

ABORDAGEM MULTIVARIADA EXPLORATÓRIA DA EMISSÃO  
DE CO<sub>2</sub> DO SOLO EM ÁREAS AGRÍCOLAS

Vivian Aparecida Brancaglioni

Dissertação apresentada à Universidade Estadual Paulista “Júlio de Mesquita Filho” para a obtenção do título de Mestre em Biometria.

BOTUCATU  
São Paulo - Brasil  
Janeiro – 2017

**ABORDAGEM MULTIVARIADA EXPLORATÓRIA DA EMISSÃO  
DE CO<sub>2</sub> DO SOLO EM ÁREAS AGRÍCOLAS**

**Vivian Aparecida Brancaglioni**

Orientador: Prof. Dr. **José Silvio Govone**

Coorientador: Prof. Dr. **Alan Rodrigo Panosso**

Dissertação apresentada à Universidade Estadual Paulista “Júlio de Mesquita Filho” para a obtenção do título de Mestre em Biometria.

BOTUCATU  
São Paulo - Brasil  
Janeiro – 2017

# Ficha Catalográfica

FICHA CATALOGRÁFICA ELABORADA PELA SEÇÃO TÉC. AQUIS. TRATAMENTO DA INFORM.  
DIVISÃO TÉCNICA DE BIBLIOTECA E DOCUMENTAÇÃO - CÂMPUS DE BOTUCATU - UNESP

BIBLIOTECÁRIA RESPONSÁVEL: ROSEMEIRE APARECIDA VICENTE-CRB 8/5651

Brancaglioni, Vivian Aparecida.

Análise multivariada exploratória da emissão de CO<sub>2</sub> do solo em áreas agrícolas / Vivian Aparecida Brancaglioni. - Botucatu, 2017

Dissertação (mestrado) - Universidade Estadual Paulista "Júlio de Mesquita Filho", Instituto de Biociências de Botucatu

Orientador: José Silvio Govone

Coorientador: Alan Rodrigo Panosso

Capes: 90100000

1. Análise multivariada. 2. Atmosfera do solo. 3. Solos - Análise. 4. Agricultura. 5. Sistemas silviculturais. 6. Cultivo.

Palavras-chave: Análise multivariada; Atributos físicos e químicos do solo; Emissão de CO<sub>2</sub> do solo; Técnicas exploratórias.

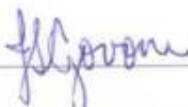
MEMBROS DA COMISSÃO JULGADORA DA DISSERTAÇÃO DE MESTRADO DE VIVIAN APARECIDA BRANCAGLIONI, INTITULADA "ABORDAGEM MULTIVARIADA EXPLORATÓRIA DE EMISSÃO DE CO<sub>2</sub> DO SOLO EM ÁREAS AGRÍCOLAS", APRESENTADA AO INSTITUTO DE BIOCÊNCIAS, UNESP, CAMPUS DE BOTUCATU, SÃO PAULO, EM 25 de janeiro de 2017.

APROVADA PELA COMISSÃO JULGADORA:

Prof(a) Dr(a) JOSE SILVIO GOVONE

Instituição: IGCE / UNESP

Assinatura:



Prof(a) Dr(a) CARLOS ROBERTO PADOVANI

Instituição: Instituto de Biociências de Botucatu

Assinatura:



Prof(a) Dr(a) CARLOS TADEU DOS SANTOS DIAS

Instituição: USP Piracicaba

Assinatura:



*Às melhores pessoas do mundo,  
minha querida família, por acre-  
ditarem em mim quando nem eu  
mesma acreditava: Antônio Car-  
los, Maria Luiza, Caren e Felipe.*

## Agradecimentos

Agradeço a Deus por sustentar e guiar minha vida.

Aos meus pais Antônio Carlos e Maria Luiza que com amor sempre compreenderam minha ausência e fizeram de tudo para que eu chegasse até aqui, apoiando-me em minhas escolhas e rezando por mim.

A minha irmã Caren que é meu exemplo de dedicação e amor aos estudos, minha companheira para a vida.

Ao meu namorado Felipe, que com paciência e amor, esteve sempre ao meu lado, me confortando e acreditando em mim, tornando-se a cada dia mais essencial.

Ao meu orientador Prof. Dr. José Silvio Govone pelas palavras de incentivo, por acreditar em minha capacidade e ter despertado em mim o interesse pela Bioestatística quando eu ainda estava na graduação e pelas contribuições para a realização deste trabalho.

Ao meu coorientador Prof. Dr. Alan Rodrigo Panosso por sua amizade, motivação desde à época da iniciação científica e pela disponibilização dos dados para o desenvolvimento deste trabalho.

Aos Professores Carlos Roberto Padovani e Carlos Tadeu dos Santos Dias membros da banca de qualificação e que muito contribuíram para o enriquecimento e desenvolvimento deste trabalho, tanto em seu mérito como em sua forma, com suas preciosas sugestões.

À Maria Elisa, que muito me auxiliou na interpretação biológica dos resultados sendo sempre solícita e gentil.

Aos amigos, que próximos ou distantes, contribuíram de alguma forma

para a realização deste trabalho, direta ou indiretamente, com palavras de apoio e incentivo. E particularmente aos amigos que fiz em Botucatu, em especial à Aiara, Cica, Farid, Fátima, Jacqueline, Livia, Magali, Márcio e sua família, Thomas e Vinícius.

Aos professores e funcionários do Departamento de Bioestatística da Unesp-Botucatu, em especial ao Professor Carlos Roberto Padovani, que sempre esteve disponível para sanar minhas dúvidas com muita gentileza e dedicação.

À CAPES pelo apoio financeiro durante o desenvolvimento deste trabalho e à FAPESP (processo n° 2013/18726-8).

# Conteúdo

|   | Página     |
|---|------------|
| <b>LISTA DE FIGURAS</b>   | <b>ix</b>  |
| <b>LISTA DE TABELAS</b>   | <b>xii</b> |
| <b>RESUMO</b>   | <b>xiv</b> |
| <b>SUMMARY</b>  | <b>xvi</b> |
| <b>1 INTRODUÇÃO</b>   | <b>1</b>   |
| <b>2 REVISÃO DE LITERATURA</b>  | <b>4</b>   |
| 2.1 Análise Multivariada . . . . .  | 4          |
| 2.2 Análise de Componentes Principais (ACP) . . . . .                                     | 6          |
| 2.2.1 Biplot . . . . .  | 10         |
| 2.3 Análise Fatorial (AF) . . . . .   | 11         |
| 2.4 Análise de Correlação Canônica (ACC) . . . . .  | 19         |
| 2.5 Relação entre a emissão de CO <sub>2</sub> e atributos físicos e químicos do solo . . | 25         |
| <b>3 MATERIAIS E MÉTODOS</b>  | <b>29</b>  |
| 3.1 Descrição das áreas experimentais . . . . .   | 30         |
| 3.2 Coleta de dados . . . . .   | 33         |
| 3.3 Análise estatística . . . . .   | 39         |
| <b>4 RESULTADOS E DISCUSSÕES</b>  | <b>41</b>  |
| 4.1 Experimento em área de cana-de-açúcar (2008) . . . . .                                | 41         |

|   |           |
|---|-----------|
|   | viii      |
| 4.1.1 Análise de Componentes Principais . . . . . | 47        |
| 4.1.2 Análise Fatorial . . . . .                  | 53        |
| 4.1.3 Análise de Correlação Canônica . . . . .    | 60        |
| 4.2 Experimento em área de soja em 2013 . . . . . | 68        |
| 4.2.1 Análise de Componentes Principais . . . . . | 74        |
| 4.3 Análise Fatorial . . . . .                    | 80        |
| 4.4 Análise de Correlação Canônica . . . . .      | 84        |
| <b>5 CONCLUSÕES</b>                               | <b>91</b> |
| <b>REFERÊNCIAS</b>                                | <b>94</b> |

## Lista de Figuras

|    | Página  |
|----|---|
| 1  | Gradeado e área amostral de Guariba-SP para o experimento em cana-de-açúcar. . . . . 31   |
| 2  | Gradeado e área amostral de Selvíria-MS para o experimento em soja. . . 32  |
| 3  | (a) Sistema LI-8100 interligado à câmara de solos; (b) sensor de temperatura do solo e (c) câmara para solo inserida sobre o colar de PVC. . . . 34 |
| 4  | TDR - “Hydrosense system”, sistema portátil utilizado para avaliação da umidade do solo. . . . . 35   |
| 5  | Histogramas e curvas de distribuição das variáveis do experimento em área de cana-de-açúcar. . . . . 44   |
| 6  | Gráfico de probabilidade <i>Q-Q Plot</i> multivariado para os dados do experimento em área de cana-de-açúcar. . . . . 45                            |
| 7  | Correlações das variáveis para o experimento em cana-de-açúcar. . . . . 46  |
| 8  | “Scree-plot” das componentes principais (CP) para a área de cana-de-açúcar. 49  |
| 9  | Biplot das associações entre CP1 e CP2 para o experimento em área de cana-de açúcar. . . . . 51   |
| 10 | Biplot das associações entre CP1 e CP3 para o experimento em área de cana-de açúcar. . . . . 52   |
| 11 | Biplot das associações entre CP2 e CP3 para o experimento em área de cana-de açúcar. . . . . 52   |
| 12 | “Scree-plot” das componentes principais (CP) para experimento em área de cana-de-açúcar. . . . . 54   |

|    |  |    |
|----|--|----|
| 13 | Gráfico de dispersão das cargas fatoriais de F1 e F2 para as variáveis avaliadas. . . . .  | 56 |
| 14 | Gráfico de dispersão das cargas fatoriais de F1 e F3 para as variáveis avaliadas. . . . .  | 56 |
| 15 | Gráfico de dispersão das cargas fatoriais de F2 e F3 para as variáveis avaliadas. . . . .  | 57 |
| 16 | Gráfico de dispersão dos escores das variáveis canônicas U1 e V1, U2 e V2 para experimento em área de cana-de-açúcar. . . . .        | 67 |
| 17 | Histogramas e curvas de distribuição das variáveis do experimento em área de soja. . . . .   | 70 |
| 18 | Gráfico de probabilidade <i>Q-Q Plot</i> multivariado para os dados do experimento em soja. . . . .                                  | 71 |
| 19 | Correlações das variáveis para o experimento em área de soja. . . . .  | 72 |
| 20 | “Scree-plot” das componentes principais (CP) para o experimento em área de soja. . . . .   | 76 |
| 21 | Biplot das associações entre CP1 e CP2 para o experimento em área de soja. . . . .   | 79 |
| 22 | Biplot das associações entre CP1 e CP3 para o experimento em área de soja. . . . .   | 79 |
| 23 | Biplot das associações entre CP2 e CP3 para o experimento em área de soja. . . . .   | 80 |
| 24 | “Scree-plot” das componentes principais (CP). . . . .  | 81 |
| 25 | Gráfico de dispersão das cargas fatoriais de F1 e F2 para as variáveis avaliadas. . . . .  | 83 |
| 26 | Gráfico de dispersão dos escores das variáveis canônicas U <sub>1</sub> e V <sub>1</sub> , U <sub>2</sub> e V <sub>2</sub> . . . . . | 90 |

## Lista de Quadros

- 1 Informações das áreas onde foram conduzidos os experimentos de variabilidade espacial da emissão de CO<sub>2</sub> e atributos físicos e químicos do solo . . . . . 33
- 2 Identificação e descrição dos atributos mensurados em cada experimento 38

## Lista de Tabelas

|   | Página |
|---|--------|
| 1 Estatística descritiva das variáveis para o experimento em área de cana-de-açúcar . . . . .   | 43     |
| 2 Variâncias, covariâncias ( <b>S</b> ) e correlações ( <b>R</b> ) das variáveis do experimento em cana-de-açúcar <sup>1</sup> . . . . .  | 48     |
| 3 Autovalores ( $\lambda_i$ ) e porcentagens da variação explicadas pelas componentes principais (VCP) . . . . .  | 49     |
| 4 Coeficientes de ponderação das variáveis e suas correlações (entre parênteses) com as componentes principais para o experimento em cana-de-açúcar.  | 50     |
| 5 Cargas fatoriais (rotação Varimax), communalidades, variâncias específicas das variáveis, autovalores e percentual acumulado de explicação da variância da análise dos dados do experimento em área de cana-de-açúcar | 55     |
| 6 Matriz residual da análise fatorial para atributos do solo em área de cana-de-açúcar . . . . .  | 60     |
| 7 Matriz de correlações ( <b>R</b> <sup>22</sup> ) das variáveis físicas e químicas do solo. . . . .  | 61     |
| 8 Matriz de correlações ( <b>R</b> <sup>12</sup> ) entre as variáveis relacionadas à dinâmica do carbono e os atributos físico-químicos do solo. . . . .  | 61     |
| 9 Autovalores, correlações canônicas e teste de significância para as correlações canônicas ( $\alpha=0,05$ ). . . . .  | 62     |
| 10 Cargas canônicas e cargas canônicas cruzadas das variáveis do experimento em cana-de-açúcar. . . . .   | 64     |
| 11 Estatística descritiva das variáveis para o experimento em área de soja . . . . .  | 69     |

|    |   |    |
|----|---|----|
| 12 | Variâncias, covariâncias e correlações das variáveis do experimento em soja <sup>1</sup>  | 75 |
| 13 | Autovalores ( $\lambda_i$ ) e porcentagem da variação explicada pelas componentes principais (VCP)  | 76 |
| 14 | Coefficientes de ponderação das variáveis e suas correlações (entre parênteses) com as componentes principais.  | 77 |
| 15 | Cargas fatoriais, comunalidades, variâncias específicas das variáveis, autovalores e percentual acumulado de explicação da variância da análise dos dados em soja | 82 |
| 16 | Matriz residual da análise fatorial para experimento em soja  | 84 |
| 17 | Matriz de correlações ( $\mathbf{R}^{22}$ ) das variáveis físico-químicas do solo.  | 85 |
| 18 | Matriz de correlações ( $\mathbf{R}^{12}$ ) entre as variáveis relacionadas à dinâmica do carbono e os atributos físico-químicos do solo.                         | 86 |
| 19 | Autovalores, correlações canônicas e teste de significância para as correlações canônicas ( $\alpha=0,05$ ).  | 86 |
| 20 | Cargas canônicas e cargas canônicas cruzadas para experimento em área de soja.  | 88 |

# ABORDAGEM MULTIVARIADA EXPLORATÓRIA DA EMISSÃO DE CO<sub>2</sub> DO SOLO EM ÁREAS AGRÍCOLAS

Autora: VIVIAN APARECIDA BRANCAGLIONI

Orientador: Prof. Dr. JOSÉ SILVIO GOVONE

Coorientador: Prof. Dr. ALAN RODRIGO PANOSSO

## RESUMO

As atividades relacionadas à agricultura, silvicultura e outros usos do solo são responsáveis por cerca de um quarto das emissões antrópicas de gases do efeito estufa. Os solos estocam grandes quantidades de carbono nos ecossistemas naturais e o processo de perda de carbono do solo para a atmosfera ocorre na forma de emissão de CO<sub>2</sub>, principal gás responsável pelo efeito estufa adicional. O objetivo deste trabalho consistiu em descobrir as possíveis relações existentes entre a emissão de CO<sub>2</sub> e atributos do solo, manejo em área agrícola e cultura cultivada, por meio da utilização de técnicas multivariadas. Para isso, considerou-se dois conjuntos de dados provenientes de experimentos conduzidos em áreas agrícolas nos estados de São Paulo e Mato Grosso do Sul, realizados em 2008 e 2013, respectivamente. O primeiro experimento foi obtido a partir de uma área comercial de cana-de-açúcar sob manejo de colheita mecanizada e o outro foi realizado em área experimental

da Fazenda de Ensino, Pesquisa e Extensão – Setor de Produção Vegetal, da Faculdade de Engenharia de Ilha Solteira (FEIS/UNESP), sob manejo de plantio direto com cultivo de soja. Ambos os experimentos são relacionados à dinâmica de carbono do solo e seus atributos físico-químicos. Dessa forma, buscou-se estudar a dinâmica do carbono em solos agrícolas, podendo indicar de forma mais informativa a importância de cada atributo físico-químico do solo em razão das emissões de dióxido de carbono ( $\text{CO}_2$ ) e/ou acúmulo de carbono, por meio das técnicas multivariadas empregadas em caráter exploratório: análise de componentes principais, análise fatorial e análise de correlação canônica. Os resultados obtidos, de maneira geral, indicaram haver maior associação da emissão de dióxido de carbono com os atributos físicos do solo, especialmente porosidade livre de água (PLA), densidade do solo (DS) e Macro (macroporosidade), em área de cana-de-açúcar; em área de cultivo direto com plantio de soja, o acúmulo de carbono apresentou relação com as mesmas propriedades físicas do solo, já a emissão de  $\text{CO}_2$  associou-se aos atributos químicos do solo, como P (Fósforo) e SB (soma de bases).

Palavras-chave: análise multivariada, técnicas exploratórias, emissão de  $\text{CO}_2$  do solo, atributos físico-químicos do solo.

# MULTIVARIATE EXPLORATORY APPROACH OF SOIL CO<sub>2</sub> EMISSION IN AGRICULTURAL AREAS

Author: VIVIAN APARECIDA BRANCAGLIONI

Adviser: Prof. Dr. JOSÉ SILVIO GOVONE

Co-adviser: Prof. Dr. ALAN RODRIGO PANOSSO

## SUMMARY

Activities related to agriculture, silviculture and other uses of the soil account for about a quarter of the anthropogenic emissions of greenhouse effect gases. Soils store large amounts of carbon in the natural ecosystems and the process of carbon loss from the soil to the atmosphere occurs in the form of CO<sub>2</sub> emission, the main gas responsible by the additional greenhouse effect. The aim of this study is to find out the possible relationship between CO<sub>2</sub> emission and soil attributes, management in agricultural area and cultivated culture, through the use of multivariate techniques. For this, were considered two sets of data from experiments conducted in agricultural areas in the states of São Paulo and Mato Grosso do Sul, Brazil, carried out in 2008 and 2013, respectively. The first experiment was obtained from a commercial area of sugarcane under management of mechanized harvest and

the other was carried out in an experimental area of the Teaching, Research and Extension Farm - Vegetal Production Sector, Faculty of Engineering of Ilha Solteira (FEIS/UNESP), under no-till management with soybean cultivation. Both experiments are related to dynamics of soil carbon and their physical-chemical attributes. Thus, we sought to study the dynamics of carbon in agricultural soils, it could indicate in a more informative way the importance of each physical-chemical attribute of the soil due to carbon dioxide ( $\text{CO}_2$ ) emissions and/or carbon accumulation, through multivariate techniques used in exploratory character: principal components analysis, factorial analysis and canonical correlation analysis. The results obtained, in general, indicate that there is a greater association of the emission of carbon dioxide with the physical attributes of the soil, especially free water porosity (PLA), soil density (DS) and Macro (macroporosity), in the area of sugarcane; in direct cultivation area with soybean planting, the carbon accumulation presented a relation with the same physical properties of the soil, since the  $\text{CO}_2$  emission was associated to the soil chemical attributes, such as P (Phosphorus) and SB (sum of bases).

Keyword: multivariate analysis, exploratory techniques, soil  $\text{CO}_2$  emission, physical-chemical attributes of the soil.

# 1 INTRODUÇÃO

As técnicas multivariadas, cujo desenvolvimento ocorreu a partir do início do século passado, não tiveram seu uso difundido, por exigirem cálculos trabalhosos, impossíveis de serem realizados sem o auxílio computacional, exceto para conjuntos de dados com pequenas dimensões. Somente nas últimas décadas, com a evolução dos computadores e desenvolvimento de softwares estatísticos, muitas análises puderam ser realizadas e, dessa forma, as limitações metodológicas operacionais deixaram de ser uma preocupação para o pesquisador.

A estatística multivariada consiste em um conjunto de métodos estatísticos aplicados quando há múltiplas variáveis obtidas simultaneamente a partir de uma mesma unidade amostral ou observacional que, em geral, são correlacionadas entre si.

Neste contexto, a utilização de técnicas univariadas, ou seja, o estudo de variáveis de maneira isolada, não é uma abordagem adequada para a análise estatística de tais conjuntos de dados. Logo, em situações em que há interdependência de variáveis, utilizar técnicas multivariadas constitui-se na melhor alternativa para a análise exploratória e/ou procedimento inferencial.

Dessa forma, a estatística multivariada tem seu propósito para ser utilizada como uma alternativa à interpretação de dados de natureza biológica, procurando medir, explicar e prever as relações de inter e intradependência de variáveis de uma mesma unidade amostral ou observacional. Um exemplo de aplicação das técnicas multivariadas em dados multidimensionais, consiste no estudo da dinâmica do carbono em solos agrícolas, podendo estas técnicas indicar de forma mais informativa a importância de cada atributo do solo em razão das emissões de dióxido de

carbono ( $\text{CO}_2$ ) e/ou acúmulo de carbono.

Atividades relacionadas à agricultura, silvicultura e outros usos do solo são responsáveis por cerca de um quarto das emissões antrópicas de gases do efeito estufa, equivalente a cerca de 10 à 12 Gt  $\text{CO}_2\text{eq}$  por ano, para os anos de 2000 à 2010, provenientes principalmente do desmatamento, manejo do solo e seus nutrientes e pecuária (IPCC, 2014). No Brasil, atividades agrícolas associadas às mudanças no uso da terra são responsáveis por 70% das emissões totais de  $\text{CO}_2$ , como por exemplo, o uso de áreas desflorestadas para o cultivo agrícola (Sartori et al., 2006).

Alterações alarmantes no clima do planeta também vêm sendo causadas devido a elevação das concentrações dos gases do efeito estufa na atmosfera, tais como, o dióxido de carbono ( $\text{CO}_2$ ), metano ( $\text{CH}_4$ ) e óxido nitroso ( $\text{N}_2\text{O}$ ). Embora os gases  $\text{CH}_4$  e  $\text{N}_2\text{O}$  tenham potencial de aquecimento global muito mais expressivo que o  $\text{CO}_2$ , sendo respectivamente 25 e 298 vezes maior, o dióxido de carbono devido a sua grande quantidade emitida, apresenta a maior contribuição em termos relativos para o efeito estufa adicional (IPCC, 2007).

Além disso, os solos estocam grandes quantidades de carbono nos ecossistemas naturais, o processo de perda de carbono do solo para a atmosfera ocorre na forma de emissão de  $\text{CO}_2$ , que se dá por meio da oxidação química da matéria orgânica do solo promovida pela atividade microbiana e respiração das raízes. Apesar de todos os esforços, ainda existem grandes incertezas na quantificação de emissão de  $\text{CO}_2$  em áreas agrícolas e a sua relação com fatores ambientais e variáveis físicas e químicas do solo que controlam sua magnitude e estrutura de variação espaço-temporal (Graf et al., 2012).

Assim, a presente pesquisa considera um conjunto de dois experimentos relacionados à dinâmica de carbono do solo e seus atributos físico-químicos, envolvendo dados multidimensionais submetidos a algumas técnicas multivariadas: análise de componentes principais, análise fatorial e análise de correlação canônica.

Tem-se como hipótese para a pesquisa, que a análise das estruturas multivariadas dos atributos do solo auxilie e propicie ganho de informações quanto

a detecção de relações da emissão de CO<sub>2</sub> com variáveis físicas e químicas do solo, com o tipo de manejo realizado na área agrícola e cultura cultivada.

Diante do exposto, o objetivo desta pesquisa consiste em descobrir as possíveis relações existentes entre a emissão de CO<sub>2</sub> e atributos do solo, manejo em área agrícola e cultura cultivada, por meio da utilização de técnicas multivariadas.

## 2 REVISÃO DE LITERATURA

A revisão de literatura deste trabalho buscou apresentar conceitos teóricos das técnicas de análise multivariada que subsidiaram as análises dos dados que foram coletados.

### 2.1 Análise Multivariada

De acordo com Hair Jr. et al. (2009), a análise multivariada pode ser definida como o conjunto de todas as técnicas estatísticas que simultaneamente analisam múltiplas medidas referentes a indivíduos ou objetos sob investigação. Rencher (2002) apresenta uma definição muito similar, afirmando que a análise multivariada é constituída por um conjunto de métodos que podem ser utilizados quando são feitas várias medições em cada indivíduo, objeto ou observação em uma ou mais amostras. Morrison (1967) afirma que a análise estatística multivariada é voltada a dados coletados sob várias dimensões de um mesmo indivíduo.

Já Tabachnick & Fidell (2007), definem a estatística multivariada como um conjunto de técnicas estatísticas cada vez mais populares utilizadas para analisar conjuntos de dados complexos, com várias variáveis independentes correlacionadas entre si, em diferentes graus e/ou muitas variáveis dependentes, todas correlacionadas umas com as outras em diferentes graus.

Segundo Hair Jr. et al. (2009), o termo multivariado não é utilizado de uma maneira consistente na literatura estatística, sendo empregado por alguns autores apenas para examinar as relações entre duas ou mais variáveis, enquanto que para outros, o termo apenas se aplica diante da pressuposição de normalidade

multivariada.

Na prática, conjuntos de dados multivariados são comuns, embora nem sempre são analisados como tal, sendo muitas vezes avaliados por meio de procedimentos univariados, o que já não é mais justificável, dada a disponibilidade de técnicas de análise multivariada e poder computacional de baixo custo para realizá-las (Rencher, 2002).

Para Mingoti (2005), a análise multivariada consiste em técnicas exploratórias de sintetização da estrutura de variabilidade dos dados, composto por métodos como a análise de componentes principais, análise fatorial, análise de correlações canônicas, análise discriminante, análise de agrupamentos e análise de correspondência. Além disso a análise multivariada também apresenta técnicas de inferência estatística, composta pelos métodos de estimação de parâmetros, testes de hipóteses, análise de variância, de covariância e de regressão multivariadas.

Hair Jr. et al. (2009) classificam as técnicas multivariadas como técnicas de dependência e interdependência. Técnicas de dependência são aquelas que se constituem de uma ou um conjunto de variáveis dependentes que podem ser previstas ou explicadas por variáveis independentes, como por exemplo, regressão múltipla, análise de correlação canônica, análise discriminante, regressão logística, análise multivariada de variância e modelagem de equações estruturais.

Técnicas que não classificam as variáveis envolvidas como dependentes e/ou independentes, são ditas técnicas de interdependência, tais como: análise de componentes principais, análise fatorial, escalonamento multidimensional, análise de correspondência e análise de agrupamentos; as quais analisam as variáveis simultaneamente, buscando encontrar uma estrutura subjacente a todo o conjunto de variáveis.

Nas próximas seções são apresentadas as técnicas abordadas neste estudo.

## 2.2 Análise de Componentes Principais (ACP)

A análise de componentes principais (ACP) é uma técnica multivariada que está fundamentada no artigo de Hotelling (1933) e foi inicialmente descrita por Karl Pearson em 1901. Esta técnica consiste em explicar a estrutura de variância e covariância de um vetor aleatório, composto de  $p$ -variáveis aleatórias, por meio da construção de combinações lineares das variáveis originais; tais combinações são chamadas de “componentes principais” e são não correlacionadas entre si (Mingoti, 2005; Ferreira, 2008). Os objetivos gerais da análise são a redução da dimensão do espaço vetorial original sem perda da estrutura de variabilidade dos dados e maior facilidade na interpretação das análises realizadas (Johnson & Wichern, 2007).

