighting - IDW</i>) para o componente principal 2 (PC ₂), Área experimental 1. (B) Parcelas experimentais georreferenciadas. (C) Sobreposição das parcelas experimentais sobre o mapa interpolado e obtenção dos valores médios dos pontos de interesse.	37
Figura 15. (A) Mapa dos 17.694 pontos de leituras de condutividade elétrica aparente do solo (CEa) obtidos pelo sensor EM38-MK2® na camada de 0,375 m, Área experimental 1. (B) Parcelas experimentais georreferenciadas. (C) Sobreposição das parcelas experimentais sobre o mapa de CEa, 11.722 pontos de leitura e obtenção dos valores médios dos pontos de interesse.	38
Figura 16. (A) Mapa dos 17.694 pontos de leituras de condutividade elétrica aparente do solo (CEa) obtidos pelo sensor EM38-MK2® na camada de 0,75 m, Área experimental 1. (B) Parcelas experimentais georreferenciadas. (C) Sobreposição das parcelas experimentais sobre o mapa de CEa, 11.722 pontos de leitura e obtenção dos valores médios dos pontos de interesse.	38
Figura 17. (A) Ortomosaico da Área experimental 1 no formato RGB (banda <i>Red</i> , <i>Green</i> e <i>Blue</i>), obtido através do voo com a RPA. (B) Mapa gerado após o cálculo do índice de vegetação da diferença normalizada (<i>NDVI</i>). (C) Sobreposição das parcelas experimentais sobre o mapa <i>NDVI</i> . (D) Obtenção dos valores médios de <i>NDVI</i> dentro de cada parcela experimental.	40
Figura 18. (A) Valores dos dados brutos de toneladas de cana por hectare (TCH, t ha ⁻¹) de cada parcela experimental (família). (B) Visualização do padrão espacial dos valores de TCH; Área experimental 1.	41
Figura 19. (A) Gráfico <i>scree</i> dos autovalores dos componentes principais e (B) Gráfico bi-dimensional (<i>biplot</i>) dos componentes principais 1 e 2 (PC ₁ e PC ₂) das 17 variáveis físicas e químicas do solo, amostradas em 68 pontos, na profundidade de 0 a 20 cm, na Área experimental 2.	46
Figura 20. (A) Correlações entre as diferentes variáveis físicas e químicas do solo: areia grossa (areia_g); areia fina (areia_f); areia total (areia_t); silte; fósforo (P); matéria orgânica (mo); pH; potássio (K); cálcio (Ca); magnésio (Mg); acidez potencial (Hal); alumínio (Al); saturação por base (SB); capacidade de troca de cátions (CTC); percentual de saturação por base (V); porcentagem de saturação por alumínio (m), com os diferentes sensores: condutividade elétrica aparente do solo a 0,75 m (Eca_1m); condutividade elétrica aparente do solo a 0,375 m (Eca_05m); com os componentes principais: componente principal 1 (PC ₁); componente principal 2 (PC ₂); componente principal 3 (PC ₃), combinado com o teste de significância considerando um nível de 0,05% (P<0.05). (B) Valores de correlação entre as diferentes variáveis estudadas. Informações referente a Área experimental 2.	47
Figura 21. (A) Mapa temático gerado pela Ponderação do Inverso das Distâncias (<i>Inverse Distance Weighting - IDW</i>) para o componente principal 1 (PC ₁), Área experimental 2. (B) Parcelas experimentais georreferenciadas. (C) Sobreposição das parcelas experimentais sobre o mapa interpolado.	50
Figura 22. (A) Mapa temático gerado pela Ponderação do Inverso das Distâncias (<i>Inverse Distance Weighting - IDW</i>) para o componente principal 2 (PC ₂), Área experimental 2. (B) Parcelas experimentais georreferenciadas. (C) Sobreposição das parcelas experimentais sobre o mapa interpolado.	51
Figura 23. (A) Mapa dos 25.457 pontos de leituras de condutividade elétrica aparente do solo (CEa) obtidos pelo sensor EM38-MK2® na camada de 0,375 m, Área experimental 2. (B) Parcelas experimentais georreferenciadas (761). (C) Sobreposição das parcelas experimentais sobre o mapa de CEa, 16.617 pontos de leitura.	53
Figura 24. (A) Mapa dos 25.457 pontos de leituras de condutividade elétrica aparente do solo (CEa) obtidos pelo sensor EM38-MK2® na camada de 0,75 m, Área experimental 2. (B) Parcelas experimentais georreferenciadas (761). (C) Sobreposição das parcelas experimentais sobre o mapa de CEa, 16.617 pontos de leitura.	54

Figura 25. (A) Valores dos dados brutos de toneladas de cana por hectare (TCH, tons/ha) de cada parcela experimental (família). (B) Visualização do padrão espacial dos valores de TCH; Área experimental 2.	55
Figura 26. <i>Boxplot</i> dos valores dos diferentes atributos físicos (Areia fina, grossa, total, argila e silte – g kg ⁻¹ ; M.O. – g dm ⁻³) e químicos (Al, Ca, CTC, H+Al, K, Mg, SB - mmol _c dm ⁻³ ; Presina - mg dm ⁻³ ; V e m - %) do solo obtidos a partir de 56 pontos amostrais, na profundidade de 0 a 20 cm, na Área experimental 1. Identificação de <i>outliers</i>	71
Figura 27. <i>Boxplot</i> dos valores dos diferentes atributos físicos (Areia fina, grossa, total, argila e silte – g kg ⁻¹ ; M.O. - g dm ⁻³) e químicos (Al, Ca, CTC, H+Al, K, Mg, SB - mmol _c dm ⁻³ ; Presina – mg dm ⁻³ ; V e m - %) do solo obtidos a partir de 55 pontos amostrais (<i>outliers</i> removidos), na profundidade de 0 a 20 cm, Área experimental 1.....	72
Figura 28. <i>Boxplot</i> dos valores dos diferentes atributos físicos (Areia fina, grossa, total, argila e silte – g kg ⁻¹ ; M.O. - g dm ⁻³) e químicos (Al, Ca, CTC, H+Al, K, Mg, SB - mmol _c dm ⁻³ ; Presina - mg dm ⁻³ ; V e m - %) do solo obtidos a partir de 68 pontos amostrais (não identificado presença de <i>outliers</i>), na profundidade de 0 a 20 cm, Área experimental 2.	73
Figura 29. Variograma experimental para o componente principal 1 (PC ₁), considerando: modelo esférico (sph), exponencial (exp), gaussiano (gaus) e plotagem conjunta, Área experimental 1....	74
Figura 30. Variograma experimental para o componente principal 2 (PC ₂), considerando: modelo esférico (sph), exponencial (exp), gaussiano (gaus) e plotagem conjunta, Área experimental 1....	75
Figura 31. Variograma experimental para o componente principal 1 (PC ₁), considerando: modelo esférico (sph), exponencial (exp), gaussiano (gaus) e plotagem conjunta, Área experimental 2....	76
Figura 32. Variograma experimental para o componente principal 2 (PC ₂), considerando: modelo esférico (sph), exponencial (exp), gaussiano (gaus) e plotagem conjunta, Área experimental 2....	77

LISTA DE TABELAS

Tabela 1. Estatística descritiva dos atributos físicos e químicos do solo na camada de 0 a 20 cm de profundidade, Área experimental 1.....	32
Tabela 2. Parâmetros dos modelos teóricos ajustados aos variogramas experimentais do conjunto de dados na profundidade de 0 a 20 cm, Área experimental 1.....	35
Tabela 3. Seleção de covariáveis para o modelo identidade (ID) e autorregressivo de primeira ordem (AR1 x AR1) utilizando o método “ <i>forward</i> ”, através da inclusão das diferentes informações de solo e dos sensores de forma individualizada. <i>NDVI</i> : Índice de Vegetação da Diferença Normalizada; <i>Eca05</i> : condutividade elétrica aparente do solo a 0,375 m; <i>Eca1</i> : condutividade elétrica aparente do solo a 0,75 m; <i>PC₁</i> : componente principal 1; <i>PC₂</i> : componente principal 2. Informações referentes a Área experimental 1.....	42
Tabela 4. Inclusão de novas variáveis para o modelo identidade (ID) e autorregressivo de primeira ordem (AR1 x AR1) já considerando a variável <i>PC₁</i> ; <i>NDVI</i> : Índice de Vegetação da Diferença Normalizada; <i>Eca05</i> : condutividade elétrica aparente do solo a 0,375 m; <i>Eca1</i> : condutividade elétrica aparente do solo a 0,75 m; <i>PC₁</i> : componente principal 1; <i>PC₂</i> : componente principal 2. Informações referentes a Área experimental 1.	43
Tabela 5. Estimativas dos componentes de variância (REML) considerando os modelos: identidade (ID); identidade, incluindo a covariável significativa, componente principal 1 (ID+ <i>PC₁</i>); autorregressivo de ordem 1 (AR1 x AR1); autorregressivo de ordem 1, incluindo a covariável significativa, componente principal 1 (AR1 x AR1+ <i>PC₁</i>). O caráter principal considerado foi toneladas de cana por hectare (TCH).....	44
Tabela 6. Estatística descritiva dos atributos físicos e químicos do solo na camada de 0 a 20 cm de profundidade, Área experimental 2.....	45
Tabela 7. Parâmetros dos modelos teóricos ajustados aos variogramas experimentais do conjunto de dados na profundidade de 0 a 20 cm, Área experimental 2.....	49
Tabela 8. Seleção de covariáveis para o modelo identidade (ID) e autorregressivo de primeira ordem (AR1 x AR1) utilizando o método “ <i>forward</i> ”, através da inclusão das diferentes informações de solo e dos sensores de forma individualizada. <i>Eca05</i> : condutividade elétrica aparente do solo a 0,375 m; <i>Eca1</i> : condutividade elétrica aparente do solo a 0,75 m; <i>PC₁</i> : componente principal 1; <i>PC₂</i> : componente principal 2. Informações referentes a Área experimental 2.	56
Tabela 9. Inclusão de novas variáveis para o modelo identidade (ID) e autorregressivo de primeira ordem (AR1 x AR1) já considerando a variável <i>Eca05</i> (condutividade elétrica aparente do solo a 0,375 m); <i>Eca1</i> : condutividade elétrica aparente do solo a 0,75 m; <i>PC₁</i> : componente principal 1; <i>PC₂</i> : componente principal 2. Informações referentes à Área experimental 2.....	56
Tabela 10. Estimativas dos componentes de variância (REML) considerando os modelos: identidade (ID); identidade, incluindo as covariáveis significativas, condutividade elétrica aparente do solo a 0,375 m e 0,75 m (ID+ <i>Eca05</i> e <i>Eca1</i>); autorregressivo de ordem 1 (AR1); autorregressivo de ordem 1, incluindo as covariáveis significativas, condutividade elétrica aparente do solo a 0,375 m e 0,75 m (AR1+ <i>Eca05</i> e <i>Eca1</i>). O caráter principal considerado foi toneladas de cana por hectare (TCH).	57
Tabela 11. Diferentes modelos para as estruturas de variância-covariância (VCOV) da matriz residual (R).....	70
Tabela 12. Valores de AIC (<i>Akaike Information Criterion</i>) e BIC (<i>Bayesian Information Criterion</i>) considerando diferentes estruturas de VCOV, com a inclusão de covariáveis significativas no modelo, para o caráter toneladas de cana por hectare (TCH). Valores obtidos para a Área experimental 1 e 2.	78

1. INTRODUÇÃO

Estima-se que em 2050 a população mundial estará na ordem de 9,7 bilhões e, até o final do século, serão mais de 10 bilhões, ritmo de crescimento demográfico mundial acima dos 70 milhões de pessoas por ano (FAO, 2019). Para alimentar esta crescente população mundial, a produção total das culturas agrícolas precisará ser significativamente aumentada, com menos terra arável e sob condições ambientais muito severas (TILMAN et al., 2002). A melhor solução para atingir a meta parece depender do aumento do rendimento das culturas agrícolas, em vez de expandir as superfícies cultivadas. Segundo Fan et al. (2012), melhorar o potencial de rendimento das diversas culturas agrícolas através do melhoramento de plantas será um componente crítico para o futuro da segurança alimentar.

Para isso, modernas plataformas e técnicas de melhoramento, juntamente com melhores práticas agronômicas, incluindo melhor fenotipagem em campo, melhor compreensão da arquitetura genética dos caracteres, são essenciais (XU et al., 2017).

Atualmente, os componentes genéticos dos genótipos podem ser estudados e avaliados através de abordagens moleculares, utilizando técnicas de sequenciamento genético (genômica), microarranjo de DNA, entre outras. Adicionalmente, o componente fenotípico também pode ser medido precisamente com o desenvolvimento de ferramentas e metodologias de fenotipagem de alto rendimento (*high-throughput*).

Entretanto, comparado com a genotipagem e a fenotipagem, a determinação e a medição de fatores ambientais estão em um segundo plano de prioridade, devido a três razões principais. Primeiro, os fatores ambientais têm sido amplamente considerados como um todo e tratados como uma caixa-preta que interage com os genótipos para afetar o crescimento e a produtividade das plantas. Em segundo lugar, apenas os principais fatores ambientais têm sido considerados e medidos em ensaios. Em terceiro lugar, a maioria dos fatores ambientais é dinâmica e muda constantemente ao longo do período de crescimento da planta (XU, 2016).

Ainda que as técnicas de melhoramento genético de plantas atualmente empregadas sejam promissoras, nos métodos tradicionais de melhoramento, experimentos de campo são essenciais no processo de seleção e recomendação de cultivares melhoradas. Nesses experimentos é desejável um alto grau de precisão experimental e, conseqüentemente, uma alta acurácia na inferência sobre as médias genotípicas, isto é, sobre os valores genotípicos dos tratamentos genéticos em avaliação (RESENDE; DUARTE, 2007). Portanto, a consideração dos fatores ambientais é de suma importância para o sucesso do melhoramento.

Normalmente, pesquisadores utilizam técnicas estatísticas para reduzir possíveis variações naturais dentro de campos experimentais (PIMENTEL-GOMES, 1984). Isso permite que os tratamentos sejam comparados sem os efeitos de confundimento das diferenças de fatores ambientais, como o tipo de solo. No entanto, muitas vezes é difícil determinar a localidade mais apropriada dos blocos experimentais dentro de um experimento quando a variação natural no local é desconhecida ou difícil de mensurar. Isto é particularmente um problema em programas de melhoramento de cana-de-açúcar, onde muitas vezes são necessárias grandes áreas experimentais para avaliar o desempenho de centenas de clones, principalmente, em fases iniciais, em que, além do elevado número de indivíduos, há restrição de material vegetativo, impossibilitando o emprego de alguns princípios básicos de experimentação, isto é, a repetição.

Segundo Wei et al. (2015), a suposição de homogeneidade do local dentro de uma repetição ou bloco pode nem sempre ser válida. Esta violação da suposição pode levar a uma seleção ineficiente dentro dos programas de melhoramento e, por conseguinte, reduzir o progresso genético.

Neste contexto, há necessidade de se utilizar técnicas que contabilizem os efeitos ambientais, não somente visando a racionalização do uso de insumos, em áreas de produção comercial, mas também objetivando a redução do erro experimental, tanto em parcelas como em experimentação em escala de campo.

Segundo Adamchuk et al. (2004), a necessidade de caracterização espacial de fatores tanto de planta como de solo tem levado ao surgimento de uma série de propostas e sistemas de sensoriamento e monitoramento; por exemplo, na tentativa de desenvolver técnicas de medida indireta de propriedades do solo, têm sido estudados sistemas ópticos, eletromagnéticos, eletroquímicos, mecânicos, fluxo de ar e acústicos. Esta capacidade de identificar variações em campo seria de grande utilidade em experimentos de seleção de cultivares, visto que a diferença identificada pelos diferentes sensores remotos poderia ser utilizada como um método de ajuste durante o processo de análise de dados do potencial genético de cada clone em experimentação.

Como o tema ainda carece de maiores informações, este trabalho tem por objetivo avaliar a acurácia na inferência sobre as médias genotípicas ajustadas de clones em fases iniciais do melhoramento da cana-de-açúcar, por meio da adoção de informações de variabilidade espacial obtidas por dados de amostragem de solo, em alta densidade, e por diferentes sensores proximais.

2. REVISÃO BIBLIOGRÁFICA

2.1. Melhoramento genético da cana-de-açúcar

Segundo Sleper e Poehlman (2006), com o melhoramento genético de plantas, torna-se possível aumentar o desempenho dos descendentes por meio de diversas maneiras. O aumento da produtividade e a melhoria na qualidade são frequentemente os principais objetivos buscados. Para isso, buscam-se resistência de plantas a doenças e pragas, a adaptação a diferentes condições ambientais tais como áreas com altas temperaturas, sujeitas a condições de seca e de salinidade, entre outros riscos ambientais para produção.

Atualmente, as cultivares modernas de cana-de-açúcar correspondem a híbridos de duas espécies geneticamente complexas, derivados da hibridação interespecífica entre *S. officinarum* e *S. spontaneum* (DANIELS; ROACH, 1987). Dentre os genótipos cultivados acredita-se que haja proporções variáveis do genoma dessas espécies e, ainda, cromossomos recombinantes (D'HONT et al., 1996).

Apesar da complexidade genômica, a obtenção de novos genótipos é relativamente facilitada devido à existência de mecanismos de propagação vegetativa, ou seja, os genótipos são reproduzidos integralmente através de gerações (BRESSIANI, 2011; RAM, 2014). Esta particularidade torna o melhoramento mais eficiente, pois faz com que todos os efeitos genéticos sejam aproveitados. Portanto, após a geração de variabilidade, por meio de processos de hibridação, basta proceder a seleção e clonagem dos indivíduos superiores durante as diferentes etapas de experimentação em campo.

Para que os experimentos forneçam informações com qualidade é importante que sejam planejados e conduzidos de modo que o erro experimental, também chamado de variação aleatória, seja minimizado (STORCK et al., 2000). Segundo Duarte (1996), o erro experimental pode ser decorrente, principalmente, da heterogeneidade ambiental não prevista pelo experimentador, como pequenas variações na fertilidade do solo de uma parcela para outra, na umidade, na profundidade de semeadura, no teor de germinação das sementes, entre outros. Elas provem também de variações inerentes ao material experimental (variabilidade dentro dos tratamentos), ou mesmo da aplicação desuniforme dos tratamentos às parcelas. Logo, se dentro de um mesmo tratamento ocorrer variação de grande magnitude, não se poderá inferir com confiança, sobre a diferença entre tratamentos.

Assim, cabe ao experimentador conviver com esse tipo de variação, buscando minimizá-la, uma vez que ela estará sempre presente. Para isso, diversos procedimentos podem ser adotados, tais como: escolha de ambientes homogêneos para condução dos experimentos; uso de bordaduras (de experimento e parcelas); aplicação uniforme dos tratamentos às parcelas; seleção criteriosa do material experimental; condução uniforme das parcelas; coleta de dados criteriosa e imparcial, com equipamentos de boa precisão; uso de observações auxiliares ou concomitantes, associado à análise de covariância, entre outros (DUARTE, 1996).

Tais recomendações, no entanto, terão pouco efeito na redução do erro experimental, caso o experimento não tenha sido concebido sob um plano experimental adequado. Para isso, a análise de experimentos deve ser regida por três princípios básicos: a repetição, que estima o erro experimental; a aleatorização, que visa proporcionar uma estimativa válida para o erro experimental; e o controle local, que tende a tornar o experimento mais eficiente, pela redução do erro experimental (BANZATTO; KRONKA, 2006).

Dessa forma, espera-se obter alto grau de precisão experimental e, conseqüentemente, uma alta acurácia na inferência sobre as médias genotípicas, isto é, sobre os valores genotípicos dos tratamentos genéticos em avaliação (RESENDE; DUARTE, 2007). No entanto, as fases iniciais do melhoramento da cana-de-açúcar são caracterizadas

pela limitação de material vegetativo, pequenas parcelas, sem repetição. Segundo Stringer (2006), tais experimentos estão sujeitos a variações decorrentes da variabilidade espacial e devido a competição entre parcelas, que podem comprometer seriamente as estimativas dos genótipos e, portanto, reduzir o progresso genético.

Buscando solucionar o problema de testar um grande número de clones de cana-de-açúcar sob escassez de material de propagação, nas fases preliminares do programa de melhoramento da estação experimental *HSPA* (*Hawaiian Sugar Planter's Association*) no Havaí, EUA, Federer (1956) propôs que se escolhesse um delineamento padrão para as cultivares testemunhas e aumentasse o número de parcelas de seus blocos (linhas e/ou colunas), para acomodar os novos clones (tratamentos adicionais ou novos). Esta nova proposta, conhecida como delineamentos aumentados, passou a ser empregada em programas de melhoramento genético vegetal, envolvendo diversas espécies.

Entretanto, alguns autores afirmam que o delineamento não se mostra apropriado à estimação de parâmetros genéticos e fenotípicos devido à baixa precisão das estimativas (SOUZA, 1997). Segundo Gomes (1995), o delineamento não apresenta eficiência para estimar coeficientes de correlações entre caracteres, com base nos tratamentos adicionais. Algumas destas limitações, todavia, revelam-se por dificuldades operacionais na realização de uma análise estatística mais elaborada, seja pelo desbalanceamento inerente desses delineamentos, seja pela complexidade dos modelos adotados, em geral, mistos.

Ainda, segundo este mesmo autor, no que concerne à seleção de genótipos em fases preliminares dos programas de melhoramento, via delineamentos aumentados, outro aspecto preocupante é o pequeno tamanho das parcelas. Nessas fases é comum a adoção de parcelas pequenas, devido a limitação de material de propagação, como mencionado anteriormente. Também é usual a alocação sistemática dos cultivares testemunhas ou mesmo de grupos dos tratamentos adicionais com alguma procedência comum. Tudo isso gera preocupações acerca da validade da adoção de uma análise clássica de variância.

Outras questões como o tamanho ideal de blocos, a proporção das parcelas do bloco com cultivares testemunhas e o número de testemunhas, também não estão ainda totalmente definidos nesse tipo de experimentação. Neste sentido, considerando a adoção usual de blocos relativamente grandes, é provável que em boa parte dos ensaios os blocos não sejam suficientemente homogêneos. Esta característica, associada à aplicação de tratamentos não repetidos e à pequena discrepância entre genótipos nos programas de seleção, também deve contribuir para uma análise estatística de baixa eficiência.

2.2. Ganho genético no melhoramento

Segundo Xu et al. (2017), melhorias adicionais para obtenção de maiores ganhos genético são necessárias para atender a demanda futura por produtos agrícolas e, para isso, modernas plataformas e técnicas de melhoramento, juntamente com melhores práticas agronômicas, incluindo melhor fenotipagem em campo, melhor compreensão da arquitetura genética dos caracteres, são essenciais.

Atualmente, os componentes genéticos dos genótipos podem ser estudados e avaliados através de abordagens moleculares, utilizando técnicas de sequenciamento genético (genômica), microarranjo de DNA, entre outras. Por outro lado, o componente fenotípico também pode ser medido precisamente com o desenvolvimento de ferramentas e metodologias de fenotipagem de alto rendimento (*high-throughput*).

Comparado com a genotipagem e a fenotipagem, a determinação e a medição de fatores ambientais estão em um segundo plano de prioridade, devido a três razões principais. Primeiro, os fatores ambientais têm sido amplamente considerados como um todo e tratados como uma caixa-preta que interage com os genótipos para afetar o crescimento e a produtividade das plantas. Em segundo lugar, apenas os principais fatores ambientais têm sido

considerados e medidos em ensaio. Em terceiro lugar, a maioria dos fatores ambientais é dinâmica e muda constantemente ao longo do período de crescimento da planta (XU, 2016).