De uma forma geral, a explicação de toda a variabilidade do sistema determinado por  $p$ -variáveis só pode ser efetuada por  $p$  componentes principais. No entanto, uma grande parte dessa variabilidade pode ser explicada por um número  $k$  de componentes, com  $k < p$ . Caso as variáveis originais sejam altamente correlacionadas entre si, então  $k$  pode assumir um valor pequeno, de forma que um grande número de variáveis originais sejam reduzidas a um pequeno número de variáveis transformadas (Mingoti, 2005; Lattin et al., 2011). No entanto se as variáveis originais forem não correlacionadas, sua redução em um número menor de variáveis transformadas nem sempre é possível (Manly, 2004).

Uma análise de componentes principais, muitas vezes, revela relações entre variáveis que não seriam facilmente visualizadas pela inspeção exploratória, permitindo assim interpretações que não seriam obtidas observando somente os dados originais (Johnson & Wichern, 2007).

Geometricamente, a ACP consiste em selecionar um novo sistema de coordenadas obtido por meio da rotação do sistema de eixos originais, fazendo com que esses novos eixos resultantes sejam posicionados na direção de máxima variabilidade e forneçam uma descrição mais simples e reduzida da estrutura de covariância, ou seja, tomando um conjunto de variáveis ( $p > 2$ ) que sejam fortemente relacionadas, tal rotação é capaz de reduzir as  $p$  dimensões originais em  $k < p$  co-

ordenadas sem que se perca muita informação contida no sistema original (Mingoti, 2005; Johnson & Wichern, 2007; Ferreira, 2008).

A obtenção das componentes principais envolve a decomposição espectral da matriz de variâncias e covariâncias das variáveis em estudo. No entanto, em situações reais, como a que se apresenta neste trabalho, utilizando-se de dois conjuntos amostrais constituídos por oito variáveis com 89 e 133 pontos amostrados respectivamente, as componentes principais são obtidas a partir da matriz de variância e covariância amostral ( $\mathbf{S}$ ) ou matriz de correlação amostral ( $\mathbf{R}$ ).

Em geral, as variáveis utilizadas para o estudo em diversas áreas do conhecimento encontram-se em unidades de medidas diferentes. Pode ocorrer também, a presença de *outliers* em algumas das variáveis. Estes fatos podem ocasionar grandes diferenças em suas variâncias. Dessa forma, em busca de tornar as variáveis aleatórias adimensionais é aconselhável a padronização dos dados, evitando que haja predominância nos primeiros componentes por parte de algumas variáveis (Jolliffe, 2002). O autor afirma ainda que, resultados obtidos a partir da matriz de correlação das análises para diferentes conjuntos de variáveis aleatórias são mais fáceis de serem interpretados do que para análises baseadas em matrizes de covariância.

Para a padronização das variáveis aleatórias considera-se, em geral, o procedimento usual de se obter variáveis com média zero e variância unitária. Dada  $\mathbf{X}_{n \times p} = (X_{ij})$  a matriz com  $n$  linhas que representam as observações multidimensionais das  $p$  variáveis que compõem suas colunas. Assim, padroniza-se por meio da transformação:

$$Z_{ij} = \frac{X_{ij} - \bar{X}_j}{s_j}$$

em que  $Z_{ij}$  indica quanto o valor original ( $X_{ij}$ ) se afasta da média amostral ( $\bar{X}_j$ ) por desvio padrão amostral ( $s_j$ ) ou por unidade de desvio padrão.

A padronização apresentada neste trabalho embora muito utilizada, não é única. Ainda existem outras maneiras de padronizar as variáveis em estudo, a fim de obter uma escala que se ajuste melhor ao problema abordado pelo pesquisador.

Desta forma, a partir da matriz  $\mathbf{R}_{p \times p}$ , de variâncias e covariâncias das

variáveis padronizadas, ou de correlação das variáveis originais, são extraídas as  $p$  raízes  $(\lambda_1, \lambda_2, \dots, \lambda_p)$  da equação característica  $\det(\mathbf{R} - \lambda I) = 0$ , denominadas autovalores em que  $I$  representa a matriz identidade de mesma ordem da matriz  $\mathbf{R}$ . Organiza-se os autovalores em ordem decrescente  $(\lambda_1 > \lambda_2 > \dots > \lambda_p)$  e posteriormente, calcula-se seus respectivos autovetores normalizados associados,  $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_p$  (Morrison, 1967; Mingoti, 2005), garantindo a unicidade da solução do sistema linear homogêneo.

Os autovetores  $\mathbf{e}_j$  satisfazem as seguintes condições:

- (i)  $\mathbf{e}'_j \mathbf{e}_{j'} = 0 \forall j \neq j'$  (ortogonalidade);
- (ii)  $\mathbf{e}'_j \mathbf{e}_j = 1 \forall j = 1, 2, \dots, p$  (normalização);
- (iii)  $\mathbf{R}_{p \times p} \mathbf{e}_j = \lambda_j \mathbf{e}_j, \quad \forall j = 1, 2, \dots, p$  (equação característica).

Quando a técnica de componentes principais é aplicada à matriz  $\mathbf{R}$ , as componentes construídas são combinações lineares das variáveis  $Z_j$  padronizadas (Mingoti, 2005). Com relação aos autovalores obtidos a partir da matriz  $\mathbf{R}$ , tem-se que:

A soma dos autovalores é igual ao traço da matriz  $\mathbf{R}$

$$tr(\mathbf{R}) = \sum_{j=1}^p \lambda_j = p$$

O determinante da matriz  $\mathbf{R}$  é igual ao produto de seus autovalores

$$\det(\mathbf{R}) = \prod_{j=1}^p \lambda_j$$

Modelando as variáveis aleatórias padronizadas a partir da matriz  $\mathbf{R}$ , a  $l$ -ésima componente principal, com  $l = 1, \dots, p$ , é dada por:

$$Y_l = \mathbf{e}'_l \mathbf{Z} = e_{l1} Z_1 + e_{l2} Z_2 + \dots + e_{lp} Z_p$$

em que  $\mathbf{Z}_{n \times p}$  é a matriz  $Z_{ij}$ .

A variância de  $Y_l$  é igual a  $\lambda_l$ ,  $l = 1, 2, \dots, p$  e  $Cov(Y_l, Y_{l'}) = 0$ , para todo  $l \neq l'$ .

De fato,

$$\begin{aligned}
Var(Y_l) &= Var(\mathbf{e}'_l \mathbf{Z}) = \mathbf{e}'_l Var(\mathbf{Z}) \mathbf{e}_l = \mathbf{e}'_l \mathbf{R} \mathbf{e}_l = \mathbf{e}'_l \lambda_l \mathbf{e}_l = \mathbf{e}'_l \mathbf{e}_l \lambda_l = \lambda_l \\
Cov(Y_l, Y_{l'}) &= Cov(\mathbf{e}'_l \mathbf{Z}, \mathbf{e}'_{l'} \mathbf{Z}) = \mathbf{e}'_l Cov(\mathbf{Z}, \mathbf{Z}) \mathbf{e}'_{l'} = \mathbf{e}'_l Var(\mathbf{Z}) \mathbf{e}'_{l'} = \mathbf{e}'_l \mathbf{R} \mathbf{e}'_{l'} = \\
&= \mathbf{e}'_l \lambda_l \mathbf{e}_{l'} = \mathbf{e}'_l \mathbf{e}_{l'} \lambda_l = 0,
\end{aligned}$$

A correlação entre a componente  $Y_l$  e a variável padronizada  $Z_j$  é igual a:

$$r_{Y_l, Z_j} = r_{Y_l, X_j} = e_{lj} \sqrt{\lambda_l}$$

Por consequência, segundo Mingoti (2005) as variáveis  $Z_j$  com os maiores coeficientes na componente principal  $Y_l$  são as mais correlacionadas com a componente. A proporção da variância total explicada pela  $l$ -ésima componente principal é igual a  $\frac{\lambda_l}{p}$ ,  $l = 1, 2, \dots, p$  (Johnson & Wichern, 2007) e a informação da variação acumulada nos primeiros  $k$  componentes, expressa em porcentagem, dada por:

$$Inf(Y_k) = \frac{\sum_{l=1}^k \lambda_l}{p} \times 100\%, \quad k = 1, 2, \dots, p$$

Por conta da ordenação decrescente dos autovalores, as componentes principais são apresentadas por ordem de importância. A primeira componente principal retém a maior porcentagem de explicação da variabilidade das variáveis  $Z_j$ . A segunda componente explica o máximo da variabilidade das  $Z_j$  que não foi explicada pela primeira componente, e assim por diante, até obter a última componente que explicará a menor porcentagem da variabilidade das variáveis  $Z_j$ .

Existem vários critérios para a seleção do número de componentes principais que devem ser utilizadas, para conjuntamente, reduzir a dimensão do espaço paramétrico explicando a maior parte da variância das variáveis e assim analisar os dados. Deve-se considerar também, que a escolha do critério a ser utilizado pode levar a diferentes resultados e tem característica subjetiva, uma vez que, constitui-se em uma decisão exploratória, de acordo com o pesquisador (Silva & Padovani, 2006).

Segundo Manly (2004), deve-se descartar componentes que expliquem somente uma pequena proporção da variação das variáveis. Para Hair Jr. et al. (2009), a técnica mais comumente utilizada é a da raiz latente, ou critério de Kaiser,

o qual sugere que para a matriz de correlação, sejam utilizadas as componentes que apresentam seus respectivos autovalores superiores a um, com o intuito de manter no novo sistema novas dimensões que representem pelo menos a informação de variância de uma variável.

Outro critério utilizado é o Gráfico dos autovalores ou “Scree-plot” visto apenas como uma alternativa gráfica, apresentando em ordem decrescente os  $j$  autovalores no eixo da abscissa e os  $\lambda_j$  na ordenada, escolhendo-se o número de componentes no ponto onde o gráfico apresenta um “cotovelo” (Ferreira, 2008) ou mostre uma estabilidade da variância da componente. Para Mingoti (2005) e Hair Jr. et al. (2009) o número de componentes “escolhidas” corresponde ao ponto em que os valores começam a se estabilizar (horizontalizar).

Na literatura existem ainda outras alternativas para a escolha do número de componentes; uma delas é por exemplo definir um ponto de corte em relação a quantidade da variação total que deve ser explicada pelo modelo reduzido, e então selecionar as componentes que acumulam esse valor.

De acordo com Reis (2001), o critério que utiliza o ponto de corte da quantidade do acúmulo da variância explicada é subjetivo, cabendo ao pesquisador definir um valor de interesse. Segundo Jolliffe (2002), citado por Ferreira (2008) esse valor deve estar entre 70% e 90% da variância total acumulada, semelhante ao indicado por Reis (2001) e Silva & Padovani (2006), que sugerem que este valor seja no mínimo 70%, por outro lado Rencher (2002) afirma que esse valor deve ser superior a 80%.

### 2.2.1 Biplot

O biplot é uma técnica multivariada desenvolvida por Gabriel (1971), e tem como objetivo principal representar graficamente de maneira aproximada uma matriz de dados  $\mathbf{Y}_{n \times p}$ , em outra matriz de baixa dimensão, tal que se possa visualizar em um plano as relações e inter-relações entre linhas e colunas de  $\mathbf{Y}$  (Klefens, 2009), ou seja, que se possa obter uma boa representação simultânea (aproximada) dos

indivíduos ou observações e das variáveis, em baixa dimensão.

A representação gráfica realizada por meio dos vetores, denominados marcadores,  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$  para as linhas de  $\mathbf{Y}$  e  $\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_p$  para suas colunas, é chamada de Biplot, e esse bi em seu nome indica justamente as duas informações obtidas, sendo uma em suas linhas, referentes às unidades amostrais ou amostras, e a outra em suas colunas, pertencentes às variáveis (Kleffens, 2009). O produto interno  $\mathbf{a}_i' \mathbf{b}_j$ , com  $i = 1, \dots, n$  e  $j = 1, \dots, p$  aproximações dos elementos  $y_{ij}$  da matriz  $\mathbf{Y}$  (Souza, 2010; Godoi et al., 2012).

Esta avaliação visual da estrutura da matriz de dados torna-se possível baseando-se na fatoração da matriz de dados originais pela decomposição em valores singulares (Souza, 2010).

De acordo com o exposto por Souza (2010) os valores da decomposição em valores singulares são distribuídos para os marcadores de linhas e colunas de diferentes maneiras, de tal forma, que se o interesse for a relação entre linhas, aplica-se o JK-Biplot, ou entre colunas, GH-Biplot, ou ainda entre linhas e colunas, SQRT-Biplot, sendo as três maneiras propostas por Gabriel, que podem ser encontradas com mais detalhes em (Souza, 2010).

### 2.3 Análise Fatorial (AF)

A Análise Fatorial é uma das técnicas multivariadas mais conhecidas e tem sido utilizada em diversas áreas. Introduzida por Spearman (1904), muitos dos exemplos iniciais apareceram nas áreas da psicologia e ciências sociais na tentativa de identificar os fatores relacionados com a inteligência humana e ligá-los, de algum modo, à etnia (Manly, 2004; Mingoti, 2005).

Classificada como uma técnica de interdependência, a AF busca reduzir o espaço paramétrico das variáveis respostas possibilitando uma melhor compreensão da estrutura de variabilidade. Seu objetivo principal é descrever a variabilidade original do vetor aleatório  $\mathbf{X}_{p \times 1}$ , em termos de um número menor  $m$  de variáveis não-observadas, denominadas fatores comuns ou variáveis latentes e que estão relacio-

nadas com o vetor original  $\mathbf{X}$  por meio de um modelo linear (Mingoti, 2005; Ferreira, 2008; Hair Jr. et al., 2009).

De acordo com Rencher (2002), a análise fatorial está relacionada com a ACP no sentido de buscar uma estrutura de variabilidade mais simples em um conjunto de variáveis. A análise de componentes principais parte de uma expressão matemática, enquanto que AF parte de um modelo estatístico, e ambas encerram-se com a fatoração da matriz de variâncias e covariâncias, ou correlação. No entanto ambas possuem diferentes abordagens, tais como o modelo mais elaborado em AF que inclui o efeito residual, podendo-se obter um número infinito de matrizes de pesos fatoriais, diferente da ACP que possui fatoração única por meio da decomposição espectral.

A análise de componentes principais pode ser caracterizada apenas como um processo matemático de redução do espaço paramétrico, já a AF gera novos fatores, possibilitando identificar o peso de cada variável em relação as variáveis latentes, sendo um componente indicativo de seleção de variáveis. Outra diferença é que a ACP busca explicar uma grande parte da variância total das variáveis em um número menor de componentes, pois para a explicação de toda a variabilidade são necessárias tantas componentes quantas forem as variáveis, enquanto que a AF busca explicar as covariâncias entre as mesmas.

Os fatores podem ser correlacionados (fatores oblíquos) ou não correlacionados (fatores ortogonais). Deve-se considerar nesta técnica, que parte da variabilidade do vetor  $\mathbf{X}$  é atribuída aos fatores comuns, sendo o restante atribuído às variáveis que não foram incluídas no modelo, ou seja, ao erro aleatório (Mingoti, 2005). Para Jobson (1992) e Ferreira (2008), os fatores comuns são responsáveis por explicar as covariâncias e variâncias, enquanto que a porção da variância das variáveis originais não explicada por estes é atribuída a termos de erros residuais chamados de fatores únicos ou específicos.

A AF pode ser vista como um procedimento estatístico para agrupar variáveis em subconjuntos (fatores) tais que, as variáveis associadas a cada fator

são altamente correlacionadas entre si, mas têm correlações relativamente pequenas com as variáveis que compõem um fator distinto (Jobson, 1992; Johnson & Wichern, 2007).

De acordo com Tabachnick & Fidell (2007), a análise fatorial é parcimoniosa quando o número de fatores para resumir os padrões de correlação entre as variáveis observadas for geralmente muito menor que o número de variáveis que participaram da análise.

Para a compreensão do modelo da análise fatorial, considera-se  $\mathbf{X}$  um vetor aleatório  $p$ -dimensional, com média amostral  $\bar{\mathbf{X}}$  e matriz de covariância  $\mathbf{S}$ . No modelo fatorial,  $\mathbf{X}$  é linearmente dependente de algumas variáveis aleatórias não observáveis  $F_1, F_2, \dots, F_m$ , chamados fatores comuns, e  $p$  fontes adicionais de variação  $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_p$ , chamados erros, ou fatores específicos. De acordo com Johnson & Wichern (2007) sendo  $\mathbf{Z}$ , o vetor das variáveis padronizadas, com média nula e variância unitária, em particular, o modelo ortogonal fatorial da análise fatorial é:

$$\mathbf{Z}_{(p \times 1)} = \mathbf{L}_{(p \times m)} \mathbf{F}_{(m \times 1)} + \boldsymbol{\varepsilon}_{(p \times 1)}$$

em que  $\mathbf{F}_{(m \times 1)}$  é um vetor aleatório com  $m$  fatores comuns não observáveis ( $1 \leq m \leq p$ ),  $\mathbf{L}_{(p \times m)}$  é uma matriz de parâmetros que serão estimados por algum procedimento e  $\boldsymbol{\varepsilon}_{(p \times 1)}$  é um vetor de erros aleatórios.

O coeficiente  $l_{jk}$ , com  $j = 1, \dots, p$  e  $k = 1, \dots, m$  que compõe a matriz  $\mathbf{L}_{(p \times m)}$  é chamado de carga fatorial ou “loading”, e é o coeficiente da  $j$ -ésima variável  $Z_j$ , ao  $k$ -ésimo fator  $F_k$ , indicando o grau de relacionamento linear entre  $Z_j$  e  $F_k$ .

As seguintes suposições são necessárias para que se possa operacionalizar a estimação do modelo:

$$\begin{aligned} &\mathbf{F}_{(m \times 1)} \text{ e } \boldsymbol{\varepsilon}_{(p \times 1)} \text{ são independentes} \\ &E(\mathbf{F}_{(m \times 1)}) = \mathbf{0}, \quad Cov(\mathbf{F}_{(m \times 1)}) = \mathbf{I}_{m \times m}, \quad E(\boldsymbol{\varepsilon}_{(p \times 1)}) = \mathbf{0}, \\ &Cov(\boldsymbol{\varepsilon}_{(p \times 1)}) = \boldsymbol{\psi}_p, \text{ em que } \boldsymbol{\psi}_p \text{ é uma matriz diagonal} \end{aligned}$$

Uma consequência imediata a estas suposições está relacionada a matriz de correlação teórica  $\mathbf{P}_{(p \times p)}$ , que pode ser reparametrizada na forma:

$$\mathbf{P}_{(p \times p)} = \mathbf{L}\mathbf{L}' + \boldsymbol{\psi}$$

De fato,

$$\begin{aligned} \mathbf{P}_{(p \times p)} &= Cov(\mathbf{Z}) = Cov(\mathbf{L}\mathbf{F} + \boldsymbol{\varepsilon}) = E[(\mathbf{L}\mathbf{F} + \boldsymbol{\varepsilon})(\mathbf{L}\mathbf{F} + \boldsymbol{\varepsilon})'] \\ &= E(\mathbf{L}\mathbf{F}\mathbf{F}'\mathbf{L}') + E(\mathbf{L}\mathbf{F}\boldsymbol{\varepsilon}') + E(\boldsymbol{\varepsilon}\mathbf{F}'\mathbf{L}') + E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \\ &= \mathbf{L}E(\mathbf{F}\mathbf{F}')\mathbf{L}' + \mathbf{L}E(\mathbf{F}\boldsymbol{\varepsilon}') + E(\boldsymbol{\varepsilon}\mathbf{F}')\mathbf{L}' + \boldsymbol{\psi} = \\ &= \mathbf{L}\mathbf{I}_m\mathbf{L}' + \mathbf{L}\mathbf{0}_{m \times p} + \mathbf{0}_{p \times m}\mathbf{L}' + \boldsymbol{\psi} = \mathbf{L}\mathbf{L}' + \boldsymbol{\psi} \end{aligned}$$

Ainda por independência de  $\mathbf{F}_{(m \times 1)}$  e  $\boldsymbol{\varepsilon}_{(p \times 1)}$ , temos que  $Cov(\boldsymbol{\varepsilon}, \mathbf{F}) = E(\boldsymbol{\varepsilon}\mathbf{F}') = \mathbf{0}$

A partir desse modelo, pode-se dizer que a variância da variável padronizada  $Z_j$  é decomposta em duas partes (Morrison, 1967; Mingoti, 2005; Johnson & Wichern, 2007): a primeira que representa a quantidade de variância explicada pelos  $m$  fatores incluídos no modelo, chamada de comunalidade, denotada por  $h_j^2 = l_{j1}^2 + l_{j2}^2 + \dots + l_{jm}^2$ ; e a segunda representando a parte da variabilidade de  $Z_j$  associada ao erro aleatório, ou seja, variabilidade não explicada pelos fatores e, comumente chamada de unicidade ou variância específica ( $\boldsymbol{\psi}$ ).

Infelizmente, nem sempre é possível realizar essa fatoração, principalmente quando  $m$  for muito menor que  $p$ , podendo-se obter soluções inconsistentes, sem interpretações estatísticas, como por exemplo, obter variância específica negativa (Johnson & Wichern, 2007).

A literatura apresenta alguns métodos para estimar as matrizes  $\mathbf{L}$  e  $\boldsymbol{\psi}$ . Um deles é o método das componentes principais a partir da matriz de correlação  $\mathbf{R}$ , comumente usado como análise exploratória de dados por não exigir suposições sobre a distribuição de probabilidade do vetor  $\mathbf{Z}$  (Mingoti, 2005). São mantidos os  $m$  autovalores ( $\lambda_k$ ) que representam maiores proporções da variância total, dessa forma, a quantidade de fatores será igual ao número de componentes retidos.

Segundo Johnson & Wichern (2007), a proporção da variância total explicada por cada fator,  $\mathbf{F}_k$  é expressa por  $\frac{\lambda_k}{p}$ . Dessa forma, pode-se obter as seguintes estimativas para as matrizes (Mingoti, 2005)

$$\hat{\mathbf{L}}_{p \times m} = [\sqrt{\hat{\lambda}_1}\hat{\mathbf{e}}_1 \quad \sqrt{\hat{\lambda}_2}\hat{\mathbf{e}}_2 \quad \dots \quad \sqrt{\hat{\lambda}_m}\hat{\mathbf{e}}_m]$$

$$\hat{\boldsymbol{\psi}}_{p \times p} = \text{diag}(\mathbf{R} - \hat{\mathbf{L}}_{p \times m} \hat{\mathbf{L}}'_{m \times p})$$

em que  $\text{diag}(\cdot)$  denota a matriz diagonal,  $\mathbf{R}$  a matriz de correlação e  $\mathbf{e}_k$  são os autovetores associados aos respectivos autovalores  $\lambda_k$ . Dessa forma, a matriz  $\mathbf{R}$  pode ser estimada por  $\mathbf{R} = \hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\boldsymbol{\psi}}$ .

Outros métodos utilizados para a estimação das matrizes são: o método dos fatores principais, método iterativo dos fatores principais e método de máxima verossimilhança.

Segundo Mingoti (2005), o método dos fatores principais e o método iterativo dos fatores principais, são descritos como um refinamento do método de componentes principais. Ferreira (2008) afirma que, ambos os métodos utilizam uma estimativa inicial ( $\boldsymbol{\psi}_0 = [\text{diag}(\mathbf{S}^{-1})]^{-1}$ ) da matriz diagonal das variâncias específicas  $\boldsymbol{\psi}$ , o que é negligenciado pelo método das componentes principais e também baseiam-se na minimização da soma de quadrados dos elementos de  $\mathbf{S} - \boldsymbol{\Sigma}$  ou  $\mathbf{R} - \mathbf{P}$ .

A ideia principal do método dos fatores principais é que se o valor utilizado para a matriz  $\boldsymbol{\psi}$  for adequado, a matriz  $\hat{\mathbf{L}}$  poderá ser estimada pela decomposição espectral da matriz  $\mathbf{S}_r = \mathbf{S} - \boldsymbol{\psi}_0 = \hat{\mathbf{L}}\hat{\mathbf{L}}'$ , que possuirá posto  $m$  e os seus  $p - m$  últimos autovalores serão nulos, devendo assim as cargas fatoriais estimadas representarem fielmente a matriz  $\mathbf{S}_r$ , mas na grande maioria dos casos esta reprodução não é exata (Ferreira, 2008); tal método, como o de componentes principais necessita que o valor  $m$  de fatores seja preestabelecido.

Buscando o aprimoramento das estimativas da comunalidade obtida a partir do método dos fatores principais, desenvolveu-se o método dos fatores iterativos. O procedimento de estimação da matriz  $\mathbf{L}$  é repetido sucessivamente, até que as estimativas das comunalidades e das variâncias específicas convirjam, o que nem sempre é possível (Rencher, 2002; Ferreira, 2008).

O método da máxima verossimilhança é indicado apenas quando o vetor aleatório das variáveis de interesse têm distribuição normal multivariada, sendo considerado um procedimento mais sofisticado e preciso que os mencionados anteriormente. O número  $m$  de fatores também deve ser estabelecido a priori, no entanto,

diferentemente do que acontece com o método de componentes principais, o aumento do valor  $m$  altera os coeficientes dos fatores obtidos anteriormente (Mingoti, 2005; Ferreira, 2008). As estimativas de máxima verossimilhança de  $\mathbf{L}$  e  $\boldsymbol{\psi}$  são as matrizes  $\hat{\mathbf{L}}$  e  $\hat{\boldsymbol{\psi}}$  que maximizam a função de verossimilhança, obtidas por procedimentos numéricos, os quais nem sempre convergem (Mingoti, 2005).

Para determinar o número de fatores a ser utilizado na análise, Rencher (2002) apresenta quatro critérios semelhantes àqueles utilizados para a escolha do número de componentes principais. Tais como: 1) escolher  $m$  para atingir uma porcentagem predeterminada da variância total explicada (em torno de 80%, segundo o autor); 2) escolher o valor de  $m$  igual ao número de autovalores maiores do que o valor médio, para a matriz de correlação,  $\mathbf{R}$ , este valor é igual a 1, já para a matriz de covariâncias,  $\mathbf{S}$ , é igual a  $\frac{\sum_{j=1}^p \lambda_j}{p}$ , 3) obter  $m$  por meio da análise do teste “scree”, baseado na análise do gráfico dos autovalores de  $\mathbf{S}$  ou  $\mathbf{R}$ , de forma que  $m$  seja igual ao número de autovalores obtidos antes da linha do gráfico se estabilizar, e por fim, 4) testar a hipótese de que  $m$  é o número adequado de fatores, sendo  $H_0 : \mathbf{S} = \hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\boldsymbol{\psi}}$ , no qual  $\hat{\mathbf{L}}$  é de ordem  $m \times p$ . A estatística de teste, é dada por meio da função da razão de verossimilhança:

$$\left(n - \frac{2p + 4m + 11}{6}\right) \ln\left(\frac{|\hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\boldsymbol{\psi}}|}{|\mathbf{S}|}\right)$$

que segue aproximadamente uma distribuição  $\chi^2$  com  $\frac{1}{2}[(p - m)^2 - p - m]$  graus de liberdade, quando  $H_0$  é verdadeira, com  $\hat{\mathbf{L}}$  e  $\hat{\boldsymbol{\psi}}$  sendo os estimadores de máxima verossimilhança. Assim, a rejeição de  $H_0$  implica que  $m$  é um número de fatores muito pequeno, sendo necessária a inclusão de mais fatores no modelo. Este método é apontado como um limite superior para o número de fatores selecionados, uma vez que, aponta mais fatores significativos quando comparado com os três outros métodos apresentados.