Compreender melhor o ambiente onde os genótipos são cultivados é fundamental para o sucesso do melhoramento, pois a expressão gênica é amplamente dependente do ambiente onde a cultura se desenvolve. O mapeamento genético e a clonagem dependem do ambiente em que a fenotipagem é realizada. Ademais, muitos procedimentos de fenotipagem, incluindo avaliação de estresse abiótico e biótico, são conduzidos sob condições ambientais controladas. Por fim, a consideração precisa dos fatores ambientais oferece uma nova oportunidade para a gestão, controle e otimização de fatores ambientais para o aumento do ganho genético (XU, 2016).

Segundo Falconer e Mackay (1996), eficiência do melhoramento de plantas, ou progresso genético, tem sido descrita pelo conceito de ganho genético (ΔG), e pode ser mensurada pelo diferencial de seleção, ou seja, a diferença entre a média da população selecionada e a média da população original.

O ganho genético esperado por ciclo de melhoramento pode ser definido como:

$$\Delta G = i r \sigma_A / t$$

em que ΔG é a resposta à seleção, i é a intensidade de seleção (desvio médio da seleção de indivíduos selecionados em unidades de desvio padrão fenotípico), r é a acurácia de seleção, σ_A é o desvio padrão dos valores genéticos ou desvio padrão genético (a raiz quadrada da variância genética aditiva), e t é o tempo do ciclo de melhoramento (FALCONER; MACKAY, 1996).

No contexto da seleção fenotípica, r é igual à raiz quadrada da herdabilidade no sentido restrito (h^2) e, portanto:

$$\Delta G = i h \sigma_A / t$$

Em genética quantitativa, a herdabilidade refere-se à proporção de variação genética (σ_g^2) na variação fenotípica (σ_f^2), sendo esta última composta por variações genéticas (σ_g^2) e ambientais (σ_e^2). Em outras palavras, quanto maior a estimativa de herdabilidade que uma característica agrônômica pode obter, com base em dados experimentais de campo em um determinado ambiente, melhor será o desempenho preditivo. Portanto, o manejo do solo da forma correta reduzirá os efeitos ambientais e os erros experimentais e, assim, melhorará a estimativa de herdabilidade e a acurácia das predições.

A variação ambiental é amplamente afetada por fatores abióticos, como instabilidade do microclima, fertilidade do solo, ventos e tempestades, e fatores bióticos como patógenos, insetos, ervas daninhas e plantas e animais indesejados ao redor das plantas cultivadas. No entanto, diversas medidas podem ser tomadas buscando manejar o ambiente de maneira eficiente, incluindo gerenciamento consistente da cultura, utilização de genótipos experimentais uniformes, seleção correta das testemunhas, bom controle dos efeitos de bordadura, escolha correta do delineamento experimental, e estabelecimento de rede de sensores para avaliação e medição das condições físico-química do solo (ARAUS; CAIRNS, 2014; KLUKAS et al., 2014).

2.3. Agricultura de precisão

A origem do termo Agricultura de Precisão (AP) está fundamentada no fato de que as lavouras não são uniformes no espaço nem no tempo (MOLIN et al., 2015). De acordo com a Comissão Brasileira de Agricultura de Precisão do Ministério da Agricultura, a AP é um conjunto de ferramentas e tecnologias aplicadas para permitir um sistema de gerenciamento agrícola baseado na variabilidade espacial e temporal da unidade produtiva, visando o aumento de retorno econômico e a redução do impacto ambiental. Para a gestão da heterogeneidade existente nas

lavouras, as seguintes etapas são consideradas: coleta de dados, por meio de amostragem de solo ou por meio de sensoriamento remoto e proximal; processamento de dados, utilizando *softwares* e *hardwares* para o processamento de imagens e para a análise e filtragem de dados; tomada de decisão, através da elaboração de mapas de prescrição e interpretação dos resultados; aplicação, através da definição da forma de ação para gestão da variabilidade (PUSCH et al., 2019).

Portanto, a integração de sensoriamento remoto com sistemas de navegação que permitem a obtenção de informações sobre as coordenadas geográficas de qualquer ponto da superfície terrestre (*GPS – Global Position System*), juntamente com os *SIG*, *softwares* compostos de vários módulos dedicados ao armazenamento e ao processamento de dados com coordenadas conhecidas (geoprocessamento), sem dúvida constitui um eficiente método de análise e avaliação de forma espacializada das condições dos solos e das culturas agrícolas.

2.3.1. Sensoriamento remoto

A necessidade de caracterização espacial de fatores de planta e de solo tem levado ao surgimento de uma série de propostas e sistemas de sensoriamento e monitoramento (MOLIN; RABELLO, 2011); por exemplo, na tentativa de desenvolver técnicas de medida indireta de propriedades do solo, têm sido estudados sistemas ópticos, eletromagnéticos, eletroquímicos, mecânicos, fluxo de ar e acústicos (ADAMCHUK et al., 2004).

O sensoriamento remoto, que tem como definição a aquisição de dados sobre um objeto sem tocá-lo fisicamente, por meio de radiação eletromagnética, pode contribuir de modo significativo no fornecimento de dados oportunos e precisos na área da agricultura (ELACHI; VAN-ZYL, 2006). De forma geral, o sensoriamento remoto pode ser realizado de duas formas: passivo, onde a radiação não é controlada pelo sistema de sensor (depende da radiação solar), e ativo, onde a radiação é parte do sistema de sensoriamento (ex., microondas do radar).

Os sensores remotos são dispositivos capazes de detectar, em determinadas faixas do espectro eletromagnético, a energia eletromagnética proveniente de um objeto, transformá-la em um sinal elétrico e registrá-la, de tal forma que esse dado possa ser armazenado ou transmitido em tempo real, para posteriormente ser convertido em informações que descrevam as feições dos objetos que compõem a superfície terrestre (MORAES, 2015).

É conhecido o fato de que toda matéria reflete, absorve, transmite e/ou emite radiação eletromagnética de forma específica, conforme suas características próprias. As regiões do espectro eletromagnético têm diferentes nomenclaturas, abrangendo desde os raios gama, passando pelos raios X, ultravioleta (UV), luz visível (V) e infravermelho (IV), e indo até as ondas de rádio, desde os menores até os maiores comprimentos de onda dessas radiações (FORMAGGIO; SANCHES, 2017).

Os sensores remotos possuem a capacidade de registrar a radiação eletromagnética em um amplo intervalo de comprimento de ondas, sendo os de maiores interesse: o espectro óptico (400 nm a 2.500 nm); o termal (considerando as emissões das superfícies entre 8.000 nm e 14.000 nm); e as micro-ondas (radar), baseadas na radiação eletromagnética entre 1 nm e 1 m (Figura 1).

Figura 1. Bandas utilizadas em sensoriamento remoto (MUNDO GEO, 2004).

Cada tipo de material da superfície terrestre apresenta um comportamento espectral específico quanto às quantidades de energia refletida, absorvida ou transmitida, em função de suas particularidades e das condições em que se encontra no momento em que é sensoriada. Dessa forma, diferentes tipos de superfícies absorvem ou refletem a radiação de forma distinta nos diferentes comprimentos de onda, sendo este comportamento denominado de assinatura espectral da característica da composição de um determinado material. A Figura 2 ilustra as diferentes assinaturas espectrais de alguns dos materiais mais comumente conhecidos, água, vegetação e solo, apresentando diferentes valores de reflectância nos variados comprimentos de onda. Segundo Kumar (2018), a água na forma líquida apresenta alta refletância na região do visível entre 400 nm e 600 nm. Comprimentos de onda superiores a 700 nm são completamente absorvidos. Por outro lado, as vegetações saudáveis são muito boas absorvedoras de energia eletromagnética na região do visível. A absorção reduz e aumenta a reflexão no limite do vermelho/infravermelho próximo a 700 nm. No comprimento de onda em 650 nm, a absorção é atribuída ao pigmento clorofila. Em solos, a refletância é quase constante de 700 a 1.300 nm e depois diminui para os comprimentos de onda mais longos. Porém, alguns fatores afetam a sua refletância, como o teor de umidade, textura do solo (proporção de areia, silte e argila), rugosidade superficial, presença de óxido de ferro e teor de matéria orgânica. Esses fatores são complexos, variáveis e inter-relacionados.

Figura 2. Exemplos de assinaturas espectrais da água (*water*), vegetação (*vegetation*) e do solo (*soil*), nos diferentes comprimentos de onda. Fonte: Siegmund e Menz (2005).

Dentre as diversas plataformas de sensoriamento proximal disponíveis atualmente, destacam-se os sensores proximais de alta resolução (EM38-MK2® – sensor de indução eletromagnética no solo), e os sensores aéreos multiespectrais (RPA, drones); plataformas que serão abordadas a seguir.

2.3.2. Sensor de condutividade elétrica aparente do solo (CEa)

Sensores de condutividade elétrica do solo (CE) têm tido destaque como indicador no monitoramento de características do solo, como salinidade, textura, umidade, densidade, matéria orgânica e derivados, CTC, lixiviação, partição de doses de herbicidas, definição de bordas em classificação de solos, classes de drenagem, recarga de lençol freático, entre outras (CORWIN; LESCH, 2005). A condutividade elétrica é uma propriedade específica de cada substância. O solo é um meio composto por diversos materiais, cada um possuindo condutividade própria. A condução de corrente elétrica através do solo se deve à presença de sais livres na solução do solo e íons trocáveis na superfície das partículas (FRIEDMAN, 2005). A utilidade da corrente elétrica no meio agrícola, provém do fato que os componentes físicos existentes no solo apresentam diferentes níveis de condutividade elétrica (LUND et al., 1998). A corrente elétrica pode ser conduzida através da água intersticial que contém eletrólitos dissolvidos e através dos cátions trocáveis que residem perto da superfície de partículas de solo carregadas e são eletricamente móveis em vários níveis (NADLER; FRENKEL, 1980). Assim, solos com baixo teor de umidade apresentam resistência elétrica muito alta. Alguns minerais presentes aparecem como isolantes, apesar de que em alguns solos pode existir uma pequena corrente sendo conduzida através da superfície das partículas. Portanto, o valor obtido para a condutividade elétrica de um solo é principalmente devido ao seu teor de água e de sais dissolvidos (FREELAND, 1989).

Segundo Luck e Eisenreich (2001), o mapeamento da condutividade elétrica é uma ferramenta eficiente na investigação do comportamento e da variabilidade espacial de propriedades do solo. Ela permite identificar áreas com propriedades semelhantes e delimitar unidades de gestão diferenciada.

Diferentes métodos podem ser empregados para obter medidas da CEa, sendo que o método resistivo e o indutivo apresentam maior expressão em função da simplicidade e praticidade, respectivamente. Além disso, a mensuração da CEa é relativamente de baixo custo, portanto, a ocorrência de relacionamento desta com outras

propriedades de interesse agrônomo de difícil mensuração é desejável, sobretudo pela possibilidade de caracterizar a variabilidade espacial de uma determinada área (GIMENEZ, 2013). Segundo Johnson (2005), por sua capacidade em identificar a variabilidade de importantes características físicas e químicas do solo, a CEa pode ser empregada para subdividir áreas onde se desejam implantar experimentos, reduzindo possíveis erros, e permitindo ainda obter maior segurança na interpretação dos dados de experimentos em áreas maiores em que não é possível realizar repetições. Esta capacidade de identificar variações em campo seria de grande utilidade em experimentos de seleção de cultivares, visto que a diferença identificada pela CEa poderia ser utilizada como um método de ajuste durante o processo de análise de dados do potencial genético de cada tratamento em experimentação.

2.3.3. Sensores aéreos multiespectrais (RPA, drones)

O uso de aeronaves remotamente pilotadas (RPA) surgiu como uma importante opção para a agricultura. Sua aplicação vem sendo favorecida e facilitada pelo atual estágio de desenvolvimento tecnológico, principalmente pela redução do custo e do tamanho dos equipamentos e pela necessidade de otimização da produção (JORGE; INAMASU, 2014). Na área agrícola, o foco tem sido principalmente na utilização de sensores baseados na espectroscopia de reflectância, ou seja, em medidas da reflexão da radiação eletromagnética em diferentes comprimentos de onda. Nesse tipo de equipamento, a radiação eletromagnética refletida pela planta é coletada por sensores especiais, sensíveis a determinados comprimentos de onda e colocados em plataformas que podem estar localizados em diferentes níveis de aquisição de dados: solo, baixas altitudes (RPA, aviões, helicópteros), e altas altitudes (satélites). A energia captada contém informações das culturas quanto ao vigor, quantidade de matéria verde, quantidade de água, ataques de pragas, entre outras (LAMPARELLI et al., 2001).

Nos últimos anos, um número significativo de índices de vegetação tem sido utilizado, a grande maioria considerando informações de reflectância espectral da vegetação nas regiões do visível e do infravermelho próximo. Essas regiões são predominantemente afetadas pela absorção da clorofila na região do visível e pela dispersão da radiação do infravermelho próximo causada pela estrutura celular da vegetação (KALAITZIDIS et al., 2010).

Índices de vegetação, tal como o índice de vegetação por diferença normalizada (*NDVI*), tem sido utilizado com sucesso para o mapeamento a priori da variabilidade espacial de lavouras, através de medições dos valores de reflectância da planta nos diferentes comprimentos de onda do infravermelho próximo (*Near Infrared - NIR*). Estes métodos de quantificação da variabilidade espacial, a priori, baseiam-se na identificação de tendências existentes dentro de um campo, os quais, subsequentemente, podem ser considerados e corrigidos por meio da gestão localizada do fator causal e, no caso de áreas experimentais, na adoção de delineamentos experimentais apropriados.

2.4. Geoestatística

A geoestatística é um conjunto de técnicas que incluem análise exploratória, análise estrutural de correlação espacial (modelagem de semivariogramas) e validação do modelo e interpolação estatística da superfície (MOLIN et al., 2015). Portanto, tem por objetivo a caracterização espacial de uma determinada variável de interesse por meio do estudo de sua distribuição e variabilidade espacial, com determinação das incertezas associadas (YAMAMOTO; LANDIM, 2013).

O francês Georges Matheron em 1963 incorporou o conceito de estatística clássica ao conceito de variáveis regionalizadas. Este termo foi determinado visando alertar para dois aspectos, aleatório e estruturado, dos fenômenos sobre os quais se ocupa. É aleatório no sentido de que os valores das medições que fazemos podem variar

consideravelmente entre si, e sua característica regionalizada, estruturada segundo uma certa lei no espaço, é evidente se considerarmos que os valores das amostras com que trabalhamos não são completamente independentes da sua localização geográfica. Assim, próximo a um valor elevado é mais provável que encontremos outro valor elevado do que um muito baixo; segundo a Teoria de Probabilidade, os valores destas duas amostras próximas estão correlacionados (ANDRIOTTI, 1988).

Neste sentido, buscando contornar problemas inerentes às fases iniciais do melhoramento, Kempton (1984) apresentou uma abordagem clássica para lidar com as tendências de fertilidade do solo em ensaios sem repetições. Segundo este autor, a inclusão de parcelas como testemunhas, distribuídas ao longo do experimento, em diversas repetições, geraria um índice de fertilidade que poderia ser utilizado como referência para avaliar a produtividade das parcelas do ensaio. Uma outra alternativa para o ajuste é a utilização de métodos analíticos em que uma parcela é ajustada para a variabilidade espacial usando informações de vizinhos mais próximos. Este método de análise, comumente conhecido como análise espacial, vem sendo utilizado com sucesso e de forma rotineira em programas de melhoramento de cereais na Austrália (GILMOUR, 1997). A abordagem clássica deste método de análise baseia-se na suposição de que a variabilidade dentro de uma repetição é relativamente menor em comparação com a variabilidade entre as repetições. Com isso, o mesmo efeito de blocos é estimado para todas as parcelas dentro de uma repetição independentemente da distância física que as separam (MATASSA, 1998).

Ainda, segundo este mesmo autor, na abordagem de análise espacial, um experimento de campo é considerado como uma matriz de linha e coluna. Quando há presença de algum tipo de tendência dentro do experimento assume-se que as parcelas mais próximas fisicamente são mais intimamente relacionadas do que aquelas parcelas mais afastadas. Esta suposição implica que a correlação entre parcelas, tanto na direção de linha ou coluna, diminua à medida que a distância entre parcelas aumente. Gilmour (1997) descreveu as estruturas de correlação entre parcelas experimentais com um modelo chamado de um processo autorregressivo separável de primeira ordem $AR(1) \times AR(1)$, o qual modela a variação das tendências tanto na direção de linhas quanto na de colunas.

As possíveis tendências dentro de um experimento podem ser subdivididas em três componentes aditivos (MATASSA, 1998):

1- Tendência local ou natural (*local or natural trend*): reflete pequenas alterações na fertilidade do solo, umidade, luz e outros fatores ambientais. Se existe algum tipo de tendência dentro de um campo de experimentação, então, erros em parcelas que estão mais próximas, serão positivamente relacionados. Se a correlação entre os erros das parcelas, tanto no sentido de linha ou de coluna, tender a zero à medida que a distância aumente, então este é característico de um processo autorregressivo (AR).

2- Variação em larga escala ou tendência global (*large scale variation or global trend*): variação em larga escala alinhado tanto com linhas ou colunas do experimento.

3 - Tendência não relacionadas (*extraneous trend*): devido as práticas agrícolas, como direção de plantio ou colheita, irregularidades no campo experimental, parcelas perdidas, entre outras. Estes procedimentos podem causar efeito dominante de linha ou coluna.

Exemplo prático de tendência global (a), não relacionada (b), e modelo próximo ao ideal (c) podem ser observados na Figura 3 (STRINGER, 2006).

Figura 3. Semivariograma (a) modelo de tendência global; (b) não relacionada; e (c) modelo com tendência ajustada.

Segundo Matassa (1998), a estratégia para a modelagem das tendências espaciais consiste na condução de etapas sucessivas. Primeiramente, a tendência natural deve ser modelada pelo processo autorregressivo separável de primeira ordem para linhas e colunas $AR(1) \times AR(1)$ ajustado para o desvio dos efeitos clonais dos valores de dados observados. Posteriormente, o semivariograma é analisado para observar a necessidade de incorporar ao modelo algum tipo de tendência adicional (global ou não relacionada). Caso o modelo contenha todas as tendências, cada parcela apresentará a seguinte forma:

$$\text{Observação} = \text{Efeito clonal} + \text{Tendência global} + \text{Tendência não relacionada} + \text{Tendência natural} + \text{Erro}$$

Os métodos geoestatísticos permitem minimizar a variação ambiental de ordem sistemática tornando as análises estatísticas mais precisas e permitindo a obtenção de médias genotípicas mais próximas do valor real, principalmente em condições de fases iniciais nas quais têm-se maior necessidade deste tipo de correção.

2.4.1. Semivariograma

O semivariograma é uma ferramenta básica de suporte às técnicas de interpolação, que permite representar quantitativamente a variação de um fenômeno regionalizado no espaço (HUIJBREGTS, 1975). Segundo Molin et al. (2015), esta ferramenta expressa o comportamento espacial de uma variável em determinada região, ou seja, indica quão díspares tornam-se os pontos quando a distância entre elas aumenta, portanto, mostra quanto da variação encontrada entre as amostras é atribuída às diferenças de distâncias entre pontos e quanto é efeito aleatório (Figura 3 e Figura 4).

Os semivariogramas apresentam três importantes parâmetros: a- o efeito pepita (C_0), que se refere ao valor da variância para a distância zero e representa o componente da variação ao acaso ou devido a incertezas analíticas; b- o patamar (C) = ($C_0 + C_1$), que é o valor da variância em que a curva estabiliza sobre um valor constante (é o máximo da variância); o patamar é atingido quando a variância dos dados se torna constante com as distâncias entre as amostras, e esse parâmetro permite a determinação da distância limite entre a dependência e a independência entre as amostras; c- o alcance (a), que é a distância da origem até onde o patamar atinge valores estáveis, e considerado o limite da dependência espacial da grandeza medida (SILVA et al., 2003) (Figura 4).

Figura 4. Semivariograma experimental, modelo teórico, mostrando os seguintes componentes: efeito pepita (C_0), semivariância estrutural (C_1), patamar ($C = C_0 + C_1$) e alcance (a) (OLIVEIRA, 2003).

Após a confecção dos semivariogramas dos valores experimentais, procura-se ajustar um modelo matemático que represente o mais próximo possível a configuração dos mesmos. Embora possa existir uma infinidade de funções que se ajustem aos semivariogramas experimentais, a prática tem mostrado que alguns modelos, fundamentados nas suposições teóricas das variáveis regionalizadas, têm satisfeito a maioria das suas aplicações (STURARO, 2015). Segundo Landim (2006), os modelos mais utilizados podem ser classificados em:

1. Modelo esférico (Sph):

$$\gamma(h) = C \left[\frac{3}{2} \left(\frac{h}{a} \right) - \frac{1}{2} \left(\frac{h}{a} \right)^3 \right] \text{ para } h < a$$

$$\gamma(h) = C, \text{ para } h \geq a,$$

neste modelo a inclinação da tangente junto à origem ($h \cong 0$) é $3C/2a$; é o modelo mais comum, podendo-se afirmar que equivale à função de distribuição normal da estatística clássica.

2. Modelo exponencial (Exp):

$$\gamma(h) = C[1 - e^{-3h/a}]$$

neste modelo a inclinação da tangente junto à origem é C/a ; C é a assíntota de uma curva exponencial e pode ser equalizada junto à soleira; “ a ” corresponde ao alcance prático igual à distância segundo a qual 95% da soleira foi alcançada.

3. Modelo gaussiano (Gau):

$$\gamma(h) = C[1 - e^{(-3h/a^2)}]$$

A curva é parabólica junto à origem e a tangente nesse ponto é horizontal, o que indica pequena variabilidade para curtas distâncias; “ a ” corresponde ao alcance prático igual à distância segundo a qual 95% da soleira foi alcançada.

2.4.2. Métodos de interpolação

O processo de reprodução das características do fenômeno espacial baseado em pontos amostrais é denominado de interpolação ou estimativa. A interpolação ou estimativa de um ponto não amostrado é feita por meio do ajuste de funções matemáticas locais (pontos mais próximos ao ponto não amostrado) ou globais (todos os pontos amostrais) (YAMAMOTO; LANDIM, 2013).

Dentre os diversos métodos de interpolação de dados espaciais, os mais utilizados são a krigagem, ponderação do inverso da distância (IDW), interpolação polinomial global (GPI), interpolação polinomial local e funções de base radial (*spline*) (NOGUEIRA; AMARAL, 2009). Neste estudo, serão abordados somente os métodos de interpolação por krigagem e IDW.

2.4.2.1. Krigagem

Myers (1994) e Laslett (1994) fizeram uma abordagem geral de métodos utilizados para interpolação espacial de parâmetros em áreas de pesquisas como ciência do solo, ecologia, engenharia florestal, hidrologia e meteorologia, destacando-se a krigagem (ou interpolação geoestatística) como um preditor em potencial de inúmeros parâmetros nestas áreas.

A krigagem, entendida como um estimador que se baseia numa série de técnicas de análise de regressão, sejam essas lineares ou não, procura minimizar a variância estimada a partir de um modelo prévio levando em consideração a dependência estocástica entre os dados distribuídos no espaço (LANDIM, 2003). É, portanto, um processo geoestatístico de estimativa de valores de variáveis distribuídas no espaço e/ou tempo, com base em valores adjacentes quando considerados interdependentes pela análise variográfica (YAMAMOTO; LANDIM, 2013). Ainda, segundo este mesmo autor, este método de interpolação pode ser comparado com os tradicionais de estimativa por médias ponderadas ou por médias móveis, mas a diferença fundamental é que somente a krigagem apresenta estimativas não tendenciosas e a mínima variância associada ao valor estimado. Existem diversas formas, sendo, a krigagem ordinária e a cokrigagem as mais usuais.