Diferente da ACP a AF não possui solução única, e muitas vezes a interpretação dos fatores obtidos inicialmente não é uma tarefa fácil. Uma medida que pode ser adotada é aplicar uma rotação nos fatores, dividida em rotação ortogonal e oblíqua, tornando-os mais interpretáveis sem alterar suas propriedades matemáti-

cas subjacentes, tais como as comunalidades e variâncias específicas (Tabachnick & Fidell, 2007).

A aplicação de rotação é sugerida quando os “loadings” (cargas fatoriais) possuem valores similares e não desprezíveis, indicando que a suposição de ortogonalidade está sendo violada (Mingoti, 2005). Dessa forma, pode-se utilizar o recurso da rotação ortogonal para obter soluções mais simples e teoricamente mais significativas (Reis, 2001), sendo o método Varimax, que se concentra na simplificação das colunas da matriz fatorial, o mais utilizado (Mingoti, 2005; Johnson & Wichern, 2007; Hair Jr. et al., 2009).

Como a carga fatorial é a correlação existente entre variável e o fator, é preciso tomar uma decisão sobre quais cargas fatoriais serão consideradas. Hair Jr. et al. (2009) apresentam dois tipos de significância de cargas fatoriais, classificadas como prática e estatística. A significância prática é aplicável quando trabalha-se com amostras de tamanho igual ou superior a 100 unidades observacionais, podendo categorizar cargas fatoriais entre 0,30 e 0,40 (em módulo) como minimamente aceitáveis para a interpretação da estrutura. Valores maiores ou iguais a 0,50 são rotulados como praticamente significantes, já valores maiores do que 0,70 são considerados indicativos de estrutura bem definida e são meta de qualquer análise fatorial.

De acordo com Hair Jr. et al. (2009), diferentemente de outras análises estatísticas, as suposições da AF estão mais relacionadas ao caráter e na composição das variáveis incluídas na análise. Inicialmente, é necessário que exista alguma estrutura de correlação subjacente no conjunto de variáveis, sendo papel do pesquisador garantir que os padrões observados sejam válidos e adequados. Rencher (2002) afirma que na análise fatorial, as suposições são essencialmente auto verificáveis, enquanto que em outros procedimentos, deve-se verificar os pressupostos com parcelas residuais, testes e assim por diante.

Hair Jr. et al. (2009) ressaltam ainda que as suposições da análise fatorial são mais conceituais do que estatísticas, mesmo que desvios da normalidade, homocedasticidade e da linearidade apenas diminuam as correlações entre as variá-

veis, os autores consideram que estas suposições fazem-se necessárias caso algum teste estatístico seja aplicado para a significância dos fatores. Afirmam também que multicolinearidades leves, podem ser até mesmo desejáveis, uma vez que um dos objetivos da análise é identificar conjuntos de variáveis inter-relacionadas.

Previamente à análise fatorial, deve-se avaliar se as variáveis são suficientemente correlacionadas umas com as outras produzindo assim fatores representativos. Para tal, Hair Jr. et al. (2009) sugerem realizar uma busca visual, verificando se as correlações obtidas são muito pequenas (menores que 0,30), ou muito semelhantes a ponto de não ser possível obter uma estrutura de agrupamento entre as variáveis.

A literatura apresenta outros critérios que permitem avaliar a viabilidade do modelo fatorial, como por exemplo a matriz de correlações parciais, o teste de esfericidade de Bartlett e as medidas de adequação da amostra (Hair Jr. et al., 2009).

Entende-se por correlação parcial aquela que não é explicada quando os efeitos de outras variáveis são levados em consideração. Se as correlações parciais forem altas ( $>0,7$ ), indica-se a ausência de fatores inerentes, portanto, pode ser um indicativo de que a análise fatorial é inadequada (Hair Jr. et al., 2009).

O teste de esfericidade de Bartlett, sob a pressuposição de normalidade multivariada, consiste em verificar se a matriz de correlação populacional é próxima ou não da matriz identidade (Mingoti, 2005), ou seja, se existem correlações significativas entre os pares de variáveis. Dessa forma as hipóteses testadas são:  $H_0 : \mathbf{P}_{p \times p} = I_{p \times p}$  contra  $H_1 : \mathbf{P}_{p \times p} \neq I_{p \times p}$ , com  $I_{p \times p}$  sendo a matriz identidade e  $\mathbf{P}_{p \times p}$  sendo a matriz de correlação populacional das  $p$ -variáveis. A estatística do teste é definida por:

$$T = -\left[n - \frac{1}{6}(2p + 1)\right] \left[ \sum_{j=1}^p \ln(\hat{\lambda}_j) \right]$$

sendo  $\hat{\lambda}_j$  os autovalores da matriz de correlações amostrais. Para  $n$  grande, a estatística tem uma distribuição assintótica de  $\chi^2$  com  $[\frac{1}{2}p(p-1)]$  graus de liberdade. Se

a hipótese nula do teste for rejeitada, o modelo fatorial pode ser ajustado (Mingoti, 2005).

Além destes critérios, há também, uma medida muito importante de adequação dos dados, que foi proposta por Kaiser-Meyer-Olkin, denominada por MSA ou KMO (critério de KMO). Tal medida de adequação é destinada a comparar as relações entre as variáveis e avaliar a adequação da análise fatorial (Reis, 2001). Esse índice varia de 0 a 1, e além de ser um critério global, pode ser ampliado à avaliação das variáveis individualmente. Hair Jr. et al. (2009) afirmam que o pesquisador deve examinar os valores de MSA para cada variável e eliminar da análise aquelas que apresentem valores inferiores a 0,5. De acordo com o apresentado por Rencher (2002), a estatística KMO é definida da seguinte maneira:

$$\text{KMO} = \frac{\sum_{j \neq j'} r_{jj'}^2}{\sum_{j \neq j'} r_{jj'}^2 + \sum_{j \neq j'} q_{jj'}^2}$$

no qual  $r_{ij}$  é o coeficiente de correlação entre as variáveis  $j$  e  $j'$  e  $q_{jj'}$  é o coeficiente de correlação parcial entre as referidas variáveis.

## 2.4 Análise de Correlação Canônica (ACC)

A análise de correlação canônica (ACC) é uma técnica estatística multivariada que foi inicialmente proposta por Hotelling (1936) e é considerada como a generalização do método de regressão múltipla (Rencher, 2002; Manly, 2004).

As intercorrelações de dois conjuntos de medidas aleatórias feitas sobre as mesmas unidades amostrais podem ser estudadas por meio desta técnica, que busca identificar e quantificar as associações entre os dois conjuntos (Cooley & Lohnes, 1962; Johnson & Wichern, 2007). Morrison (1967) ressalta que os referidos conjuntos devem ser formados a partir da natureza das respostas ou outros meios externos e nunca a partir da inspeção dos dados ou padrão das correlações das amostras.

O objetivo principal da ACC é determinar simultaneamente as relações entre dois conjuntos de variáveis aleatórias quantitativas, resumindo a informação de cada conjunto de variáveis observadas em funções lineares (Jobson, 1992), isto é,

tomando  $\mathbf{X}^{(1)}$  e  $\mathbf{X}^{(2)}$  dois conjuntos distintos, procura-se avaliar o grau de associação entre uma combinação linear das variáveis de  $\mathbf{X}^{(1)}$  e uma combinação linear das variáveis em  $\mathbf{X}^{(2)}$  (Johnson & Wichern, 2007).

Para cada conjunto de variáveis são determinadas várias funções lineares, chamadas de variáveis canônicas, de tal forma que a busca de seus coeficientes considere a maximização da correlação entre os grupos, representando uma tentativa de concentrar uma alta relação dimensional entre os conjuntos de variáveis em alguns pares de combinações lineares (Cooley & Lohnes, 1962; Johnson & Wichern, 2007).

A ideia é primeiramente determinar o par de combinações lineares que apresenta a maior correlação e, em seguida, determinar o par de combinação linear que tem a segunda maior correlação entre todos os pares não correlacionados com o par inicialmente selecionado, e assim, sucessivamente. As correlações dos pares de combinações lineares do primeiro e segundo conjuntos de variáveis canônicas são chamadas de correlações canônicas (Johnson & Wichern, 2007).

Para melhor entendimento do modelo teórico considere o vetor  $\mathbf{X}_{((p+q) \times 1)}$ , com matriz de variâncias e covariâncias amostral  $\mathbf{S}$  e vetor de médias amostrais  $\bar{\mathbf{X}}_{((p+q) \times 1)}$ . Sejam os vetores  $\mathbf{X}_{(p \times 1)}^{(1)}$  e  $\mathbf{X}_{(q \times 1)}^{(2)}$  definidos como sendo partições do vetor original  $\mathbf{X}$ , representando um grupo com  $p$  variáveis e outro com  $q$ , respectivamente (Morrison, 1967). O vetor de médias  $\bar{\mathbf{X}}_{((p+q) \times 1)}$  e a matriz de variância e covariância  $\mathbf{S}_{(p+q) \times (p+q)}$  de  $\mathbf{X}_{((p+q) \times 1)}$  também podem ser representados da seguinte forma:

$$\bar{\mathbf{X}}_{((p+q) \times 1)} = \begin{bmatrix} \bar{\mathbf{X}}^{(1)} \\ \bar{\mathbf{X}}^{(2)} \end{bmatrix}$$

$$\mathbf{S}_{(p+q) \times (p+q)} = \left[ \begin{array}{c|c} \mathbf{S}_{(p \times p)}^{11} & \mathbf{S}_{(p \times q)}^{12} \\ \hline \mathbf{S}_{(q \times p)}^{21} & \mathbf{S}_{(q \times q)}^{22} \end{array} \right]$$

As covariâncias entre os pares de variáveis pertencentes aos dois grupos,  $\mathbf{X}^{(1)}$  e  $\mathbf{X}^{(2)}$ , estão contidas em  $\mathbf{S}_{(p \times q)}^{12}$  ou  $\mathbf{S}_{(q \times p)}^{21}$  sendo uma matriz a transposta da

outra, e dessa forma, os  $pq$  elementos dessas matrizes medem par a par, a associação entre as variáveis dos dois grupos. O estudo dessas associações dá-se por meio de variáveis canônicas, que são construídas a partir das referidas combinações lineares das variáveis de ambos os grupos, obtendo-se  $r = \min(p, q)$  de variáveis desta natureza (Johnson & Wichern, 2007).

Sejam as variáveis  $U$  e  $V$  combinações lineares das variáveis  $\mathbf{X}^{(1)}$  e  $\mathbf{X}^{(2)}$ , respectivamente, definidas por:

$$U = \mathbf{a}'\mathbf{X}^{(1)} \quad V = \mathbf{b}'\mathbf{X}^{(2)}$$

em que  $\mathbf{a}_{p \times 1}$  e  $\mathbf{b}_{q \times 1}$  são vetores não nulos dos coeficientes dessas combinações lineares, respectivamente, escolhidos de forma a maximizar a correlação entre as variáveis canônicas. As variâncias das variáveis canônicas  $U$  e  $V$  são dadas por

$$\begin{aligned} Var(U) &= \mathbf{a}'Cov(\mathbf{X}^{(1)})\mathbf{a} = \mathbf{a}'\mathbf{S}_{(p \times p)}^{11}\mathbf{a} \\ Var(V) &= \mathbf{b}'Cov(\mathbf{X}^{(2)})\mathbf{b} = \mathbf{b}'\mathbf{S}_{(q \times q)}^{22}\mathbf{b} \end{aligned}$$

Da mesma forma, a covariância entre as variáveis canônicas é dada por

$$Cov(U, V) = \mathbf{a}'Cov(\mathbf{X}^{(1)}, \mathbf{X}^{(2)})\mathbf{b} = \mathbf{a}'\mathbf{S}_{(p \times q)}^{12}\mathbf{b}$$

Segundo Cooley & Lohnes (1962), o problema é encontrar  $\mathbf{a}$  e  $\mathbf{b}$ , tais que maximizem a correlação entre as variáveis canônicas. Para tal deve-se impor as seguintes restrições sobre a escala dos componentes lineares (Morrison, 1967)

$$\mathbf{a}'\mathbf{S}_{(p \times p)}^{11}\mathbf{a} = 1 \quad \mathbf{b}'\mathbf{S}_{(q \times q)}^{22}\mathbf{b} = 1$$

Tais restrições não alteram as correlações, pois estas são invariantes à mudança de escala, e foram impostas para se obter a unicidade das soluções, exceto pelos sinais, e também assegurar que essas variáveis canônicas sejam não-correlacionadas com outras dentro de cada um dos dois grupos de variáveis originais (Morrison, 1967; Ferreira, 2008). Além disso, as referidas restrições garantem a padronização da variância das variáveis canônicas.

De acordo com Tabachnick & Fidell (2007), existem vários caminhos para escrever a equação fundamental para a correlação canônica, alguns mais interessantes intuitivamente do que outros, no entanto, todos são variantes da mesma equação. Assim, os coeficientes  $a_k$  e  $b_k$  são obtidos a partir da solução não trivial do seguinte sistema linear de equações homogêneas:

$$\begin{cases} (\mathbf{S}^{12}(\mathbf{S}^{22})^{-1}\mathbf{S}^{21} - \lambda_k\mathbf{S}^{11})\mathbf{a}_k = \mathbf{0} \\ (\mathbf{S}^{21}(\mathbf{S}^{11})^{-1}\mathbf{S}^{12} - \lambda_k\mathbf{S}^{22})\mathbf{b}_k = \mathbf{0} \end{cases}$$

sendo  $\lambda_k$  o  $k$ -ésimo autovalor por ordem de importância, obtido a partir da matriz,  $(\mathbf{S}^{11})^{-1}\mathbf{S}^{12}(\mathbf{S}^{22})^{-1}\mathbf{S}^{21}$ , ou equivalentemente, da matriz  $(\mathbf{S}^{22})^{-1}\mathbf{S}^{21}(\mathbf{S}^{11})^{-1}\mathbf{S}^{12}$ .

Dessa forma é possível determinar os  $k$  pares de variáveis canônicas  $(U_k, V_k)$  para  $k = 1, \dots, \min(p, q)$  e a correlação existente entre essas variáveis canônicas, dada pelo coeficiente de correlação canônico (Mingoti, 2005; Johnson & Wichern, 2007) expresso por:

$$\rho_k = \text{Corr}(U_k, V_k) = \frac{(a'_k\mathbf{S}^{12}b_k)}{\sqrt{a'_k\mathbf{S}^{11}a_k}\sqrt{b'_k\mathbf{S}^{22}b_k}} = \sqrt{\lambda_k}$$

Segundo Morrison (1967), correlações idênticas podem ser obtidas a partir da matriz  $\mathbf{R}$ . Se os cálculos forem efetuados a partir da matriz  $\mathbf{R}$ , os autovalores obtidos são iguais aos adquiridos por meio da matriz  $\mathbf{S}$ , as variáveis canônicas são adimensionais devendo ser avaliadas em termos de escores padronizados das observações originais. Entretanto, quando os dados padronizados são utilizados, os autovetores obtidos não são os mesmos (Jobson, 1992).

De acordo com Mingoti (2005) e Johnson & Wichern (2007), se os vetores  $\mathbf{X}^{(1)}$  e  $\mathbf{X}^{(2)}$  forem independentes entre si, ou não correlacionados, a análise de correlação canônica é inútil. Ferreira (2008) ressalta que a qualidade da análise de correlação canônica depende de garantir que a matriz de covariância ou de correlação entre as variáveis dos dois vetores distintos seja diferente da matriz nula.

Tabachnick & Fidell (2007) afirmam que, enquanto a análise de correlação canônica for utilizada de forma descritiva, ou seja, apenas com o intuito de

estimar as correlações canônicas, sem testar sua significância, nenhuma exigência é feita em relação a distribuição das variáveis. Enquanto técnica exploratória a ACC, pode ser aplicada visando a simplificação estrutural das variáveis (Ferreira, 2008).

No entanto, um teste estatístico deve ser aplicado antes de se iniciar uma análise mais refinada das variáveis canônicas e de suas correlações. Para validar a análise de correlação canônica Johnson & Wichern (2007) apresentam o teste de Bartlett (Bartlett, 1939), que exige distribuição normal multivariada do conjunto de variáveis. O teste tem por hipóteses:

$$H_0 : \mathbf{P}^{12} = \mathbf{0} \quad (\rho_1 = \rho_2 = \dots = \rho_r = 0) \quad vs \quad H_1 : \mathbf{P}^{12} \neq \mathbf{0}$$

A estatística de teste é definida por

$$- \left( n - 1 - \frac{1}{2} (p + q + 1) \right) \ln \left( \prod_{i=1}^r (1 - \rho_i^2) \right)$$

com  $n$  o tamanho da amostra e  $\rho_i^2 = \lambda_i$ , a qual sob  $H_0$ , tem a distribuição  $\chi_{pq}^2(\alpha)$ . Se o valor da estatística do teste calculado para a amostra, for maior que o valor tabelado para o nível de significância  $\alpha$ , rejeita-se  $H_0$ , ou seja, a matriz de covariância, ou correlação, não é nula. De acordo com Manly (2004), um resultado não significativo indica que mesmo a maior correlação canônica pode ser explicada somente por variação de amostragem.

Se a hipótese nula  $H_0 : \mathbf{P}^{12} = \mathbf{0}$  for rejeitada, é natural buscar quais correlações canônicas são significativas (Rencher, 2002), ou seja, qual é o número de pares de variáveis canônicas que simplificam a representação da associação entre os dois conjuntos de variáveis. Deve-se considerar que as correlações canônicas estão ordenadas da maior para a menor, e pode-se começar assumindo a hipótese de que o primeiro coeficiente de correlação canônico é diferente de zero e os restantes  $r - 1$  são zero. Se esta hipótese for rejeitada, assume-se que as duas primeiras correlações canônicas são diferentes de zero e as restante  $r - 2$  são nulas, e assim por diante (Johnson & Wichern, 2007).

As seguintes hipóteses são consideradas, tomando  $m < r = \min(p, q)$ :

$$H_0^m : \rho_1 \neq 0, \rho_2 \neq 0, \dots, \rho_m \neq 0, \rho_{m+1} = \dots = \rho_r = 0$$

$$H_1^m : \rho_i \neq 0, \text{ para algum } i \geq m + 1$$

A hipótese nula é rejeitada caso a estatística de teste definida por:

$$- \left( n - 1 - \frac{1}{2} (p + q + 1) \right) \ln \left( \prod_{i=m+1}^r (1 - \rho_i^2) \right)$$

for maior que a estatística de teste associada, sob  $H_0$ , a uma distribuição  $\chi_{(p-m)(q-m)}^2$  com nível de significância  $\alpha$ .

Após verificar quais correlações canônicas encontradas são significativas, algumas medidas podem ser utilizadas para auxiliar na interpretação e na análise das variáveis canônicas. As correlações das variáveis canônicas com as variáveis originais e o percentual da variância total explicada podem ser utilizados (Mingoti, 2005).

As correlações entre as variáveis canônicas  $U_k$  e  $V_k$ , para  $k = 1, \dots, \min(p, q)$ , e as variáveis originais, utilizando a matriz de correlação estimada  $\mathbf{R}$ , são apresentadas a seguir

$$R_{U_k, X^{(1)}} = \mathbf{R}^{11} \mathbf{a}_k \quad R_{V_k, X^{(2)}} = \mathbf{R}^{22} \mathbf{b}_k$$

Outra medida é a correlação cruzada, conhecida como carga canônica cruzada, que consiste nas correlações entre a variável canônica  $U_k$  e as variáveis originais pertencentes ao vetor  $\mathbf{X}^{(2)}$ , e as correlações da variável canônica  $V_k$  com as variáveis do vetor  $\mathbf{X}^{(1)}$ , dadas por:

$$R_{U_k, X^{(2)}} = \mathbf{R}^{21} \mathbf{a}_k \quad R_{V_k, X^{(1)}} = \mathbf{R}^{12} \mathbf{b}_k$$

Para determinar a proporção da variância total dos dados originais explicada por cada variável canônica  $U_k$  e  $V_k$ , para  $k = 1, \dots, \min(p, q)$ , faz-se:

$$PVTE_{U_k} = \frac{\sum_{i=1}^p \text{corr}(U_k, X_i^{(1)})^2}{p} \times 100\%$$

$$PVTE_{V_k} = \frac{\sum_{i=1}^q \text{corr}(V_k, X_i^{(2)})^2}{q} \times 100\%$$

Esses valores medem a importância relativa das variáveis canônicas na reprodução das correlações amostrais dos dois conjuntos, podendo ser utilizadas como medidas relativas da qualidade da análise de correlação canônica efetuada (Ferreira, 2008).

## **2.5 Relação entre a emissão de CO<sub>2</sub> e atributos físicos e químicos do solo**

A emissão de CO<sub>2</sub> do solo é resultado de processos físicos, químicos e biológicos, que interferem na produção e transporte de CO<sub>2</sub> do solo para atmosfera, sendo a produção relacionada à ação dos microrganismos durante o processo de decomposição da matéria orgânica do solo e à respiração das raízes das plantas, enquanto que o transporte relaciona-se à estrutura física do solo, principalmente a porosidade que condiciona fluxo de gases no solo (Tavares, 2014).

De acordo com Martins (2016), a dinâmica do Carbono (C) no solo está essencialmente ligada à presença da Matéria Orgânica (MO). Segundo o autor, o solo é o compartimento terrestre que mais armazena carbono e pode ser uma fonte ou um sumidouro de CO<sub>2</sub> atmosférico. Ademais, afirma também que sistemas florestais têm uma enorme responsabilidade na dinâmica do carbono, estando seu balanço dependente das práticas de gestão associadas a estes sistemas, bem como as variações climáticas. Ambos podem conduzir a perdas dos contribuintes mais ativos, de nutrientes, de MO e carbono associado e alterar o balanço de C, entre “inputs” (respiração do ecossistema, pastoreio, erosão) e os “outputs” (fotossíntese e fornecimento de matéria orgânica) e, conseqüentemente o estoque de C no solo.

Mudanças no uso da terra provocam perturbações do ecossistema e podem influenciar os estoques e fluxos de carbono, como por exemplo a conversão da floresta em ecossistemas agrícolas, que afeta várias propriedades do solo, mais especialmente, a concentração e estoque de carbono no solo (Lal, 2005).

Panosso et al. (2011) analisaram as relações entre o FCO<sub>2</sub> e as pro-

priedades do solo em áreas de cana-de-açúcar sob colheita com queima da palha e colheita mecanizada, e a correlação da emissão de CO<sub>2</sub> com propriedades do solo em cada um dos sistemas de colheita. O estudo apontou que a conversão para colheita verde resultou em uma menor emissão de CO<sub>2</sub> do solo quando comparada com o sistema de queima, indicando estar relacionada com as alterações observadas em algumas propriedades físicas e químicas, especialmente ao índice de matéria orgânica e densidade do solo.

Em estudo semelhante os resultados em Tavares et al. (2015) indicaram que a emissão de CO<sub>2</sub> foi 36% maior no sistema de 10 anos de colheita verde, após encerrar as queimas naquela área. Dos resultados obtidos concluiu-se que a maioria das emissões de CO<sub>2</sub> foram em áreas de colheita verde, devido a maior concentração de carbono quando comparada com área de queima da cana, sendo a densidade do solo, a macroporosidade, o pH e o V% os parâmetros que mais influenciaram a emissão de CO<sub>2</sub>.

Ao avaliar a incerteza na predição da variabilidade espacial da emissão de CO<sub>2</sub> do solo (FCO<sub>2</sub>) e demais variáveis utilizando simulação sequencial gaussiana (SSG - método que busca reproduzir a variabilidade espacial dos dados eliminando o efeito de suavização da Krigagem Ordinária), Pereira et al. (2012) indicaram que as variações espaciais da FCO<sub>2</sub> são controladas principalmente por propriedades físicas, químicas e biológicas do solo, tais como densidade do solo, textura, teor de matéria orgânica, estoque de carbono, biomassa microbiana, pH, capacidade de troca catiônica e mineralogia do solo. Em escala temporal a FCO<sub>2</sub> é controlada por propriedades que variam ao longo do tempo, como a umidade do solo, porosidade livre de água (PLA) e oxigenação do solo.

Brito et al. (2009) relataram que em área que apresentou menor densidade e resistência do solo à penetração, maior quantidade de agregados e macroporosidade do solo; a emissão de CO<sub>2</sub> do solo foi maior, resultado que reforça a importância do espaço poroso do solo para o transporte gasoso e, conseqüentemente, para a atividade microbiana, encontrados por Xu & Qi (2001), no qual boa parte da

variação espacial foi explicada pela biomassa de raízes finas, biomassa microbiana e propriedades físicas.

De acordo com Teixeira et al. (2012), estudos desenvolvidos em áreas de cana-de-açúcar, no interior de São Paulo, mostraram que a densidade do solo (DS), juntamente com a porosidade livre de água (PLA) têm impacto direto no transporte de gás no interior do solo, afetando tanto a entrada do oxigênio necessário para a realização da atividade dos microrganismos aeróbicos, quanto a saída do CO<sub>2</sub>, que é um subproduto final dessa atividade. A matéria orgânica por sua vez, é a principal fonte de energia utilizada pelos microrganismos, explicando assim sua relação positiva com a respiração do solo.

Graf et al. (2012) utilizaram a análise de componentes principais (ACP) para extrair os mais importantes padrões da variabilidade espaço-temporal subjacente ao fluxo de CO<sub>2</sub>. O padrão espacial mais importante para o fluxo de CO<sub>2</sub> foi claramente relacionado com a umidade do solo e as propriedades físicas do solo de condução do gás. Temperatura, por outro lado, foi o fator mais importante no controle da variabilidade temporal da média espacial de respiração do solo.

Além dos fatores relacionados ao processo de transporte do carbono, ou seja, as variáveis físicas do solo, os atributos químicos também têm grande importância na dinâmica do carbono. O teor de fósforo é um elemento químico importante para o metabolismo de microrganismos, de acordo com o estudo realizado por Nordgren (1992); este atributo pode ser considerado um fator limitante à dinâmica da atividade microbiana do solo.

O pH pode afetar a quantidade e diversidade de microrganismos no solo (Moreira & Siqueira, 2006). Na agricultura, a calagem tem o papel de corrigir o pH e melhorar a absorção dos nutrientes pelas raízes das plantas. Xu & Qi (2001) em estudo do fluxo de CO<sub>2</sub> da superfície do solo em plantação de pinhos, na Califórnia, encontraram que a densidade do solo e valor de pH foram negativamente correlacionados com o fluxo CO<sub>2</sub>.

A capacidade de troca catiônica (CTC) e a saturação por bases (V%)

são importantes propriedades relacionadas à matéria orgânica do solo (MO). A CTC representa a capacidade do solo de liberação de vários nutrientes, favorecendo a manutenção da MO e da fertilidade do solo, já a V% é um excelente indicativo das condições gerais de fertilidade do solo, utilizada até como complemento na nomenclatura dos solos (Ronquim, 2010).

Em estudo realizado por Tavares et al. (2015), que consistiu em avaliar diferentes sistemas de manejo em relação ao ciclo do dióxido de carbono e parâmetros do solo (químicos, físicos e biológicos), observou-se que V% e o pH afetaram a variabilidade da emissão de CO<sub>2</sub> em 18% e 14%, respectivamente, especialmente na área de queima durante a pré-colheita da cana.

Dessa forma, práticas de manejo que resultam em modificações na matéria orgânica do solo e influenciam os atributos físicos e químicos do solo afetam diretamente a atividade microbiana e conseqüentemente a emissão de CO<sub>2</sub> (Tavares, 2014).

### 3 MATERIAIS E MÉTODOS

Este trabalho utilizou-se de dois conjuntos de dados provenientes de experimentos conduzidos em áreas agrícolas nos estados de São Paulo e Mato Grosso do Sul, cujo objetivo dos pesquisadores foi estudar a variabilidade espacial da emissão de CO<sub>2</sub> e atributos físicos e químicos do solo, por meio de malhas geoestatísticas com diferentes distribuições espaciais em áreas de 0,25 ha, no decorrer dos anos de 2008 e 2013, com 89 e 133 pontos amostrais, respectivamente, (Panosso, 2011; Brancaglioni et al., 2015). Aplicou-se, nestes dois conjuntos de dados, as seguintes técnicas multivariadas: Análise de Componentes Principais, Análise Fatorial e Análise de Correlação Canônica.

Uma técnica muito utilizada em trabalhos desta natureza, a análise de componentes principais foi aplicada neste trabalho com o intuito de encontrar combinações lineares das variáveis em estudo, através de um número mínimo de componentes, que pudessem explicar a maior parte da variabilidade do conjunto original de dados.

A análise fatorial, foi empregada com o propósito de tentar explicar as variáveis em estudo por meio de combinações de variáveis não observadas (fatores comuns) e um erro aleatório (fator específico, não relacionado com qualquer outro fator subjacente do modelo), permitindo-se observar em cada um dos fatores, agrupamentos de variáveis que possuem correlações mais fortes entre si, e a partir daí estudar tais associações. Tal técnica, é muito confundida com a ACP, sendo menos popularizada que esta, possuindo menos aplicações na literatura em conjunto desta natureza.

A aplicação da análise de correlação canônica teve como objetivo exibir

a relação existente entre dois conjuntos distintos de variáveis por meio de um pequeno número variáveis canônicas, maximizando as correlações possíveis entre esses grupos, ou seja, relações existentes entre as variáveis relacionadas a dinâmica do carbono e características físicas e químicas do solo. Tal aplicação foi pouco encontrada na literatura.

### 3.1 Descrição das áreas experimentais

**Experimento em cana-de-açúcar (2008):** foi conduzido em área comercial com coordenadas 21° 24' S e 48° 09' O e elevação de 550 m acima do nível do mar, localizada no município de Guariba, SP, destinada à produção de cana-de-açúcar (*Saccharum spp.*) por 38 anos sob manejo de cana crua, com colheita mecanizada, deixando resíduos vegetais na superfície do solo (Figura 1). O clima da região foi classificado como Aw, de acordo com Köppen (Ometto, 1981), tropical de verão chuvoso e inverno seco. A precipitação média foi de 1.425 mm, concentrada entre os meses de outubro e março, com 22,2 °C de temperatura média anual registrada na região, nos últimos 30 anos. A vegetação natural era constituída por floresta tropical subcaducifólia. O solo da área, segundo a classificação proposta pela EMBRAPA (2006), era um Latossolo Vermelho eutroférico, textura muito argilosa (LVef). A malha amostral utilizada era composta por 89 pontos amostrais (Figura 1), instalada em uma área de 50m × 50m, com distâncias mínimas de separação entre pontos de 0,50 m.

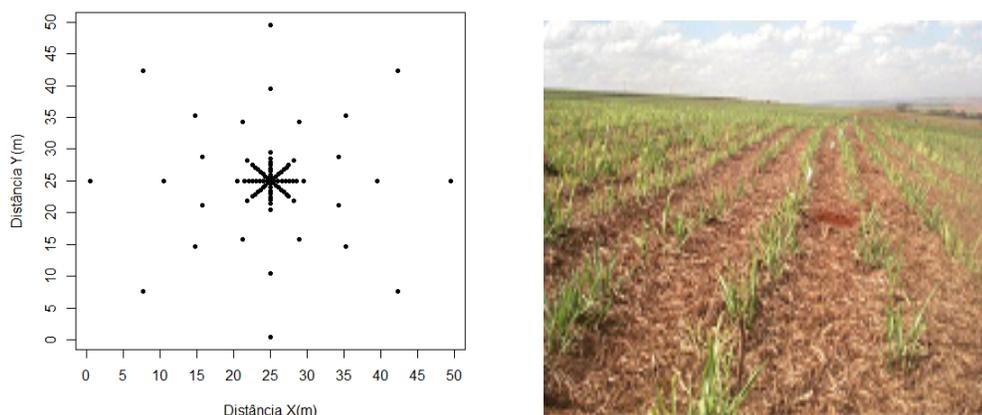


Figura 1 - Gradeado e área amostral de Guariba-SP para o experimento em cana-de-açúcar.

Fonte: Panosso (2011).

**Experimento em soja (2013):** conduzido na área experimental da Fazenda de Ensino, Pesquisa e Extensão – Setor de Produção Vegetal, da Faculdade de Engenharia de Ilha Solteira (FEIS/UNESP), localizada no município de Selvíria, MS (Figura 2). A cultura agrícola cultivada durante a condução do experimento foi a soja (*Glycine max* - L. Merrill), em sistema de plantio direto há 10 anos e histórico de pelo menos 25 anos de sucessão das culturas do milho (verão) e feijão (inverno). Com coordenadas geográficas 20° 18' 05" S e 52° 40' 28" O e elevação de 360 m acima do nível do mar. O tipo climático era Aw, segundo Köppen (Ometto, 1981), caracterizado como tropical úmido com estação chuvosa no verão e seca no inverno. A precipitação média anual foi de 1.300 mm, com temperatura média de 23,7 °C. O solo foi classificado como Latossolo Vermelho distroférico típico argiloso, com declive homogêneo de 0,025 mm<sup>-1</sup> (EMBRAPA, 2006). A vegetação natural era do tipo cerrado, caracterizada por apresentar árvores de porte médio (3-5 m) e gramíneas na superfície. A malha amostral (Figura 2) foi de 50m × 50m contendo 133 pontos amostrais, simétricos e com distâncias mínimas de separação entre pontos de 0,5 m.

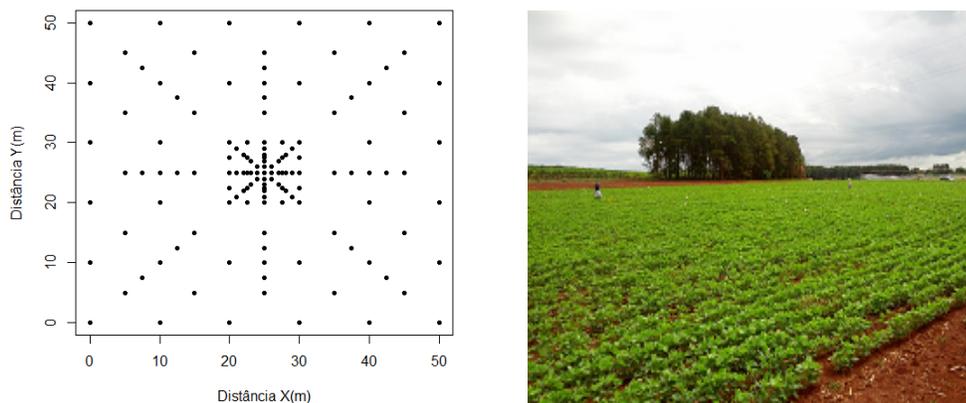


Figura 2 - Gradeado e área amostral de Selvíria-MS para o experimento em soja.

Fonte: Brancaglioni et al. (2015).

Os experimentos possuem basicamente a mesma distribuição dos pontos amostrados, no entanto em quantidades diferentes. Isso aconteceu pois, por meio do experimento desenvolvido em área de cana-de-açúcar em 2008, buscou-se testar a melhor forma de gradeado para se calcular a dimensão fractal do conjunto de dados. No entanto, com o aumento do número de equipamentos adquiridos pelo grupo de pesquisa, foi possível aumentar o número de pontos para melhorar a amostragem, como aconteceu com o experimento em soja, no ano de 2013.

O Quadro 1 fornece uma descrição resumida das áreas experimentais descritas, onde foram conduzidos os experimentos de variabilidade espacial da emissão de  $\text{CO}_2$  e atributos físicos e químicos do solo.

Quadro 1. Informações das áreas onde foram conduzidos os experimentos de variabilidade espacial da emissão de CO<sub>2</sub> e atributos físicos e químicos do solo

| Cidade/Estado | Cultura        | Manejo                    | Período Amostrado     | Número de Amostras |
|---------------|----------------|---------------------------|-----------------------|--------------------|
| Guariba/SP    | cana-de-açúcar | cana crua há 7 anos       | 25/10 a 17/11 de 2008 | 89                 |
| Selvíria/MS   | soja           | plantio direto há 10 anos | 19/11 a 10/12 de 2013 | 133                |

### 3.2 Coleta de dados

A emissão de CO<sub>2</sub> do solo (FCO<sub>2</sub>) e a temperatura do solo (TS) foram registradas concomitantemente por meio dos sistemas LICOR (LI-8100) (Figura 3). Para FCO<sub>2</sub>, o sistema monitora as mudanças na concentração de CO<sub>2</sub> dentro da câmara por meio de espectroscopia na região do infravermelho (Figura 3(a)). A câmara para solos (Figura 3(c)) tem um volume interno de 854,2 cm<sup>3</sup> com área de contato circular de 83,7 cm<sup>2</sup>. Essa câmara foi colocada sobre colares de PVC previamente inseridos no solo em cada ponto da malha amostral na profundidade de 3 cm. O fluxo de CO<sub>2</sub> foi computado em cada ponto por um ajuste da concentração de CO<sub>2</sub> do ar dentro da câmara em função de uma regressão parabólica no tempo após o fechamento da mesma (Silva, 2016). As avaliações de emissão de CO<sub>2</sub>, foram realizadas durante os estágios iniciais de crescimento das culturas, quando a emissão de CO<sub>2</sub> do solo relaciona-se, principalmente, à atividade microbiana do solo.



Figura 3 - (a) Sistema LI-8100 interligado à câmara de solos; (b) sensor de temperatura do solo e (c) câmara para solo inserida sobre o colar de PVC.

Fonte: Brancaglioni et al. (2015).

A temperatura do solo (TS) foi monitorada utilizando-se um sensor (Figura 3(b)) de temperatura que é parte integrante do sistema ao LI-8100. Tal sensor consiste em uma haste de 20 cm que foi inserida no interior do solo em região próxima aos colares de PVC. A umidade do solo (US) foi determinada por meio de um equipamento de TDR (“Time Domain Reflectometry” - Hydrosense TM, Campbell Scientific, Austrália). O aparelho de TDR (Figura 4) é constituído por uma sonda, apresentando duas hastes de 12 cm, inseridas no interior do solo em região próxima aos colares de PVC.



Figura 4 - TDR - “Hydrosense system”, sistema portátil utilizado para avaliação da umidade do solo.

Fonte:(Brancaglioni et al., 2015).

As medidas de FCO<sub>2</sub>, TS e US foram coletadas ao longo do tempo durante a condução do experimento (Quadro 1), no período da manhã. Para o experimento em cana-de-açúcar, as avaliações foram repetidas quatro vezes e a amostragem do solo (0,0 a 0,1 m) realizada em todos os pontos amostrais após o término das avaliações no campo. Já para o experimento em soja foram sete dias de coletas e a amostragem do solo (0,0 m a 0,1 m) foi realizada após a operação da colheita.

As amostras de solo coletadas foram secas ao ar, destorroadas e passadas na peneira de malha 2,0 mm, posteriormente foram feitas as seguintes análises de rotina: determinação do teor de matéria orgânica (MO), de P (Fósforo) disponível, K (Potássio), Ca (Cálcio), Mg (Magnésio) e H+Al (acidez potencial). A partir dos resultados, calculou-se a soma de bases (SB), a capacidade de troca de cátions (CTC) e a saturação por bases (V). Cálcio, magnésio e potássio trocáveis e fósforo disponível foram extraídos, utilizando-se do método da resina trocadora de íons (Raij, 2001).

A soma de bases é um excelente indicativo das condições gerais de fertilidade do solo e representa a soma dos teores de cátions permutáveis, sendo

obtida por meio da seguinte expressão

$$SB = K^+ + Ca^{2+} + Mg^{2+}$$

A troca de cátions nos solos é de grande importância para regular a retenção e a liberação de elementos químicos em forma disponível para plantas (Raij, 2001). Neste trabalho utilizou-se a CTC total, calculada da seguinte forma

$$CTC = K^+ + Ca^{2+} + Mg^{2+} + H+Al$$

A saturação por bases (V), muito utilizada na calagem, foi dada por:

$$V = \frac{SB}{CTC} \times 100\%$$

A densidade do solo (DS) foi determinada em amostras indeformadas (o solo é amostrado com as menores alterações possíveis em sua estrutura natural, não sofrendo deformações e/ou perturbações físicas) coletadas com amostrador adaptado a cilindros com 0,05 m de diâmetro interno e 0,04 m de altura (EMBRAPA, 1997). De acordo com Raij (2011), a densidade do solo representa o peso seco do volume do solo no campo, incluindo a matéria sólida e o volume de poros sem água, conforme a equação:

$$DS = \frac{PS}{VS}$$

com PS: peso seco do solo e VS: volume do solo incluindo poros. Calculou-se a porosidade total do solo (PT), com base no valor de densidade. Anteriormente a obtenção da PT, foi necessário calcular a densidade de partículas (DP) (Raij, 2011):

$$DP = \frac{PS}{VP}$$

com PS: peso seco do solo e VP: volume das partículas ou da matéria sólida do solo. Calculou-se a PT por meio da expressão:

$$PT = \left(1 - \frac{DS}{DP}\right) \times 100\%$$

A distribuição de poros por tamanho, macroporosidade (Macro) e microporosidade (Micro), foi determinada utilizando-se de funil de placa porosa sob a tensão de 0,6 m de altura de coluna d'água, em amostras previamente saturadas. O volume de água retido na amostra, nesta condição, corresponde aos microporos e os macroporos calculados por diferença (EMBRAPA, 1997).

$$\text{Macro} = \text{PT} - \text{Micro}$$

sendo a macroporosidade os poros de maior diâmetro, que permitem a livre drenagem de água e a movimentação de ar (Raij, 2011). A porosidade livre de água (PLA), também chamada de porosidade de aeração, é a fração do volume de uma amostra ocupada por ar. Calculou-se pela diferença entre a porosidade total (PT) e a fração da porosidade preenchida por água, equivalente à umidade do solo (US), definida pela expressão

$$\text{PLA} = \text{PT} - \text{US}$$

O estoque de carbono (EstC) foi calculado com base na massa de solo equivalente (Bayer et al., 2000), dado pela expressão

$$\text{EstC} = \frac{(\text{CO} \times \text{DS} \times \text{E})}{10}$$

em que EstC é o estoque de carbono ( $\text{Mg ha}^{-1}$ ); CO é o teor de carbono orgânico oxidável ( $\text{g kg}^{-1}$ ); Ds é a densidade do solo ( $\text{kg dm}^{-3}$ ), e E a espessura da camada estudada (cm), que neste trabalho foi de 10 cm.

Todos os atributos mensurados estão apresentados e descritos no Quadro 2, com suas respectivas classificações e unidades de medida.

Quadro 2. Identificação e descrição dos atributos mensurados em cada experimento

| Tipo de Atributo    | Identificação                                | Descrição                          |
|---------------------|--|------------------------------------|
| Dinâmica do carbono | FCO2 ( $\mu\text{mol m}^{-2}\text{s}^{-1}$ ) | Emissão de CO <sub>2</sub> do solo |
|                     | EstC (Mg ha <sup>-1</sup> )                  | Estoque de Carbono                 |
| Abiótico            | TS (°C)                                      | Temperatura do Solo                |
|                     | US (% volume)                                | Umidade do Solo                    |
| Físico              | DP (g cm <sup>-3</sup> )                     | Densidade de Partícula             |
|                     | DS (g cm <sup>-3</sup> )                     | Densidade do Solo                  |
|                     | PLA (%)                                      | Porosidade livre de água           |
|                     | Macro (%)                                    | Macroporosidade                    |
|                     | Micro (%)                                    | Microporosidade                    |
|                     | PT (%)                                       | Porosidade total                   |
| Químico             | pH   | Potencial Hidrogênico              |
|                     | MO (g dm <sup>-3</sup> )                     | Teor de Matéria Orgânica           |
|                     | P (mg dm <sup>-3</sup> )                     | Teor de Fósforo disponível         |
|                     | K (mmol <sub>c</sub> dm <sup>-3</sup> )      | Potássio                           |
|                     | Ca (mmol <sub>c</sub> dm <sup>-3</sup> )     | Cálcio                             |
|                     | Mg (mmol <sub>c</sub> dm <sup>-3</sup> )     | Magnésio                           |
|                     | H + Al (mmol <sub>c</sub> dm <sup>-3</sup> ) | Acidez Potencial                   |
|                     | SB (mmol <sub>c</sub> dm <sup>-3</sup> )     | Soma de Bases                      |
|                     | CTC (mmol <sub>c</sub> dm <sup>-3</sup> )    | Capacidade de Troca de Cátions     |
|                     | V (%)  | Saturação por Bases                |

### 3.3 Análise estatística

Os dados foram organizados em planilhas para a aplicação das técnicas multivariadas. Para os atributos FCO<sub>2</sub>, US e PLA foram utilizados os valores médios obtidos a partir das avaliações ao longo do tempo. Todas as variáveis envolvidas no estudo são quantitativas e pela conformação do experimento são regionalizáveis, portanto atributos que apresentaram valores faltantes foram estimados por meio da média de seus vizinhos mais próximos. Tal procedimento foi necessário para completar o vetor de dados, uma vez que as técnicas multivariadas apresentam sensibilidade a dados faltantes.

A emissão de CO<sub>2</sub> e os atributos físico-químicos do solo, foram inicialmente analisados por meio da estatística descritiva e exploratória dos dados na qual se incluem os métodos gráficos e avaliação dos coeficientes de assimetria e curtose. Em seguida, pela matriz de correlações amostral e da relação Nc (razão do maior autovalor pelo menor), buscou-se detectar possíveis multicolinearidades em cada conjunto de dados. Para tal as técnicas utilizadas foram: fator de inflação de variância e relação do número de condição, Nc (razão do maior autovalor pelo menor).

O fator de inflação de variância, de acordo com Hair Jr. et al. (2009), é calculado como o inverso do valor de tolerância, dado por:

$$VIF(j) = \frac{1}{1 - R_j^2}$$

no qual  $R_j^2$  corresponde a quantia da variável em questão que é explicada por todas as demais variáveis independentes no modelo de regressão.

Quanto maior for o  $VIF$ , mais severa é a multicolinearidade. Entende-se por multicolinearidade, quando qualquer variável independente é altamente correlacionada com um conjunto de variáveis independentes. Uma regra prática aceitável é a de que, se  $VIF(j) > 10$ , a colinearidade existente entre a variável “ $j$ ” e as demais é notável (Panosso et al., 2011).

O  $VIF$  foi utilizado para examinar a multicolinearidade existente entre as variáveis em estudo e escolher o conjunto de propriedades do solo para compor

as análises multivariadas, uma vez que muitas variáveis apresentavam dependência entre si. Dessa forma, buscou-se utilizar esse valor como parâmetro para excluir da análise variáveis que poderiam ser explicadas por outras.

Outro critério utilizado foi a relação  $N_c$  ou número de condição, que consiste na razão do maior autovalor obtido da matriz de correlações pelo menor autovalor, buscando identificar a presença de multicolinearidade. Para valores de  $N_c$  menores do que 100, aponta-se a existência de multicolinearidade fraca, sendo considerada moderada para valores entre 100 e 1000, e acentuada para valores de  $N_c > 1000$  (Toebe & Cargnelutti Filho, 2013). Neste sentido, as variáveis mais correlacionadas entre si foram retiradas da análise uma a uma (Maestre et al., 2010). A cada retirada, calculou-se novamente a relação  $N_c$  com a nova combinação de variáveis, até que se obteve o menor valor para  $N_c$  possível, utilizando assim esta combinação como base para a aplicação das técnicas estudadas.

O conjunto inicial dos dados era composto por 20 variáveis. No entanto, por meio das técnicas de diagnóstico de multicolinearidade, optou-se por manter nas análises as variáveis: FCO<sub>2</sub>, EstC, PLA, DS, Macro, pH, P e SB.

Por tratar-se de um trabalho com caráter unicamente exploratório, as características dos dados foram visualmente avaliadas, sem enfatizar a hipótese de normalidade multivariada dos dados.

Os conjuntos de dados referentes aos dois experimentos, foram analisados envolvendo as técnicas de componentes principais, análise fatorial e análise de correlação canônica realizadas por meio do software R (R Core Team, 2016), por se tratar de um software livre, com seus códigos abertos, reproduzíveis e adaptáveis, além de possibilitar a utilização de pacotes já desenvolvidos e certificados, para a aplicação das técnicas.

## 4 RESULTADOS E DISCUSSÕES

### 4.1 Experimento em área de cana-de-açúcar (2008)

As análises preliminares das estatísticas descritivas dos dados de cana-de-açúcar são apresentadas na Tabela 1. A emissão média de CO<sub>2</sub> do solo (2,203  $\mu\text{mol m}^{-2}\text{s}^{-1}$ ) foi maior do que aquela registrada por Teixeira et al. (2012) em experimento também realizado em área de cultura de cana-de-açúcar sob o manejo de colheita mecanizada. Tal diferença pode ter ocorrido devido a alta compactação do solo, indicando valores médios de 1,50 g cm<sup>-3</sup>; e baixo teor de matéria orgânica (4,75 g dm<sup>-3</sup>), quando comparados com os valores médios apresentados para a densidade do solo e matéria orgânica nesse trabalho.

O coeficiente de variação (CV) para FCO<sub>2</sub> amostrado foi de 36,95% (Tabela 1), valor menor do que aquele apresentado por Teixeira et al. (2012), mas muito semelhante ao encontrado por Teixeira et al. (2011) para experimento conduzido em área da fazenda experimental da Faculdade de Ciências Agrárias e Veterinárias - FCAV/UNESP, que obteve os coeficientes de variação 34,55% e 34,77% para a emissão de CO<sub>2</sub> do solo durante o período da manhã e da tarde, respectivamente, sendo classificados como alto (CV>24%), de acordo com o critério de classificação do coeficiente de variação proposto por Warrick & Nielsen (1980). Esses resultados se enquadram com aqueles apresentados por diversos autores em áreas de cana-de-açúcar (Brito et al., 2009; Silva et al., 2014; Tavares et al., 2015).

Pelo critério de Warrick & Nielsen (1980), podem ser considerados de baixa variabilidade (CV < 12%) os valores observados para EstC, US, DP, DS, Micro, PT, pH, MO e CTC e variabilidade moderada (12% < CV < 24%) em P, Ca, H+Al,

SB e V. Juntamente com FCO<sub>2</sub>, as variáveis PLA, Macro, K e Mg tiveram seus CVs classificados como altos ( $CV > 24\%$ ).

Com o intuito de evitar redundância dessas variáveis, aplicou-se a técnica do diagnóstico de multicolinearidade. Os resultados exibiram a relevância de cada atributo estudado, permitindo assim a seleção do subconjunto na aplicação das técnicas multivariadas. Logo, foram eleitos os seguintes atributos: FCO<sub>2</sub>, EstC, PLA, DS, Macro, pH, P e SB.

Em virtude da dependência existente, as demais variáveis foram removidas por apresentarem fortes correlações lineares, e  $VIF > 10$ . Vale ressaltar que o descarte de um número considerável de variáveis já era esperado, uma vez que muitos atributos são oriundos da combinação linear dos outros, como por exemplo  $\text{Macro} = \text{PT} - \text{Micro}$ . Dessa forma, reduziu-se o valor de Nc de  $1,88 \times 10^{16}$  para aproximadamente 3,43. Esse valor elevado de Nc deve-se ao fato que, quando uma das variáveis for função linear da outra, o menor autovalor assume um valor próximo de zero, elevando o valor do quociente.

Tabela 1. Estatística descritiva das variáveis para o experimento em área de cana-de-açúcar

|       | Média   | Mediana | EP    | Mínimo | Máximo  | Q1     | Q3      | CV     | Desv P | Assim. | Curt.  |
|-------|---------|---------|-------|--------|---------|--------|---------|--------|--------|--------|--------|
| FCO2  | 2,203   | 2,120   | 0,086 | 0,750  | 4,540   | 1,630  | 2,730   | 36,954 | 0,814  | 0,683  | 0,125  |
| EstC  | 8,031   | 7,962   | 0,082 | 6,668  | 9,717   | 7,450  | 8,598   | 9,619  | 0,773  | 0,203  | -0,843 |
| US    | 27,268  | 27,450  | 0,317 | 20,750 | 33,525  | 25,000 | 29,750  | 10,969 | 2,991  | -0,205 | -0,542 |
| PLA   | 15,512  | 14,666  | 0,444 | 7,397  | 26,708  | 12,543 | 18,822  | 27,028 | 4,193  | 0,494  | -0,301 |
| DP    | 2,044   | 2,043   | 0,008 | 1,860  | 2,229   | 2,000  | 2,084   | 3,467  | 0,071  | 0,029  | 0,092  |
| DS    | 1,170   | 1,185   | 0,007 | 1,014  | 1,278   | 1,125  | 1,216   | 5,482  | 0,064  | -0,509 | -0,507 |
| Macro | 6,598   | 6,224   | 0,331 | 1,792  | 16,114  | 4,172  | 8,375   | 47,303 | 3,121  | 0,922  | 1,013  |
| Micro | 36,183  | 36,308  | 0,209 | 30,227 | 40,074  | 34,976 | 37,585  | 5,443  | 1,969  | -0,556 | 0,504  |
| PT    | 42,780  | 42,761  | 0,249 | 37,827 | 47,630  | 40,818 | 44,322  | 5,494  | 2,350  | 0,270  | -0,766 |
| pH    | 4,596   | 4,600   | 0,027 | 4,000  | 5,300   | 4,400  | 4,700   | 5,479  | 0,252  | 0,476  | 1,010  |
| MO    | 24,213  | 24,000  | 0,240 | 20,000 | 30,000  | 23,000 | 26,000  | 9,370  | 2,269  | 0,112  | -0,583 |
| P     | 17,243  | 17,000  | 0,413 | 10,000 | 29,000  | 15,000 | 19,000  | 22,592 | 3,896  | 0,320  | 0,195  |
| K     | 5,790   | 5,800   | 0,193 | 2,700  | 12,500  | 4,700  | 6,500   | 31,459 | 1,821  | 1,120  | 2,838  |
| Ca    | 27,742  | 27,000  | 0,697 | 16,000 | 49,000  | 23,000 | 31,000  | 23,703 | 6,576  | 0,732  | 1,131  |
| Mg    | 11,618  | 12,000  | 0,325 | 6,000  | 21,000  | 10,000 | 13,000  | 26,419 | 3,069  | 0,505  | 0,780  |
| H_Al  | 58,596  | 58,000  | 1,253 | 38,000 | 88,000  | 52,000 | 64,000  | 20,178 | 11,823 | 0,566  | 0,424  |
| SB    | 45,149  | 43,900  | 0,958 | 25,800 | 76,400  | 39,500 | 50,400  | 20,020 | 9,039  | 0,894  | 1,830  |
| CTC   | 103,970 | 103,400 | 1,002 | 83,800 | 133,100 | 97,100 | 109,400 | 9,091  | 9,452  | 0,645  | 0,318  |
| V     | 43,585  | 43,540  | 0,898 | 25,170 | 66,783  | 36,897 | 48,029  | 19,428 | 8,468  | 0,455  | 0,328  |

EP: erro padrão; Q1: primeiro quartil; Q3: terceiro quartil; CV: coeficiente de variação; Desv P: desvio padrão; Assim.: coeficiente de assimetria; Curt.: coeficiente de curtose.