2.4.2.2. Ponderação do Inverso da Distância (IDW)

O método de interpolação IDW é baseado na suposição de que o valor em um ponto não amostrado pode ser aproximado como uma média ponderada de valores em pontos dentro de uma certa distância de corte ou a partir de um determinado número m dos pontos mais próximos (normalmente 10 a 30). Os pesos geralmente são inversamente proporcionais a uma potência de distância (BURROUGH, 1986; WATSON, 1992) o qual, em uma localização não amostrada r , leva a um estimador.

Os melhores resultados da interpolação IDW são obtidos quando a amostragem é densa em relação à variação local que se está tentando simular, ao passo que se a amostragem for em número insuficiente ou com distribuição espacial inadequada à característica da superfície a ser modelada, os resultados podem não representar suficientemente a superfície desejada (ESRI, 2016).

O interpolador IDW pode ser descrito através das seguintes equações (MARCUIZZO et al., 2011):

$$Z(x) = \frac{\sum_{i=1}^n \omega_i \cdot Z(x_i)}{\sum_{i=1}^n \omega_i} \quad (2.1)$$

em que $Z(x)$ é o valor do ponto que se deseja interpolar; n é a quantidade de pontos próximos utilizados na interpolação do ponto x , $Z(x_i)$ é o valor do ponto x_i ; ω_i é o peso do valor de x_i sobre o ponto x .

E o parâmetro ω_i é obtido por:

$$\omega_i = \frac{1}{h(x, x_i)^p} \quad (2.2)$$

em que $h(x, x_i)$ é a distância entre o ponto x e o ponto x_i ; p é o parâmetro de potência, geralmente igual a dois.

3. OBJETIVO

3.1. Geral

Avaliar a acurácia na inferência sobre as médias genotípicas ajustadas de clones em fases iniciais do melhoramento da cana-de-açúcar, por meio da adoção de informações de variabilidade espacial obtidas por dados de amostragem de solo, em alta densidade, e por diferentes sensores proximais.

3.2. Específicos

Avaliar de forma comparativa diferentes métodos para ajuste da dependência espacial:

- (i) utilização de métodos analíticos baseados em covariáveis ambientais para identificação de padrões de variação ambiental;
- (ii) Integração de diferentes geotecnologias (GIS) para inferência da variabilidade espacial presente nos campos de experimentação.

4. MATERIAL E MÉTODOS

4.1. Caracterização da área experimental

Para a utilização dos diferentes sensores proximais visando a identificação e quantificação do efeito da variabilidade espacial em experimentação, foram implantados dois experimentos referentes as fases iniciais do melhoramento genético da cana-de-açúcar. Por se tratarem de etapas onde se tem um elevado número de genótipos, restrição de material vegetativo, não sendo possível a implantação de experimentos com repetições e, portanto, elevado efeito ambiental, optou-se por focar os estudos especificamente nestas fases.

As áreas experimentais consideradas no presente trabalho são pertencentes ao Programa de Melhoramento Genético da Cana-de-açúcar, da Universidade Federal de São Carlos (UFSCar), uma das dez universidades federais integrantes da Rede Interuniversitária para o Desenvolvimento do Setor Sucroenergético (RIDESA). A população base utilizada para compor os experimentos foi amostrada dentre 475 cruzamentos de irmãos-completos e incompletos, realizados no ano de 2017, na Estação de Floração e Cruzamentos da Serra do Ouro, município de Murici-AL. A estação situa-se à latitude 9° 13' S, longitude 35° 50' W e a 450-500 m de altitude e tem suas atividades coordenadas pela Universidade Federal de Alagoas (UFAL), também vinculada à RIDESA. No total, foram produzidas e distribuídas cerca de 140.000 plântulas para compor dois experimentos em diferentes regiões do estado de São Paulo, denominados de Área experimental 1 e 2, como detalhadas a seguir.

4.1.1. Área experimental 1

A implantação e a condução das avaliações na Área experimental 1 (6,5 ha) foram realizadas nas dependências do Centro de Ciências Agrárias (CCA) da UFSCar, localizado no município de Araras, estado de São Paulo, cujas coordenadas geográficas são: latitude 22° 21' 25" S e longitude 47° 23' 03" W de Greenwich e altitude média de 650 m (Figura 5). O clima, segundo a classificação de Köppen, é do tipo Cwa mesotérmico, com verões quentes e úmidos e invernos secos, com precipitação anual média de 1.300 mm e temperatura média anual de 21,1°C. De acordo com Peres et al. (2010), o solo predominante nesta estação experimental é classificado como Latossolo Vermelho-Escuro, distrófico, A moderado, de textura argilosa. A adubação e tratos culturais foram realizados de acordo com aqueles recomendados para a cultura da cana-de-açúcar.

4.1.2. Área experimental 2

A implantação e a condução das avaliações na Área experimental 2 (9,7 ha) foram realizadas nas dependências da Estação Experimental de Valparaíso, localizada no Noroeste Paulista, cujas coordenadas geográficas são: latitude 21° 13' 20" S e longitude 50° 52' 00" W de Greenwich e altitude média de 460 m (Figura 5). Segundo a classificação de Köppen, a região é classificada como clima tropical com estação seca (Aw), megatérmico, com precipitação anual média de 1.168 mm e temperatura média anual de 21,9°C. De forma contrastante, esta estação se caracteriza por possuir solos predominantemente arenosos, classificados em Podzólicos Vermelho-Amarelo (DIAS, et al., 1999). A adubação e tratos culturais foram realizados de acordo com aqueles recomendados para a cultura da cana-de-açúcar.

Figura 5. Localização das duas áreas experimentais utilizadas para o estudo. Área experimental 1, localizada no município de Araras/SP, nas dependências da Universidade Federal de São Carlos (UFSCar), no Centro de Ciências Agrárias (CCA); região administrativa de Campinas. Área experimental 2, localizada no município de Valparaíso, no Noroeste Paulista, também nas dependências da Estação Experimental da UFSCar; região administrativa de Araçatuba.

4.1.3. Delineamento experimental

Em ambos os locais, os experimentos foram implantados considerando a estrutura de família, amplamente adotado em diferentes programas de melhoramento ao redor do mundo (JACKSON, 1995; BARBOSA et al., 2004; ZHOU et al., 2013). Esta estrutura, consiste no agrupamento dos indivíduos, provenientes de um mesmo cruzamento (família), portanto, que compartilham informações genéticas mais próximas, em uma mesma parcela. Neste caso, tem-se a possibilidade de implantação de um experimento sob o delineamento estatístico em blocos incompletos, com repetições. No presente estudo, a unidade experimental foi composta por dois sulcos com 27 m cada, espaçados a 1,4 m entre linhas, com 54 plântulas por sulco (espaçadas a 0,5 m), perfazendo uma área total da parcela de aproximadamente 40 m², com duas repetições. No total, cada família (parcela) foi representada por 108 plântulas, ou 216, considerando as duas repetições. Por se tratar de um experimento de fase inicial, onde normalmente alguns cruzamentos apresentam número limitado de plântulas, dados desbalanceados, algumas famílias foram compostas de apenas uma parcela (repetição). De forma geral, a Área experimental 1 e 2 foram compostas por 443 e 432 famílias respectivamente.

4.2. Esquema de amostragem de solo

Antes da implantação dos experimentos, ambas as áreas experimentais foram previamente georreferenciadas e as amostragens de solo realizadas, majoritariamente, em grade regular, esquema este onde se coletam amostras em pontos igualmente espaçados e homogeneamente distribuídos por toda a região experimental; a outra parte, devido ao posicionamento dos talhões e curvas de nível, tiveram seus pontos amostrais ajustados. Buscando investigar uma possível estrutura de variabilidade espacial, optou-se por estabelecer uma alta densidade amostral, visto que para se obter boas estimativas do modelo para explicação do padrão de variabilidade de cada parâmetro ou atributo do solo se faz necessário número suficiente de amostras. Para tanto, determinou-se um número médio de sete amostras por hectare.

Na área experimental 1 foram amostrados 56 pontos na profundidade de 0 a 20 cm, sendo cada ponto amostral composto por seis subamostras, coletadas em um raio médio de 3 a 5 m em torno do ponto principal (Figura

6A), com o objetivo de corrigir possíveis erros de posicionamentos de receptores de código C/A (Timbó, 2000). O mesmo ocorreu para a área experimental 2, porém, por se tratar de um experimento de maior dimensão, coletou-se 68 pontos amostrais (Figura 6B).

Parte das amostras de solo foram enviadas para o laboratório de solos do Departamento de Recursos Naturais e Proteção Ambiental da UFSCar, para análise e determinação dos atributos químicos; macronutrientes: pH, matéria orgânica (M.O.), fósforo em resina (P), alumínio (Al), porcentagem de saturação por alumínio (m%), hidrogênio mais alumínio (H+Al), potássio (K), cálcio (Ca), magnésio (Mg), soma de base (SB), capacidade de troca de cátions (CTC), e porcentagem de saturação por bases (V%) (RAIJ et al., 2001); e as características físicas do solo: teores de areia (fina, grossa e total), silte e argila (CAMARGO et al., 1987). A outra parte das amostras, após secagem a 45°C por 48 horas, foram encaminhadas para o Laboratório de Agricultura de Precisão da Faculdade de Engenharia Agrícola da Universidade Estadual de Campinas (UNICAMP), onde utilizou-se o sensor terrestre FieldSpec® para obtenção dos dados hiperespectrais.

Figura 6. (A) Área experimental 1, com 6,5 ha, localizada no município de Araras/SP, nas dependências do Centro de Ciências Agrárias (CCA), da Universidade Federal de São Carlos (UFSCar); (B) Área experimental 2, com 9,7 ha, localizada no município de Valparaíso/SP, nas dependências da estação experimental, também da UFSCar; no total, foram amostrados e georreferenciados 56 (A) e 68 pontos (B), respectivamente.

4.3. Leitura de condutividade elétrica aparente do solo (CEa)

Após a implantação dos experimentos, as leituras de condutividade elétrica aparente do solo (CEa) foram realizadas utilizando o sensor EM38-MK2® (Geonics, Mississauga, Ontário, CANADA), que tem como princípio a Indução Eletromagnética (IEM). Em ambas as áreas experimentais, utilizou-se o sensor conectado a um receptor GPS, do tipo L1 (*Trimble*), para registro das coordenadas geográficas, e a um coletor de dados (*Juniper Arber Field PC*), para registro das informações (Figura 7); em ambas as áreas experimentais, a condução do equipamento foi feita de forma manual, não sendo utilizado nenhum veículo para transporte. O espaçamento entre passadas foi estabelecido de forma paralela, na distância aproximada de 1,5 m, i.e., em todas as entre linhas da cultura. Portanto, todas as leituras foram realizadas na área central das parcelas experimentais. Antes do início das leituras de CEa, procedeu-se a calibração do equipamento de forma manual, conforme instruções do fabricante (Geonics Limited, 2011). Tanto na etapa de

calibração quanto no processo de leitura, evitou-se a utilização de metais (anéis, relógios, correntes, cintos, entre outros) para preservar o equipamento de possíveis interferências no campo magnético.

Segundo informações do fabricante, o equipamento EM38-MK2® fornece medições de condutividade elétrica dentro de duas faixas distintas de profundidade, de forma simultânea. Isto é possível, pois o equipamento possui duas bobinas receptores, separadas por 1 m e 0,5 m do transmissor, fornecendo dados de profundidades efetivas de 1,5 m e 0,75 m, respectivamente, quando posicionadas na orientação vertical do dipolo, e 0,75 m e 0,375 m, respectivamente, quando na orientação do dipolo horizontal. No presente estudo, utilizou-se leituras das camadas de 0,75 m e 0,375 m, uma vez que, segundo Junk et al. (2005), esta orientação fornece a melhor relação entre a CE_a e as propriedades do solo nas camadas de até 30 cm, profundidade essa contemplada por meio das amostragens de solo.

Figura 7. Equipamento utilizado para mensuração da condutividade elétrica aparente do solo, na orientação do dipolo horizontal (0,75 m e 0,375 m), área experimental 1 (A) e 2 (B); coletor de dados para registro das informações coletadas em campo (C).

4.4. Tratamento e análise dos dados

Para o processamento das informações provenientes da análise de solo, realizou-se a análise exploratória do conjunto de dados, buscando compreender as principais características inerentes aos dados coletados em campo e advindo de laboratório. Esse procedimento é de suma importância para que dados discrepantes sejam identificados e removidos. Para tanto, os dados foram submetidos a tratamentos estatísticos associado a utilização de ferramentas gráficas, i.e., histogramas e *box plots*, para avaliar a forma e dispersão do conjunto de dados. Segundo Yamamoto e Landim (2013), histogramas são amplamente utilizados para análise da distribuição espacial de pontos, ou seja, se a distribuição é unimodal, bimodal ou plurimodal; sendo unimodal, se a distribuição é simétrica ou assimétrica; se, no caso de presença de assimetria, há indicação ou não de valores atípicos (*outliers*). De forma complementar, o *box plot* permite avaliar a distribuição empírica dos dados, sendo possível obter os valores de limite inferior e superior dos dados, assim como, primeiro quartil (Q_1), segundo quartil, ou mediana (Q_2), e terceiro quartil (Q_3).

Em seguida, visando a redução, eliminação de sobreposição e escolha das formas mais representativas de dados a partir de combinações lineares das variáveis provenientes da análise de solo, optou-se pela redução da dimensionalidade dos dados por meio da metodologia de análise de componentes principais - ACP ou PCA (*Principal Component Analysis*). De forma geral, o PCA calcula um conjunto não correlacionado de variáveis (componentes), os quais são ordenados para que os primeiros retenham a maior parte da variação presente em todas as variáveis originais, neste caso, variáveis de solo.

Posteriormente, para o tratamento geoestatístico dos diferentes componentes principais, utilizou-se o pacote estatístico *geoR* versão 1.8-1 (RIBEIRO JUNIOR; DIGGLE, 2001), que fornece funções para a leitura e preparo dos dados, análise exploratória, inferência nos parâmetros do modelo, incluindo métodos baseados em semivariograma e baseados em probabilidade e interpolação espacial. A presença de dependência espacial foi expressa através do semivariograma, estimado pela seguinte equação.

$$\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(X_i) - Z(X_i + h)]^2$$

em que, $N(h)$ é o número de pares de valores medidos $Z(X_i)$, $Z(X_i + h)$, separado pela distância h , se a variável for escalar. O gráfico de $\hat{\gamma}(h)$ versus os valores correspondentes de h é chamado de semivariograma.

Após a construção dos semivariogramas para as diferentes componentes principais, testou-se os seguintes modelos teóricos para melhor ajuste ao conjunto de dados: esférico, exponencial e gaussiano. A escolha do modelo ocorreu a partir da avaliação de seus respectivos resíduos, os quais foram observados na etapa de validação cruzada, sendo adotado os seguintes critérios: Raiz do Erro Médio Quadrático (*Root Mean Square Error* - RMSE), mede a qualidade do ajuste de um modelo, equação 4.1; Coeficiente de Determinação (R^2), indica quanto o modelo foi capaz de explicar os dados coletados, equação 4.2; Coeficiente do Desvio Quadrático Médio (*Mean Squared Deviation Rate* - MSDR), avalia quão bem o modelo de previsão estima a incerteza de previsão, equação 4.3. Baixo RMSE está associado com alta capacidade preditiva do modelo, enquanto que valores de R^2 e MSDR próximos a um significam bom ajuste do modelo em relação ao conjunto de dados.

$$RMSE = \sqrt{\frac{1}{n} \sum_{\alpha=1}^n (z_{\alpha}^* - z_{\alpha})^2} \quad (4.1)$$

$$R^2 = 1 - \frac{\sum (z_{\alpha}^* - z_{\alpha})^2}{\sum (z_{\alpha}^* - \bar{z}_{\alpha})^2} \quad (4.2)$$

$$MSDR = \frac{1}{n} \sum_{\alpha=1}^n \frac{[(z_{\alpha}^* - z_{\alpha})]^2}{\sigma^2(\alpha)} \quad (4.3)$$

$$r = \frac{cov(z_{\alpha}^*, z_{\alpha})}{\sqrt{var(z_{\alpha}^*) \cdot var(z_{\alpha})}} \quad (4.4)$$

em que, n é o número de pontos de validação; z_{α}^* - z_{α} são os valores preditos e observados na localização α ; cov a covariância; var a variância.

O ajuste dos dados a partir dos semivariogramas possibilitou definir os seguintes parâmetros: efeito pepita (C0); patamar ((C) = C0 + C1); e alcance (a).

Com base nos coeficientes extraídos dos semivariogramas, determinou-se o índice de dependência espacial (*spatial dependence index* - SDI) para cada componente principal solo avaliado, sendo adotada a classificação proposta por Seidel e Oliveira (2016), em que, para o semivariograma esférico: $SDI_{Spherical} (\%) \leq 7\%$ (dependência espacial fraca), $7\% < SDI_{Spherical} (\%) \leq 15\%$ (dependência espacial moderada) e $SDI_{Spherical} (\%) > 15\%$ (forte dependência espacial);

para o semivariograma exponencial: $SDI_{Exponential} (\%) \leq 6\%$ (fraca dependência espacial), $6\% < SDI_{Exponential} (\%) \leq 13\%$ (dependência espacial moderada), $SDI_{Exponential} (\%) > 13\%$ (forte dependência espacial); e para o semivariograma gaussiano: $SDI_{Gaussian} (\%) \leq 9\%$ (fraca dependência espacial), $9\% < SDI_{Gaussian} (\%) \leq 20\%$ (dependência espacial moderada) e $SDI_{Gaussian} (\%) > 20\%$ (forte dependência espacial).

No presente estudo, optou-se pelo interpolador estatístico Inverso do Quadrado da Distância IDW, conforme equações 2.1 e 2.2, uma vez que os melhores resultados neste tipo de interpolação são obtidos quando a amostragem é realizada em alta densidade. Adicionalmente, a formação de “*bull's eyes*”, ou “olho-de-boi”, característica essa normalmente observada em mapas interpolados por esse método, não seria uma característica indesejável, visto que o presente estudo tem por objetivo identificar padrões de variação ambiental em pequenas escalas, no âmbito da parcela experimental. Embora este método de interpolação não tenha dependência dos modelos e dos parâmetros acima citados, optou-se por obter a classificação proposta por Seidel e Oliveira (2016), de modo que fosse possível a compreensão do grau de dependência espacial dos diferentes componentes principais, assim como, para a identificação da intensidade de variabilidade presente nas áreas experimentais.

Obtidos os mapas temáticos provenientes das informações dos diferentes componentes principais, determinou-se a posição e a localização exata de cada parcela experimental (Figura 8). Após a sobreposição das parcelas sobre os mapas temáticos, extraíram-se os valores de cada pixel (interpolação em pixels de 0,5 m x 0,5 m) dentro de cada parcela experimental e, após submetidos a uma análise estatística descritiva, o valor médio foi considerado e utilizado como covariável no modelo genético-estatístico, conforme detalhado no item 4.6.

Figura 8. Esquema da posição e localização de cada parcela experimental para extração do valor médio de todos os pixels, Área experimental 1. A mesma rotina foi empregada para a Área experimental 2.

Para uma possível análise de correlação entre os valores obtidos dos diferentes componentes principais com os resultados da análise de solo, geraram-se polígonos com raio igual a 5 m, utilizando como referência os pontos de amostragem de solo. Dentro desses polígonos consideraram-se os valores de cada pixel da imagem interpolada (0,5 m x 0,5 m), os quais foram extraídos e submetidos a uma análise estatística descritiva. Valores superiores a ± 3 vezes o desvio padrão (SD), foram excluídos. Em seguida, atribuiu-se o valor médio do componente principal ao ponto amostral (Figura 10). Este valor foi utilizado apenas para a análise de correlação (eq. 4.4) com os diferentes atributos físicos e químicos do solo.

Com relação ao sensor EM38-MK2, consideraram-se todos os dados brutos coletados em ambas as áreas experimentais, sendo que os mesmos foram diretamente plotados sobre as parcelas georreferenciadas. Em seguida, calculou-se o valor médio de condutividade elétrica aparente do solo de todos os pontos obtidos dentro de cada parcela experimental. Assim como realizado para as informações provenientes dos componentes principais, o valor médio de CEa de cada parcela foi considerado e utilizado como covariável no modelo genético-estatístico, conforme detalhado no item 4.6.

Para a análise de correlação (eq. 4.4) entre os dados obtidos via sensor de CEa e os resultados da análise de solo, diferentemente do procedimento realizado para os componentes principais, os polígonos com raio igual a 5 m, utilizando como referência os pontos de amostragem de solo, contornaram apenas os valores brutos obtidos pelo sensor. Para este conjunto de dados, optou-se por não remover valores superiores a ± 3 vezes o desvio padrão, uma vez que os valores obtidos dentro de cada polígono não são estimados, e sim, reais.

4.5. Índice de vegetação

Uma cultura de rotação (*Crotalaria juncea* L.), antecedendo à implantação do experimento de fase inicial do melhoramento, foi considerada para o mapeamento a priori da variabilidade espacial da área experimental 1. Este tipo de técnica é comumente conhecido como ensaio de uniformidade ou experimento “em branco”. Segundo Swearingin e Holt (1976), este tipo de técnica é geralmente realizado na mesma área que deverá ser utilizada para implantação de um experimento subsequente. Os resultados obtidos nestes tipos de ensaios podem ser utilizados para aumentar a precisão em experimentos subsequentes, ajustando as variações espaciais encontradas para explicar a variação natural revelada no estudo. Normalmente, métodos de covariância ou regressão são usados para fazer este ajuste.

Nesse sentido, o *NDVI* foi calculado considerando os valores obtidos da reflectância espectral da cultura de rotação, 80 dias após o plantio (antes do período de florescimento), através da aquisição das imagens (*RGB – red, green and blue* e *NIR*) a partir de uma aeronave remotamente pilotada - RPA (*Drone DJI Inspire Pro*) contendo uma câmera modificada (Câmera Zenmuse X3) com filtro próprio para captação de imagens para estudo com *NDVI*, aplicados em Agricultura de Precisão. Padronizou-se a altitude de voo igual a 50 m e sobreposição longitudinal e lateral das imagens em 75% e 65%, respectivamente, com pixel de resolução igual a 1,5 cm, conforme Figura 9; a velocidade de voo foi definida em 5 m/s. Além disso, outra etapa de *checklist* pré-voo consistiu na verificação das condições climáticas do local, estabelecendo os voos em horário de sol a pino, cenário este altamente recomendado para mapeamento aéreo. As coletas de imagens foram realizadas em um único dia.

Figura 9. Esquema do plano de voo na área experimental 1 para obtenção das imagens *RGB* e *NIR*, considerando velocidade de voo definida em 5 m/s, altitude de voo de 50 m e sobreposição longitudinal (A) e lateral (B) das imagens em 75% e 65%, respectivamente; pixel com resolução igual a 1,5 cm.

As imagens obtidas (RGB e NIR) foram então processadas utilizando o modo automático do *software* Photoscan (Agrisoft). Geraram-se o modelo digital do terreno (MDT), modelo digital de superfície (MDS), nuvens de pontos e o ortomosaico. Em seguida, os dados foram importados pelo software SIG QGIS® para cálculo do índice de interesse, *NDVI*. Para tanto, considerou-se a seguinte equação:

$$NDVI = \frac{NIR - RED}{NIR + RED}$$

em que *NDVI* é o índice de vegetação por diferença normalizada; *RED* é a refletância da banda do vermelho (0,62 a 0,70 μm) e *NIR* a refletividade da banda no infravermelho próximo (0,7 a 1,5 μm).