Os resultados apontaram valores próximos a zero para os coeficientes de assimetria e curtose das variáveis que foram selecionadas. Atributos como a Macro e SB, apresentaram coeficientes levemente expressivos de assimetria 0,922 e 0,894; e curtose 1,013 e 1,830, respectivamente, o que indica uma distribuição de probabilidade com assimetria à direita ( $>0$ ) e forma de distribuição leptocúrtica, como a variável pH que também apresentou curtose leptocúrtica ( $1,010 > 0$ ). A variável FCO2 apresentou leve assimetria à direita ( $0,683 > 0$ ) com valor semelhante aqueles encontrados por Teixeira et al. (2011) e Teixeira et al. (2012).

Os histogramas (Figura 5) permitem a observação dos moderados va-

lores de assimetria e curtose na distribuição do atributo selecionado. Os resultados sugerem leve assimetria à esquerda para as variáveis FCO2, Macro e SB. Quanto ao atributo DS, o qual exibiu assimetria à direita, destacou-se por apresentar visualmente alterações mais expressivas.

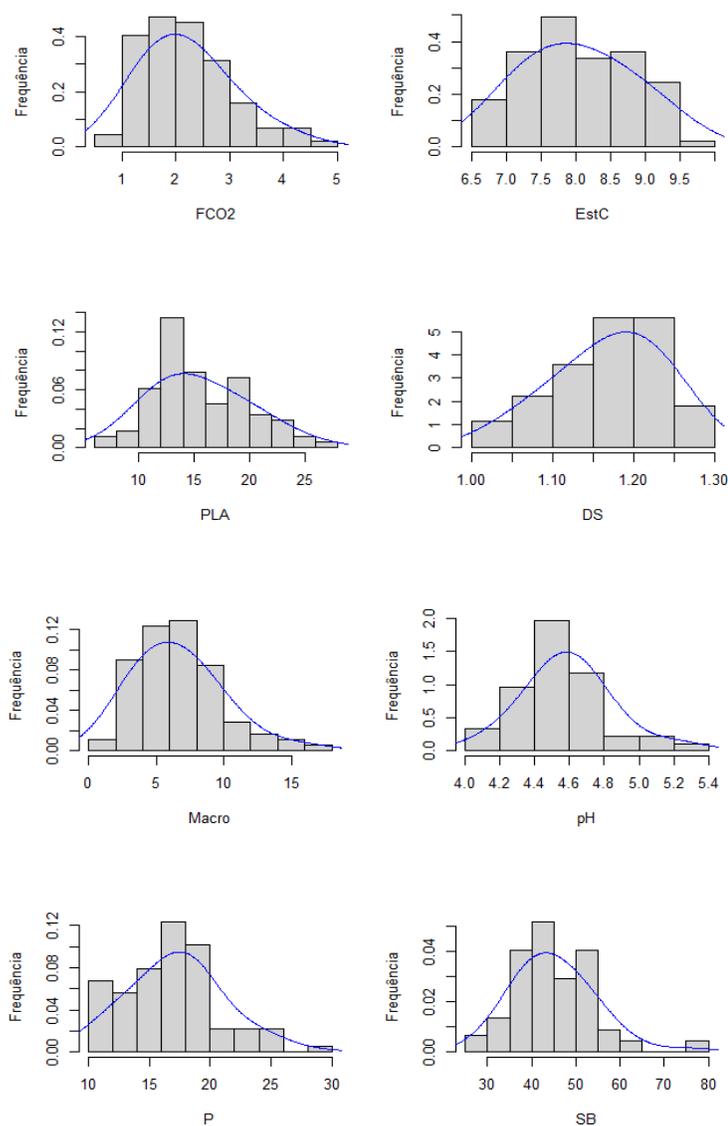


Figura 5 - Histogramas e curvas de distribuição das variáveis do experimento em área de cana-de-açúcar.

Em relação a curtose, os atributos EstC e PLA apresentaram maior achatamento em suas curvas, diferenciando-se do pH e SB que exibiram curvas menos suavizadas. No entanto, por inspeção gráfica, pode-se sugerir aderência a normalidade univariada das variáveis.

Para verificar a existência da normalidade multivariada dos atributos do solo, utilizou-se o gráfico de probabilidade *Q-Q Plot* (*Chi-Square plot*) multivariado, apresentado na Figura 6. Por ser uma ferramenta muito útil para verificação da adequação de distribuição de frequência dos dados a uma determinada distribuição de probabilidade, verificou-se no contexto global a tendência das distâncias de Mahalanobis alinharem-se ao redor de uma reta. Assim, por inspeção visual, pode-se observar que os padrões do gráfico *Q-Q Plot* obedecem as características de multinormalidade, embora pelos valores de assimetria e curtose da Tabela 1 e pelos histogramas da Figura 5, nem todas as variáveis parecem se aderir a normalidade, quando analisado de forma univariada.

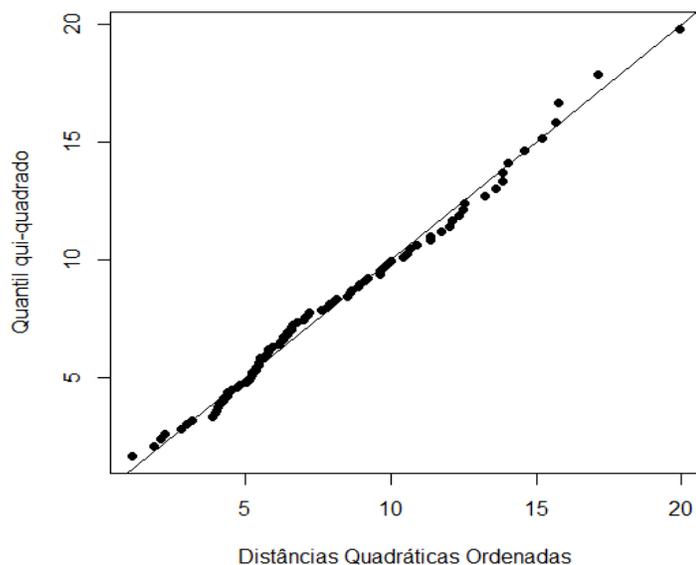


Figura 6 - Gráfico de probabilidade *Q-Q Plot* multivariado para os dados do experimento em área de cana-de-açúcar.

Decidiu-se, aplicar o teste t-Student para verificar a significância da correlação, para cada par de variáveis.

A matriz de correlação amostral com o indicativo do grau de correlação por meio de quadrados coloridos, com vivacidades diferentes de azul e vermelho, é destacada na Figura 7. Os quadrados em tons de azul indicam correlações positivas e os quadrados vermelhos, correlações negativas. Quanto mais intensa a tonalidade da cor, maior o grau de associação, tanto positiva quanto negativa, e quanto mais clara, menor é o grau de associação linear.

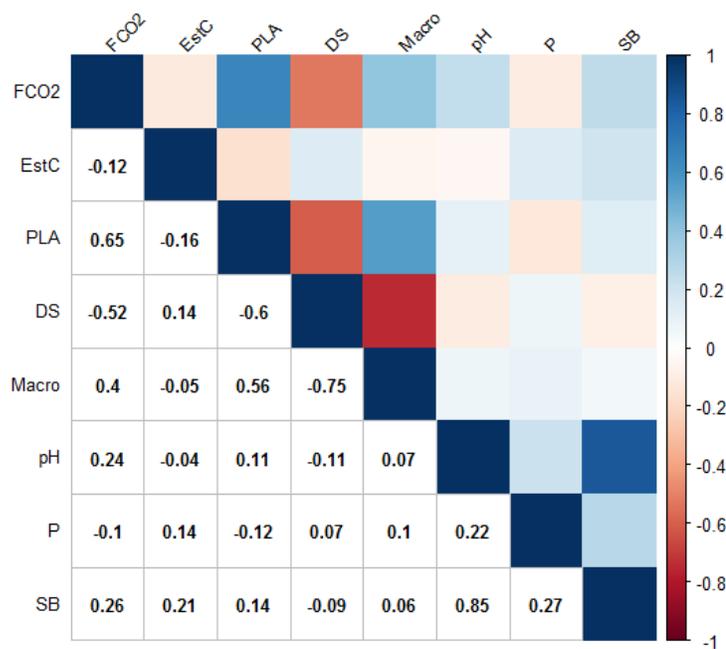


Figura 7 - Correlações das variáveis para o experimento em cana-de-açúcar.

Os resultados apontaram a existência de correlações positivas significativas entre a variável FCO2 e as variáveis PLA (0,65,  $p < 0,05$ ), Macro (0,40,  $p < 0,05$ ) e negativa em relação a DS (-0,52,  $p < 0,05$ ). Tais correlações corroboram com aquelas encontradas por Bicalho et al. (2014), em estudo que buscou investigar a estrutura de variabilidade espacial do FCO2 do solo e seus atributos em uma área de colheita me-

canizada e com os resultados encontrados por vários outros autores (Xu & Qi, 2001; Teixeira et al., 2012, 2013). Bicalho et al. (2014) afirmaram ainda que, tais correlações lineares sugerem a importância dos atributos físicos do solo para a atividade microbiana do solo e trocas de gases com a atmosfera.

As variáveis FCO<sub>2</sub> e EstC apresentaram correlação negativa muito baixa, sugerindo que a emissão de gases para a atmosfera não se correlaciona com o estoque de carbono existente na área. Este fato, contraria o que foi proposto por Scharlemann et al. (2014), citado por Leal et al. (2015), que afirmaram existir dependência entre a emissão de CO<sub>2</sub> do solo e o estoque de carbono. Este resultado pode sugerir que o aumento de estoque de carbono na área não provoque o aumento de emissões de CO<sub>2</sub> do solo naquele local.

Outras correlações que se destacaram, foram a soma de base (SB) e índice de pH do solo (0,85,  $p < 0,05$ ), SB e P (0,27,  $p < 0,05$ ). Em relação às variáveis químicas do solo, a FCO<sub>2</sub> apresentou maiores correlações com pH e SB, com valores iguais a 0,24 e 0,26, respectivamente. Os demais pares de variáveis apresentaram baixas correlações.

#### **4.1.1 Análise de Componentes Principais**

Como estabelecido anteriormente, elegeu-se por meio da análise de multicolinearidade as variáveis: FCO<sub>2</sub>, EstC, PLA, Macro, pH, P e SB. A análise de componentes principais foi realizada a partir da matriz de correlação (Tabela 2). Em razão das diferentes unidades de medidas das variáveis e variâncias distintas, (maior variância igual a 81,706 e a menor igual a 0,001), optou-se por evitar o efeito de escala, por meio da padronização das variáveis, de modo a tornarem as variáveis analisadas adimensionais.

Tabela 2. Variâncias, covariâncias (**S**) e correlações (**R**) das variáveis do experimento em cana-de-açúcar<sup>1</sup>

|       | FCO2         | EstC         | PLA           | DS           | Macro        | pH           | P             | SB            |
|-------|--------------|--------------|---------------|--------------|--------------|--------------|---------------|---------------|
| FCO2  | <b>0,663</b> | -0,118       | 0,654*        | -0,522*      | 0,398*       | 0,243*       | -0,102        | 0,258         |
| EstC  | -0,074       | <b>0,597</b> | -0,164        | 0,144        | -0,054       | -0,045       | 0,144         | 0,206*        |
| PLA   | 2,232        | -0,532       | <b>17,578</b> | -0,603*      | 0,558*       | 0,109        | -0,123        | 0,137         |
| DS    | -0,027       | 0,007        | -0,162        | <b>0,004</b> | -0,746*      | -0,110       | 0,071         | -0,086        |
| Macro | 1,012        | -0,131       | 7,302         | -0,149       | <b>9,740</b> | 0,071        | 0,098         | 0,060         |
| pH    | 0,050        | -0,009       | 0,115         | -0,002       | 0,056        | <b>0,063</b> | 0,218*        | 0,848*        |
| P     | -0,323       | 0,433        | -2,002        | 0,018        | 1,195        | 0,214        | <b>15,176</b> | 0,271*        |
| SB    | 1,902        | 1,438        | 5,176         | -0,050       | 1,682        | 1,930        | 9,548         | <b>81,706</b> |

<sup>1</sup> As variâncias estão na diagonal principal, as covariâncias e as correlações estão apresentadas abaixo e acima da diagonal principal, respectivamente.

\* Valores significativos do coeficiente de correlação de Pearson ( $p < 0,05$ ).

Utilizou-se o critério de Kaiser, que consiste em considerar apenas as componentes relacionadas aos autovalores maiores ou iguais a unidade (Kaiser, 1958). Os resultados sugerem a utilização de três componentes principais, que acumularam 73,85% da variabilidade total dos dados. As componentes principais e suas respectivas porcentagens de explicações da variância total da amostra estão na Tabela 3, e os autovalores ordenados foram representados graficamente a partir do *scree-plot* (Figura 8).

A primeira componente principal explica 36,27% da variância total; a segunda componente captou 24,22%; enquanto que a terceira componente explica cerca de 13,35% da variabilidade total.

Tabela 3. Autovalores ( $\lambda_i$ ) e porcentagens da variação explicadas pelas componentes principais (VCP)

| CP  | $\lambda_i$ | VCP (%) | VCP Acumulada (%) |
|-----|-------------|---------|-------------------|
| CP1 | 2,90        | 36,27   | 36,27             |
| CP2 | 1,94        | 24,22   | 60,49             |
| CP3 | 1,06        | 13,35   | 73,85             |
| CP4 | 0,88        | 11,02   | 84,86             |
| CP5 | 0,56        | 7,01    | 91,87             |
| CP6 | 0,32        | 4,04    | 95,91             |
| CP7 | 0,22        | 2,72    | 98,63             |
| CP8 | 0,11        | 1,37    | 100,00            |

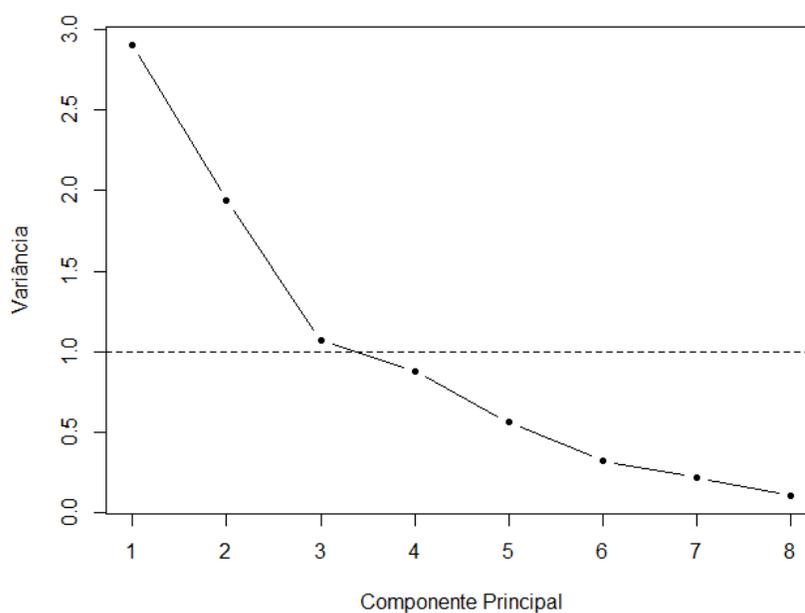


Figura 8 - “Scree-plot” das componentes principais (CP) para a área de cana-de-açúcar.

Na Tabela 4 estão os coeficientes de ponderação das variáveis e suas correlações com as componentes principais. Dessa forma, os atributos associados a cada componente foram escolhidos mediante sua correlação com cada umas delas.

Tabela 4. Coeficientes de ponderação das variáveis e suas correlações (entre parênteses) com as componentes principais para o experimento em cana-de-açúcar.

| Variável | CP1                  | CP2                  | CP3                  |
|----------|----------------------|----------------------|----------------------|
| FCO2     | <b>-0,46</b> (-0,78) | 0,03 (0,05)          | 0,17 (0,18)          |
| EstC     | 0,10 (0,17)          | -0,24 (-0,33)        | <b>-0,63</b> (-0,65) |
| PLA      | <b>-0,48</b> (-0,82) | 0,17 (0,23)          | 0,04 (0,04)          |
| DS       | <b>0,49</b> (0,83)   | -0,19 (-0,26)        | 0,17 (0,18)          |
| Macro    | <b>-0,45</b> (-0,76) | 0,15 (0,21)          | -0,40 (-0,41)        |
| pH       | -0,24 (-0,40)        | <b>-0,58</b> (-0,80) | 0,32 (0,33)          |
| P        | < 0,01 (< 0,01)      | -0,37 (-0,52)        | <b>-0,52</b> (-0,54) |
| SB       | -0,23 (-0,39)        | <b>-0,62</b> (-0,86) | 0,13 (0,13)          |

Os atributos mais expressivos retidos em CP1 relacionam o processo de emissão de CO<sub>2</sub> e a estrutura porosa do solo (DS, PLA e Macro). A relação direta entre as variáveis PLA e Macro com FCO2 é compreendida, uma vez que as características destes atributos propõem poros com maior diâmetro, afetando o processo de drenagem e transporte de gases em camadas subsuperficiais no solo. De acordo com Smart & Penuelas (2005), a infiltração de água nos espaços porosos, em razão de chuvas ou irrigações, beneficiam a emissão de grandes quantidades de CO<sub>2</sub> do solo para a atmosfera. Dessa maneira, essa componente pode ser caracterizada como a relação entre o transporte e emissão do dióxido de carbono, uma vez que solos mais densos influenciam o transporte de gases no solo.

A segunda componente principal (CP2) tem maior associação com as variáveis químicas do solo, indicadas pelo pH e a soma de bases, indicando uma relação direta entre essas variáveis. Essa relação direta ocorre principalmente em

virtude da calagem do solo, promovendo aumento do pH e da saturação por bases.

Os dois atributos que complementam o estudo, EstC e P, destacaram-se como mais contributivas para a composição da terceira componente (CP3). Essas variáveis também apresentam uma relação direta, que pode ser explicada pela relação que possuem na natureza, uma vez que, de acordo com Fernandes et al. (1997), ocorre maior quantidade de P nas camadas superficiais do solo, correlacionado com maior teor de matéria orgânica, que tem forte relação com o EstC. Assim CP3 caracteriza o acúmulo de carbono no solo.

As Figuras 9, 10 e 11 correspondem às representações gráficas dos pares de componentes principais para os pontos amostrais. A relação entre os atributos do solo foi indicada pelo cosseno do ângulo formado entre as setas e a importância destas variáveis, pelo comprimento das setas, que representa a variância de cada variável. O grau com que a variável se correlaciona com o componente principal é dado pela angulação a partir do eixo e a seta de determinada variável.

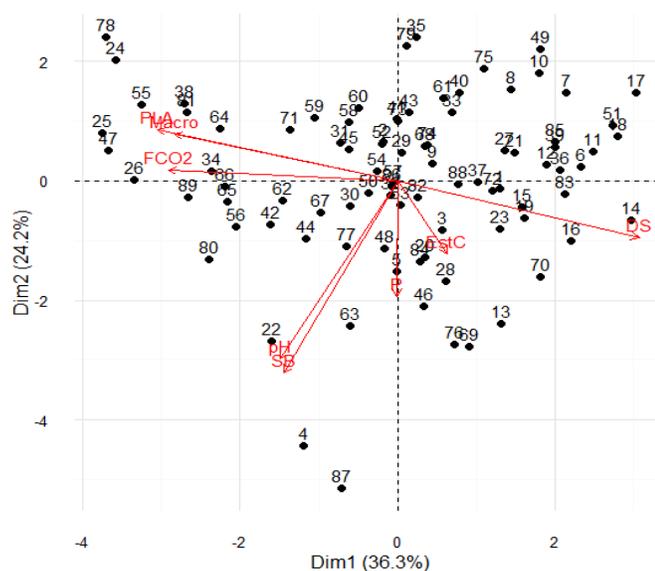


Figura 9 - Biplot das associações entre CP1 e CP2 para o experimento em área de cana-de-açúcar.

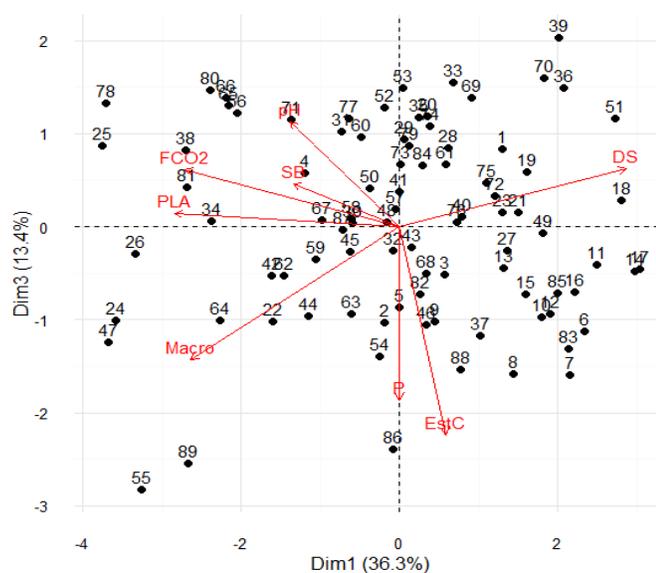


Figura 10 - Biplot das associações entre CP1 e CP3 para o experimento em área de cana-de açúcar.

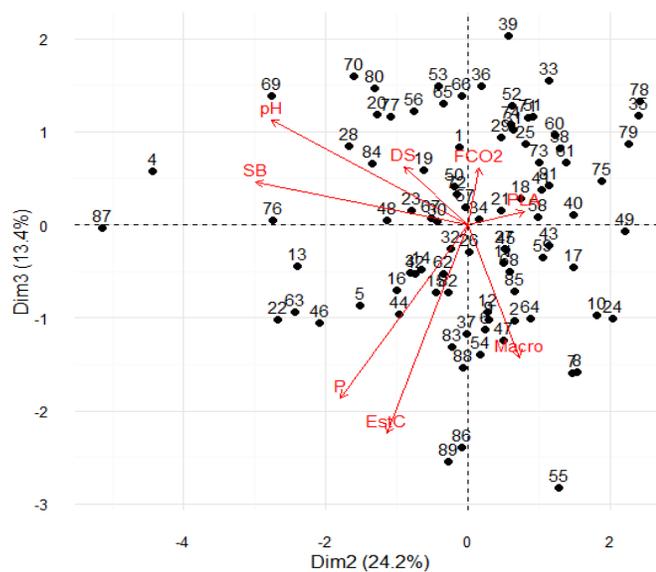


Figura 11 - Biplot das associações entre CP2 e CP3 para o experimento em área de cana-de açúcar.

Pode-se observar por meio das Figuras 9, 10 e 11 que os pontos amostrais 4 e 87 destacam-se em relação aos demais, sugerindo que nesses pontos para as variáveis que compõem a primeira componente principal, FCO<sub>2</sub>, PLA, DS e Macro, são apresentados valores próximos a média. Em relação a segunda componente, esses pontos indicam ter altos valores para os atributos pH e SB. Já de acordo com a CP3, indicam valores próximos a média para o atributo Fósforo e estoque de carbono. Os pontos amostrais 55 e 89, indicam valores médios em relação a CP2 (pH e P), e altos valores em relação a CP1 (FCO<sub>2</sub>, PLA, DS e Macro) e CP3 (EstC e P), destacando-se dos demais, de acordo com as variáveis que compõem esta componente.

Vale ressaltar que o novo conjunto de variáveis latentes, composto pelas três primeiras componentes principais foi capaz de destacar as relações existentes entre as variáveis que compõem cada componente, possibilitando a redução do espaço paramétrico em estudo.

#### 4.1.2 Análise Fatorial

Inicialmente, verificou-se a aplicabilidade da técnica de análise fatorial para as variáveis FCO<sub>2</sub>, EstC, PLA, DS, Macro, pH, P e SB, por meio da medida de adequação amostral global KMO, que apresentou um resultado igual 0,60, indicando uma adequação aceitável ( $KMO > 0,5$ ) (Hair Jr. et al., 2009). No entanto, as medidas de adequação amostral para cada variável, apresentaram resultados insatisfatórios para pH (0,47), SB (0,47) e EstC (0,22).

Por apresentar medida de adequação amostral muito baixa, decidiu-se retirar a variável EstC na aplicação da técnica de análise fatorial. Assim, o valor do KMO global passou a ser igual a 0,66 com medidas satisfatórias de adequação amostral para as demais variáveis, com: FCO<sub>2</sub> (0,77), PLA (0,77), Ds (0,71), Macro (0,66), pH (0,55), P (0,50) e SB (0,54), sendo este o novo subconjunto a ser analisado.

A Figura 12 apresenta o “scree-plot”, onde verifica-se que as três primeiras componentes apresentam variâncias iguais ou superiores a unidade.

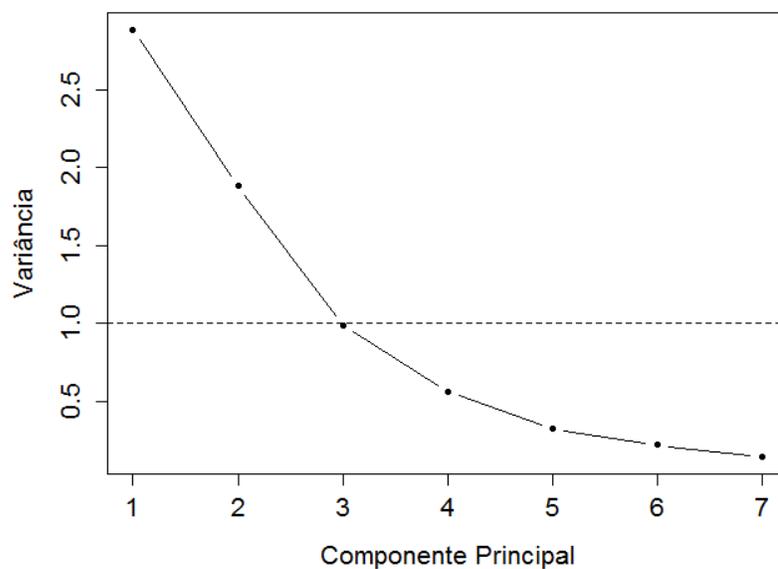


Figura 12 - “Scree-plot” das componentes principais (CP) para experimento em área de cana-de-açúcar.

Para a extração dos fatores, foi utilizado o método de componentes principais, obtendo-se três fatores associados aos maiores autovalores que juntos explicam 82,14% da variância total das variáveis originais. Os resultados das estimativas das cargas fatoriais (*loadings*), utilizando a rotação varimax, das comunalidades e das variâncias específicas são apresentados na Tabela 5. Percebe-se que todas as variáveis apresentaram comunalidades elevadas em relação às variâncias específicas (comunalidade > variância específica), ou seja, os três fatores de uma forma geral representam adequadamente o comportamento biológico das variáveis no solo.

Tabela 5. Cargas fatoriais (rotação Varimax), comunalidades, variâncias específicas das variáveis, autovalores e percentual acumulado de explicação da variância da análise dos dados do experimento em área de cana-de-açúcar

| Variável         | Carga Fatorial |             |             | Comunalidades | Variâncias específicas |
|------------------|----------------|-------------|-------------|---------------|------------------------|
|                  | Fator 1        | Fator 2     | Fator 3     |               |                        |
| FCO2             | <b>-0,70</b>   | 0,32        | -0,31       | 0,70          | 0,30                   |
| PLA              | <b>-0,83</b>   | 0,12        | -0,22       | 0,74          | 0,26                   |
| DS               | <b>0,89</b>    | <0,01       | -0,03       | 0,79          | 0,21                   |
| Macro            | <b>-0,86</b>   | -0,07       | 0,28        | 0,82          | 0,18                   |
| pH               | -0,07          | <b>0,94</b> | 0,09        | 0,90          | 0,10                   |
| P                | -0,04          | 0,21        | <b>0,92</b> | 0,88          | 0,18                   |
| SB               | -0,07          | <b>0,95</b> | 0,13        | 0,91          | 0,09                   |
| Autovalor        | 2,88           | 1,88        | 0,99        |               |                        |
| VT Acumulada (%) | 41,18          | 68,06       | 82,14       |               |                        |

Por meio das cargas fatoriais (Tabela 5) é possível identificar os atributos do solo que possuem maiores associações com os respectivos fatores. Representações bidimensionais permitem a visualização dessas associações (Figuras 13, 14 e 15), indicando o cruzamento das cargas fatoriais de dois fatores quaisquer.

As representações gráficas, juntamente com a observação das cargas fatoriais (Tabela 5), permitem caracterizar os fatores. Foram consideradas cargas fatoriais maiores do que 0,55 de acordo com Hair Jr. et al. (2009), indicadas para amostras com até 100 unidades amostrais. Cargas grandes ou moderadas indicam como as variáveis estão relacionadas com os fatores (Manly, 2004).

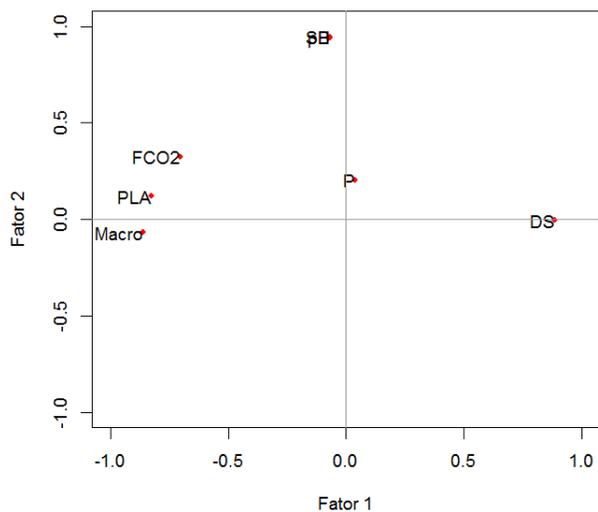


Figura 13 - Gráfico de dispersão das cargas fatoriais de F1 e F2 para as variáveis avaliadas.

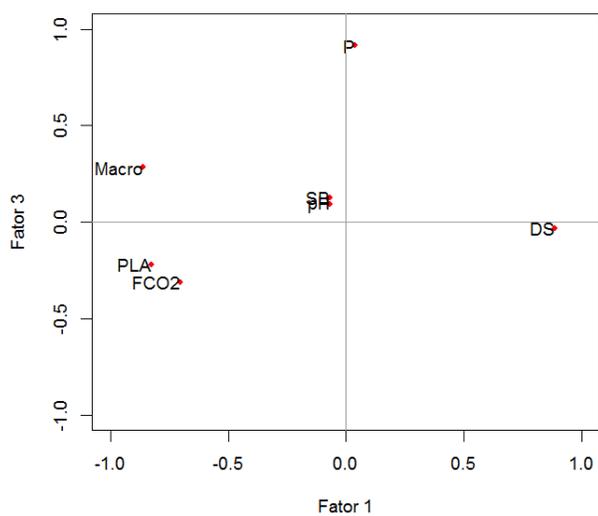


Figura 14 - Gráfico de dispersão das cargas fatoriais de F1 e F3 para as variáveis avaliadas.

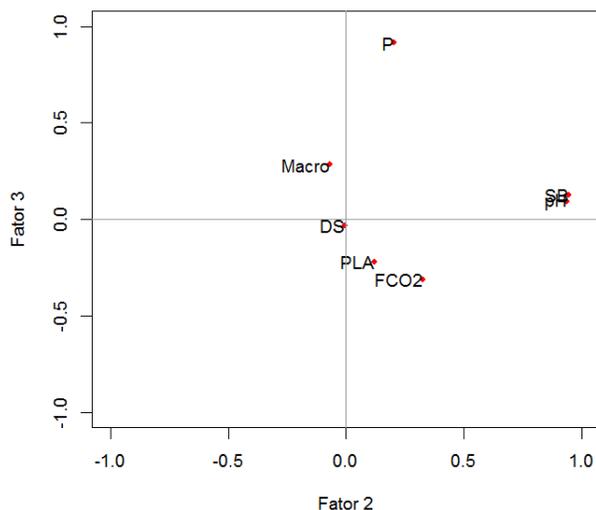


Figura 15 - Gráfico de dispersão das cargas fatoriais de F2 e F3 para as variáveis avaliadas.

O Fator 1 reteve 41,18% da variabilidade total das variáveis, captando os atributos físicos do solo e o FCO<sub>2</sub>, podendo ser denominado fator relacionado ao transporte de gases no solo, que envolvem a saída da molécula de CO<sub>2</sub> e a entrada da molécula de O<sub>2</sub> no solo, apresentando um antagonismo entre a densidade do solo (DS) e as demais variáveis que compõem o primeiro fator.

A DS é um dos parâmetros mais empregados na determinação da compactação do solo. Por tratar-se de um atributo variável, depende das condições estruturais e do estado de compactação do solo, que é um grave problema para a qualidade do solo e conseqüentemente o desenvolvimento de uma agricultura sustentável, pois modifica os fluxos de água e ar no solo resultando em uma baixa produtividade agrícola (Freddi et al. (2007); Reinert et al. (2008); Jorge et al. (2012)).

Para Brito et al. (2009), a variação temporal e espacial da emissão de CO<sub>2</sub> solo-atmosfera é influenciada por inúmeros atributos do solo relacionados à produção de CO<sub>2</sub> e à difusão do gás no solo. Os autores avaliaram as alterações espaciais e temporais do CO<sub>2</sub> do solo e sua relação com os atributos físicos do solo, em área de cultivo de cana-de-açúcar, sob um solo Latossolo eutrófico típico, em

Jaboticabal – SP, assim, concluíram que áreas com menor densidade aparente do solo, menor resistência à penetração e maior macroporosidade, apresentaram maior emissão de CO<sub>2</sub>.

Neste trabalho, o solo não apresentou compactação, uma vez que, o valor médio da variável DS foi inferior ao valor crítico estabelecido (1,45 g cm<sup>-3</sup>) (Nogueira, 2013). Eden et al. (2011) descrevem que dentre outras funções, o sistema poroso do solo serve como uma via de transporte de água e ar. Desta forma o fluxo de gases no solo é facilitado.

Em estudos conduzidos por Panosso et al. (2012), em condições semelhantes ao presente estudo em área de cana-de-açúcar, os atributos físicos do solo, macroporosidade, porosidade total e porosidade livre de água, apresentaram maior poder discriminatório (36,6 %) na análise dos componentes principais.

O Fator 2, pode ser apontado como um índice de fertilidade do solo, representando 26,88% da variabilidade total. Segundo Raij (2011), solos que apresentam valores de pH inferiores a 5,5, possuem baixa capacidade de troca catiônica e elevada saturação por alumínio, desta forma comprometem a disponibilidade de nutrientes para as plantas. Assim, segundo Caires et al. (2004), por meio da prática da calagem é possível promover a elevação do pH, da saturação por bases (SB) e dos teores de cálcio (Ca), além de reduzir os teores do Al tóxico do solo.

Para Trevisan et al. (2002), solos com pH elevado favorecem a atividade microbiana, podendo refletir no acúmulo ou no decréscimo de matéria orgânica do solo. Entretanto, Fernandes & da Silva (1999), relatam que os níveis baixos ou muito altos de pH, inibem a atividade microbiana. A taxa de mineralização da matéria orgânica está condicionada a alguns fatores como a variação da temperatura, teor de água e o pH. Embora a matéria orgânica esteja em pequeno percentual no solo, ela influencia nos atributos físicos químicos e biológicos do solo (Pariente & Lavee, 2000).

Em área de canavial manejado sob sistema de cana crua (CC) e cana queimada (CQ), Tavares et al. (2016), verificaram que o pH do solo contribui para um

melhor desempenho dos microrganismos no processo de decomposição da palhada no solo. Ensaios conduzidos por Hall et al. (1998), com plantas forrageiras de Azevém (*Lolium Multiflorum Lam.*), cultivadas sob solo com três diferentes valores de pH, observaram que o pH tem efeito sobre o fluxo de CO<sub>2</sub> no solo.

Por fim, o terceiro fator, Fator 3, está associado apenas ao P, sendo denominado como índice de Fósforo no solo, retendo cerca de 14% da variabilidade total. Tal atributo no solo se origina da mineralização da matéria orgânica, adição de fertilizantes fosfatados e solubilização dos minerais fosfatados. É um importante macronutriente, componente estrutural de macromoléculas.

O Fósforo é considerado elemento essencial para as plantas, e também faz parte do metabolismo dos microrganismos do solo (Ourives et al., 2010). Segundo dos Santos et al. (2008), a dinâmica do fósforo no solo está associada a fatores ambientais que controlam a atividade dos microrganismos e às propriedades físico-químicas e mineralógicas do solo.

Segundo Ferreira et al. (2008), a atividade microbiana é influenciada por fatores como a disponibilidade de nutrientes, incluindo o P. Estudos conduzidos pelos autores em um Latossolo Vermelho distrófico em região de cerrado, observaram que a adição de P aumentou a atividade respiratória avaliada pelo C (carbono) liberado.

Para Schwendenmann et al. (2003), em experimento envolvendo a variação espacial e temporal do fluxo de CO<sub>2</sub> em dois tipos de solos em áreas de floresta na Costa Rica, observaram que há influência da biomassa das raízes, teor de carbono do solo e concentração de P na emissão de CO<sub>2</sub>. Entretanto, Zanini et al. (2005), estudaram a aplicação de fertilizantes fosfatados por meio de fertirrigação, em um Latossolo Vermelho desprovido de vegetação em Jaboticabal - SP. Os autores concluíram que não houve efeito significativo da aplicação do ácido fosfórico sobre a taxa de emissão de CO<sub>2</sub> do solo. Tal resultado foi semelhante ao obtido por este estudo em relação a influência do teor de fósforo sobre a FCO<sub>2</sub>, uma vez que, como esperado, apenas o P ficou retido no Fator 3, indicando a falta de correlação dessa

com as demais variáveis.

Portanto, a partir do modelo de análise fatorial com três fatores, é possível analisar a qualidade do ajuste por meio da matriz residual (Tabela 6). Tal matriz indica que a matriz de correlação amostral original  $\mathbf{R}$  foi reproduzida adequadamente pela matriz estimada  $\hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\psi}$ , uma vez que seus elementos estão próximos a zero.

Tabela 6. Matriz residual da análise fatorial para atributos do solo em área de cana-de-açúcar

|       | FCO2  | PLA   | DS    | Macro | pH    | P     | SB    |
|-------|-------|-------|-------|-------|-------|-------|-------|
| FCO2  | 0,00  | -0,03 | 0,10  | -0,10 | -0,08 | 0,14  | -0,06 |
| PLA   | -0,03 | 0,00  | 0,12  | -0,08 | -0,04 | 0,08  | -0,01 |
| DS    | 0,10  | 0,12  | 0,00  | 0,03  | -0,04 | 0,07  | -0,02 |
| Macro | -0,10 | -0,08 | 0,03  | 0,00  | 0,05  | -0,12 | 0,03  |
| pH    | -0,08 | -0,04 | -0,04 | 0,05  | 0,00  | -0,06 | -0,06 |
| P     | 0,14  | 0,08  | 0,07  | -0,12 | -0,06 | 0,00  | -0,04 |
| SB    | -0,06 | -0,01 | -0,02 | 0,03  | -0,06 | -0,04 | 0,00  |

#### 4.1.3 Análise de Correlação Canônica

Apresenta-se, nesta seção, os resultados do estudo da associação entre as variáveis relacionadas a dinâmica do carbono e aquelas relacionadas aos atributos físicos e químicos do solo, em área de cultivo de cana-de-açúcar, realizado no município de Guariba - SP, no ano de 2008.

O grupo das variáveis relacionadas a dinâmica do carbono composto pelo atributos FCO2 e EstC (emissão de CO<sub>2</sub> e estoque de carbono do solo, respectivamente), já o grupo de atributos físicos e químicos do solo, foi formado por PLA, DS, Macro, pH, P e SB. A análise de correlação canônica é uma extensão da análise de regressão múltipla (Manly, 2004) e tem o papel de associar grupos, que não são

classificados por variáveis dependentes e/ou independentes onde cada um contenha duas ou mais variáveis.

A correlação linear do grupo de variáveis ligadas a dinâmica do carbono,  $\mathbf{R}^{11}$ , é igual a -0,12. As correlações entre as variáveis físicas e químicas do solo, matriz  $\mathbf{R}^{22}$ , estão representadas na Tabela 7.

Tabela 7. Matriz de correlações ( $\mathbf{R}^{22}$ ) das variáveis físicas e químicas do solo.

|       | PLA    | DS     | Macro  | pH    | P     | SB    |
|-------|--------|--------|--------|-------|-------|-------|
| PLA   | 1,00   | -0,60* | 0,56*  | 0,11  | -0,12 | 0,14  |
| DS    | -0,60* | 1,00   | -0,75* | -0,11 | 0,07  | -0,09 |
| Macro | 0,56*  | -0,75* | 1,00   | 0,07  | 0,10  | 0,06  |
| pH    | 0,11   | -0,11  | 0,07   | 1,00  | 0,22* | 0,85* |
| P     | -0,12  | 0,07   | 0,10   | 0,22* | 1,00  | 0,27* |
| SB    | 0,14   | -0,09  | 0,06   | 0,85* | 0,27* | 1,00  |

\* significativa à probabilidade de 0,05.

As correlações amostrais entre as variáveis relacionadas a dinâmica do carbono (primeiro conjunto) e atributos físicos e químicos do solo (segundo conjunto) são apresentadas na Tabela 8.

Tabela 8. Matriz de correlações ( $\mathbf{R}^{12}$ ) entre as variáveis relacionadas à dinâmica do carbono e os atributos físico-químicos do solo.

|      | PLA    | DS     | Macro | pH    | P     | SB    |
|------|--------|--------|-------|-------|-------|-------|
| FCO2 | 0,65*  | -0,52* | 0,40* | 0,24* | -0,10 | 0,26* |
| EstC | -0,16* | 0,14   | -0,05 | -0,04 | 0,14  | 0,21* |

\* significativa à probabilidade de 0,05.

A maior correlação significativa (com valor-p < 0,05) existente entre qualquer variável do conjunto ligado à dinâmica do carbono e as variáveis físicas e químicas do solo correspondeu a 0,65 (emissão de CO<sub>2</sub> do solo e porosidade livre

de água). A variável FCO2 apresentou, de uma forma geral, maior correlação com as variáveis do grupo de atributos físicos e químicos do solo, já EstC apresentou correlação significativa a 5% apenas com a variável soma de bases e PLA.

A partir da matriz de correlação do vetor  $\mathbf{X}' = (\mathbf{X}^{(1)'} \quad \mathbf{X}^{(2)'})$ , no qual  $\mathbf{X}^{(1)}$  corresponde ao vetor de variáveis ligadas a dinâmica do carbono do solo e  $\mathbf{X}^{(2)}$  ao conjunto de variáveis físicas e químicas do solo, foram obtidas as variáveis canônicas: definidas pela combinação linear das variáveis em  $\mathbf{X}^{(1)}$  e pela combinação linear em  $\mathbf{X}^{(2)}$ .

Os coeficientes das correlações canônicas estão expressos na Tabela 9. Também apresenta-se o teste de significância dos coeficientes de correlação canônica, que tem por hipótese:

$$H_0^1 : \rho_1^2 \neq 0, \rho_2^2 = 0 \quad vs \quad H_1^1 : \rho_2^2 \neq 0$$

ou seja, na  $H_0^1$  apenas a primeira correlação canônica é significativa e em  $H_1^1$  ambas as correlações são significativas.

Tabela 9. Autovalores, correlações canônicas e teste de significância para as correlações canônicas ( $\alpha=0,05$ ).

| <b>Pares de Variáveis<br/>Canônicas</b> | <b>Autovalores</b> | <b>Correlações<br/>Canônicas</b> | <b>Estatística<br/>de Teste</b> | <b>gl</b> | <b>Conclusão</b>   |
|---|--------------------|----------------------------------|---------------------------------|-----------|--------------------|
| (U <sub>1</sub> , V <sub>1</sub> )      | 0,706              | 0,840                            | 133,52                          | 5         | Rejeita-se $H_0^1$ |
| (U <sub>2</sub> , V <sub>2</sub> )      | 0,495              | 0,704                            |                                 |           |                    |

Assim, por meio da rejeição de  $H_0^1$  conclui-se que os dois coeficientes de correlações encontrados são significativos a 5%, apresentando valores iguais a 0,706 e 0,495. Pode-se afirmar ainda, com esses resultados, que as variáveis do primeiro grupo expressam características distintas relacionadas à dinâmica do carbono (U<sub>1</sub> e U<sub>2</sub>), bem como as duas variáveis canônicas V<sub>1</sub> e V<sub>2</sub> que expressam diferentes relações entre os atributos físicos e químicos do solo, e ainda a existência de associação entre os pares de variáveis canônicas, (U<sub>1</sub>,V<sub>1</sub>) e (U<sub>2</sub>,V<sub>2</sub>).

Analisa-se o primeiro par de variáveis canônicas, o qual explica cerca de 59% da relação entre as variáveis, de maneira que  $U_1$  é a melhor combinação linear na associação de  $V_1$  e vice-versa. Já o segundo par explica aproximadamente 41% da relação entre as variáveis canônicas de ambos os grupos.

A suposição de normalidade multivariada para aplicação do teste foi avaliada visualmente por meio do gráfico de probabilidade *Q-Q Plot* (*Chi-Square plot*) multivariado apresentado na Figura 6. Percebeu-se uma tendência das distâncias alinharem-se ao redor da bissetriz, sugerindo, aderência dos dados à distribuição normal multivariada de probabilidade.

As correlações entre as variáveis originais com os pares de variáveis canônicas (conhecidas como cargas canônicas e cargas canônicas cruzadas) são apresentadas na Tabela 10. Tais valores tem a função de ajudar a compreender o significado das variáveis canônicas; quanto maior o valor absoluto de uma carga canônica, maior a associação entre a variável original e o respectivo componente canônico.

Tabela 10. Cargas canônicas e cargas canônicas cruzadas das variáveis do experimento em cana-de-açúcar.

| Variáveis                         | Cargas Canônicas |                | Cargas Cruzadas |                |
|-----------------------------------|------------------|----------------|-----------------|----------------|
|                                   | U <sub>1</sub>   | U <sub>2</sub> | V <sub>1</sub>  | V <sub>2</sub> |
| Dinâmica do carbono do solo       |                  |                |                 |                |
| FCO2                              | -0,987           | -0,164         | -0,696          | -0,086         |
| EstC                              | 0,288            | -0,958         | 0,203           | -0,474         |
| PVTE(%)                           | 52,66            | 47,34          |                 |                |
|                                   | V <sub>1</sub>   | V <sub>2</sub> | U <sub>1</sub>  | U <sub>2</sub> |
| Atributos físico-químicos do solo |                  |                |                 |                |
| PLA                               | -0,933           | -0,054         | -0,659          | -0,027         |
| DS                                | 0,748            | 0,018          | 0,528           | 0,009          |
| Macro                             | -0,557           | -0,125         | -0,394          | -0,062         |
| pH                                | -0,342           | -0,052         | -0,242          | -0,026         |
| P                                 | 0,174            | -0,228         | 0,122           | -0,113         |
| SB                                | -0,302           | -0,564         | -0,213          | -0,279         |
| PVTE(%)                           | 32,99            | 6,53           |                 |                |

Observa-se que a proporção da variância total, que é explicada separadamente pelas variáveis canônicas (Tabela 10), é de 52,66% por U<sub>1</sub> e de 32,99% por V<sub>1</sub>, ou seja, a variável U<sub>1</sub> representa 52,66% da variação total do grupo de variáveis ligadas a dinâmica do carbono e a variável V<sub>1</sub> representa 32,99% da variação total do grupo de variáveis físico-químicas. O segundo par, U<sub>2</sub> e V<sub>2</sub> explica separadamente 47,34% e 6,53% da proporção da variância total.

Pode-se afirmar que a variável FCO2 (-0,987) apresenta maior carga canônica em U<sub>1</sub>. Já a variável canônica V<sub>1</sub> pode ser representada pelo antagonismo entre a densidade do solo, DS (0,748), com a PLA (-0,933), Macro (-0,557) e pH (-0,342). Esse padrão sugere que U<sub>1</sub> pode ser interpretada pela emissão de dióxido de

carbono do solo e  $V_1$  é uma variável de atributos físicos e pH do solo. A partir disso, pode-se indicar que solos com maiores valores de FCO2 tendem a possuir valores altos de PLA, Macro e pH e baixos valores para DS, uma vez que as propriedades físicas do solo tem relação direta com o fluxo de gases no interior do solo e o pH tem a característica de permitir a disponibilidade de nutrientes no solo.

Além disso, observando as cargas canônicas cruzadas entre as variáveis originais da dinâmica do carbono com  $V_1$ , destaca-se maior associação de FCO2 (-0,696) com esta variável, indicando que  $V_1$  alto associa-se com altas emissões de  $CO_2$ . Também nota-se que as cargas canônicas cruzadas mais elevadas entre as variáveis físico-químicas e  $U_1$  são: PLA (-0,659), DS (0,528) e Macro (-0,394). Estes resultados indicam que o componente canônico, representando a emissão de  $CO_2$ , assume valores altos associados com valores altos de porosidade livre de água e macroporosidade e solo menos denso.

Avaliou-se da mesma forma, o segundo par canônico, verificando-se que a variável estoque de carbono foi predominante em  $U_2$ , e  $V_2$  representada por um único atributo químico do solo, SB. Assim, os solos que apontaram alto valor de EstC apresentaram valores de SB mais elevados. O manejo adequado do solo promove o maior acúmulo e ciclagem da matéria orgânica e pode aumentar significativamente os estoques de C ao longo do tempo (Souza et al., 2012).

A matéria orgânica contribui para a melhoria dos atributos físicos, químicos e biológicos do solo e atua como poder de tamponamento da acidez do solo. Logo, altera a dinâmica dos nutrientes, aumentando a capacidade de troca de cátions e influenciando desta forma na SB.

Segundo Ronquim (2010), a CTC representa a capacidade de liberação de vários nutrientes, favorecendo a manutenção da fertilidade do solo e é fortemente dependente de pH, a CTC está intimamente ligada a SB, pois representa a soma dos teores de cátions permutáveis, com exceção do  $H^+$  e  $Al^{3+}$ .

As mesmas variáveis se destacaram quando observa-se as cargas canônicas cruzadas dos conjuntos de variáveis  $X^{(1)}$  e  $V_2$  e  $X^{(2)}$  e  $U_2$ , ou seja, obtêm-se

resultados semelhantes aos da análise das correlações canônicas (não cruzadas).

A variável P, que representa o teor de fósforo no solo, não apresentou associação a quaisquer das variáveis canônicas, devido a fraca correlação com as demais variáveis para este estudo.

A Figura 16 representa os pares de variáveis canônicas e os valores padronizados que tais pares assumem em cada ponto amostral. O gráfico de dispersão para os pontos do primeiro par de variáveis canônicas apresenta um comportamento linear positivo entre a variável canônica  $U_1$  e  $V_1$ , neste sentido, pode-se dizer que a variável canônica  $U_1$  apresenta valores altos associados a valores altos da variável  $V_1$ , principalmente nos pontos 14 e 17, enquanto que valores baixos para as duas variáveis foi observado para o ponto 25.

Os pontos amostrais para o primeiro par de variáveis canônicas ajustam-se melhor a uma reta quando comparados ao segundo par, uma vez que  $U_1$  e  $V_1$  é responsável pela maior representação da variabilidade total dos dados (Mingoti, 2005).

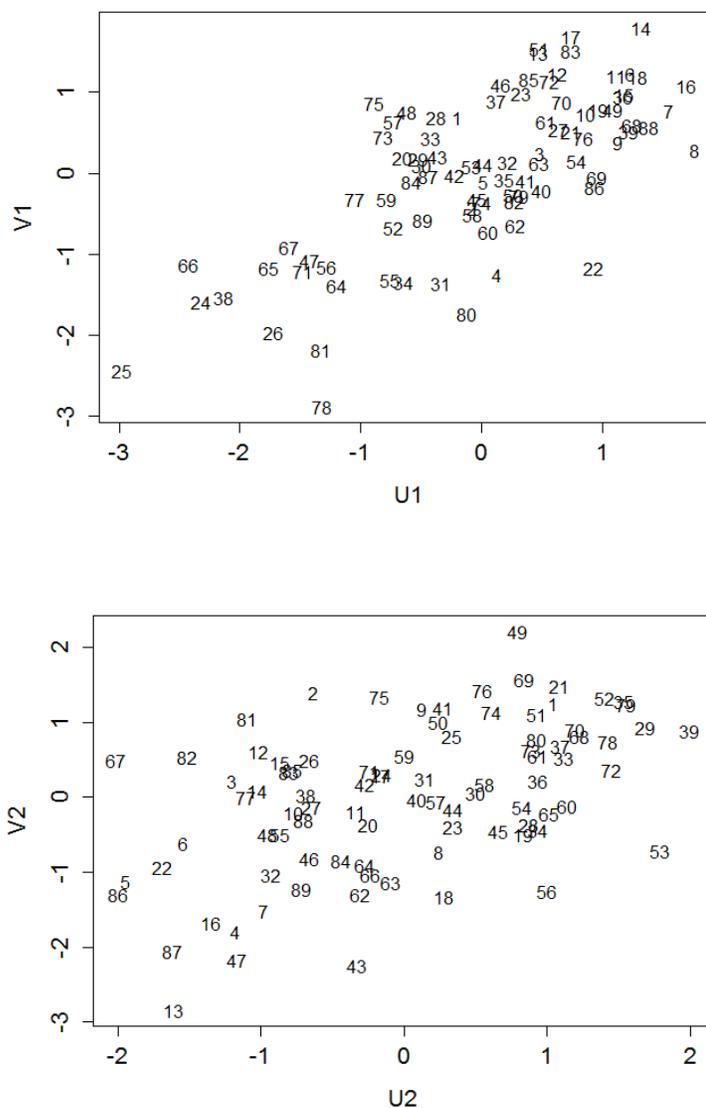


Figura 16 - Gráfico de dispersão dos escores das variáveis canônicas U1 e V1, U2 e V2 para experimento em área de cana-de-açúcar.

Portanto, de uma maneira geral, os resultados discutidos indicam que é possível resumir as informações das variáveis ligadas a dinâmica do carbono e os atributos físicos e químicos do solo em dois pares de variáveis canônicas expressando suas relações.

## 4.2 Experimento em área de soja em 2013

O conjunto de dados obtido em área de soja no ano de 2013 também foi preliminarmente analisado por meio das técnicas de estatísticas descritivas (Tabela 11). A média para a emissão de CO<sub>2</sub> do solo apresentou valor igual a 2,157  $\mu\text{mol m}^{-2}\text{s}^{-1}$ , valor semelhante ao encontrado para o experimento realizado em cana-de-açúcar apresentado neste trabalho.