Para uma possível análise de correlação entre os valores obtidos pelo *NDVI* com os resultados da análise de solo, geraram-se polígonos com raio igual a 5 m, utilizando como referência os pontos de amostragem de solo, contornando os valores de *NDVI* obtidos em cada pixel da imagem (1,5 cm). Todos os valores dentro de cada polígono foram extraídos e submetidos a uma análise estatística descritiva. Valores superiores a ± 3 vezes o desvio padrão (SD), foram excluídos. Em seguida, atribui-se o valor de *NDVI* médio ao ponto amostral (Figura 10). Este valor foi utilizado para a análise de correlação (eq. 4.4) e utilizado como covariável no modelo genético-estatístico, conforme detalhado no item 4.6.

Figura 10. Esquema para análise de correlação entre os valores médios dos componentes principais (PCA) e/ou *NDVI* com os atributos do solo nos diferentes pontos amostrais; valores obtidos a partir do valor médio dos pixels dentro do raio de 5 m contornando cada ponto amostral.

4.6. Avaliação experimental e ajuste dos modelos genético-estatísticos

Ambos os experimentos foram previamente avaliados (fenotipados) e selecionados seguindo os mesmos critérios adotados de forma rotineira nas etapas de seleção em fases iniciais do melhoramento da RIDESEA/UFSCar. Isto é, a performance de cada família em produzir toneladas de cana por hectare (TCH) foi avaliada durante os meses de abril e maio de 2019, no estágio de cana-planta (12 meses). O caráter TCH foi estimado por meio do corte mecanizado e pesagem total de cada parcela (família) com o auxílio de um caminhão balança. Com isso, as estimativas de TCH foram calculadas através da seguinte equação:

$$TCH = \frac{(PP \times 10)}{TP}$$

em que PP é o peso total da parcela (em kg) e TP é o tamanho da parcela em m², neste caso igual a 40 m².

Para cada experimento, dados de TCH foram analisados por meio de diferentes modelos lineares mistos. Primeiramente, dentre as diferentes classes de estrutura que podem ser consideradas para a matriz de variância-covariância (VCOV), utilizou-se o modelo independente (ID), o mais simples para modelar a matriz de efeito residual, equação 4.5. Este modelo considera homogeneidade de variâncias ao longo do experimento e ausência de associação entre os blocos (Tabela 11) (Apêndice A). Posteriormente, analisaram-se os dados utilizando a estrutura auto-regressiva separável de primeira ordem (AR1 x AR1) como modelo de dependência espacial entre os erros, equação 4.7 (Tabela 11) (Apêndice A). Para ambos os modelos, utilizou-se a metodologia de seleção “*forward*” (KUTNER et al., 2004) para incorporação das informações dos diferentes sensores. Esta metodologia consiste na adição de cada variável de forma independente, de modo que a primeira variável selecionada é aquela com maior significância. Este procedimento consiste nas seguintes etapas: calcula-se a estatística F para testar se a primeira variável é realmente significativa para o modelo; se o valor de F for igual ou superior a 0,05, a variável testada é incluída no modelo; o segundo passo é testar a próxima variável que possua maior correlação com a variável resposta, considerando a presença da primeira variável no modelo; este procedimento é repetido com todas as variáveis que se deseja testar, i.e., informações obtidas via sensor de condutividade elétrica aparente do solo (CEa), nas diferentes profundidades, índice de vegetação da diferença normalizada (NDVI) e componentes principais, até que não seja mais incluída mais nenhuma variável explicativa no modelo. Para o modelo ID e AR1 x AR1, com a inclusão de covariável, utilizou-se a equação 4.6 e 4.8, respectivamente.

$$y_{ij} = \mu + F_i + b_j + e_{ij} \quad \text{Eq. 4.5}$$

$$y_{ij} = \mu + F_i + b_j + \beta_k d_{ij} + e_{ij} \quad \text{Eq. 4.6}$$

$$y_{ijuv} = \mu + F_i + b_j + Row_u + Col_v + Row:Column_{ij} + e_{ijuv} \quad \text{Eq. 4.7}$$

$$y_{ijuv} = \mu + F_i + b_j + Row_u + Col_v + Row:Column_{ij} + \beta_k d_{ijuv} + e_{ijuv} \quad \text{Eq. 4.8}$$

em que, y_{ij} é a unidade de observação; μ é a média geral; F_i é o efeito aleatório da i-ésima família; b_j é o efeito fixo do j-ésimo bloco; e_{ij} é o erro; $\beta_k d_{ij}$ é o efeito da k-ésima covariável significativa na i-ésima família, e no j-ésimo bloco; y_{ijuv} é a unidade de observação; Row_u é o efeito da u-ésima linha; Col_v é o efeito v-ésima coluna; $Row:Column_{ij}$ é o efeito de linha dentro de coluna e $\beta_k d_{ijuv}$ é o efeito da k-ésima covariável significativa na i-ésima família, no j-ésimo bloco, na u-ésima linha, e na v-ésima coluna; e_{ijuv} o erro.

Por fim, analisaram-se os dados utilizando os modelos ID e AR1 x AR1, e também com a inclusão das diferentes variáveis significativas no modelo, selecionando-se a melhor estrutura de variância-covariância (VCOV), para a matriz de efeito residual, com base nos critérios de informação de Akaike (AIC) (AKAIKE, 1974) e/ou o critério de informação Bayesiana (BIC) (SCHWARZ, 1978). Suas expressões são apresentadas a seguir:

$$AIC = -2\log(L) + 2n_{IC}$$

$$BIC = -2\log(L) + \log(N) \times n_{IC}$$

em que $\log(L)$ é o logaritmo da verossimilhança residual, N é o número total de observações (indivíduos), e n_{IC} é o número de parâmetros da matriz de variância-covariância (VCOV). As estruturas de VCOV mais prováveis à luz das nossas observações deverão apresentar o menor valor de AIC ou BIC na análise (maior verossimilhança).

Para cada modelo e experimento, o coeficiente de variação residual (CV_e , %) foi computado como a proporção do erro residual sobre a média experimental para TCH, assim como, a herdabilidade no sentido amplo (h^2_a), por meio da seguinte equação:

$$h^2_{\bar{x}} = \left(\frac{\sigma_g^2}{\sigma_g^2 + \frac{\sigma_e^2}{R}} \right)$$

em que $h^2_{\bar{x}}$ é herdabilidade no sentido amplo em nível de média de família; σ_g^2 a variância dos efeitos genéticos entre famílias; σ_e^2 a variância dos efeitos residuais (ambiental); R é o número de repetições das famílias.

A acurácia seletiva (AC_s) entre famílias foi obtida por meio da seguinte equação:

$$AC_s = \sqrt{h^2_{\bar{x}}}$$

em que AC_s é a acurácia seletiva entre famílias, e $h^2_{\bar{x}}$ a herdabilidade no sentido amplo no âmbito da média de famílias.

Antes da aplicação de modelos mistos para análise dos dados, verificou-se o padrão de distribuição espacial dos dados brutos de cada parcela/família. Para isso, utilizou-se o pacote do R *spatstat* versão 1.63-3 (BADDELEY et al., 2015), que tem como funcionalidade, análise exploratória dos dados, ajuste de modelo e simulação.

Para a realização das análises estatísticas utilizaram-se os softwares GenStat (PAYNE, 2009) para predição dos valores genotípicos dos tratamentos (famílias), considerando os diferentes modelos (ID e AR1), e o sistema estatístico R (R, 2015) para as análises geoestatísticas, gráficas, correlações, entre outras.

5. RESULTADOS

5.1. Análise exploratória dos dados da Área experimental 1

Os resultados da análise de solo para os atributos químicos, macronutrientes: pH, M.O., P em resina, Al, m%, H+Al, K, Ca, Mg, SB, CTC e V%, e as características físicas do solo: teores de areia (fina, grossa e total), silte e argila, obtidos a partir de 56 pontos amostrais, na profundidade de 0 a 20 cm, estão apresentados na Tabela 1. De acordo com os resultados obtidos, identificou-se a presença de valores atípicos (*outliers*) no conjunto de dados (Figura 26, Apêndice B), os quais foram removidos para as análises posteriores; no total, 55 pontos amostrais foram considerados, Figura 27 (Apêndice C).

Valores de desvio padrão (D.P.) expressam o grau de dispersão de um conjunto de dados, quanto mais próximo de zero, mais homogêneo são os dados. Portanto, analisando o conjunto de dados para os diferentes atributos do solo, nota-se que os valores mais distantes de zero foram obtidos para a argila, silte e areia total, respectivamente (Tabela 1). Por outro lado, baixos valores, portanto, pequena variação em relação à média, foram obtidos para o pH (D.P. = 0,25), potássio (D.P. = 0,34) e alumínio (D.P. = 0,73). Considerando o coeficiente de variação (CV), i.e., medida padronizada de dispersão de uma distribuição de frequências, os maiores valores encontrados foram para a porcentagem de saturação por alumínio (55,75%), potássio (46,82%), alumínio (44,18%) e fósforo (39,49%).

Após a eliminação de valores atípicos presentes no conjunto de dados, foram obtidos os resultados da análise de componentes principais (PCA) considerando as 17 variáveis físicas e químicas do solo. A Figura 11 (A) exibe o número do componente principal (eixo x) versus seu autovalor correspondente (eixo y). Com base nos autovalores, nota-se que o primeiro componente principal explica 41% da variabilidade dos dados e este valor cai drasticamente ao observar a segunda componente (15,4%), isto é, 56,4% da variabilidade é capturada pelos dois primeiros componentes; se o terceiro componente for considerado, o valor passa para 68,6%. Os componentes principais podem ser interpretados como combinações lineares das variáveis originais, de forma que o primeiro componente principal (PC₁) foi essencialmente explicado pelos maiores valores positivos das variáveis saturação por base, cálcio, magnésio e percentual de saturação por base, em contraste com as variáveis de maiores valores negativos, porcentagem de saturação por alumínio, argila, alumínio e acidez potencial. O segundo componente (PC₂) foi explicado, principalmente, pelos maiores valores positivos das variáveis pH, fósforo e silte, em contraste com os maiores valores negativos das variáveis areia total, grossa e fina. Por outro lado, o terceiro componente (PC₃) foi explicado pelo maior valor positivo da variável percentual de saturação por base, em contraste com os maiores valores negativos das variáveis capacidade de troca de cátions e acidez potencial.

Tabela 1. Estatística descritiva dos atributos físicos e químicos do solo na camada de 0 a 20 cm de profundidade, Área experimental 1.

Atributo	Profundidade 0 a 20 cm					
	Mínimo	Máximo	Mediana	Média	D.P.	CV (%)
Argila	600,00	682,00	641,00	642,30	18,74	2,92
Areia grossa	89,00	118,00	102,00	102,30	7,13	6,97
Areia fina	51,00	79,00	70,00	68,58	6,62	9,65
Areia total	148,00	195,00	173,00	170,90	11,35	16,56
Silte	155,00	230,00	184,00	186,80	16,04	8,59
P	5,00	35,00	15,00	16,58	6,55	39,49
M.O.	17,00	28,00	22,00	22,35	2,21	9,90
pH	4,00	5,90	4,90	4,89	0,25	5,14
K	0,20	2,00	0,70	0,73	0,34	46,82
Ca	16,00	38,00	23,00	23,49	5,06	21,55
Mg	5,00	14,00	9,00	8,64	1,90	21,99
H+Al	16,00	55,00	42,00	43,47	7,25	16,67
Al	0,50	3,60	1,40	1,66	0,73	44,18
SB	21,90	52,50	32,80	32,86	6,76	20,56
CTC	38,90	93,60	76,60	76,33	7,83	10,26
V%	29,60	59,30	42,50	43,11	7,61	17,65
m%	1,10	12,80	4,30	5,17	2,88	55,75

Fósforo (P); Matéria orgânica (M.O.); pH; potássio (K); cálcio (Ca); magnésio (Mg); acidez potencial (H+Al); alumínio (Al); saturação por base (SB); capacidade de troca de cátions (CTC); percentual de saturação por base (V); porcentagem de saturação por alumínio (m); D.P. – desvio padrão; CV – coeficiente de variação.

Adicionalmente, os resultados da análise de componentes principais podem ser mais facilmente interpretados de forma gráfica, por meio da visualização bidimensional dos dados, Figura 11 (B). Neste tipo de gráfico as diferentes amostras de solo são apresentadas na forma de pontos, enquanto as variáveis físicas e químicas do solo apresentadas como vetores. De forma geral, é possível identificar o agrupamento das diferentes variáveis apresentando distintos padrões de similaridades. Um ponto interessante em se destacar é o menor ângulo formado entre os vetores, que sugere o grau de correlação entre as variáveis. A correlação máxima ($r = 1$) é obtida quando o ângulo é igual a 0, assim como a correlação mínima ($r = -1$) ocorre quando o ângulo é de 180°. Para casos em que formam-se ângulos ortogonais (90°), a correlação é nula. Portanto, pode-se evidenciar que todos os teores de areia (fina, grossa e total) apresentam alta correlação entre si. Ademais, outras correlações positivas podem ser destacadas, como é o caso da saturação por base (SB) com cálcio (Ca), e argila com acidez potencial (Hal). Por outro lado, fósforo (P. resina) e percentagem de saturação por alumínio (m%), provavelmente, não são correlacionadas por apresentarem um ângulo aproximadamente perpendicular. Além disso, em casos onde existem a formação de ângulos superiores a 90°, isto indica que ambas as variáveis são negativamente correlacionadas entre si, i.e., argila e saturação por base (V).

Para reforçar as evidências observadas por meio da visualização bidimensional dos dados, as estimativas dos coeficientes de correlação de Pearson (r) entre as diferentes variáveis de solo estudadas em combinação com os valores obtidos por meio dos diferentes sensores proximais e componentes principais, na Área experimental 1, são apresentadas nas Figuras 12 (A) e (B). Nota-se que dentre os componentes físicos do solo, apenas a argila possui forte correlação negativa e significativa ($P < 0,05$) com os demais atributos físicos, i.e., areia fina, grossa, total e silte ($r = -0,80$; $r = -0,65$; $r = -0,75$; $r = -0,79$, respectivamente). Por outro lado, apesar do atributo silte apresentar correlação positiva com os teores de areia, esses são baixos e não significativos ($r = 0,30$; $r = 0,07$; $r = 0,19$), portanto, desprezíveis.

Dentre os atributos químicos do solo, destacam-se os valores positivos altamente correlacionados ($r > 0,90$) entre os elementos cálcio e magnésio ($r = 0,99$), cálcio e saturação por base ($r = 1,0$), cálcio e percentual de saturação por base ($r = 0,91$), magnésio e saturação por base ($r = 0,99$), magnésio e percentual de saturação por base ($r = 0,92$), alumínio e percentagem de saturação por alumínio ($r = 0,99$), saturação por base e percentual de saturação por base ($r = 0,91$), todos significativos ($P < 0,05$). Em contrapartida, valores negativos altamente correlacionados também são observados, como é o caso dos atributos cálcio e alumínio ($r = -0,89$), cálcio e percentagem de saturação por alumínio ($r = -0,95$), magnésio e alumínio ($r = -0,92$), percentagem de saturação por alumínio ($r = -0,96$), acidez potencial e percentual de saturação por base ($r = -0,89$), alumínio e saturação por base ($r = -0,90$), alumínio e percentual de saturação por base ($r = -0,89$), saturação por base e percentagem de saturação por alumínio ($r = -0,95$), percentual de saturação por base e percentagem de saturação por alumínio ($r = -0,91$).

Figura 11. (A) Gráfico *scree* dos autovalores dos componentes principais e (B) Gráfico bi-dimensional (*biplot*) dos componentes principais 1 e 2 (PC1 e PC2) das 17 variáveis físicas e químicas do solo, amostradas em 55 pontos, na profundidade de 0 a 20 cm, na Área experimental 1.

Em relação aos componentes principais, valores positivos altamente correlacionados ($r > 0,90$) podem ser observados entre o componente principal 1 (PC1) e percentual de saturação por base ($r = 0,94$), magnésio ($r = 0,99$), cálcio ($r = 0,99$) e saturação por base ($r = 0,99$). De outro modo, valores negativos altamente correlacionados foram obtidos entre a PC1 e argila ($r = -0,92$) e percentagem de saturação por alumínio ($r = -0,94$). Destacam-se ainda os valores positivos com forte correlação entre a PC1 e areia fina ($r = 0,74$), grossa ($r = 0,76$), total ($r = 0,79$) e matéria orgânica ($r = 0,80$), PC2 e fósforo ($r = 0,80$) e pH ($r = 0,83$). Correlações negativas com forte correlação também foram obtidas entre a PC1 e alumínio ($r = -0,89$) e condutividade elétrica aparente do solo a 0,75 m ($r = -0,71$), PC3 e capacidade de troca de cátions ($r = -0,78$) e acidez potencial ($r = -0,77$). Complementarmente, valores positivos de correlação moderada foram obtidos entre a PC1 e capacidade de troca de cátions ($r = 0,51$) e silte ($r = 0,63$), PC2 e condutividade elétrica aparente do solo a 0,75 m ($r = -0,65$). Por fim, valores negativos de correlação moderada também foram obtidos entre PC1 e fósforo ($r = -0,53$) e acidez potencial ($r = -0,68$) e PC2 com areia fina ($r = -0,63$), grossa ($r = -0,67$) e total ($r = -0,69$). Devido à baixa correlação encontrada entre o componente principal 3 e os diferentes atributos do solo, além da baixa contribuição na explicação da variabilidade dos dados, optou-se por proceder as etapas posteriores do estudo considerando apenas os componentes 1 e 2.

Quanto ao sensor de condutividade elétrica aparente do solo, diferentes valores de correlações com os atributos físicos e químicos do solo são observados quando as leituras são realizadas em profundidades distintas. Nota-

se que leituras a 0,375 m se correlacionam fracamente com todos os atributos do solo. No entanto, a 0,75 m de profundidade, correlação positiva moderada é observada com a argila ($r = 0,65$) e porcentagem de saturação por alumínio ($r = 0,54$), e correlação forte com o fósforo ($r = 0,73$). Ainda, nesta mesma profundidade, valores de correlações negativas moderadas também podem ser destacadas, como é o caso da matéria orgânica ($r = -0,54$), cálcio ($r = -0,66$), magnésio ($r = -0,64$), saturação por base ($r = -0,66$) e percentual de saturação por base ($r = -0,65$) e forte correlação com os atributos físicos, areia fina, grossa e total ($r = -0,87$; $r = -0,89$; $r = -0,92$, respectivamente).

Para o índice de vegetação da diferença normalizada (*NDVI*), destacam-se as correlações positivas e moderadas observadas com a areia fina ($r = 0,65$), areia grossa ($r = 0,54$), areia total ($r = 0,62$), matéria orgânica ($r = 0,53$), percentual de saturação por base ($r = 0,52$), componente principal 1 ($r = 0,60$), magnésio ($r = 0,54$), cálcio ($r = 0,54$) e saturação por base ($r = 0,54$), todas significativas ($P < 0,05$). Por outro lado, observam-se correlações negativas e moderadas com a condutividade elétrica aparente do solo a 0,75 m ($r = -0,62$), argila ($r = -0,58$) e porcentagem de saturação por alumínio ($r = -0,53$). Valores negativos de fraca correlação são evidenciados com a condutividade elétrica aparente do solo a 0,375 m ($r = -0,43$), fósforo ($r = -0,46$), alumínio ($r = -0,48$) e o componente principal 3 ($r = -0,02$).

Figura 12. (A) Correlações entre as diferentes variáveis físicas e químicas do solo: areia grossa (areia_g); areia fina (areia_f); areia total (areia_t); silte; fósforo (P); matéria orgânica (mo); pH; potássio (K); cálcio (Ca); magnésio (Mg); acidez potencial (Hal); alumínio (Al); saturação por base (SB); capacidade de troca de cátions (CTC); percentual de saturação por base (V); porcentagem de saturação por alumínio (m), com os diferentes sensores: condutividade elétrica aparente do solo a 0,75 m (Eca_1m); condutividade elétrica aparente do solo a 0,375 m (Eca_05m); com os componentes principais: componente principal 1 (PC₁); componente principal 2 (PC₂); componente principal 3 (PC₃); e o Índice de Vegetação da Diferença Normalizada (*NDVI*), combinado com o teste de significância considerando um nível de 0,05% ($P < 0,05$). (B) Valores de correlação entre as diferentes variáveis estudadas. Informações referente a Área experimental 1.

5.1.1. Análise geoestatística do conjunto de dados

Para a identificação e classificação do tipo de dependência espacial e posterior construção dos mapas temáticos para os componentes principais consideradas (PC₁ e PC₂), obtiveram-se os cálculos do variograma

experimental para os ajustes dos modelos teóricos ao conjunto de dados (Tabela 2). Para a PC₁, o maior valor de alcance observado foi de 66,64 m, quando considerado o modelo esférico (sph). Por outro lado, o menor alcance foi de 28,89 m para o modelo exponencial (exp). Para este mesmo componente principal, o modelo teórico de melhor ajuste ao conjunto de dados foi o gaussiano, uma vez que o baixo valor apresentado pelo critério RMSE (1,9733) está altamente associado com a alta capacidade preditiva do modelo. Adicionalmente, tanto o coeficiente de determinação ($R^2 = 0,4305$) como o coeficiente do desvio quadrático médio (MSDR = 0,9801) apresentaram os maiores valores, indicando também que este modelo foi capaz de explicar os dados coletados. Os diferentes variogramas experimentais para o componente principal 1, considerando os modelos esférico (sph), exponencial (exp) e gaussiano (gau) podem ser melhor observados na Figura 29 (Apêndice E). De acordo com o índice de dependência espacial (SDI = 8,4280%), proposto por Seidel e Oliveira (2014), para o modelo gaussiano, a PC₁ foi categorizada como sendo de fraca dependência (SEIDEL; OLIVEIRA, 2016).

Para a PC₂, o maior valor de alcance obtido foi de 10306 m considerando o modelo exponencial (exp). No entanto, o menor valor observado para este mesmo parâmetro foi de 193,09 m para o modelo esférico (sph). Diferentemente da PC₁, o modelo teórico de melhor ajuste para a PC₂ foi o esférico (sph), pois foi o que apresentou o menor valor de RMSE (1,3304) e o maior valor de R^2 (0,3131). Apesar do coeficiente do desvio quadrático médio não ter sido o de maior valor (MSDR = 0,9923), observa-se que a diferença entre os demais é na ordem milesimal, portanto, insignificante. Os diferentes variogramas experimentais para o componente principal 2, considerando os modelos esférico (sph), exponencial (exp) e gaussiano (gau) podem ser melhor observados na Figura 30 (Apêndice F). De acordo com o índice de dependência espacial (SDI = 17,5121%), proposto por Seidel e Oliveira (2014), para o modelo esférico, a PC₂ foi categorizada como sendo de forte dependência (SEIDEL; OLIVEIRA, 2016).

Tabela 2. Parâmetros dos modelos teóricos ajustados aos variogramas experimentais do conjunto de dados na profundidade de 0 a 20 cm, Área experimental 1.

PC	Modelo	Efeito Pepita	Patamar (C1)	Alcance	RMSE	R^2	MSDR	SDI (%)	SDI Clas.
PC ₁	exp	0,0000	7,3247	28,8911	2,2549	0,2574	0,9546	4,7559	weak
	sph	0,0000	6,9903	66,6365	2,0195	0,4111	0,8981	12,9764	moderate
	gau	0,0000	7,7255	32,2019	1,9733	0,4305	0,9801	8,4280	weak
PC ₂	exp	1,4869	53,1228	10306,0027	1,3387	0,3047	0,9957	1650,3332	strong
	sph	1,3378	1,1662	193,0948	1,3304	0,3131	0,9923	17,5121	strong
	gau	1,6695	1,8142	238,5381	1,3416	0,3017	0,9962	32,5123	strong

PC: Componente principal; PC₁: Componente principal 1; PC₂: Componente principal 2; Modelo exp: exponencial; Modelo sph: esférico; Modelo gau: gaussiano; RMSE: Raiz do Erro Médio Quadrático; R^2 : Coeficiente de Determinação; MSDR: Coeficiente do Desvio Quadrático Médio; SDI: Índice de dependência espacial (SEIDEL; OLIVEIRA, 2014); SDI Clas.: Classificação da dependência espacial (SEIDEL; OLIVEIRA, 2016).

Conforme o mapa obtido aplicando-se o método de interpolação pela Ponderação do Inverso das Distâncias (*Inverse Distance Weighting - IDW*) (Figura 13), para o componente principal 1, observa-se a presença de diferentes padrões de variabilidade, basicamente, divididos em cinco classes, com valores variando de -3,07 a 3,04, Figura 13 (A). Valores negativos tendem a representar regiões onde alguns parâmetros da PC₁ apresentam maior expressão positiva, face sul do experimento, enquanto que regiões com valores positivos, variáveis com maiores efeitos negativos, face norte. Na etapa seguinte, obteve-se o georreferenciamento das 555 parcelas presentes no campo experimental (Figura 13, B), e então, procedeu-se com a sobreposição junto as informações de variabilidade obtidas por meio da interpolação

(Figura 13, C). Nota-se que com a remoção das áreas sem a sobreposição das parcelas experimentais a amplitude dos valores da PC1 foi alterada, passando a variar de -4,98 a 6,29 (Figura 13, C). Após este procedimento o valor médio da PC1, para cada parcela, foi obtido para então proceder com as análises posteriores.

Figura 13. (A) Mapa temático gerado pela Ponderação do Inverso das Distâncias (*Inverse Distance Weighting - IDW*) para o componente principal 1 (PC₁), Área experimental 1. (B) Parcelas experimentais georreferenciadas. (C) Sobreposição das parcelas experimentais sobre o mapa interpolado e obtenção dos valores médios dos pontos de interesse.

Resultados semelhantes foram obtidos considerando o componente principal 2 (PC₂). De acordo com o mapa interpolado, observa-se a presença de diferentes padrões de variabilidade, também divididos em cinco classes, com valores variando de -1,74 a 2,05 (Figura 14, A). Valores negativos (face norte) e positivos (face sul) possuem o mesmo significado conforme especificado anteriormente. As mesmas 555 parcelas georreferenciadas e sobrepostas sobre o mapa interpolado da PC₁ também foram considerados para a PC₂ (Figura 14, B). Neste novo cenário, nota-se que com a remoção das áreas sem a sobreposição das parcelas experimentais a amplitude dos valores da PC₂ foi alterada, passando a variar de -4,74 a 3,92 (Figura 14, C). Após este procedimento o valor médio da PC₂ para cada parcela foi obtido para então proceder com as análises posteriores.

Figura 14. (A) Mapa temático gerado pela Ponderação do Inverso das Distâncias (*Inverse Distance Weighting - IDW*) para o componente principal 2 (PC_2), Área experimental 1. (B) Parcelas experimentais georreferenciadas. (C) Sobreposição das parcelas experimentais sobre o mapa interpolado e obtenção dos valores médios dos pontos de interesse.

5.1.2. Condutividade elétrica aparente do solo (CEa)

Para a condutividade elétrica aparente do solo, obtiveram-se 17.694 pontos de leitura dentro da Área experimental 1, tanto na profundidade de 0,375 m como para 0,75 m, Figuras 15 (A) e 16 (A). Devido à grande densidade de pontos de leituras, optou-se pela não realização do processo de interpolação e geração de mapas temático, uma vez que esse procedimento não é recomendado quando se tem esse cenário de alta densidade de informações. Portanto, considerou-se os dados brutos das leituras coletadas em campo. Conforme os dados obtidos para a profundidade de 0,375 m, observa-se a presença de diferentes padrões de variabilidade, basicamente, divididos em cinco classes, com valores variando de -89 a 458 (Figura 15, A). Valores negativos representam regiões de baixa condutividade elétrica, enquanto que regiões com valores positivos, indicam o oposto. Após a remoção das áreas sem a sobreposição das parcelas experimentais e, por consequência, dos pontos de leituras, a amplitude dos valores da CEa foi alterada, passando a variar de -89 a 354; o total de pontos de leitura na área experimental foi reduzido para 11.722 (Figura 15, C). Após este procedimento o valor médio da CEa, na profundidade de 0,375 m, para cada parcela experimental foi obtido para então proceder com as análises posteriores.

Figura 15. (A) Mapa dos 17.694 pontos de leituras de condutividade elétrica aparente do solo (CEa) obtidos pelo sensor EM38-MK2® na camada de 0,375 m, Área experimental 1. (B) Parcelas experimentais georreferenciadas. (C) Sobreposição das parcelas experimentais sobre o mapa de CEa, 11.722 pontos de leitura e obtenção dos valores médios dos pontos de interesse.

Figura 16. (A) Mapa dos 17.694 pontos de leituras de condutividade elétrica aparente do solo (CEa) obtidos pelo sensor EM38-MK2® na camada de 0,75 m, Área experimental 1. (B) Parcelas experimentais georreferenciadas. (C) Sobreposição das parcelas experimentais sobre o mapa de CEa, 11.722 pontos de leitura e obtenção dos valores médios dos pontos de interesse.

Para os dados obtidos na profundidade de 0,75 m, observa-se também a presença de diferentes padrões de variabilidade, porém, em diferentes proporções e intensidades (Figura 16, A). Divididos em cinco classes, os valores apresentam variações de -161 a 98. Assim como apresentado para a profundidade de 0,375 m, valores negativos representam regiões de baixa condutividade elétrica, enquanto que regiões com valores positivos, indicam o oposto.

Após a remoção das áreas sem a sobreposição de parcelas experimentais e, por consequência, os pontos de leituras, a amplitude dos valores da CEa foi alterada, passando a variar de -23,4 a 97,8. Assim como para a profundidade de 0,375 m, o número total de pontos de leituras foi reduzido para 11.722 (Figura 16, C). Após este procedimento o valor médio da CEa, na profundidade de 0,75 m, para cada parcela experimental foi obtido para então proceder com as análises posteriores.

5.1.3. Índice de vegetação

Para o índice de vegetação da diferença normalizada (*NDVI*), inicialmente, obteve-se o ortomosaico da Área experimental 1, no formato RGB (banda *Red*, *Green* e *Blue*), através do voo com a RPA (Figura 17, A). De modo geral, em uma avaliação visual, observa-se alto vigor vegetativo das plantas presentes no campo experimental, com a não observação de padrões de variabilidade. Após o processo de renderização da banda RGB para a cor falsa (*singleband pseudocolor*), e o cálculo do *NDVI*, obteve-se o mapa com coloração RdYlGn (*Red*, *Yellow* e *Green*) dividido em cinco classes (Figura 17, B). Valores próximos a 1 obtiveram cores verde, enquanto que, classes intermediárias, apresentaram tons de amarelo, com valores ao redor de 0, e colorações avermelhadas para valores próximos a -1. Por se tratar de um índice de alta correlação com a intensidade da atividade de clorofila nas plantas, pode-se inferir que regiões de coloração mais próximas ao verde apresentam plantas com maior vigor vegetativo quando comparado com áreas avermelhadas. Neste sentido, observa-se que a intensidade da coloração avermelhada aumenta, principalmente, na face sul do experimento. Portanto, diferente do observado na imagem RGB, fica evidente a presença de regiões com padrões de variabilidade.

Após o procedimento de sobreposição das 555 parcelas georreferenciadas sobre o mapa de *NDVI* gerado, determinaram-se as regiões de interesse onde obtiveram-se os valores médios do índice para cada parcela, os quais foram, posteriormente, utilizados para as análises seguintes (Figura 17, C e D).

Figura 17. (A) Ortomosaico da Área experimental 1 no formato RGB (banda *Red*, *Green* e *Blue*), obtido através do voo com a RPA. (B) Mapa gerado após o cálculo do índice de vegetação da diferença normalizada (*NDVI*). (C) Sobreposição das parcelas experimentais sobre o mapa *NDVI*. (D) Obtenção dos valores médios de *NDVI* dentro de cada parcela experimental.

5.1.4. Ajuste dos modelos genético-estatísticos

Antes do procedimento de ajuste do modelo genético-estatístico, obtiveram-se os resultados da análise exploratória dos dados da Área experimental 1, de modo que foi possível observar a presença de diferentes padrões de variabilidade dos dados brutos de TCH coletados em campo. De acordo com a Figura 18 (A), observam-se variações

nos valores de TCH para as diferentes parcelas experimentais (famílias) ao longo do experimento, sendo as maiores produtividades indicadas pela coloração azul escuro (máximo = 160, tons/ha) e as menores pela coloração avermelhada (mínimo = 0, tons/ha). Vale destacar que apenas três parcelas experimentais obtiveram resultados nulos devido a perda de dados. Na Figura 18 (B), é possível analisar os padrões de pontos espaciais para TCH. O valor deste parâmetro corresponde à tonalidade de cor associado à escala da direita. Regiões de coloração amarelada indicam áreas de maiores produtividades, enquanto que colorações azuladas indicam o oposto. Fica evidente que produtividades maiores se concentram na face norte, ao passo que as menores produtividades são observadas com maiores frequências na região sul da área experimental, indicando uma possível influência ambiental na expressão do caráter fenotípico.

Figura 18. (A) Valores dos dados brutos de toneladas de cana por hectare (TCH, t ha⁻¹) de cada parcela experimental (família). (B) Visualização do padrão espacial dos valores de TCH; Área experimental 1.

A próxima etapa de análises consistiu na obtenção das médias genotípicas removendo as variações observadas no experimento. Para tanto, utilizando a abordagem de modelos mistos foi possível estudar as estruturas de VCOV entre os diferentes fatores, o que na prática reflete a modelagem mais elaborada dos efeitos que ocorrem em campo. Logo, o resíduo foi testado assumindo variância ambiental constante, i.e., modelo identidade (ID), versus modelo com dependência espacial bidimensional (linhas e colunas). Além destas estruturas de VCOV, as variáveis ambientais criadas nos passos anteriores foram utilizadas como variáveis adicionais visando remover variações não controladas pelo delineamento experimental. Neste caso, adotou-se o processo de seleção de variáveis do tipo “forward”, onde todas as informações foram incluídas de forma individualizada, i.e., *NDVI*: Índice de Vegetação da Diferença Normalizada; *Eca05*: condutividade elétrica aparente do solo a 0,375 m; *Eca1*: condutividade elétrica aparente do solo a 0,75 m; *PC1*: componente principal 1; *PC2*: componente principal 2 (Tabela 3).

Considerando o modelo ID, dentre as cinco variáveis estudadas a de maior associação com a variável resposta foi o componente principal 1 (*PC1*). Esta variável foi a única com nível de significância abaixo de 0,05 (0,011) e, portanto, a primeira e única selecionada. As demais variáveis, *NDVI*, *Eca05* (profundidade de 0,375 m) e *Eca1* (profundidade de 0,75 m), e o componente principal 2 (*PC2*), após testadas apresentaram valores não significativos de chi-quadrado de Pearson, 0,783, 0,555, 0,930 e 0,398, respectivamente (Tabela 3). O mesmo ocorreu para o modelo AR1 x AR1, dentre as cinco variáveis estudadas a de maior associação com a variável resposta foi o componente principal 1 (*PC1*), demonstrando consistência da variável, mesmo considerando estruturas residuais distintas. Esta variável foi a única com nível de significância abaixo de 0,05 (0,012) e, portanto, a primeira e única selecionada. As demais variáveis, *NDVI*, *Eca05* (profundidade de 0,375 m) e *Eca1* (profundidade de 0,75 m), e o componente

principal 2 (PC₂), após testadas apresentaram valores de chi-quadrado de Pearson de 0,770, 0,626, 0,877 e 0,295, respectivamente (Tabela 3).

Tabela 3. Seleção de covariáveis para o modelo identidade (ID) e autorregressivo de primeira ordem (AR1 x AR1) utilizando o método “*forward*”, através da inclusão das diferentes informações de solo e dos sensores de forma individualizada. *NDVI*: Índice de Vegetação da Diferença Normalizada; ECa05: condutividade elétrica aparente do solo a 0,375 m; ECa1: condutividade elétrica aparente do solo a 0,75 m; PC₁: componente principal 1; PC₂: componente principal 2. Informações referentes a Área experimental 1.

Modelo	Variável	Estatística wald	d.f.	chi pr
ID	<i>NDVI</i>	0,08	1	0,783
	ECa05	0,35	1	0,555
	ECa1	0,01	1	0,930
	PC₁	6,49	1	0,011*
	PC ₂	0,72	1	0,398
AR1 x AR1	<i>NDVI</i>	0,09	1	0,770
	ECa05	0,24	1	0,626
	ECa1	0,02	1	0,877
	PC₁	6,34	1	0,012*
	PC ₂	1,10	1	0,295

*Variável significativa considerando um nível de significância de 0,05.

Na etapa seguinte, foram obtidos os resultados da inclusão das demais variáveis já considerando a presença da PC₁, para ambos os modelos (Tabela 4); nota-se que nenhuma das variáveis foram explicativas. Para o modelo ID, quando incluído a variável *NDVI*, o valor de chi-quadrado (chi pr) foi de 0,713, enquanto que para ECa05, ECa1 e PC₂, os valores foram de 0,728, 0,812 e 0,945, respectivamente. Considerando o modelo AR1 x AR1, quando incluído a variável *NDVI*, o valor de chi-quadrado (chi pr) foi de 0,541, enquanto que para ECa05, ECa1 e PC₂, os valores foram de 0,759, 0,807 e 0,983, respectivamente.

Tabela 4. Inclusão de novas variáveis para o modelo identidade (ID) e autorregressivo de primeira ordem (AR1 x AR1) já considerando a variável PC₁; *NDVI*: Índice de Vegetação da Diferença Normalizada; ECa05: condutividade elétrica aparente do solo a 0,375 m; ECa1: condutividade elétrica aparente do solo a 0,75 m; PC₁: componente principal 1; PC₂: componente principal 2. Informações referentes a Área experimental 1.

Modelo	Variável	Estatística <i>wald</i>	d.f.	chi pr
ID	PC₁	6,34	1	0,012*
	<i>NDVI</i>	0,14	1	0,713
	PC₁	6,41	1	0,011*
	ECa05	0,12	1	0,728
	PC₁	6,44	1	0,011*
	ECa1	0,06	1	0,812
	PC₁	6,38	1	0,012*
	PC ₂	0,00	1	0,945
AR1 x AR1	PC₁	6,32	1	0,012*
	<i>NDVI</i>	0,37	1	0,541
	PC₁	6,31	1	0,012*
	ECa05	0,09	1	0,759
	PC₁	6,31	1	0,012*
	ECa1	0,06	1	0,807
	PC₁	6,32	1	0,012*
	PC ₂	0,00	1	0,983

*Variável significativa considerando um nível de significância de 0,05.

As estimativas dos componentes de variância e dos parâmetros genéticos considerando os modelos: identidade (ID); identidade, incluindo a covariável significativa componente principal 1 (ID+PC₁); autorregressivo de ordem 1 (AR1 x AR1); autorregressivo de ordem 1, incluindo a covariável significativa componente principal 1 (AR1 x AR1+PC₁) para o caráter toneladas de cana por hectare (TCH), estão apresentados na Tabela 5.

Para todos os modelos observa-se que os coeficientes de variação residual (CV_e) apresentaram valores entre 10 e 15%. Reduções nos valores desse parâmetro foram obtidos quando a covariável significativa (PC₁) foi incluída, tanto para o modelo ID como para AR1 x AR1. Com base nos valores obtidos de coeficiente de variação genética total (CV_g), somente os modelos AR1 x AR1 e AR1 x AR1+PC₁ apresentaram resultados acima de 10% para TCH, 10,70 e 11,02%, respectivamente. Em contrapartida, valores inferiores a 10% podem ser observados quando considerados os modelos ID (9,40%) e ID+PC₁ (9,62%).

As estimativas dos valores de herdabilidade no sentido amplo (h^2_a) para TCH, considerando os modelos ID e ID+PC₁, apresentaram valores próximos a 0,50. Por outro lado, para ambos os modelos AR1 x AR1 e AR1 x AR1+PC₁ os valores foram superiores a 0,60. Fica evidente que a estimativa de h^2_a é aumentada com a inclusão da covariável PC₁. Para a acurácia seletiva (AC_s), obtiveram-se valores entre 0,70 e 0,80 considerando todos os modelos. Vale destacar que, assim como para a h^2_a , a inclusão da variável PC₁ elevou as estimativas de AC_s.

Considerando as diferentes estruturas de VCOV, observa-se que os menores valores obtidos para os critérios de AIC e BIC, independente da inclusão ou não das variáveis ambientais, foram para o modelo AR1 x AR1+PC₁, Tabela 12 (Apêndice).

Tabela 5. Estimativas dos componentes de variância (REML) considerando os modelos: identidade (ID); identidade, incluindo a covariável significativa, componente principal 1 (ID+PC₁); autorregressivo de ordem 1 (AR1 x AR1); autorregressivo de ordem 1, incluindo a covariável significativa, componente principal 1 (AR1 x AR1+PC₁). O caráter principal considerado foi toneladas de cana por hectare (TCH).

Modelo	BLUP (médio)	V _e	V _g	V _f	CV _e (%)	CV _g (%)	CV _f (%)	h ² _a	AC _s
ID	119,53	264,20	125,30	389,50	13,60	9,40	16,50	0,49	0,70
ID+PC ₁	119,65	256,60	132,50	289,10	13,40	9,62	14,21	0,51	0,71
AR1 x AR1	120,20	223,69	165,50	389,19	12,44	10,70	16,41	0,60	0,77
AR1 x AR1+PC ₁	120,12	211,60	175,30	386,90	12,11	11,02	16,36	0,62	0,79

V_e: variância ambiental; V_g: variância genética; V_f: variância fenotípica; CV_e(%): coeficiente de variação ambiental; CV_g(%): coeficiente de variação genética total; CV_f(%): coeficiente de variação fenotípica; h²_a: herdabilidade no sentido amplo em nível de média de família; AC_s: acurácia seletiva.

5.2. Análise exploratória dos dados da Área experimental 2

Os resultados da análise de solo para os atributos químicos, macronutrientes: pH, M.O., P em resina, Al, m%, H+Al, K, Ca, Mg, SB, CTC e V%, e as características físicas do solo: teores de areia (fina, grossa e total), silte e argila, obtidos a partir de pontos amostrais, na profundidade de 0 a 20 cm, estão apresentados na Tabela 6. Diferente da Área experimental 1, não se identificou a presença de valores atípicos (*outliers*) no conjunto de dados, portanto, considerou-se todos os 68 pontos amostrais (Figura 28) (Apêndice D).

De todos os parâmetros avaliados, os maiores valores desvio padrão (D.P.) foram obtidos para os atributos físicos do solo. Em ordem de maior dispersão dos dados, destacam-se: areia grossa (32,55), areia fina (29,35), argila (19,19), silte (17,38) e areia total (15,89). Por outro lado, assim como observado para a Área experimental 1, baixos valores de desvio padrão, portanto, pequena variação em relação à média, foram obtidos para o alumínio (0,12), potássio (0,16), pH (0,41) e, adicionalmente para a Área experimental 2, porcentagem de saturação por alumínio (0,66). Considerando o coeficiente de variação (CV), os maiores valores encontrados foram para o silte (48,29%), porcentagem de saturação por alumínio (47,83%), alumínio (39,79%) e acidez potencial (34,88%) (Tabela 6).

Tabela 6. Estatística descritiva dos atributos físicos e químicos do solo na camada de 0 a 20 cm de profundidade, Área experimental 2.

Atributo	Profundidade 0 a 20 cm					
	Mínimo	Máximo	Mediana	Média	D.P.	CV (%)
Argila	72,00	158,00	112,00	112,88	19,19	17,00
Areia grossa	526,00	665,00	605,00	605,60	32,55	5,37
Areia fina	176,00	311,00	241,50	245,60	29,35	11,95
Areia total	807,00	886,00	852,50	851,10	15,89	6,47
Silte	6,00	85,00	38,50	35,99	17,38	48,29
P	7,00	29,00	17,00	17,12	4,49	26,25
M.O.	5,00	12,00	8,00	8,22	1,54	18,78
pH	4,70	6,50	5,50	5,49	0,41	7,43
K	0,60	1,30	0,90	0,92	0,16	17,76
Ca	10,00	37,00	15,00	15,71	4,02	25,61
Mg	3,00	10,00	6,00	5,62	1,22	21,75
H+Al	10,00	52,00	13,50	14,21	4,96	34,88
Al	0,10	0,60	0,30	0,29	0,12	39,79
SB	13,90	41,80	21,75	22,25	4,63	20,83
CTC	29,00	75,70	35,35	36,45	6,53	17,92
V%	31,30	76,30	61,25	61,06	7,20	11,80
m%	0,30	3,20	1,30	1,38	0,66	47,83

Fósforo (P); Matéria orgânica (M.O.); pH; potássio (K); cálcio (Ca); magnésio (Mg); acidez potencial (H+Al); alumínio (Al); saturação por base (SB); capacidade de troca de cátions (CTC); percentual de saturação por base (V); porcentagem de saturação por alumínio (m); D.P. – desvio padrão; CV – coeficiente de variação.

Do mesmo modo como apresentado para a Área experimental 1, de acordo com os resultados da análise de componentes principais (PCA), considerando as 17 variáveis físicas e químicas do solo, observa-se que a quantidade de variação explicada cai consideravelmente após o primeiro componente (Figura 19, A). Isso sugere que apenas dois, ou até mesmo três, componentes podem ser suficientes para resumir o conjunto de dados. Considerando apenas os componentes 1 (PC₁) e 2 (PC₂), observa-se que ambas foram as que capturaram maiores porcentagens da variação, sendo responsáveis por 26,9% e 17,2%, respectivamente. Portanto, 44,1% das diferentes influências das características originais são explicadas somente pelas PC₁ e PC₂. Se o terceiro componente for considerado, o valor passa para 56%. O primeiro componente principal (PC₁) foi essencialmente explicado pelos maiores valores positivos das variáveis porcentagem de saturação por alumínio, alumínio e argila, em contraste com as variáveis de maiores valores negativos, saturação por base, cálcio, percentual de saturação por base, magnésio, pH, silte e capacidade de troca de cátions. O segundo componente (PC₂) foi explicado, principalmente, pelos maiores valores positivos das variáveis areia fina, percentual de saturação por base, matéria orgânica e pH, em contraste com os maiores valores negativos das variáveis areia grossa, acidez potencial e capacidade de troca de cátions. Por outro lado, o terceiro componente (PC₃) foi explicado pelo maior valor positivo da variável areia total, em contraste com o maior valor negativo da variável argila.