De acordo com Pinto Jr et al. (2009), em estudo que objetivou estimar o fluxo de CO<sub>2</sub> do solo e verificar a influência da serrapilheira acumulada, umidade e temperatura do solo em uma Floresta de Transição Amazônica Cerrado e em uma área de Pastagem, localizadas no norte do Mato Grosso, os valores médios do fluxo de CO<sub>2</sub> do solo na Floresta e na área de Pastagem foram iguais à 5,45 e 4,95  $\mu\text{mol m}^{-2}\text{s}^{-1}$ , respectivamente, valores muito acima do encontrado no presente conjunto de dados.

Com relação a área de cana-de-açúcar, a emissão de CO<sub>2</sub> do solo apresentou coeficiente de variação pouco menor, quando comparado a área de cana-de-açúcar, igual a 32,953%, sendo classificado como alto de acordo com o critério estabelecido por Warrick & Nielsen (1980). Buscando identificar a dependência espacial de atributos físicos e químicos do solo e sua correlação espacial entre variáveis e a emissão de gases do efeito estufa (CO<sub>2</sub> e N<sub>2</sub>O) Nogueira et al. (2013), em três diferentes sistemas de manejo agrícola: pastagem em degradação, cana-de-açúcar com queima e produção de soja em plantio direto apresentou CVs iguais a 105,82%, 31,15% e 20,32%, respectivamente.

Aplicou-se as técnicas para o diagnóstico de multicolinearidade, com o objetivo de evitar a redundância de explicação da variabilidade do conjunto em questão. De maneira semelhante ao conjunto de dados da cana-de-açúcar, as variáveis eleitas foram: FCO<sub>2</sub>, EstC, PLA, DS, Macro, pH, P e SB. Os demais atributos do solo foram descartados, uma vez que, trataram-se de combinações lineares umas das outras, apresentando valor de variância inflacionada superiores a 10, comprometendo métodos que envolvem inversão de matrizes, tais como análise fatorial e de correlação

canônica. Dessa forma, reduziu-se o valor de  $N_c$  de  $2,09 \times 10^{16}$  para aproximadamente 15,58.

Tabela 11. Estatística descritiva das variáveis para o experimento em área de soja

|       | Média  | Mediana | EP    | Mínimo | Máximo | Q1     | Q3     | CV     | D_P   | Assim. | Curt.  |
|-------|--------|---------|-------|--------|--------|--------|--------|--------|-------|--------|--------|
| FCO2  | 2,157  | 2,079   | 0,062 | 0,946  | 4,620  | 1,580  | 2,479  | 32,953 | 0,711 | 0,950  | 1,359  |
| EstC  | 5,382  | 5,310   | 0,066 | 4,064  | 9,108  | 4,882  | 5,762  | 14,082 | 0,758 | 0,658  | 0,463  |
| US    | 17,505 | 17,524  | 0,097 | 14,759 | 20,857 | 16,775 | 18,190 | 6,420  | 1,124 | -0,022 | 0,525  |
| PLA   | 29,991 | 30,450  | 0,408 | 16,310 | 40,620 | 26,850 | 32,950 | 15,678 | 4,702 | -0,270 | 0,314  |
| DP    | 2,605  | 2,601   | 0,016 | 1,198  | 3,019  | 2,529  | 2,687  | 6,951  | 0,181 | 0,583  | 0,822  |
| DS    | 1,368  | 1,384   | 0,008 | 1,158  | 1,556  | 1,304  | 1,428  | 6,668  | 0,091 | -0,351 | -0,485 |
| Macro | 10,572 | 9,693   | 0,372 | 1,826  | 24,272 | 7,760  | 12,207 | 40,610 | 4,293 | 1,105  | 1,399  |
| Micro | 37,069 | 37,142  | 0,257 | 30,693 | 45,117 | 35,446 | 39,123 | 8,004  | 2,967 | -0,094 | -0,326 |
| PT    | 47,641 | 47,801  | 0,355 | 35,181 | 57,064 | 45,209 | 50,188 | 8,603  | 4,099 | -0,425 | 0,788  |
| pH    | 4,690  | 4,700   | 0,015 | 4,200  | 5,100  | 4,600  | 4,800  | 3,756  | 0,176 | -0,153 | -0,277 |
| MO    | 17,594 | 17,000  | 0,165 | 14,000 | 24,000 | 16,000 | 19,000 | 10,794 | 1,899 | 0,755  | 0,471  |
| P     | 18,489 | 17,000  | 0,600 | 8,000  | 38,000 | 13,000 | 23,000 | 37,435 | 6,921 | 0,835  | 0,299  |
| K     | 2,221  | 2,000   | 0,089 | 0,700  | 7,000  | 1,500  | 2,600  | 45,989 | 1,021 | 1,718  | -0,122 |
| Ca    | 11,293 | 11,000  | 0,202 | 6,000  | 18,000 | 10,000 | 13,000 | 20,617 | 2,328 | 0,227  | 4,038  |
| Mg    | 10,278 | 10,000  | 0,178 | 5,000  | 18,000 | 9,000  | 11,000 | 19,981 | 2,054 | 0,635  | 1,575  |
| H-Al  | 34,654 | 34,000  | 0,365 | 25,000 | 47,000 | 33,000 | 36,000 | 12,150 | 4,210 | 0,128  | 0,496  |
| SB    | 23,749 | 23,500  | 0,400 | 14,000 | 38,000 | 20,300 | 26,600 | 19,405 | 4,608 | 0,514  | 0,601  |
| CTC   | 58,469 | 58,800  | 0,468 | 45,000 | 72,000 | 54,400 | 61,900 | 9,236  | 5,400 | -0,015 | -0,267 |
| V     | 40,602 | 41,000  | 0,527 | 23,000 | 56,000 | 36,000 | 44,000 | 14,978 | 6,081 | 0,133  | 0,197  |

EP: erro padrão; Q1: primeiro quartil; Q3: terceiro quartil; CV: coeficiente de variação; Desv P: desvio padrão; Assim.: coeficiente de assimetria; Curt.: coeficiente de curtose.

Analisando-se os coeficientes de assimetria e curtose, constatou-se que os valores para tais coeficientes foram próximos a zero, exceto para os atributos FCO2 e Macro, com coeficientes de assimetrias iguais a 0,950 e 1,105, respectivamente. Para o coeficiente de curtose, tais variáveis apresentaram os valores 1,359 e 1,399, respectivamente. Esses valores indicam distribuições de probabilidade com relativa assimetria à direita de forma leptocúrtica. No entanto, vale ressaltar que a assimetria tem maior influência sobre a distribuição das variáveis em estudo.

De acordo com os histogramas (Figura 17) permitiu-se a observação da

fuga moderada da simetria e da curtose mesocúrtica na distribuição de cada atributo selecionado. Os resultados sugerem assimetria a direita para os atributos FCO2, EstC, Macro e P. Em relação ao coeficiente de curtose, observando os histogramas para cada variável, destacou-se relativa fuga da forma de sino para Macro, que apresentou curva menos suavizada (leptocúrtica).

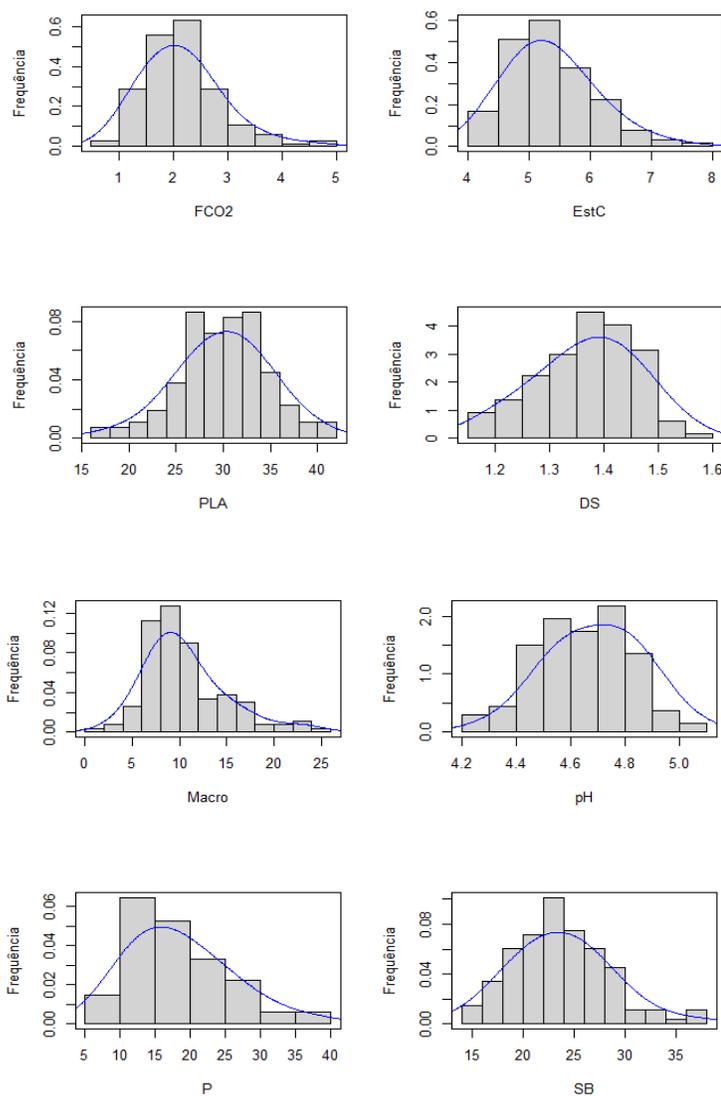


Figura 17 - Histogramas e curvas de distribuição das variáveis do experimento em área de soja.

Buscando estudar a possível normalidade multivariada dos atributos do solo, utilizou-se o gráfico de probabilidade *Q-Q Plot (Chi-Square plot)* multivariado apresentado na Figura 18. Inicialmente, por inspeção gráfica, verificou-se que salvo duas maiores distâncias que indicaram leve fuga da bissetriz, as demais distâncias se ajustaram confortavelmente a reta, sugerindo a normalidade multivariada. Assim, unindo essa informação ao ajuste de normalidade univariada e relativa simetria dos atributos e o tamanho amostral (133 pontos), sugere-se adequabilidade a multinormalidade.

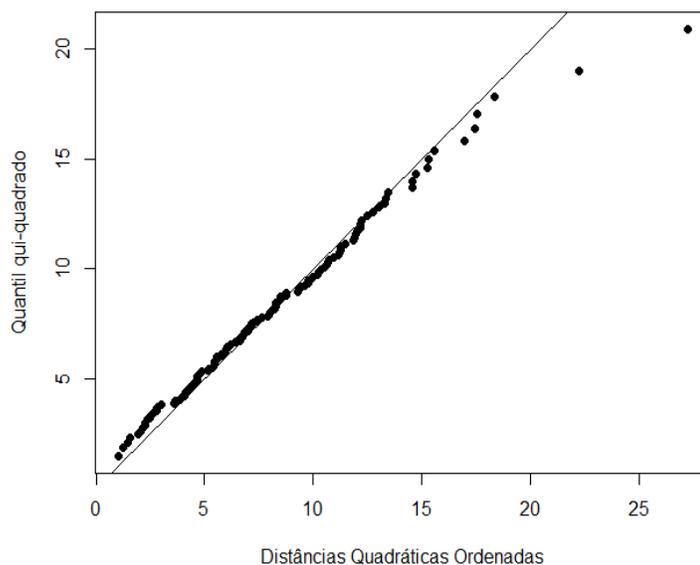


Figura 18 - Gráfico de probabilidade *Q-Q Plot* multivariado para os dados do experimento em soja.

Investigou-se os pontos que apresentaram maiores distâncias, com o objetivo de classificá-los como possíveis *outliers* multivariados, examinou-se as observações em cada variável individualmente, por meio gráfico (*box-plot*), com o propósito de identificar *outliers* univariados, ou seja, pontos que estivessem acima do limite superior,  $LS = Q3 + 1,5 \times DQ$ , ou abaixo do limite inferior,  $LI = Q1 - 1,5 \times DQ$ ,

com DQ sendo a distância interquartil ( $DQ = Q3 - Q1$ ). No entanto, nenhum dos pontos amostrais excedeu os limites em mais do que uma variável. Dessa forma, optou-se por manter tais pontos na análise.

A matriz de correlações amostrais, apresentada na Figura 19 de maneira intuitiva, por meio da intensidade das cores, possibilita identificar o grau da associação existente entre os pares de variáveis. Quadrados em tom de azul indicam correlações positivas e em tom de vermelho correlações negativas. Quanto mais intensa a tonalidade da cor, maior o grau de associação, e quanto mais fraca, menor foi o grau de associação linear.



Figura 19 - Correlações das variáveis para o experimento em área de soja.

Os resultados obtidos a partir de experimento em área de soja revelaram a existência de baixas, mas significativas correlações positivas entre a variável FCO2 e PLA (0,23  $p < 0,05$ ), Macro (0,17  $p < 0,05$ ), P (0,27  $p < 0,05$ ) e SB (0,26  $p < 0,05$ ) e correlação negativa baixa, mas significativa com DS (-0,20  $p < 0,05$ ). Pode-

se notar que, para o conjunto de dados em soja sob o plantio direto, as correlações entre a emissão de CO<sub>2</sub> do solo e os atributos químicos do solo apresentaram correlações um pouco maiores quando comparadas com a área de cana-de-açúcar.

Nogueira et al. (2013) também detectaram baixas e negativas correlações lineares entre a emissão de CO<sub>2</sub> e densidade do solo para áreas de pastagem e de cultivo de soja sob manejo de plantio direto; ainda para esta última área os autores encontraram baixas correlações positivas com a matéria orgânica (MO) e capacidade de troca de cátions (CTC). Tais resultados corroboram com os obtidos neste trabalho, uma vez que a MO está intimamente ligada ao EstC e a CTC com a SB.

De acordo com a matriz de correlações, notou-se expressivas associações entre a variável PLA e as variáveis EstC (-0,52,  $p < 0,05$ ), DS (-0,77,  $p < 0,05$ ), Macro (0,75,  $p < 0,05$ ); EstC e P (0,33,  $p < 0,05$ ), e entre a variável SB e os atributos químicos pH (0,47,  $p < 0,05$ ) e P (0,38,  $p < 0,05$ ).

Novamente, de acordo com os resultados obtidos, a emissão de CO<sub>2</sub> do solo não apresentou correlação com o estoque de carbono do solo. Outra observação pertinente das correlações entre as variáveis é que se pode notar que são mais intensas as correlações existentes entre os atributos do solo e o EstC, do que o apresentado em área de cana-de-açúcar, com maiores correlações com FCO<sub>2</sub>.

Segundo Follett (2001), sistemas conservacionistas de plantio promovem o aumento do teor de carbono orgânico no solo e a redução desse carbono orgânico total está diretamente relacionada com a sua degradação física (Wendling et al., 2012). Para Calonego et al. (2012), o alto valor de matéria orgânica (MO) no solo reduz a sua DS devido ao baixo peso específico da MO e à elevada porosidade total, natural da MO humificada. Vários autores observaram que maiores valores de EstC e conseqüentemente de MO, promovem uma menor densidade do solo (Araújo et al., 2004; Rodrigues et al., 2012), permitem maior estabilidade do sistema poroso, proporcionando maior estabilidade dos agregados (Franzluebbers, 2002), retenção de água, aeração, CTC e balanço de nitrogênio (Six et al., 2000).

Siqueira Neto et al. (2009) apresentaram o sistema plantio direto (SPD) com a utilização de leguminosas como uma estratégia que deve ser considerada tanto para o aumento da quantidade de MO bem como para seu efeito na redução das emissões dos gases de efeito estufa. Para Amado et al. (2006), sistemas de rotação de cultura com a presença de plantas leguminosas possuem maiores estoques de C e N.

No entanto, os restos da cultura da soja, quando comparados com os resíduos da colheita de cana-de-açúcar, possuem uma baixa relação C/N e lignina, tendo assim, maior velocidade de decomposição devido à maior facilidade dos microrganismos degradarem esses resíduos orgânicos, favorecendo a liberação de nitrogênio (N). Para que os microrganismos possam realizar a decomposição dos compostos orgânicos utilizam o N presente no solo, sendo esse um elemento essencial para a constituição das células dos microrganismos heterotróficos, sendo um dos responsáveis pela mineralização da matéria orgânica.

#### **4.2.1 Análise de Componentes Principais**

Foram avaliadas as mesmas variáveis para os dados obtidos a partir do experimento realizado em soja, sendo elas: FCO<sub>2</sub>, EstC, PLA, DS, Macro, pH, P e SB. A análise de componentes principais foi realizada a partir da matriz de correlação dada na Tabela 12. Em razão das diferentes unidades de medidas das variáveis e variâncias muito distintas, (maior variância igual a 47,903 e menor igual a 0,008), optou-se em evitar o efeito de escala pela padronização das variáveis, de modo que foram atribuídas condições de adimensionalidade para os atributos analisados.

Tabela 12. Variâncias, covariâncias e correlações das variáveis do experimento em soja<sup>1</sup>

|       | FCO2         | EstC         | PLA           | DS           | Macro         | pH           | P             | SB            |
|-------|--------------|--------------|---------------|--------------|---------------|--------------|---------------|---------------|
| FCO2  | <b>0,505</b> | 0,102        | 0,226*        | -0,201*      | 0,174*        | 0,086        | 0,270*        | 0,277*        |
| EstC  | 0,050        | <b>0,479</b> | -0,522*       | 0,384*       | -0,402*       | 0,177*       | 0,328*        | 0,411*        |
| PLA   | 0,756        | -1,700       | <b>22,107</b> | -0,769*      | 0,748*        | -0,079       | 0,057         | 0,002         |
| DS    | -0,013       | 0,024        | -0,330        | <b>0,008</b> | -0,673*       | 0,052        | -0,107        | -0,110        |
| Macro | 0,530        | -1,195       | 15,093        | -0,264       | <b>18,431</b> | 0,025        | -0,008        | -0,046        |
| pH    | 0,011        | 0,022        | -0,065        | 0,001        | 0,019         | <b>0,031</b> | -0,073        | 0,466*        |
| P     | 1,327        | 1,570        | 1,841         | -0,067       | -0,238        | -0,060       | <b>47,903</b> | 0,376*        |
| SB    | 0,908        | 1,311        | 0,040         | -0,046       | -0,907        | 0,379        | 12,007        | <b>21,238</b> |

<sup>1</sup> Variâncias destacadas na diagonal principal, covariâncias e correlações foram apresentadas abaixo e acima da diagonal principal, respectivamente.

\* Coeficiente de correlação linear de Pearson significativo à probabilidade de 0,05.

Utilizou-se o critério de Kaiser, que consiste em manter no sistema apenas as componentes relacionadas aos autovalores maiores ou iguais a unidade (Kaiser, 1958). As componentes principais e suas respectivas explicações da variância total da amostra são apresentadas na Tabela 13 e os autovalores ordenados são representados graficamente a partir do *Scree-plot* (Figura 20).

Tabela 13. Autovalores ( $\lambda_i$ ) e porcentagem da variação explicada pelas componentes principais (VCP)

| CP  | $\lambda_i$ | VCP (%) | VCP Acumulada (%) |
|-----|-------------|---------|-------------------|
| CP1 | 2,84        | 35,51   | 35,51             |
| CP2 | 2,01        | 25,08   | 60,60             |
| CP3 | 1,12        | 14,05   | 74,65             |
| CP4 | 0,73        | 9,17    | 83,82             |
| CP5 | 0,44        | 5,56    | 89,38             |
| CP6 | 0,41        | 5,18    | 94,56             |
| CP7 | 0,25        | 3,16    | 97,72             |
| CP8 | 0,18        | 2,28    | 100,00            |

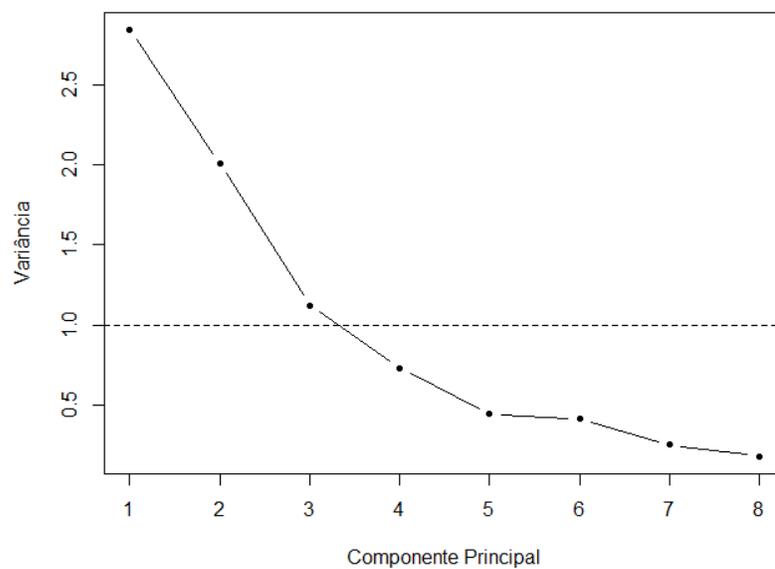


Figura 20 - “Scree-plot” das componentes principais (CP) para o experimento em área de soja.

Os resultados sugeriram a utilização de três componentes principais,

que juntas representaram 74,65% da variabilidade dos dados, sendo 35,51% da variância total explicada pela primeira componente principal; a segunda explicou 25,08%; enquanto que a terceira, explicou cerca de 14,05%.

Na Tabela 14 mostra-se os coeficientes de ponderação das variáveis e suas correlações com as componentes principais. Os atributos que apresentaram maiores coeficientes de correlação foram: para a primeira componente, PLA (-0,92), DS (0,86), Macro (-0,86) e EstC (0,66); a segunda, SB (0,84), P (0,64) e FCO2 (0,58) e; para a terceira, as variáveis pH (-0,80) e P (0,59). Foram selecionados para compor cada componente principal variáveis com coeficiente de correlação maiores do que 0,50.

Tabela 14. Coeficientes de ponderação das variáveis e suas correlações (entre parênteses) com as componentes principais.

| Variável | CP1                  | CP2                | CP3                  |
|----------|----------------------|--------------------|----------------------|
| FCO2     | -0,13 (-0,22)        | <b>0,41</b> (0,58) | 0,21 (0,22 )         |
| EstC     | <b>0,39</b> (0,66)   | 0,35 (0,50)        | 0,13 (0,14 )         |
| PLA      | <b>-0,55</b> (-0,92) | 0,09 (0,12)        | 0,01 (0,01)          |
| DS       | <b>0,51</b> (0,86 )  | -0,16 (-0,22)      | -0,02 (-0,02 )       |
| Macro    | <b>-0,51</b> (-0,86) | 0,08 (0,11)        | -0,11 (-0,11)        |
| pH       | 0,08 (0,14)          | 0,33 (0,47 )       | <b>-0,76</b> (-0,80) |
| P        | 0,02 (0,03)          | <b>0,45</b> (0,64) | <b>0,56</b> (0,59)   |
| SB       | 0,07 (0,13)          | <b>0,59</b> (0,84) | -0,21 (-0,22)        |

Os atributos mais expressivos retidos em CP1 estão relacionados ao processo do acúmulo de carbono (EstC) e a estrutura porosa do solo (DS, PLA e Macro). Tal componente captou uma relação direta entre as variáveis DS e EstC, e relação antogônica existente entre as variáveis EstC com PLA e Macro.

A matéria orgânica é um dos principais componentes do solo que interfere no seu comportamento físico, em razão de seus efeitos sobre a união de partículas primárias do solo, secundárias, exsudados de raízes e microrganismos que formam os

agregados do solo e na distribuição do tamanho de poros (Braidá et al., 2011). Atua como agente cimentante, tanto das partículas primárias quanto das partículas secundárias do solo. Proporciona redução da densidade do solo, uma vez que a densidade da MO é menor que a do solo; conseqüentemente contribui para melhor aeração do solo e retenção do armazenamento de água.

Já CP2 salientou sua associação com as variáveis FCO<sub>2</sub>, P e SB, apontando um efeito sinérgico dessa relação, uma vez que a calagem aumenta a disponibilidade de P no solo, promovendo o crescimento do sistema radicular, favorecendo o aproveitamento da água e dos nutrientes existentes no solo (Van Raij, 1991). O fósforo participa do metabolismo dos microrganismos que mineralizam a matéria orgânica (Nordgren, 1992) e conseqüentemente ocorre a emissão de CO<sub>2</sub>. Segundo Fernandes et al. (1997) a ocorrência de maior quantidade de P nas camadas superficiais do solo e está correlacionado com maior teor de matéria orgânica.

A terceira componente evidenciou maiores correlações com os atributos químicos do solo, pH e P, indicando a existência de uma relação inversa entre essas variáveis uma vez que, com o aumento do pH, a carga superficial de partículas do solo tornam-se cada vez mais negativa, aumentando a repulsão (menor adsorção) entre fosfato e superfície adsorvente (Posner & Bowden, 1980; Haynes, 1984; Barrow, 1985), citados por Novais (2016)).

As representações gráficas dos pares de componentes principais são exibidas na Figura 23. A relação entre os atributos do solo foi indicada pelo cosseno do ângulo formado entre as setas e a importância destas variáveis, pelo comprimento destas setas, que expressam a variância de cada variável. O grau com que a variável se correlaciona com o componente principal é dado pelo ângulo entre o eixo e a seta daquela variável.

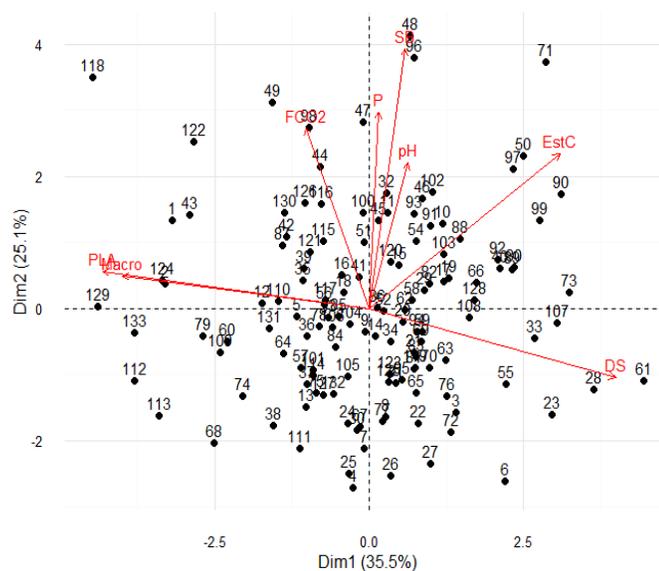


Figura 21 - Biplot das associações entre CP1 e CP2 para o experimento em área de soja.

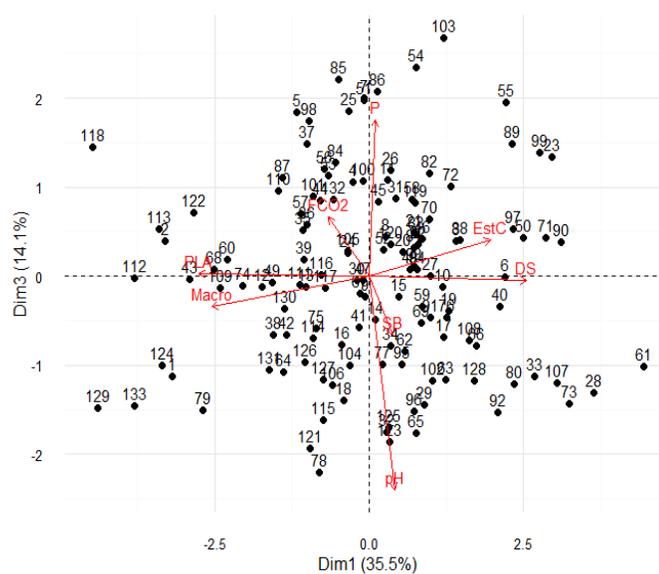


Figura 22 - Biplot das associações entre CP1 e CP3 para o experimento em área de soja.