Por meio da visualização bidimensional dos dados é possível identificar o agrupamento das diferentes variáveis apresentando distintos padrões de similaridades (Figura 19, B). Dentre os parâmetros que apresentam alta correlação entre si, i.e., menores ângulos formados entre os vetores, destacam-se: alumínio (Al) com porcentagem de saturação por alumínio (m), pH com percentual de saturação por base (V), magnésio (Mg) com silte, cálcio (Ca) com saturação por base (SB) e fósforo (P) com capacidade de troca de cátions (CTC). Por outro lado, observa-se a formação de ângulo de 90° entre alguns vetores, como é o caso da areia fina com porcentagem de saturação por alumínio (m) e

alumínio (Al) e matéria orgânica (M.O.) com capacidade de troca de cátions (CTC), logo, essas variáveis apresentam correlação nula. Ainda, variáveis negativamente correlacionadas entre si são evidenciadas visto a formação de ângulos próximos a 180°, i.e., alumínio (Al) e porcentagem de saturação por alumínio (m) com magnésio (Mg), silte, saturação por base (SB) e cálcio (Ca).

Figura 19. (A) Gráfico *scree* dos autovalores dos componentes principais e (B) Gráfico bi-dimensional (*biplot*) dos componentes principais 1 e 2 (PC₁ e PC₂) das 17 variáveis físicas e químicas do solo, amostradas em 68 pontos, na profundidade de 0 a 20 cm, na Área experimental 2.

As estimativas dos coeficientes de correlação de Pearson (r) entre as diferentes variáveis de solo estudadas, em combinação com os valores obtidos por meio dos diferentes sensores proximais e componentes principais, na Área experimental 2, são apresentadas na Figura 20 (A) e (B). Dentre os componentes físicos do solo, não se observam valores positivos altamente correlacionados ($r > 0,90$), apenas negativo entre a areia fina e areia grossa ($r = -0,90$). Adicionalmente, destacam-se os valores negativos de moderada correlação entre silte e argila ($r = -0,68$), argila e areia total ($r = -0,56$) e matéria orgânica e acidez potencial ($r = -0,55$). As demais correlações são fracas ou desprezíveis.

Dentre os atributos químicos do solo, destaca-se o valor positivo altamente correlacionado ($r > 0,90$) entre os elementos cálcio e saturação por base ($r = 0,99$). Ademais, valores positivos com forte correlação são observados, como é o caso dos atributos pH e percentual de saturação por base ($r = 0,83$), percentual de saturação por base e magnésio ($r = 0,77$), percentual de saturação por base e saturação por base ($r = 0,80$), magnésio e cálcio ($r = 0,81$), magnésio e saturação por base ($r = 0,88$), cálcio e capacidade de troca de cátions ($r = 0,73$), saturação por base e capacidade de troca de cátions ($r = 0,72$) e capacidade de troca de cátions e acidez potencial ($r = 0,80$). Em contrapartida, valores negativos com forte correlação também foram obtidos, i.e., pH e alumínio ($r = -0,77$), percentual de saturação por base e porcentagem de saturação por alumínio ($r = -0,73$), magnésio e porcentagem de saturação por alumínio ($r = -0,71$), cálcio e porcentagem de saturação por alumínio ($r = -0,82$) e saturação por base e porcentagem de saturação por alumínio ($r = -0,83$). Valores positivos, porém, de moderada correlação, foram obtidos para pH e magnésio ($r = 0,60$), pH e cálcio ($r = 0,62$), pH e saturação por base ($r = 0,64$), pH e porcentagem de saturação por alumínio ($r = 0,60$), magnésio e capacidade de troca de cátions ($r = 0,59$), cálcio e fósforo ($r = 0,54$), saturação por base e fósforo ($r = 0,54$) e fósforo e capacidade de troca de cátions ($r = 0,60$). Por fim, valores negativos com moderada correlação foram observados entre pH e porcentagem de saturação por alumínio ($r = -0,60$), percentual de saturação por base e alumínio ($r = -0,56$), percentual de saturação por base e acidez potencial ($r = -0,68$), magnésio e alumínio ($r = -0,62$), cálcio e

alumínio ($r = -0,62$), saturação por base e alumínio ($r = -0,62$) e percentual de saturação por base e capacidade de troca de cátions ($r = -0,52$).

Figura 20. (A) Correlações entre as diferentes variáveis físicas e químicas do solo: areia grossa (areia_g); areia fina (areia_f); areia total (areia_t); silte; fósforo (P); matéria orgânica (mo); pH; potássio (K); cálcio (Ca); magnésio (Mg); acidez potencial (Hal); alumínio (Al); saturação por base (SB); capacidade de troca de cátions (CTC); percentual de saturação por base (V); porcentagem de saturação por alumínio (m), com os diferentes sensores: condutividade elétrica aparente do solo a 0,75 m (Eca_1m); condutividade elétrica aparente do solo a 0,375 m (Eca_05m); com os componentes principais: componente principal 1 (PC₁); componente principal 2 (PC₂); componente principal 3 (PC₃), combinado com o teste de significância considerando um nível de 0,05% ($P < 0,05$). (B) Valores de correlação entre as diferentes variáveis estudadas. Informações referente a Área experimental 2.

Em relação aos componentes principais, valores positivos altamente correlacionados ($r > 0,90$) podem ser observados entre o componente principal 1 (PC₁) e porcentagem de saturação por alumínio ($r = 0,92$), componente principal 2 (PC₂) e areia fina ($r = 0,91$), e o componente principal 3 (PC₃) e areia total ($r = 0,83$). De outro modo, valores negativos altamente correlacionados foram obtidos entre a PC₁ e cálcio ($r = -0,94$) e saturação por base ($r = -0,95$), e PC₂ com areia grossa ($r = -0,90$). Destaca-se ainda o valor positivo com forte correlação obtido apenas entre a PC₁ e alumínio ($r = 0,78$). Correlações negativas com forte correlação também foram obtidas entre a PC₁ e silte ($r = -0,76$), pH ($r = -0,82$), percentual de saturação por base ($r = -0,87$) e magnésio ($r = -0,86$), PC₂ e acidez total ($r = -0,83$) e PC₃ e argila ($r = -0,78$). Complementarmente, valores positivos de correlação moderada foram obtidos entre a PC₂ e matéria orgânica ($r = 0,61$) e com a condutividade elétrica aparente do solo a 0,75 m ($r = -0,57$). Por fim, valores negativos de correlação moderada também foram obtidos entre PC₁ e capacidade de troca de cátions ($r = -0,57$) e PC₂ com fósforo ($r = -0,59$) e capacidade de troca de cátions ($r = -0,68$). Devido à baixa correlação encontrada entre o componente principal 3 e os diferentes atributos do solo, além da baixa contribuição na explicação da variabilidade dos dados, optou-se por proceder as próximas etapas do estudo considerando apenas os componentes 1 e 2, da mesma forma como conduzido para a Área experimental 1.

Quanto ao sensor de condutividade elétrica aparente do solo, diferentes valores de correlações com os atributos físicos e químicos do solo são observados quando as leituras são realizadas em profundidades distintas. Da

mesma forma como obtido para a Área experimental 1, nota-se que leituras a 0,375 m se correlacionam fracamente com todos os atributos do solo, com exceção da acidez potencial e pH que apresentaram correlação positiva ($r = 0,54$) e negativa ($r = -0,51$) moderada, respectivamente. No entanto, a 0,75 m de profundidade, correlação positiva moderada é observada com a areia fina ($r = 0,62$) e negativa com areia grossa ($r = -0,53$).

5.2.1. Análise geoestatística do conjunto de dados

Para a identificação e classificação do tipo de dependência espacial e posterior construção dos mapas temáticos para os componentes principais consideradas (PC₁ e PC₂), obtiveram-se os cálculos do variograma experimental para os ajustes dos modelos teóricos ao conjunto de dados (Tabela 7). Para a PC₁, o maior valor de alcance observado foi de 52,69 m, quando considerado o modelo esférico (sph). Por outro lado, o menor alcance foi de 6,66 m para o modelo exponencial (exp), que por sua vez, foi o modelo teórico de melhor ajuste ao conjunto de dados, uma vez que apresentou o menor valor pelo critério RMSE (2,1537). Adicionalmente, tanto o coeficiente de determinação ($R^2 = 0,5986$) como o coeficiente do desvio quadrático médio (MSDR = 0,9999) apresentaram os maiores valores. Os diferentes variogramas experimentais para a componente principal 1, considerando os modelos esférico (sph), exponencial (exp) e gaussiano (gaus) podem ser melhor observados na Figura 31 (Apêndice G). De acordo com o índice de dependência espacial (SDI = 1,0024%), proposto por Seidel e Oliveira (2014), para o modelo exponencial, a PC₁ foi categorizada como sendo de fraca dependência (SEIDEL; OLIVEIRA, 2016).

Para a PC₂, o maior valor de alcance obtido foi de 19881,59 m considerando o modelo exponencial (exp). No entanto, o menor valor observado para este mesmo parâmetro foi de 56,43 m para o modelo gaussiano (gaus). Diferentemente da PC₁, o modelo teórico de melhor ajuste para a PC₂ foi o esférico (sph), pois foi o que apresentou o menor valor de RMSE (1,5166) e o maior valor de R^2 (0,2076). Apesar do coeficiente do desvio quadrático médio não ter sido o de maior valor (MSDR = 1,0604), observa-se que a diferença entre os demais é insignificante. Os diferentes variogramas experimentais para a componente principal 2, considerando os modelos esférico (sph), exponencial (exp) e gaussiano (gaus) podem ser melhor observados na Figura 32 (Apêndice H). De acordo com o índice de dependência espacial (SDI = 1956,17%), proposto por Seidel e Oliveira (2014), para o modelo esférico, a PC₂ foi categorizada como sendo de forte dependência (SEIDEL; OLIVEIRA, 2016).

Conforme o mapa obtido aplicando-se o método de interpolação pela Ponderação do Inverso das Distâncias (*Inverse Distance Weighting - IDW*) (Figura 21), para o componente principal 1, observa-se a presença de diferentes padrões de variabilidade, basicamente, divididos em cinco classes, com valores variando de -2,47 a 2,59, Figura 21 (A). Assim como para a Área experimental 1, valores negativos tendem a representar regiões onde alguns parâmetros da PC₁ apresentam maior expressão positiva, enquanto que regiões com valores positivos, variáveis com maiores efeitos negativos. Na etapa seguinte, obteve-se o georreferenciamento das 761 parcelas presentes no campo experimental (Figura 21, B), e então, procedeu-se com a sobreposição junto às informações de variabilidade obtidas por meio da interpolação (Figura 21, C). Nota-se que com a remoção das áreas sem a sobreposição das parcelas experimentais a amplitude dos valores da PC₁ foi alterada, passando a variar de -5,22 a 4,66 (Figura 21, C). Após este procedimento o valor médio da PC₁, para cada parcela, foi obtido para então proceder com as análises posteriores.

Tabela 7. Parâmetros dos modelos teóricos ajustados aos variogramas experimentais do conjunto de dados na profundidade de 0 a 20 cm, Área experimental 2.

Local	PC	Modelo	Efeito Pepita	Patamar (C1)	Alcance	RMSE	R ²	MSDR	SDI	SDI Clas.
Área experimental 2	PC ₁	exp	0,0000	4,5726	6,6577	2,1537	0,5986	0,9999	1,0024	weak
		sph	4,3171	0,2563	52,6892	2,1538	0,5331	1,0000	0,5260	weak
		gau	0,0000	4,5714	14,8481	2,1526	0,2767	0,9995	3,5545	weak
	PC ₂	exp	1,4913	182,9979	19881,5906	1,5167	0,2075	1,0602	2969,3416	strong
		sph	1,4914	68,7097	11220,8529	1,5166	0,2076	1,0604	1956,1667	strong
		gau	1,0719	2,3574	56,4333	1,5265	0,1979	1,0728	9,2868	moderate

PC: Componente principal; PC₁: Componente principal 1; PC₂: Componente principal 2; Modelo exp: exponencial; Modelo sph: esférico; Modelo gau: gaussiano; RMSE: Raiz do Erro Médio Quadrático; R²: Coeficiente de Determinação; MSDR: Coeficiente do Desvio Quadrático Médio; SDI: Índice de dependência espacial (SEIDEL; OLIVEIRA, 2014); SDI Clas.: Classificação da dependência espacial (SEIDEL; OLIVEIRA, 2016).

Resultados semelhantes foram obtidos considerando o componente principal 2 (PC₂). De acordo com o mapa interpolado, observa-se a presença de diferentes padrões de variabilidade, também divididos em cinco classes, com valores variando de -2,66 a 1,87 (Figura 22, A). Valores negativos (face sul-sudeste) e positivos (face oeste) possuem o mesmo significado conforme especificado anteriormente. As mesmas 761 parcelas georreferenciadas sobrepostas sobre o mapa interpolado da PC₁ também foram considerados para a PC₂ (Figura 22, B). Neste novo cenário, nota-se que com a remoção das áreas sem a sobreposição das parcelas experimentais a amplitude dos valores da PC₂ foi alterada, passando a variar de -8,54 a 2,86 (Figura 22, C). Após este procedimento o valor médio da PC₂ para cada parcela foi obtido para então proceder com as análises posteriores.

Figura 21. (A) Mapa temático gerado pela Ponderação do Inverso das Distâncias (*Inverse Distance Weighting - IDW*) para o componente principal 1 (PC_1), Área experimental 2. (B) Parcelas experimentais georreferenciadas. (C) Sobreposição das parcelas experimentais sobre o mapa interpolado.

Figura 22. (A) Mapa temático gerado pela Ponderação do Inverso das Distâncias (*Inverse Distance Weighting - IDW*) para o componente principal 2 (PC₂), Área experimental 2. (B) Parcelas experimentais georreferenciadas. (C) Sobreposição das parcelas experimentais sobre o mapa interpolado.

5.2.2. Condutividade elétrica aparente do solo (CEa)

Para a condutividade elétrica aparente do solo, obtiveram-se 25.457 pontos de leitura dentro da Área experimental 2, tanto na profundidade de 0,375 m como para 0,75 m, Figuras 23 (A) e 24 (A). Assim como para a Área experimental 1, consideraram-se os dados brutos das leituras coletadas em campo em detrimento da interpolação dos dados. Conforme os dados obtidos para a profundidade de 0,375 m, observa-se a presença de diferentes padrões de variabilidade, basicamente, divididos em cinco classes, com valores variando de -17 a 58 (Figura 23, A). Valores negativos representam regiões de baixa condutividade elétrica, enquanto que regiões com valores positivos, indicam o

oposto. Após a remoção das áreas sem a sobreposição das parcelas experimentais e, por consequência, dos pontos de leituras, a amplitude dos valores da CEa foi alterada, passando a variar de -12,87 a 57,7; o total de pontos de leitura na área experimental foi reduzido para 16.617 (Figura 23, C). Após este procedimento, o valor médio da CEa, na profundidade até 0,375 m, para cada parcela experimental, foi obtido para então proceder com as análises posteriores.

Para os dados obtidos na profundidade de 0,75 m, observa-se também a presença de diferentes padrões de variabilidade, porém, em diferentes proporções e intensidades. Divididos em cinco classes, os valores apresentam variações de -11,0 a 70,5 (Figura 24, A). Assim como apresentado para a profundidade de 0,375 m, valores negativos representam regiões de baixa condutividade elétrica, enquanto que regiões com valores positivos, indicam o oposto. Após a remoção das áreas sem a sobreposição de parcelas experimentais e, por consequência, os pontos de leituras, não houve alteração na amplitude dos valores da CEa (Figura 24, C). Assim como para a profundidade de 0,375 m, o número total de pontos de leituras foi reduzido para 16.617. Após este procedimento o valor médio da CEa, na profundidade de 0,75 m, para cada parcela experimental foi obtido para então proceder com as análises posteriores.

Figura 23. (A) Mapa dos 25.457 pontos de leituras de condutividade elétrica aparente do solo (CEa) obtidos pelo sensor EM38-MK2® na camada de 0,375 m, Área experimental 2. (B) Parcelas experimentais georreferenciadas (761). (C) Sobreposição das parcelas experimentais sobre o mapa de CEa, 16.617 pontos de leitura.

Figura 24. (A) Mapa dos 25.457 pontos de leituras de condutividade elétrica aparente do solo (CEa) obtidos pelo sensor EM38-MK2® na camada de 0,75 m, Área experimental 2. (B) Parcelas experimentais georreferenciadas (761). (C) Sobreposição das parcelas experimentais sobre o mapa de CEa, 16.617 pontos de leitura.

5.2.3. Avaliação experimental e ajuste dos modelos genético-estatísticos

Assim como para a Área experimental 1, antes do procedimento de ajuste do modelo genético-estatístico, obtiveram-se os resultados da análise exploratória dos dados da Área experimental 2, de modo que foi possível observar a presença de diferentes padrões de variabilidade dos dados brutos de TCH coletados em campo. De acordo com a

Figura 25 (A), observam-se variações nos valores de TCH para as diferentes parcelas experimentais (famílias) ao longo do experimento, sendo as maiores produtividades indicadas pela coloração azul escuro (máximo = 350, tons/ha) e as menores pela coloração avermelhada (mínimo = 50, tons/ha). Vale destacar que, neste experimento, não houve dados perdidos. Na Figura 25 (B), é possível analisar os padrões de pontos espaciais para TCH. O valor deste parâmetro corresponde à tonalidade de cor associado à escala da direita. Regiões de coloração amarelada indicam áreas de maiores produtividades, enquanto que colorações azuladas indicam o oposto. Fica evidente que os menores níveis de produtividade se concentram, majoritariamente, na face sul-sudeste do experimento, ao passo que as maiores produtividades são observadas com maiores frequências na região norte da área experimental, indicando uma possível influência ambiental na expressão do caráter fenotípico.

Figura 25. (A) Valores dos dados brutos de toneladas de cana por hectare (TCH, tons/ha) de cada parcela experimental (família). (B) Visualização do padrão espacial dos valores de TCH; Área experimental 2.

A próxima etapa de análises foi a mesma considerada na Área experimental 1. Logo, as variáveis ambientais criadas nos passos anteriores foram utilizadas como variáveis adicionais visando remover variações não controladas pelo delineamento experimental. Neste caso, adotou-se, também, o processo de seleção de variáveis do tipo “*forward*”, onde todas as informações foram incluídas de forma individualizada, i.e., ECa05: condutividade elétrica aparente do solo a 0,375 m; ECa1: condutividade elétrica aparente do solo a 0,75 m; PC₁: componente principal 1; PC₂: componente principal 2 (Tabela 8).

Considerando o modelo ID, dentre as quatro variáveis estudadas a de maior associação com a variável resposta foi a condutividade elétrica aparente do solo a 0,375 m (ECa05). Esta variável foi a única com nível de significância abaixo de 0,05 (0,001) e, portanto, a primeira e única selecionada. As demais variáveis, ECa1 (profundidade de 0,75 m), e os componentes principais 1 (PC₁) e 2 (PC₂), após testadas apresentaram valores não significativos de chi-quadrado de Pearson, 0,474, 0,537, 0,212, respectivamente (Tabela 8). O mesmo ocorreu para o modelo AR1 x AR1, dentre as quatro variáveis estudadas a de maior associação com a variável resposta foi a condutividade elétrica aparente do solo a 0,375 m (ECa05), demonstrando consistência da variável, mesmo considerando estruturas residuais distintas. Esta variável foi a única com nível de significância abaixo de 0,05 (0,002) e, portanto, a primeira e única selecionada. As demais variáveis, ECa1 (profundidade de 0,75 m), componente principal 1 (PC₁) e o componente principal 2 (PC₂), após testadas apresentaram valores de chi-quadrado de Pearson de 0,976, 0,825 e 0,092, respectivamente (Tabela 8).

Tabela 8. Seleção de covariáveis para o modelo identidade (ID) e autorregressivo de primeira ordem (AR1 x AR1) utilizando o método “*forward*”, através da inclusão das diferentes informações de solo e dos sensores de forma individualizada. ECa05: condutividade elétrica aparente do solo a 0,375 m; ECa1: condutividade elétrica aparente do solo a 0,75 m; PC1: componente principal 1; PC2: componente principal 2. Informações referentes a Área experimental 2.

Modelo	Variável	Estatística <i>wald</i>	d.f.	chi pr
ID	ECa05	12,82	1	0,001*
	ECa1	0,51	1	0,474
	PC ₁	0,38	1	0,537
	PC ₂	1,56	1	0,212
AR1 x AR1	ECa05	10,06	1	0,002*
	ECa1	0,00	1	0,976
	PC ₁	0,05	1	0,825
	PC ₂	2,85	1	0,092

* Variável significativa considerando um nível de significância de 0,05.

Tabela 9. Inclusão de novas variáveis para o modelo identidade (ID) e autorregressivo de primeira ordem (AR1 x AR1) já considerando a variável ECa05 (condutividade elétrica aparente do solo a 0,375 m); ECa1: condutividade elétrica aparente do solo a 0,75 m; PC1: componente principal 1; PC2: componente principal 2. Informações referentes à Área experimental 2.

Modelo	Variável	Estatística <i>wald</i>	d.f.	chi pr
ID	ECa05	12,81	1	0,001*
	ECa1	4,63	1	0,031*
	ECa05	12,82	1	0,001*
	PC ₁	0,19	1	0,665
	ECa05	12,74	1	0,001*
	PC ₂	1,25	1	0,264
AR1 x AR1	ECa05	10,21	1	0,001*
	ECa1	6,48	1	0,011*
	ECa05	10,03	1	0,002*
	PC ₁	0,07	1	0,791
	ECa05	10,10	1	0,001*
	PC ₂	2,82	1	0,093

* Variável significativa considerando um nível de significância de 0,05.

Na etapa seguinte, obtiveram-se os resultados da inclusão das demais variáveis já considerando a presença da variável ECa05, para ambos os modelos (Tabela 9). Nota-se que para o modelo ID a variável ECa1 tornou-se explicativa, apresentando valor de chi-quadrado inferior a 0,05 (0,031). Por outro lado, as demais variáveis não foram explicativas, com valores de 0,665 e 0,264 para PC₁ e PC₂, respectivamente. O mesmo cenário foi observado para o modelo AR1 x AR1, em que a única variável significativa com a inclusão da ECa05 foi a ECa1 (0,011). Os componentes 1 (PC₁) e 2 (PC₂) apresentaram valores de 0,791 e 0,093, respectivamente.

As estimativas dos componentes de variância e dos parâmetros genéticos considerando os modelos: identidade (ID); identidade, incluindo as covariáveis significativas ECa05 e ECa1 (ID+ECa05 e ECa1); autorregressivo

de ordem 1 (AR1 x AR1); autorregressivo de ordem 1, incluindo as covariáveis significativas ECa05 e ECa1 (AR1 x AR1+ECa05 e ECa1) para o caráter toneladas de cana por hectare (TCH), estão apresentados na Tabela 10.

Para todos os modelos observam-se que os coeficientes de variação residual (CV_e) apresentaram valores entre 10 e 22%. Considerando o modelo ID, pequena diminuição nos valores deste parâmetro foi obtido com a inclusão das covariáveis ambientais. Por outro lado, diminuição significativa é obtida quando o modelo AR1 x AR1 é considerado. Ademais, a inclusão de ambas as covariáveis neste modelo permitiu uma redução expressiva para este parâmetro ($CV_e = 10,66\%$). Para o coeficiente de variação genética total (CV_g), todos os modelos apresentaram resultados superiores a 10% para TCH. Ainda, ressalta-se o aumento significativo neste parâmetro quando considerados os modelos AR1 x AR1 e AR1 x AR1+ECa05 e ECa1 (Tabela 10).

A herdabilidade no sentido amplo (h^2_a) para TCH foi inferior a 0,50 somente para os modelos ID e ID+ECa05 e ECa1. No entanto, um aumento significativo é observado quando considerados os modelos AR1 x AR1 e AR1 x AR1 com a inclusão das covariáveis ECa05 e ECa1 (0,72 e 0,76, respectivamente). Para este último modelo, observou-se, também, o maior valor de acurácia seletiva ($AC_s = 0,87$) (Tabela 10).

Considerando as diferentes estruturas de VCOV, com a inclusão ou não das variáveis ambientais, observa-se que os menores valores obtidos para os critérios de AIC e BIC foram para o modelo AR1 x AR1+ECa05 e ECa1, Tabela 12 (Apêndice I).

Tabela 10. Estimativas dos componentes de variância (REML) considerando os modelos: identidade (ID); identidade, incluindo as covariáveis significativas, condutividade elétrica aparente do solo a 0,375 m e 0,75 m (ID+ECa05 e ECa1); autorregressivo de ordem 1 (AR1); autorregressivo de ordem 1, incluindo as covariáveis significativas, condutividade elétrica aparente do solo a 0,375 m e 0,75 m (AR1+ECa05 e ECa1). O caráter principal considerado foi toneladas de cana por hectare (TCH).

Modelo	BLUP (médio)	V_e	V_g	V_f	CV_e (%)	CV_g (%)	CV_f (%)	h^2_a	AC_s
ID	173,60	1379,00	371,00	1750,00	21,39	11,10	24,10	0,35	0,59
ID+ECa05 e ECa1	172,60	1340,00	359,00	1699,00	21,20	10,98	23,89	0,35	0,59
AR1 x AR1	173,27	391,3	513,20	904,50	11,40	13,06	17,36	0,72	0,85
AR1 x AR1+ECa05 e ECa1	173,50	341,8	533,7	875,50	10,66	13,32	17,05	0,76	0,87

V_e : variância ambiental; V_g : variância genética; V_f : variância fenotípica; $CV_e(\%)$: coeficiente de variação ambiental; $CV_g(\%)$: coeficiente de variação genética total; $CV_f(\%)$: coeficiente de variação fenotípica; h^2_a : herdabilidade no sentido amplo em nível de média de família; AC_s : acurácia seletiva.

6. DISCUSSÃO

No presente estudo, o uso de análises físico químicas de solo, realizada a partir de alta densidade amostral, permitiu a identificação de significativa heterogeneidade ambiental ao longo das áreas experimentais, demonstrando a importância deste tipo de estratégia para minimizar a influência ambiental na experimentação agrônômica. A variabilidade dos dados de solos foi basicamente capturada por duas variáveis de maiores escores na análise de componentes principais, tanto na Área experimental 1 (56,4%), como na Área experimental 2 (44,1%). Montanari (2015) estudando uma abordagem linear, geoestatística e multivariada em atributos do solo, também obteve dois componentes explicando 56,6% da variabilidade total dos dados. Para ambas as áreas experimentais, tanto os teores de Ca e Mg como o de pH, se agruparam em um mesmo quadrante da análise bi-dimensional, corroborando com os resultados obtidos por Freitas et al. (2014), em estudo com análises multivariadas de atributos químicos do solo para caracterização de ambientes. Vale destacar que esses nutrientes estão entre as variáveis mais explicativas da PC₁ e da PC₂ na Área experimental 1.

De acordo com Raij et al. (1997), os valores médios dos macronutrientes primários, obtidos no presente estudo, podem ser classificados em teores alto para fósforo (16 - 40 mg dm⁻³), para ambos os locais, muito baixo para potássio (0 - 0,70 mmol_c dm⁻³), Área experimental 1, e baixo para a Área experimental 2 (0,80 - 1,50 mmol_c dm⁻³), Tabelas 1 e 6. Segundo Rocha et al. (2005), solos tropicais, como é o caso das diferentes classes de solos do presente estudo, se caracterizam pelo elevado grau de intemperismo e pelos baixos teores de fósforos na forma disponível às plantas. Em contrapartida, os altos teores de fósforo observados devem-se ao histórico de adubações com constante reposição deste nutriente na solução do solo. Embora a absorção seja em pequenas quantidades pela cana-de-açúcar, se comparado com o nitrogênio e o potássio, exerce função-chave no metabolismo da planta, dentre elas, destaca-se a formação de sacarose (KORNDORFER, 2004). Vale ressaltar que apesar do alto teor de fósforo encontrado, este pode ter sua disponibilidade afetada pela presença de óxidos de ferro e alumínio, além do pH e do teor de matéria orgânica. Por outro lado, os teores de K foram baixos. Segundo Otto et al. (2010), em solos de região tropical os valores desse macronutriente são, normalmente, inferiores a 1,5 mmol_c dm⁻³, fato esse observado no presente estudo. Este nutriente, seguido pelo N, é o mais absorvido pela cana-de-açúcar (MALAVOLTA, 1994) e está diretamente relacionado com diversas funções na planta, dentre elas, resistência a geadas, seca, doenças e ao acamamento (MALAVOLTA, 1980).

Para os macronutrientes secundários da Área experimental 1, i.e., Ca e Mg, os valores foram classificados em nível médio (21 - 40 mmol_c kg⁻¹) e alto (>8 mmol_c kg⁻¹), respectivamente (RAIJ et al., 1997). Neste caso, o fornecimento de Ca está em nível ideal para garantir o estímulo e o desenvolvimento das raízes, folhas novas, colmos e meristemas (VITTI et al., 2005). O Mg, por sua vez, faz parte da molécula de clorofila, portanto, está envolvido na fotossíntese, ajudando no metabolismo do fosfato, respiração da planta e ativação de sistemas enzimáticos, assim como, na produção de açúcares (MALAVOLTA, 1997). Para a Área experimental 2, os valores de Ca e Mg foram classificados em nível baixo (0 - 20 mmol_c kg⁻¹) e médio (5 - 8 mmol_c kg⁻¹), respectivamente (RAIJ et al., 1997).

Com base na classificação de coeficiente de variação (CV), proposta por Pimentel-Gomez e Garcia (2002), a variabilidade dos parâmetros argila, pH, areia fina, grossa e total, silte, e matéria orgânica, foi baixa, com valores inferiores a 10%, para ambas as áreas experimentais. Resultados semelhantes foram obtidos por Rodriguez et al. (2012), os quais observaram baixa variabilidade para argila (6%) e areia (3%) em estudo para compreensão da relação espacial entre atributos do solo e rendimento de milho no sistema de plantio direto. Por outro lado, a variabilidade do fósforo, alumínio, potássio e porcentagem de saturação por alumínio, para a Área experimental 1, e acidez potencial, alumínio,

porcentagem de saturação por alumínio e silte, para a Área experimental 2, foram muito altas (>30%), sendo classificadas como de muito baixa precisão experimental. Valores de CV superiores a 30% para o potássio e o fósforo foram reportados por Sanches (2015) e Cerri e Magalhães (2012), ambos estudando a correlação de atributos físicos e químicos do solo em cana-de-açúcar. Ademais, Rodrigues et al. (2012) relataram valores nessa ordem para a acidez potencial e porcentagem de saturação por alumínio, corroborando com os dados obtidos em ambas as áreas experimentais deste estudo.

Os valores obtidos da reflectância espectral da cultura de rotação (*Crotalaria juncea* L.), antecedendo a implantação do experimento de fase inicial do melhoramento, mostrou-se, parcialmente, eficiente para o mapeamento a priori da variabilidade espacial da Área experimental 1, uma vez que os valores de correlação com os diferentes atributos físicos e químicos do solo, em alguns casos, foram positivos e moderados. Valores semelhantes foram reportados por Moraes (2020), em estudo visando a estimativa de fitomassa em diferentes culturas utilizadas como adubos verdes, dentre elas, a *Crotalaria juncea*, por meio da utilização do índice de vegetação por diferença normalizada (NDVI). Uma peculiaridade atribuída ao NDVI é sua rápida saturação, que o torna insensível ao aumento da biomassa vegetal a partir de determinado estágio de desenvolvimento, ou seja, o índice estabiliza em um patamar apresentando um mesmo valor, embora com o aumento da densidade do dossel (ZANZARINI et al., 2013). Este fato foi evidenciado por Moraes (2020), em que a saturação máxima do NDVI estimada para a *Crotalaria juncea* ocorreu aos 65 dias após a emergência da cultura. No presente estudo, as imagens foram coletadas aos 80 dias após o plantio ou, aproximadamente, 70 dias após a estabilização na emergência das plântulas (GALDINO et al., 2019). Ainda, segundo Kuiaowski et al. (2017), estimativas fenotípicas temporal são necessárias para melhor compreender a forma que os parâmetros genéticos e fenotípicos interagem com o ambiente. Estes autores, estudando o uso de índices de vegetação para o estabelecimento de zonas de manejo na cultura da soja, observaram que em função do estágio de crescimento e desenvolvimento da cultura, variações no comportamento espectral são esperados. Esses fatores podem ter sido determinantes para que a variável NDVI não fosse capaz de remover as variações ambientais não controladas pelo delineamento experimental no presente estudo, portanto, não apresentando nível de significância adequado quando incorporada no modelo genético-estatístico. Nesse sentido, apesar da identificação de diferentes classes de variabilidade no presente estudo, conhecimentos mais detalhados sobre o comportamento espectral e temporal da *Crotalaria juncea* é fundamental para compreensão do estágio fenológico mais adequado para obtenção do índice de vegetação. Além disso, estudos adicionais sobre a aplicabilidade de outras culturas de rotação, assim como, diferentes índices de vegetação são necessários para a diminuição de ruídos nas estimativas.

Ademais, controle a posteriori dos efeitos residuais utilizando um modelo que forneça um bom ajuste aos dados pode se tornar uma eficiente alternativa, uma vez que caracteres correlacionados podem ser usados como covariáveis durante a análise (FEDERER; CROSSA, 2012). Nesse sentido, a composição de um índice de fertilidade buscando investigar uma possível estrutura de variabilidade espacial, com base em combinações lineares de variáveis de solos (componentes principais), coletadas em alta densidade amostral, mostrou-se eficiente para a Área experimental 1. Resultados semelhantes foram reportados por Silva et al. (2010), onde o método de análise multivariada baseado nos componentes principais forneceu componentes interpretáveis e correlacionados com diferentes atributos físicos e químicos do solo. Segundo estes mesmos autores, esse tipo de análise, em associação com a geoestatística, possibilitou a avaliação da variabilidade dos diferentes componentes do solo. No presente estudo, embora o primeiro componente principal tenha apresentado fraca dependência espacial, de acordo com a classificação proposta por Seidel e Oliveira (2016), foi a variável de maior associação com a variável resposta (TCH) quando incorporada no modelo genético-estatístico. Estes resultados evidenciam que o método IDW demonstrou boa competência na modelagem dos

parâmetros desejados, corroborando com estudos conduzidos por Vargas et al. (2018). A inclusão do componente principal 1 como covariável no modelo independente (ID+PC₁), o mais simples para modelar a matriz de efeito residual, possibilitou redução nos valores de CV_e (Tabela 5), apesar das diferenças serem em pequena escala quando comparado com o modelo ID sem a incorporação da covariável PC₁. Contudo, de acordo com a classificação proposta por Pimentel-Gomes (1985), ambos os CV_e foram de média magnitude, uma vez que apresentaram valores entre 10 e 20%; vale ressaltar que quanto maior o CV_e menor a precisão experimental (PIMENTEL-GOMES, 1985). Por outro lado, apenas com a consideração de um modelo genético-estatístico mais robusto, i.e., utilizando a estrutura auto-regressiva separável de primeira ordem (AR1 x AR1) como modelo de dependência espacial entre os erros, observou-se redução no erro experimental (CV_e = 12,44). Segundo Gilmour et al. (1997), esta abordagem de aplicação de um modelo espacial em ensaios é bastante eficiente e desejável, pois aprimora a precisão do experimento. Melhoria adicional, neste modelo, foi ainda obtida com a inclusão da variável PC₁, permitindo uma melhor estimativa dos valores genotípicos da população em estudo, uma vez que o erro experimental foi reduzido (CV_e = 12,11). Valores de coeficiente de variação genética total (CV_g) indicam a presença ou não de variabilidade genética na população, para determinado caractere em estudo (RESENDE, 2002a). Com base nos resultados obtidos, somente os modelos AR1 x AR1 e AR1 x AR1+PC₁ apresentaram resultados acima de 10% (10,70 e 11,02%, respectivamente). Em contrapartida, valores inferiores a 10% podem ser observados quando os modelos ID (9,40%) e ID+PC₁ (9,62) são considerados. As estimativas dos valores de herdabilidade no sentido amplo (h²_a), considerando os modelos ID e ID+PC₁, foram de alta magnitude (≥0,50) (RESENDE, 2002b). Ademais, observa-se que a inclusão da variável PC₁ no modelo ID aumenta, ligeiramente, o valor de h²_a, mostrando a contribuição desta variável na explicação do modelo. O mesmo ocorre para o modelo AR1 x AR1, em que a estimativa de h²_a é aumentada com a inclusão da variável PC₁ (h²_a = 0,62), i.e., grande parte da variação fenotípica avaliada pode ser devida à variação dos efeitos do genótipo, com menor confundimento ambiental (LEITE et al., 2006).

Segundo Deon e Resende (2007), um dos parâmetros mais relevantes para a avaliação da qualidade experimental é a acurácia seletiva. Este, por sua vez, tem a propriedade de informar sobre a eficácia da inferência acerca do valor genotípico. Tal parâmetro não depende apenas da magnitude da variação residual e do número de repetições, mas também da proporção entre as variações de natureza genética e residual associadas ao caractere em avaliação. Neste sentido, de acordo com a classificação proposta pelos citados autores, os valores obtidos de acurácia seletiva (AC_s) para todos os modelos foram de alta magnitude, pois apresentaram valores entre 0,70 e 0,85. Vale destacar que tanto para o modelo ID como para o AR1 x AR1, os valores de AC_s foram aumentados quando adicionado a variável PC₁. Portanto, os valores preditos a partir da inclusão das informações do componente principal 1 no modelo AR1 x AR1 são aproximados do valor genotípico verdadeiro dos tratamentos genéticos. Ainda, considerando as diferentes estruturas de VCOV, observa-se que os menores valores obtidos para o critério de AIC e BIC (AKAIKE, 1974; SCHWARZ, 1978) foram para o modelo AR1 x AR1+PC₁ (Tabela 12) (Apêndice). Portanto, é o modelo mais recomendado para a predição dos valores genotípicos da população em estudo.

Embora o PC₁ tenha sido capaz de remover parte das variações não controladas pelo delineamento experimental da Área 1, não apresentou contribuição significativa quando considerado para a Área experimental 2. Em contrapartida, a variável que melhor ajustou o modelo genético-estatístico para este experimento e, por consequência, melhorou as estimativas dos valores genotípicos da população em estudo, foi a condutividade elétrica aparente do solo (CE_a), obtida para ambas as profundidades de leitura. Esta variável apresentou valores de correlação positiva e negativa com diferentes parâmetros físicos e químicos do solo para ambas as áreas experimentais. Resultados semelhantes são reportados em diversos estudos conduzidos na área de agricultura de precisão (JOHNSON et al., 2001; FAROOQUE

et al., 2012; MOLIN; FAULIN, 2013). No entanto, segundo Medeiros et al. (2018), a condutividade elétrica aparente do solo não é viável para fazer estimativas generalizadas dos atributos do solo, mas pode ser aplicada como uma ferramenta para mapear a variabilidade espacial, i.e., separar diferentes tipos de solos presentes em um campo. Neste sentido, James et al. (2012) conseguiram utilizar medidas de CE_a , obtidas a partir do mapeamento de todas as parcelas experimentais, para a caracterização do padrão de salinidade presente no campo experimental. Tal caracterização permitiu a separação da área em três blocos com diferentes padrões de salinidade. Em condições de alta CE_a (ou alta salinidade), observou-se a redução no rendimento dos genótipos em estudo. Buscando corrigir os diferentes padrões de variabilidade espacial presentes no campo experimental, estes mesmos autores utilizaram a estrutura auto-regressiva separável de primeira ordem ($AR1 \times AR1$) com a inclusão das informações provenientes do sensor de CE_a como covariável no modelo genético-estatístico. Essa estratégia permitiu a melhor compreensão e identificação das informações de interesse, com redução no viés das estimativas. Resultados semelhantes foram obtidos no presente estudo, em que o coeficiente de variação residual (CV_e) apresentou redução significativa quando o modelo $AR1 \times AR1$ foi considerado. Esta redução foi ainda mais acentuada com a inclusão de ambas as covariáveis obtidas a partir das diferentes leituras de CE_a (E_{Ca05} e E_{Ca1}) ($CV_e = 10,66$). Com relação ao CV_g , todos os modelos apresentaram resultados superiores a 10% que, segundo Resende (2002a), indica a presença de variabilidade genética na população. Ainda, destaca-se o aumento significativo neste parâmetro quando considerado o modelo $AR1 \times AR1$ e, ainda mais evidente, com a inclusão de ambas as covariáveis (E_{Ca05} e E_{Ca1}). Isto demonstra que as diferentes informações obtidas via CE_a permitem reduzir o efeito ambiental e, por conseguinte, explorar de forma mais eficiente a variabilidade genética presente na população. Com exceção dos modelos ID e ID+ E_{Ca05} + E_{Ca1} , todos os demais apresentaram estimativas de herdabilidade no sentido amplo (h^2_a) de alta magnitude ($\geq 0,50$) (RESENDE, 2002). Observa-se que os valores de h^2_a para o modelo ID, com a inclusão ou não das covariáveis, se mantiveram os mesmos ($h^2_a = 0,35$). No entanto, um aumento pronunciado é observado para o modelo $AR1 \times AR1$ ($h^2_a = 0,72$) e ainda maior quando as covariáveis ambientais (CE_a) são incorporadas ($h^2_a = 0,76$). Para este último modelo, observou-se também o maior valor de acurácia seletiva ($AC_s = 0,87$). Portanto, os valores preditos a partir da inclusão das covariáveis ambientais (CE_a) no modelo $AR1 \times AR1$ são mais aproximados do valor genotípico verdadeiro dos tratamentos genéticos, corroborando com as informações obtidas por James et al. (2012). Ainda, considerando as diferentes estruturas de VCOV, observa-se que os menores valores obtidos para os critérios de AIC e BIC (AKAIKE, 1974; SCHWARZ, 1978) foram para o modelo $AR1 \times AR1 + E_{Ca05}$ e E_{Ca1} , Tabela 12 (Apêndice). Portanto, é o modelo mais recomendado para a predição dos valores genotípicos da população em estudo.

Embora os resultados obtidos, utilizando os dados brutos de CE_a , tenham sido promissores na análise dos dados da Área experimental 2, quando incorporados por meio de covariáveis no modelo genético-estatístico, estudos conduzidos por Wei et al. (2015) relatam que o uso dessas informações para a decomposição da área em diferentes zonas de variabilidade para então definição dos blocos experimentais, com base em zonas homogêneas, permite um melhor ajuste dos dados de TCH. Para esses autores, tal relato pode ser devido a um relacionamento mais complexo entre CE_a e TCH, em vez de um relacionamento linear simples.

De qualquer forma, fica evidente que experimentos de fases iniciais do melhoramento estão sujeitos a variações decorrentes da variabilidade espacial. Portanto, as estimativas dos genótipos e suas comparações podem ser seriamente comprometidas e, conseqüentemente, o progresso genético ser reduzido. Por outro lado, o controle desse tipo de adversidade por meio de métodos estatísticos, tais como, o uso de controle local (testemunhas) e, mais modernamente, por meio da utilização de modelos mistos (GILMOUR et al., 1997) e/ou geoestatística (YAMAMOTO; LANDIM, 2013), se mostram fundamentais. Complementarmente, tecnologias que permitem

remover fontes de variação sistemática do resíduo apresentam grande potencial de aplicação, proporcionando aumento na acurácia das estimativas, seja por meio de um controle a priori ou posterior da variabilidade espacial (FEDERER; CROSSA, 2012).

Para Xu (2016), apesar da grande complexidade envolvida com esse tipo de estudo, as informações geradas podem também ser eficientemente incorporadas em outras áreas do melhoramento, e.g., aos modelos de predição em seleção genômica e estudos de interação genótipo x ambiente. Este mesmo autor propõe o conceito “*envirotyping*”, como uma terceira tecnologia “*typing*”, complementada pela genotipagem e fenotipagem. No futuro, o conceito “*envirotyping*” precisará focar em parcelas experimentais e plantas individuais, juntamente com o desenvolvimento de plataformas “*envirotyping*” de alto rendimento e precisão, para integrar informações genotípicas, fenotípicas e ambientais, de modo que um sistema de melhoramento de alta eficiência e precisão seja estabelecido.

7. CONCLUSÕES

1. O uso de componentes principais, a partir de dados de amostragem de solo em alta densidade, permitiu capitalizar parte da variabilidade total dos dados para a Área experimental 1. Portanto, mostra-se como um eficiente índice a ser incorporado como covariável no modelo genético-estatístico, possibilitando redução do erro experimental (CV_e). Entretanto, o custo elevado para obtenção desse tipo de informação pode ser um fator limitante.
2. O uso do comportamento espectral da cultura de rotação (*Crotalaria juncea* L.), por meio da utilização do índice de vegetação por diferença normalizada ($NDVI$), permitiu a identificação, *a priori*, de diferentes classes de variabilidade na Área experimental 1. Porém, não é capaz de remover as variações ambientais não controladas pelo delineamento experimental. Estudos mais detalhados sobre o comportamento espectral e temporal desta cultura é fundamental para compreensão do estágio fenológico mais adequado para o cálculo do índice.
3. O modelo genético-estatístico autorregressivo, separável de primeira ordem para linhas e colunas $AR(1) \times AR(1)$, mostra-se altamente eficiente no ajuste dos efeitos da variabilidade espacial.
4. O uso do sensor de CE_a pode ser recomendado para o ajuste da dependência espacial presente nos experimentos de fases iniciais do melhoramento implantados em regiões de solos arenosos. Além disso, esse tipo de geotecnologia pode ser amplamente empregado em experimentação agrônômica e nas diversas áreas de estudos dentro do melhoramento genético de plantas, e.g., fases de experimentação, seleção genômica e estudos de interação genótipo x ambiente.

REFERÊNCIAS

- ADAMCHUK, V.I.; HUMMEL, J.W.; MORGAN, M.T.; UPADHYAYA, S.K. On-the go soil sensors for precision agriculture. **Computers and Electronics in Agriculture**. Amsterdam, v.44, p.71-91, 2004.
- ANDRIOTTI, J.L.S. Introdução à geoestatística. **Acta Geologica Leopoldensia**, n.27, v.11, p. 7-82. 1988.
- ARAUJO, J.C.; VETTORAZZI, C.A.; MOLIN, J.P. Estimativa da produtividade e determinação de zonas de manejo, em cultura de grãos, por meio de videografia aérea multispectral. **Acta Scientiarum**, Maringá, v.27, n.3, p.437-447, 2005.
- BADDELEY, A.; RUBAK, E. TURNER, R. Spatial point patterns: methodology and applications with R. Chapman and Hall/CRC Press, 2015. Publisher's web page for book Companion website for the book.
- BANZATTO, D.A.; KRONKA, S.N. Experimentação agrícola. 4. Ed. **Jaboticabal: Funep**, 2006. 237p.
- BARBOSA, M.H.P.; RESENDE, M.D.V.D.; PETERNELLI, L.A.; BRESSIANI, J.A.; SILVEIRA, L.C.I.D.; SILVA, F.L.; RODRIGUES, F.I.C. Use of REML/BLUP for the selection of sugarcane families specialized in biomass production. **Crop Breeding and Applied Biotechnology**, p. 218-226, 2004.
- BRESSIANI, J.A. Seleção sequencial em cana-de-açúcar. **Tese de Doutorado** – Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba, 2001.
- BURROUGH, P.A. Principles of geographical information systems for land resources assessment. Oxford, **Clarendon Press**. 1986.
- CAMARGO, M.N.; KLAMT, E.; KAUFFMAN, J.H. Classificação de solos usada em levantamento pedológico no Brasil. **Boletim Informativo da Sociedade Brasileira de Ciência de Solo**, v.12, n.1, p.11-33, 1987.
- CORWIN, D.L.; LESCH, S.M. Apparent soil electrical conductivity measurements in agriculture. **Computers and Electronics in Agriculture**, Amsterdam, v.46, p.11-43, 2005.
- COX, M. Optimum Family Selection for Net Merit Grade in Stage 2 Trials. **Technical report, SRA – Sugar Research Australia**, Bundaberg - Australia. 1996.
- DANIELS, J.; ROACH, B.T. Taxonomy and Evolution. In: Sugarcane improvement through breeding, edited by HEINZ, D.J., chapter 1, p.7-84, **Elsevier B.V.**, Amsterdam.
- DEON, M.; RESENDE, V.D. Precisão e controle de qualidade em experimentos de avaliação de cultivares 1. **Pesquisa Agropecuária Tropical**, v.37, p.182-194, 2007.
- DIAS, F.L.F.; MAZZA, J.A.; MATSUOKA, S.; PERECIN, D.; MAULE, R.F. Produtividade da cana-de-açúcar em relação a clima e solos da região noroeste do estado de São Paulo. **Revista Brasileira de Ciências do Solo**, v.23, p.627-634, 1999.
- DUARTE, J.B. Princípios sobre delineamentos em experimentação agrícola. **Monografia do Curso de Especialização em Estatística**. Universidade Federal de Goiás. 66p. 1996.
- D'HONT, A.; GRIVET, L.; FELDMANN, P.; RAO, S.; BERDING, N.; GLASZMANN, C. Characterization of the double genome structure of modern sugarcane varieties (*Saccharum* spp.) by molecular cytogenetics. **Molecular Genet**, p. 405-413, 1996.
- ELACHI, C.; VAN-ZYL, J.J. Introduction to the physics and techniques of remote sensing. 28, **John Wiley & Sons**, Hoboken, NJ. 2006.
- ENVIRONMENTAL SYSTEMS RESEARCH INSTITUTE (ESRI). In: ArcGIS Desktop v10.2 Help, 3D Analyst, 2016.

- FALCONER, D.S.; MACKAY, T.F.C. Introduction to quantitative genetics. **Longman Group Limited**, Endinburgh, fourth edition, 1996.
- FAN, M.; SHEN, J.; YUAN, L.; JIANG, R.; CHEN, X.; DAVIES, W.J.; ZHANG, F. Improving crop productivity and resource use efficiency to ensure food security and environmental quality in China. **Journal of Experimental Botany**, v.63, n.1, p.13-24, 2012.
- FAO. Food and Agriculture Organization of the United Nations. Disponível em: <<http://https://population.un.org/wpp/>>. Acesso em: 11 dezembro 2019.
- FAROOQUE, A.A.; ZAMAN Q.U.; SCHUMANN, A.W.; MADANI, A.; PERCIVAL, D. Delineating management zones for site specific fertilization in wild blueberry fields. **Applied Engineering in Agriculture**, v.28, n.1, p.57-70, 2012.
- FEDERER, W.T. Augmented (hoonuiaku) designs. **Hawaiian Planter's Record**, Aiea, v.55, p.191-208, 1956.
- FEDERER, W.T.; CROSSA, J. I.4 screening experimental designs for quantitative trait loci, association mapping, genotype by environment interaction, and other investigations. **Frontiers in Physiology**, v.3, 156p. 2012.
- FORMAGGIO, R.; SANCHES, I.D. Sensoriamento remoto em agricultura. São Paulo: **Oficina de Textos**, 2017.
- FRIEDMAN, S.P. Soil properties influencing apparent electrical conductivity: a review. **Computers and Electronics in Agriculture**. Amsterdam, v.46, p.45-70, 2005.
- FREELAND, R.S. Review of soil moisture sensing using soil electrical conductivity. **Transaction of the ASAE**, v.32, n.6, p.2190-2194, 1989.
- GALDINO, A.G.; CAMARA, F.T.; SANTANA, L.D.; PINTO, A.A.; SILVA, J.M. Potencial de emergência das espécies de *crotalaria juncea* e *crotalaria chroleuca* em função da profundidade de semeadura. **Nucleus**, v.16, n.1, 2019.
- GILMOUR, A.R.; CULLIS, B.R.; VERBYLA, A.P. Accounting for natural and extraneous variation in the analysis of field experiments. **Journal of Agricultural, Biological, and Environmental Statistics**, Alexandria, v.2, p.269-293, 1997.
- GILMOUR, A.R.; CULLIS, B.R.; WELHAM, S.J. THOMPSON, R. ASREML reference manual. **Orange: NSW Agriculture**, 2000. 218p.
- GIMENEZ, L.M. Relação da condutividade elétrica aparente com algumas propriedades físico-hídricas e com a variabilidade espacial dos solos. 120 f. **Tese de Doutorado** – Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba, 2013.
- GOMES, R.F.L. Análise genética de progênies F6 e F7:6 de soja obtidas de cruzamentos dialélicos. 140p. **Tese de Doutorado** - Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba. 1995.
- HUIJBREGTS, C.J. Regionalized variables and quantitative analysis of spatial data. In: DAVIS, J.C. & MC CULLAGH, M.J. (ed.) **Display and analysis of spatial data**. John Wiley, p.38-53. 1975.
- JACKSON, P.; MACRAE, T.; HOGARTH, M. Selection of sugarcane families across variable environments II. Patterns of response and association with environmental factors. **Field Crops Research**, v.43, p.119-130, 1995.
- JAMES, R.A.; BLAKE, C.; ZWART, A.B.; HARE, R.A.; RTHJEN, A.J.; MUNNS, R. Impact of ancestral wheat sodium exclusion genes *Nax1* and *Nax2* on grain yield of durum wheat on saline soils. **Functional Plant Biology**, 39, p.609-618. 2012.
- JOHNSON, C.K., DORAN, J.W., DUKE, H.R., WEINHOLD, B.J., ESKRIDGE, K.M., SHANAHAN, J.F. Field-scale electrical conductivity mapping for delineating soil condition. **Soil Science Society of America Journal**, v.65, p.1829-1837. 2001.

- JOHNSON, C.K.; MORTENSEN, D.A.; WIENHOLD, B.; SHANAHAN, J.; DORAN, J.W. Site-specific management zones based upon soil electrical conductivity in a semiarid cropping system. **Agronomy Journal**, v.95, n.2, p.303-315, 2003.
- JORGE, L.A.C.; INAMASU, R.Y. Uso de veículos aéreos não tripulados (VANT) em agricultura de precisão. **Agricultura de precisão: resultados de um novo olhar**. Embrapa, p.109-134, 2014.
- JUNK, W.K.; KITCHEN, N.R.; SUDDUTH, K.A.; KREMER, R.J.; and MOTAVALLI, P.P. Relationship of Apparent soil electrical conductivity to claypan soil properties. **Soil Science Society of America Journal**, v.69, p.883-892, 2005.
- KALAITZIDIS, C.; HEINZEL, V.; ZIANIS, D. A review of multispectral vegetation indices for biomass estimation. **IOS Press**, 2010.
- KEMPTON, R.A. The design and analysis of unreplicated field trials. **Vortrage Pflanzenzuchtung**, v.7, p.219-242, 1984.
- KUIAWSKI, A.C.M.B; SAFANELLI, J.L.; BOTTEGA, E.L.; DE OLIVEIRA NETO, A.M.; GUERRA, N. Vegetation indexes and delineation of management zones for soybean. **Pesquisa Agropecuária Tropical**, Goiânia, v.74, n.2, p.168-177, 2017.
- KUMAR, N. Remote sensing: introduction and basic concepts. Disponível em: <<https://nptel.ac.in/courses/105108077/module1/lecture5.pdf>>. Acessado em julho de 2018.
- KUTNER, M.H.; NACHTSHEIM, C.J.; NETER, J.; LI, W. Applied linear statistical models. **McGraw-Hill/Irwin**, 5 edition, August 2004.
- LAMPARELLI, R.A.; ROCHA, J.V.; BORGHI, E. Geoprocessamento e agricultura de precisão: fundamentos e aplicações. **Guaíba: Agropecuária**, 118p., 2001.
- LANDIM, P.M.B. Análise estatística de dados geológicos. São Paulo: Editora UNESP, 2003. 2. ed. Rev. e ampl. 253p.
- LANDIM, P.M.B. Sobre geoestatística e mapas. **Terrae Didatica**, 2(1):19-33. Disponível em: <<http://www.ige.unicamp.br/terraedidatica>>. 2006.
- LASLETT, G.M. Kriging and splines: an empirical comparison of their predictive performance in some applications. **Journal of the American Statistical Association**, v.89, p.391-400, 1994.
- LEITE, M.S.O.; PETERNELLI, L.A.; BARBOSA, M.H. Effects of plot size on the estimation of genetic parameters in sugarcane families. **Crop Breeding and Applied Biotechnology**, v.6, p.40-46, 2006.
- LUCK, E.; EISENREICH, M. Electrical conductivity mapping for precision agriculture (compact disc). In: **European conference on precision agriculture**, 3. Montpellier, 2001. Proceedings, Montpellier: Ecole Nationale Supérieure Agronomique, 2001.
- LUND, E.D.; COLIN, P.E.; CHRISTY, C.; DRUMMOND, P.E. Applying soil electrical conductivity technology to precision agriculture. In: **International Conference on Precision Agriculture**, 4., 1998, Madison. Proceedings. Madison: ASA-CSSA-SSSA, 1999. p.1089-1100.
- MARCUZZO, F.N.; ANDRADE, L.R.; MELO, D.C.R. Métodos de interpolação matemática no mapeamento de chuvas do estado do Mato Grosso. In: **Revista Brasileira de Geografia Física**, 2011.
- MATASSA, V.J.; BASFORD, K.E.; STRINGER, J.K.; HOGARTH, D.M. The application of spatial analysis to sugarcane variety trials. **Proceedings of the Australian Society of Sugar Cane Technology**, v.20, p.162-168.
- MEDEIROS, W.N.; VALENTE, D.S.M.; QUEIROZ, D.M.; PINTO, F.A.C.; ASSIS, I.R. Apparent soil electrical conductivity in two different soil types. **Revista Ciência Agronômica**, v.49, n.1, p.43-52, 2018.

- MORAES, E.C. Fundamentos de sensoriamento remoto. Apostila. Instituto Nacional de Pesquisas Espaciais (INPE/MCTI). Disponível em: <http://www.dsr.inpe.br/vcsr/files/capitulo_1.pdf>. Acesso em: 30 julho 2018.
- MORAES, S.T.P. Índice de vegetação por diferença normalizada (*NDVI*) visando estimativa de fitomassa de adubos verdes. Piracicaba, 2020. 51p. **Tese de Doutorado** – Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba, 2020.
- MOLIN, J.P.; AMARAL, L.R.; COLAÇO, A.F. Agricultura de Precisão. 1 ed. São Paulo: **Oficina de Textos**, 2015.
- MOLIN, J. P.; FAULIN, G. D. C. Spatial and temporal variability of soil electrical conductivity related to soil moisture. **Scientia Agricola**, v.70, n.1, p.1-5, 2013.
- MUNDOGEO. Sensoriamento hiperespectral. Disponível em: <<http://mundogeo.com/blog/2004/08/23/sensoriamento-remoto-hiperespectral/>>. Acesso em: 30 julho 2018.
- NADLER, A.; FRENKEL, H. Determination of soil solution electrical conductivity from bulk soil electrical conductivity measurements by the four-electrode method. **Soil Science Society of American Journal**, v.44, n.5, p. 1216-1221, 1980.
- NOGUEIRA, J.D.L.; AMARAL, R.F. Comparação entre os métodos de interpolação (Krigagem e Topo to Raster) na elaboração da batimetria na área da folha Touros - RN. **Anais XIV Simpósio Brasileiro de Sensoriamento Remoto**, Natal, Brasil, INPE, p.4117-4123. 2009.
- OLIVEIRA, F.D. Mapeamento dos teores de óleo e proteína de grãos de soja e análise de sua correlação com atributos do solo. 2003. 115 f. **Dissertação (Mestrado em Agronomia)** – Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Piracicaba, 2003.
- PAYNE, R.W.; MURRAY, D.A.; HARDING, S.A.; BAIRD, D.B.; SOUTAR, D.M. An introduction to genstat for windows. **VSN International**, Hemel Hempstead, UK. 2011.
- PIMENTEL-GOMES, F. Curso de estatística experimental. 12. ed. Piracicaba: **Livraria Nobel**, 1985. 467p.
- PIMENTEL-GOMES, F. A estatística moderna na pesquisa agropecuária. Piracicaba, **Potafos**. 162 p. 1984.
- PINCELLI, A.L.S. Desenvolvimento e ensaios de um sistema de mensuração de condutividade elétrica do solo. 2004. 96 f. **Dissertação (Mestrado em Agronomia)** – Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba, 2004.
- PUSCH, M.; MACHADO, T.P.; AMARAL, L.R. Introdução a agricultura de precisão. 2019. Disponível em: <www.feagri.unicamp.br/gitap>. Acesso em: 13 julho 2020.
- RAIJ, B. van; ANDRADE, J.C.; CANTARELLA, H.; QUAGGIO, J.A. Análise química para avaliação da fertilidade de solos tropicais. Campinas, **Instituto Agronômico**, 284p. 2001.
- RAM, M. Plant breeding methods. **Prentice-Hall of India Pvt.ltda**, Índia. 2014.
- R Development Core Team. 2015. R: A language and environment for statistical computing. **R Foundation for Statistical Computing**, Vienna, Austria.
- RESENDE, M.D.V.; DUARTE, J.B. Precisão e controle de qualidade em experimentos de avaliação de cultivares. **Pesquisa Agropecuária Tropical**, v.37, n.3, p.182-194, 2007.
- RESENDE, M. Genética biométrica e estatística no melhoramento de plantas perenes. **Embrapa informações Tecnológica**, Brasília. 2002a.
- RESENDE, M. Genética biométrica e estatística no melhoramento de plantas perenes. **Embrapa informações Tecnológica**, Brasília. 2002b.
- RIBEIRO JUNIOR, P.J.; DIGGLE, P.J. GeoR: A package for geostatistical analysis. **R-News**, v.1, n.2. ISSN 1609-3631. 2001.

- SEIDEL, E.J.; OLIVEIRA, M.S. A classification of a geostatistical index of spatial dependence. **Revista Brasileira de Ciência do Solo**, v.40, e0160007, 2016.
- SILVA, V.R.; REICHERT, J.M.; STORCK, L.; FEIJO, S. Variabilidade especial das características químicas do solo e produtividade de milho em um argissolo vermelho-amarelo distrófico arênico. **Revista Brasileira de Ciência do Solo**, v.27, p.1013-1020, 2003.
- SILVIA, S.A.; LIMA, J.S.S.; XAVIER, A.C.; TEIXEIRA, M.M. Variabilidade espacial de atributos químicos de um latossolo vermelho-amarelo húmico cultivado com café. **Revista Brasileira de Ciência do Solo**, v.34, p.15-22, 2010.
- SLEPER, D.A.; POEHLMAN, M. Breeding field crops. **Blackwell**, fifth edition. 2006.
- STORCK, L.; GARCIA, D.C.; LOPES, S.J.; ESTEFANEL, V. Experimentação vegetal. **Santa Maria: UFSM**, 2000, 198p.
- STRINGER, J.K. Joint modelling of spatial variability and interplot competition to improve the efficiency of plant improvement. **PhD. Thesis**, University of Queensland, Brisbane. 2006.
- STURARO, J.R. Apostila de geoestatística básica. UNESP/Campus de Rio Claro. Departamento de Geologia Aplicada – IGCE. Disponível em: <<https://igce.rc.unesp.br/Home/Departamentos47/geologiaaplicada/apostila-basica.pdf>> 2015.
- SOUZA, E.A. Alternativas experimentais na avaliação de progênes em programas de melhoramento vegetal. Piracicaba, 1997. 122p. **Tese de Doutorado** – Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba, 1997.
- SOUZA JUNIOR, J.G.; DEMATTÊ, J.A.M.; ARAÚJO, S.R. Modelos espectrais terrestres e orbitais na determinação de teores de atributos dos solos: potencial e custos. **Bragantia**, Campinas, v.70, n.3, p.610-621, 2011.
- SWEARINGIN, M.L.; HOLT, D.A. Using a “Blank” trial as a teaching tool. **Journal of Agronomic Education**. 1976.
- TILMAN, D.; CASSMAN, K.G.; MATSON, P.A.; NAYLOR, R.; POLASKY, S. Agricultural sustainability and intensive production practices. **Nature**, v.418, p.671-677, 2002.
- TIMBÓ, M.A. Levantamento através do sistema GPS. **Departamento de Cartografia**. Universidade Federal de Minas Gerais (UFMG). 2000.
- TSCHIEDEL, M.; FERREIRA, M.F. Introdução a agricultura de precisão: conceitos e vantagens. **Ciência Rural**, Santa Maria, v.32, n.1, 2002.
- VARGAS, T.; GOMES, M.G.; BELLADONA, R.; ADAMI, M.V.D. Aplicação do Interpolador IDW para elaboração de mapas hidrogeológicos paramétricos na região da serra gaúcha. **Scientia Cum Industria**, v.6, n.3, p.38-43, 2018.
- VIEIRA JÚNIOR, P.A.; MOLIN, J.P.; DOURADO NETO, D.; MANFRON, P.A.; MASCARIN, L.S.; FAULIN, G.C.; DETOMINI, E.R. População de plantas e alguns atributos do solo relacionados ao rendimento de grãos de milho. **Acta Scientiarum Agronomy**, Maringá, v.28, p.483-492, 2006.
- ZANZARINI, F.V.; PISSARRA, T.C.T.; BRANDÃO, F.J.C.; TEIXEIRA, D.D.B. Correlação espacial do índice de vegetação (NDVI) de imagem Landsat/ETM+ com atributos do solo. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v.17, n.6, p.608-614, 2013.
- ZHOU, M.M.; KIMBENG, C.A.; ANDRU, S.; TEW, T.L.; PONTIF, M.J. GRAVOIS, K.A. Evaluating sugarcane families for yield and repeatability using random coefficient models. **Crop Science**, v.53, p.2352-2362, 2013.
- WATSON, D.F. Contouring: a guide to the analysis and display of spatial data. Oxford, **Pergamon**. 1992.

- WEI, X.; STRINGER, J.; SALTER, B.; PIPERIDIS, G.; SCHROEDER, B. Improving selection accuracy in clonal assessment trials by accounting for site variability. **Proceedings of the Australian Society of Sugar Cane Technology**, v.35, 2015.
- XU, Y.; LI, P.; CHENG, Z.; YANLI, L.; CHUANXIAO, X.; XUECAI, Z.; BODDUPALLI, M.P.; OSLEN, M. Enhancing genetic gain in the era of molecular breeding. **Journal of Experimental Botany**, v.68, n.11 p.2641-2666, 2017.
- XU, Y. Envirotyping for deciphering environmental impacts on crop plants. **Theoretical and Applied Genetics**, v.129, p.653-673, 2016.

APÊNDICES

APÊNDICE A

Tabela 11. Diferentes modelos para as estruturas de variância-covariância (VCOV) da matriz residual (R).

Modelo	Descrição	Matriz R
ID	Modelo identidade	$\begin{bmatrix} \sigma_{g+}^2 & \sigma_{ge}^2 & 0 & \dots & 0 \\ 0 & \sigma_{g+}^2 & \sigma_{ge}^2 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_{g+}^2 & \sigma_{ge}^2 \end{bmatrix}$
AR1	Modelo autorregressivo de primeira ordem com variação homogênea	$\begin{bmatrix} \sigma_{g+}^2 & \sigma_{ge}^2 & \sigma_g^2 \rho_g & \dots & \sigma_g^2 \rho_g^{d(1,J)} \\ \sigma_g^2 \rho_g & \sigma_{g+}^2 & \sigma_{ge}^2 & \dots & \sigma_g^2 \rho_g^{d(2,J)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma_g^2 \rho_g^{d(J,1)} & \sigma_g^2 \rho_g^{d(J,2)} & \dots & \sigma_{g+}^2 & \sigma_{ge}^2 \end{bmatrix}$

APÊNDICE B

Figura 26. *Boxplot* dos valores dos diferentes atributos físicos (Areia fina, grossa, total, argila e silte – g kg^{-1} ; M.O. – g dm^{-3}) e químicos (Al, Ca, CTC, H+Al, K, Mg, SB – mmolc dm^{-3} ; Presina – mg dm^{-3} ; V e m – %) do solo obtidos a partir de 56 pontos amostrais, na profundidade de 0 a 20 cm, na Área experimental 1. Identificação de *outliers*.

APÊNDICE C

APÊNDICE D

Figura 28. Boxplot dos valores dos diferentes atributos físicos (Areia fina, grossa, total, argila e silte – g kg⁻¹; M.O. – g dm⁻³) e químicos (Al, Ca, CTC, H+Al, K, Mg, SB – mmolc dm⁻³; Presina – mg dm⁻³; V e m – %) do solo obtidos a partir de 68 pontos amostrais (não identificado presença de *outliers*), na profundidade de 0 a 20 cm, Área experimental 2.

APÊNDICE E

Figura 29. Variograma experimental para o componente principal 1 (PC₁), considerando: modelo esférico (sph), exponencial (exp), gaussiano (gau) e plotagem conjunta, Área experimental 1.

APÊNDICE F

Figura 30. Variograma experimental para o componente principal 2 (PC₂), considerando: modelo esférico (sph), exponencial (exp), gaussiano (gau) e plotagem conjunta, Área experimental 1.

APÊNDICE G

Figura 31. Variograma experimental para o componente principal 1 (PC_1), considerando: modelo esférico (sph), exponencial (exp), gaussiano (gau) e plotagem conjunta, Área experimental 2.

APÊNDICE H

Figura 32. Variograma experimental para o componente principal 2 (PC₂), considerando: modelo esférico (sph), exponencial (exp), gaussiano (gau) e plotagem conjunta, Área experimental 2.

APÊNDICE I

Tabela 12. Valores de AIC (*Akaike Information Criterion*) e BIC (*Bayesian Information Criterion*) considerando diferentes estruturas de VCOV, com a inclusão de covariáveis significativas no modelo, para o caráter toneladas de cana por hectare (TCH). Valores obtidos para a Área experimental 1 e 2.

Local	Caráter	Modelo	AIC	BIC
Área experimental 1	TCH	a) ID	4832,10	4840,71
		b) ID+PC ₁	4816,57	4829,48
		c) AR1 x AR1	4817,24	4838,76
		d) AR1 x AR1+PC₁	4810,63	4822,14
Área experimental 2	TCH	a) ID	7328,67	7337,79
		b) ID+E _{Ca05} +E _{Ca1}	7301,28	7310,40
		c) AR1 x AR1	7300,72	7328,09
		d) AR1xAR1+E_{Ca05}+E_{Ca1}	7288,90	7306,16

ID: modelo identidade; AR1: modelo autorregressivo de ordem 1; PC₁: componente principal 1; E_{Ca05}: condutividade elétrica aparente do solo a 0,375 m; E_{Ca1}: condutividade elétrica aparente do solo a 0,75 m.