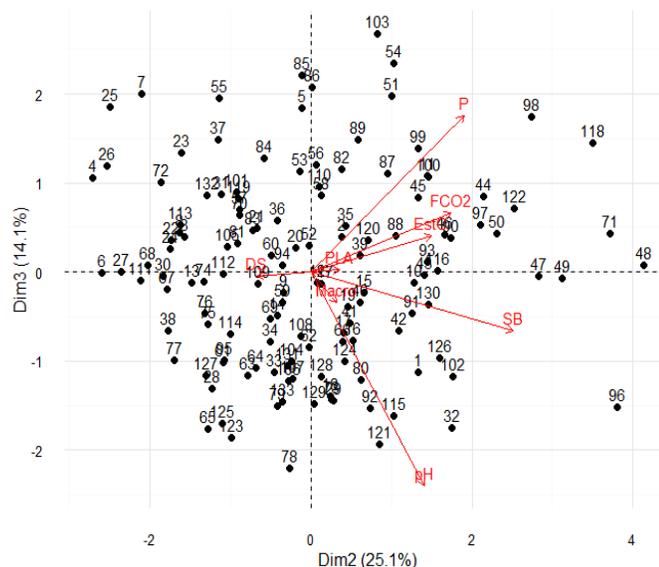


Figura 23 - Biplot das associações entre CP2 e CP3 para o experimento em área de soja.

A partir da Figura 23 pode-se observar a dispersão dos pontos amostrados por meio da representação bidimensional de acordo com o escore de cada componente no ponto. Observando o biplot para CP1 e CP2, que na figura são apresentados como Dim1 e Dim2, respectivamente, destacam-se os pontos 71 (primeiro quadrante) e 118 (segundo quadrante). Tais discrepâncias representam pontos com alta emissão de dióxido de carbono e maior presença de Fósforo e soma de bases. No entanto, o ponto amostrado 71 apresentou maiores valores para o acúmulo de carbono, diferente do 118, que indicou possuir baixo estoque de carbono.

Destaca-se também que foi possível caracterizar a área em estudo por meio de três componentes principais, reduzindo o espaço paramétrico.

### 4.3 Análise Fatorial

Por meio da medida amostral global KMO, que apresentou valor igual 0,69, indicando uma adequação aceitável ( $KMO > 0,5$ ), verificou-se a aplicabilidade

da técnica de análise fatorial. No entanto, observando as medidas de adequação amostral para cada variável, notou-se resultado insatisfatório para pH (0,41), optando-se por retirá-la da análise.

Assim, o valor de KMO global passou a ser igual a 0,74, com medidas satisfatórias de adequação amostral para todas as variáveis, com: FCO2 (0,78), EstC (0,70), PLA (0,71), Ds (0,79), Macro (0,81), P (0,69) e SB (0,63), sendo este o novo subconjunto a ser analisado.

O gráfico “Scree-plot” é apresentado na Figura 24, onde nota-se que apenas duas componentes apresentam variâncias maiores que a unidade.

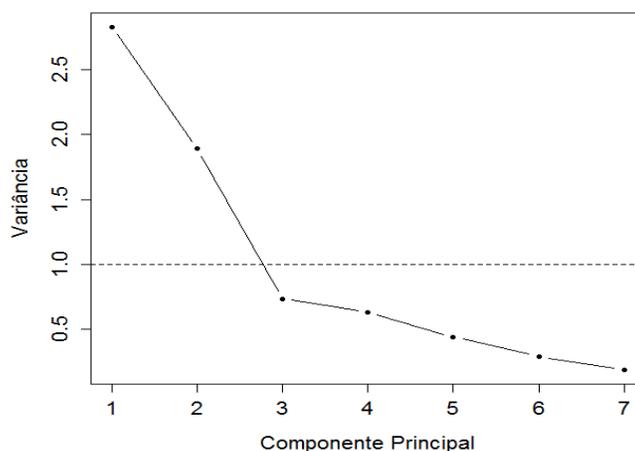


Figura 24 - “Scree-plot” das componentes principais (CP).

Para a obtenção dos fatores, foi utilizado o método de componentes principais, selecionando-se dois fatores associados aos autovalores maiores que a unidade (critério de Kaiser) que juntos explicam cerca de 67,43% da variabilidade total. Os resultados das estimativas das cargas fatoriais (loadings), das comunalidades e das variâncias específicas são apresentados na Tabela 15.

Percebe-se que apenas a variável FCO2 não apresentou comunalidade elevada relativamente à variância específica (comunalidade < variância específica), indicando que apenas 43% da variabilidade desse atributo foi explicada pelos fatores,

sendo o restante (67%) associada ao erro aleatório (variância específica).

A variável que teve maior variabilidade representada pelos fatores foi a PLA, sendo igual a 86%. De uma maneira geral, os dois fatores representaram satisfatoriamente as variáveis envolvidas nesse estudo.

Tabela 15. Cargas fatoriais, comunalidades, variâncias específicas das variáveis, autovalores e percentual acumulado de explicação da variância da análise dos dados em soja

| Variável         | Cargas Fatoriais |             | Comunalidades | Variâncias específicas |
|------------------|------------------|-------------|---------------|------------------------|
|                  | Fator 1          | Fator 2     |               |                        |
| FCO2             | -0,24            | <b>0,61</b> | 0,43          | 0,57                   |
| EstC             | <b>0,64</b>      | 0,55        | 0,72          | 0,28                   |
| PLA              | <b>-0,93</b>     | 0,09        | 0,86          | 0,14                   |
| DS               | <b>0,86</b>      | -0,20       | 0,78          | 0,22                   |
| Macro            | <b>-0,86</b>     | 0,04        | 0,74          | 0,26                   |
| P                | 0,02             | <b>0,75</b> | 0,56          | 0,44                   |
| SB               | 0,08             | <b>0,78</b> | 0,62          | 0,38                   |
| Autovalor        | 2,83             | 1,89        |               |                        |
| VT Acumulada (%) | 40,43            | 67,43       |               |                        |

A partir das cargas fatoriais (Tabela 15) foi possível identificar as variáveis que possuem maiores associações com os respectivos fatores. Representações bidimensionais permitem a visualização dessas associações (Figura 25), indicando o comportamento das amostras em função das cargas fatoriais de dois fatores quaisquer.

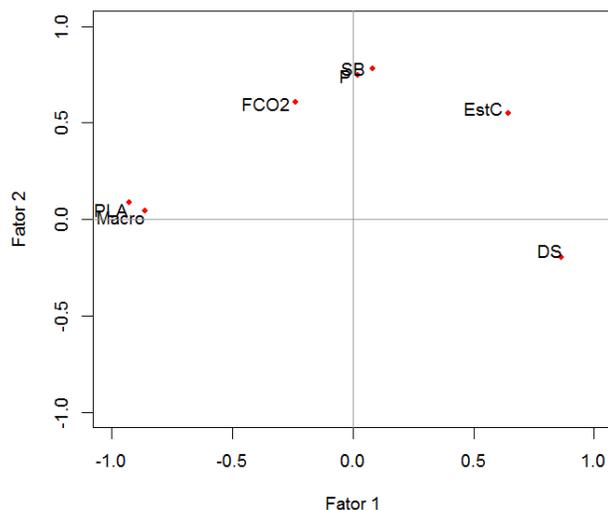


Figura 25 - Gráfico de dispersão das cargas fatoriais de F1 e F2 para as variáveis avaliadas.

As representações gráficas, juntamente com a observação das cargas fatoriais (Tabela 15), permitem caracterizar os fatores. Foram consideradas cargas fatoriais maiores do que 0,50 de acordo com Hair Jr. et al. (2009), indicadas para valores entre 120 e 149 unidades amostrais.

O Fator 1 reteve 40,43% da variabilidade total dos dados, retendo os atributos físicos do solo e o EstC, podendo ser tomado como um índice de Estoque de Carbono e atributos físicos do solo, apresentando antagonismo entre EstC e densidade do solo (DS) com as variáveis PLA e Macro. Quanto as demais variáveis que compõem o primeiro fator, nota-se que tiveram pouquíssima influência sobre este.

Segundo Pires et al. (2011), DS tem sido usada como um parâmetro importante de caracterização física da estrutura do solo e como um indicador de sua compactação. Oliveira et al. (2003) observaram que a DS sob plantio direto pode diminuir com o passar dos anos, devido ao aumento da matéria orgânica na camada superficial, possibilitando uma melhora na estrutura do solo. Os autores Tebrügge & Düring (1999), relataram que geralmente existe maior acúmulo de carbono sob sistema de plantio direto (PD) em relação ao sistema de preparo convencional (PC)

do solo.

O Fator 2, representando 27% da variabilidade total, caracteriza-se como índice da emissão de CO<sub>2</sub> e atributos químicos do solo P e SB.

Estudos conduzidos por D'Andrea et al. (2010), que avaliaram os padrões da variabilidade espacial de FCO<sub>2</sub> em área de mata nativa, sob um Latossolo Vermelho distrófico, constataram que o fluxo de CO<sub>2</sub> não apresentou dependência espacial com a porosidade total do solo (PT) e densidade do solo (DS), no entanto, houve dependência espacial com os atributos químicos do solo como : alumínio (Al), pH e SB.

A partir do modelo de análise fatorial com dois fatores, é possível analisar a qualidade do ajuste por meio da matriz residual (Tabela 16). Tal matriz indica que a matriz de correlação amostral ( $\mathbf{R}$ ) foi reproduzida adequadamente pela matriz estimada  $\hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\psi}$ , uma vez que seus elementos estão próximos a zero.

Tabela 16. Matriz residual da análise fatorial para experimento em soja

|       | FCO2  | PLA   | DS    | Macro | MO    | pH    |       |
|-------|-------|-------|-------|-------|-------|-------|-------|
| FCO2  | 0,00  | -0,08 | -0,05 | 0,12  | -0,06 | -0,18 | -0,18 |
| EstC  | -0,08 | 0,00  | 0,03  | -0,06 | 0,13  | -0,10 | -0,07 |
| PLA   | -0,05 | 0,03  | 0,00  | 0,05  | -0,06 | 0,01  | 0,01  |
| DS    | 0,12  | -0,06 | 0,05  | 0,00  | 0,08  | 0,02  | -0,03 |
| Macro | -0,06 | 0,13  | -0,06 | 0,08  | 0,00  | -0,02 | -0,01 |
| P     | -0,18 | -0,10 | 0,01  | 0,02  | -0,02 | 0,00  | -0,21 |
| SB    | -0,18 | -0,07 | 0,01  | -0,03 | -0,01 | -0,21 | 0,00  |

#### 4.4 Análise de Correlação Canônica

Os resultados do estudo da associação entre as variáveis ligadas a dinâmica do carbono e atributos físicos e químicos do solo, em área de cultivo de soja no município de Selvíria - MS, no ano de 2013, estão apresentados nessa seção.

Novamente, o grupo das variáveis ligadas à dinâmica do carbono foi composto pelo FCO2 e EstC (emissão de CO<sub>2</sub> e estoque de carbono do solo, respectivamente), já o grupo de atributos físicos e químicos do solo, por PLA, DS, Macro, pH, P e SB, utilizando-se as mesmas variáveis estudadas no experimento em cana-de-açúcar.

Estudou-se inicialmente, as relações existentes dentro de cada conjunto. A correlação linear existente entre as variáveis do grupo ligado a dinâmica do carbono,  $\mathbf{R}^{11}$ , foi igual a -0,10. As correlações entre as variáveis físico-químicas do solo, representadas pela matriz  $\mathbf{R}^{22}$ , encontram-se na Tabela 17.

Tabela 17. Matriz de correlações ( $\mathbf{R}^{22}$ ) das variáveis físico-químicas do solo.

|       | PLA    | DS     | Macro  | pH    | P     | SB    |
|-------|--------|--------|--------|-------|-------|-------|
| PLA   | 1,00   | -0,77* | 0,75*  | -0,08 | 0,06  | 0,00  |
| DS    | -0,77* | 1,00   | -0,67* | 0,05  | -0,11 | -0,11 |
| Macro | 0,75*  | -0,67* | 1,00   | 0,03  | -0,01 | -0,05 |
| pH    | -0,08  | 0,05   | 0,03   | 1,00  | -0,06 | 0,47* |
| P     | 0,06   | -0,11  | -0,01  | -0,06 | 1,00  | 0,38* |
| SB    | 0,00   | -0,11  | -0,05  | 0,47* | 0,38* | 1,00  |

\* significativa à probabilidade de 0,05.

As correlações amostrais entre as variáveis de ambos os grupos ligadas tanto a dinâmica do carbono (primeiro conjunto) quanto aos atributos físicos e químicos do solo (segundo conjunto) são apresentadas na Tabela 18.

Tabela 18. Matriz de correlações ( $\mathbf{R}^{12}$ ) entre as variáveis relacionadas à dinâmica do carbono e os atributos físico-químicos do solo.

|      | PLA    | DS     | Macro  | pH    | P     | SB    |
|------|--------|--------|--------|-------|-------|-------|
| FCO2 | 0,23*  | -0,20* | 0,17*  | 0,09  | 0,27* | 0,28* |
| EstC | -0,52* | 0,38*  | -0,40* | 0,18* | 0,33* | 0,41* |

\* significativa à probabilidade de 0,05.

A maior correlação existente entre os conjuntos de variáveis deu-se entre o EstC e a porosidade livre de água do solo (-0,52,  $p < 0,05$ ). Para a área de plantio direto, tanto FCO2 quanto EstC apresentaram correlações lineares com os atributos físicos e químicos do solo.

Os coeficientes das correlações canônicas estão expressos na Tabela 19. Mostrou-se também o teste de significância dos coeficientes de correlação canônica, que tem por hipóteses:

$$H_0^1 : \rho_1^2 \neq 0, \rho_2^2 = 0 \quad vs \quad H_1^1 : \rho_2^2 \neq 0$$

ou seja, em  $H_0^1$  apenas a primeira correlação canônica é significativa e as demais não significativas. Com a rejeição de  $H_0^1$ , assume-se  $H_1^1$  verdadeira, dessa forma afirma-se que ambas as correlações são significativas.

Tabela 19. Autovalores, correlações canônicas e teste de significância para as correlações canônicas ( $\alpha=0,05$ ).

| Pares de Variáveis Canônicas | Autovalores | Correlações Canônicas | Estatística de Teste | gl | Conclusão          |
|------------------------------|-------------|-----------------------|----------------------|----|--------------------|
| ( $U_1, V_1$ )               | 0,496       | 0,704                 | 108,85               | 5  | Rejeita-se $H_0^1$ |
| ( $U_2, V_2$ )               | 0,156       | 0,395                 |                      |    |                    |

Assim, como para o experimento em área de cana-de-açúcar, por meio da rejeição de  $H_0^1$  concluiu-se que os dois coeficientes de correlação encontrados são

significativos a 5%. Analisa-se o primeiro par de variáveis canônicas, o qual explica cerca de 76% da relação entre as variáveis, de maneira que  $U_1$  é a melhor combinação linear na associada a  $V_1$  e vice-versa. Já o segundo par explica aproximadamente 24% da relação entre as variáveis canônicas.

Por meio de análise visual do gráfico de probabilidade *Q-Q Plot* (*Chi-Square plot*) multivariado, apresentado na Figura 18, observa-se uma tendência das distâncias alinharem-se ao redor da bissetriz, sugerindo, na prática, aderência dos dados à distribuição normal multivariada de probabilidade.

As correlações entre as variáveis originais com os pares de variáveis canônicas, denominadas cargas canônicas e cargas canônicas cruzadas, são apresentadas na Tabela 20.

Tabela 20. Cargas canônicas e cargas canônicas cruzadas para experimento em área de soja.

| Variáveis                         | Cargas Canônicas |                | Cargas Can. Cruzadas |                |
|-----------------------------------|------------------|----------------|----------------------|----------------|
|                                   | U <sub>1</sub>   | U <sub>2</sub> | V <sub>1</sub>       | V <sub>2</sub> |
| Dinâmica do carbono do solo       |                  |                |                      |                |
| FCO2                              | 0,053            | 0,999          | 0,037                | 0,394          |
| EstC                              | 0,999            | 0,050          | 0,703                | 0,020          |
| PVTE(%)                           | 50,01            | 49,99          |                      |                |
|                                   | V <sub>1</sub>   | V <sub>2</sub> | U <sub>1</sub>       | U <sub>2</sub> |
| Atributos físico-químicos do solo |                  |                |                      |                |
| PLA                               | -0,761           | 0,645          | -0,536               | 0,255          |
| DS                                | 0,562            | -0,563         | 0,395                | -0,222         |
| Macro                             | -0,586           | 0,495          | -0,412               | 0,196          |
| pH                                | 0,246            | 0,195          | 0,173                | 0,077          |
| P                                 | 0,448            | 0,642          | 0,316                | 0,253          |
| SB                                | 0,566            | 0,650          | 0,399                | 0,257          |
| PVTE(%)                           | 30,33            | 30,85          |                      |                |

Constata-se que a proporção da variância total, que é explicada separadamente pelas variáveis canônicas (Tabela 20), é de 50,01% por U<sub>1</sub> e de 30,33% por V<sub>1</sub>, ou seja, a variável U<sub>1</sub> representa 50,01% da variação total do grupo de variáveis ligadas à dinâmica do carbono e a variável V<sub>1</sub> representa 30,33% da variação total do grupo de variáveis físico-químicas. O segundo par, U<sub>2</sub> e V<sub>2</sub> explica separadamente 49,99% e 30,85% da proporção da variância total.

A variável EstC apresenta maior carga canônica em U<sub>1</sub>. Composto o primeiro par canônico, V<sub>1</sub> é representada pelo comportamento antagônico entre as variáveis PLA e Macro com as variáveis DS e SB e os demais atributos que se destacam em V<sub>1</sub>, tais como DS e SB. Tais variáveis foram selecionadas de acordo

com o valor absoluto de sua carga canônica, para este caso, valores maiores do que 0,5. Por meio desses resultados, aponta-se que para essa área, pontos que possuem valores altos de DS e SB, e valores baixos de PLA e Macro, tendem a possuir valores elevados de EstC.

Em relação as cargas canônicas cruzadas, a variável EstC destaca-se por possuir maior associação com  $V_1$ ; já entre as variáveis físicas e químicas do solo, PLA (-0,536) é a que possuiu maior associação com  $U_1$ .

O segundo par canônico é composto por  $U_2$  representando a emissão de  $CO_2$  do solo; e  $V_2$  sendo uma variável que indica o antagonismo entre DS e as demais variáveis que a compõem. As cargas canônicas cruzadas, não puderam ser interpretadas, devido aos baixos valores das cargas canônicas, não atingindo o ponto de corte estabelecido neste trabalho.

Por meio da análise de correlação canônica pôde-se observar que a variável pH não apresentou relação nenhuma das variáveis canônicas, evidenciando as correlações não significativas desta com as demais variáveis.

A Figura 26 representa os pares de variáveis canônicas e os valores padronizados que os dois pares assumem em cada ponto amostral. O gráfico de dispersão para os pontos do primeiro par apresenta um comportamento mais próximo ao linear, do que o segundo par, uma vez que o primeiro par de variáveis canônicas é responsável pela maior representação da variabilidade total (Mingoti, 2005).

Neste sentido, pode-se dizer que a variável canônica  $U_1$  apresenta valores altos associados a valores altos da variável  $V_1$ , principalmente para o ponto amostrado 71.



## 5 CONCLUSÕES

Os resultados das análises estatísticas multivariadas permitem as seguintes conclusões a respeito da relação existente entre a emissão de  $\text{CO}_2$  e as propriedades físicas e químicas do solo, para dois conjuntos de dados, de regiões brasileiras distintas, sob culturas e manejos diferentes:

Para o experimento realizado em área de cana-de-açúcar tem-se que:

A análise de componentes principais caracterizou a área por meio de três componentes, que juntas explicaram 74,65% da variabilidade total das variáveis, identificando que a emissão de  $\text{CO}_2$  pode ser relacionada aos atributos físicos do solo. Dessa forma houve uma redução na dimensão paramétrica para três componentes.

Na análise fatorial envolvendo as variáveis relativas a dinâmica do carbono e os atributos físicos e químicos do solo foram estabelecidos três fatores que explicaram acumuladamente cerca de 82,14% da variabilidade total dos dados. O primeiro fator envolve as variáveis emissão de  $\text{CO}_2$  do solo, porosidade livre de água, densidade do solo e macroporosidade; o segundo, pH e soma de bases e o terceiro fator compreende apenas a variável fósforo.

Dessa forma, por meio da análise fatorial também concluiu-se que a emissão de  $\text{CO}_2$  relacionou-se com as variáveis físicas do solo. Ambas as áreas apresentaram emissões  $\text{CO}_2$  baixas quando comparadas com os valores apresentados nos trabalhos presentes na literatura, e para este caso pode-se indicar uma restrição relativa a macroporosidade, que possuiu baixos valores para o experimentos em cana-de-açúcar.

Em estudo da associação dos grupos de variáveis, por meio da análise de correlação canônica, foram obtidas duas correlações canônicas significativas, de

acordo com o teste aplicado. Relativamente às cargas canônicas, para o primeiro par canônico, a variável que se destacou na componente relacionada à dinâmica do carbono no solo foi a emissão de CO<sub>2</sub> do solo e para a componente físico-química teve-se destaque para as variáveis PLA, DS, Macro e pH. Os pontos amostrais com altos valores para porosidade livre de água, macroporosidade, maiores concentrações de pH em amostras com solos menos densos estão associados a níveis mais elevados de emissão de CO<sub>2</sub> do solo.

Para o experimento realizado em área de soja tem-se que:

A análise de componentes principais caracterizou a área por meio de três componentes, que acumularam 73,85% da variabilidade total das variáveis, indicando que a emissão de CO<sub>2</sub> relacionou-se aos atributos químicos do solo (P e SB). Dessa forma houve uma redução na dimensão paramétrica para três componentes.

De acordo com a análise fatorial foram obtidos dois fatores explicando 67,43% da variabilidade dos dados. O primeiro fator associou-se aos atributos físicos do solo e ao estoque de carbono, já o segundo fator representou a relação existente entre a emissão de CO<sub>2</sub> do solo, fósforo e soma de bases. Assim, para essa área a emissão de CO<sub>2</sub> do solo foi limitada pelos atributos químicos do solo, ou seja, aquelas variáveis relacionadas ao processo de produção do gás no interior do solo.

A análise de correlação canônica novamente forneceu duas correlações significativas, por meio do teste aplicado. No contexto das cargas canônicas, para o primeiro par canônico, a variável que foi evidenciada pela componente relacionada à dinâmica do carbono no solo foi a variável estoque de carbono do solo e para a componente físico-química teve-se destaque apenas os atributos físicos do solo, tais como: PLA, DS e Macro. Dessa forma, pontos amostrais com baixos valores para porosidade livre de água e macroporosidade em amostras de solos mais densos estão associados a níveis mais elevados de estoque de carbono do solo. A emissão relacionou-se a variável que compõe o segundo par canônico, indicando que maiores valores a esta variável estão associados a maiores valores para PLA, Macro, P e SB e solos menos densos.

Por meio das análises em ambas as áreas, conclui-se que não foi possível verificar relação direta entre a emissão de CO<sub>2</sub> e o estoque de carbono no solo, podendo assim, áreas com maior acúmulo de carbono não serem responsáveis por maiores emissões de dióxido de carbono, sendo um possível indicativo de que o manejo do solo possa ter influência na dinâmica do carbono do solo.

Comparando os resultados fornecidos pelas análises, pode-se verificar que embora ACP e AF sejam análises distintas aplicadas com objetivos diferentes, neste trabalho ambas forneceram resultados muito semelhantes. Este fato é reforçado a partir da pesquisa empírica que tem demonstrado resultados análogos em muitos casos. Na maioria das aplicações as análises chegam a resultados essencialmente idênticos se o número de variáveis exceder 30 ou se as comunalidades excederem 0,60 para a maioria das variáveis (Hair Jr. et al., 2009).

Portanto, foi possível detectar a importância de cada variável físico-química em relação a emissão de CO<sub>2</sub>, bem como, quais variáveis que não mostraram influência. Deve-se salientar que as duas aplicações realizadas referem-se a conjuntos de dados coletados para outros objetivos de pesquisa, sendo utilizados, no presente trabalho, como estudo de casos, não sendo possível comparar os resultados das duas culturas (cana-de-açúcar e soja), nem mesmo inferir as conclusões para todas as culturas de cana-de-açúcar ou de soja. Destaca-se que a amostragem deu-se em dois diferentes tipos e manejos de solo, diferentes anos e condições climáticas.

Sendo assim, conclui-se que as técnicas multivariadas são ferramentas muito importantes, no sentido de evidenciarem relações não visualizadas por meio de outras técnicas estatísticas, sugerindo-se o uso mais intenso dos métodos multivariados em diferentes aplicações.

## REFERÊNCIAS

AMADO, T. J. C.; BAYER, C.; CONCEIÇÃO, P. C.; SPAGNOLLO, E.; DE CAMPOS, B.-H. C.; DA VEIGA, M. Potential of carbon accumulation in no-till soils with intensive use and cover crops in southern Brazil. **Journal of Environmental Quality**, v.35, n.4, p.1599–1607, 2006.

ARAÚJO, M.; TORMENA, C.; SILVA, A. D. Propriedades físicas de um Latossolo Vermelho distrófico cultivado e sob mata nativa. **Bras. Ci. Solo**, v.28, p.337–345, 2004.

BARROW, N. Comparing the effectiveness of fertilizers. **Fertilizer Research**, v.8, n.1, p.85–90, 1985.

BARTLETT, M. S. A Note on test of significance in multivariate analysis. **Proceedings of the Cambridge Philosophical Society**, v.35, p.180–185, 1939.

BAYER, C.; MIELNICZUK, J.; AMADO, T. J. C.; MARTIN-NETO, L.; FERNANDES, S. V. Organic matter storage in a sandy clay loam Acrisol affected by tillage and cropping systems in southern Brazil. **Soil and tillage research**, v.54, n.1, p.101–109, 2000.

BICALHO, E. S.; PANOSSO, A. R.; TEIXEIRA, D. D. B.; MIRANDA, J. G. V.; PEREIRA, G. T.; N LA SCALA JR. Spatial variability structure of soil CO<sub>2</sub> emission and soil attributes in a sugarcane area. **Agriculture Ecosystems Environment**, v.189, p.206–215, 2014.

BRAIDA, J. A.; BAYER, C.; ALBUQUERQUE, J. A.; REICHERT, J. M. Matéria orgânica e seu efeito na física do solo. v.7, p.221–278, 2011.

BRANCAGLIONI, V. A.; TERÇARIOL, M. C.; MONTANARI, R.; LA SCALA JR, N.; FERRAUDO, A. S.; PEREIRA, G. T.; PANOSSO, A. R. Geoestatística multivariada da emissão de CO<sub>2</sub> em área agrícola. **Blucher Mathematical Proceedings**, v.1, n.1, p.357–364, 2015.

BRITO, L. F.; MARQUES JR, J.; PEREIRA, G. T.; SOUZA, Z. M.; LA SCALA JR, N. Soil CO<sub>2</sub> emission of sugarcane fields as affected by topography. **Scientia Agricola**, v.66, n.1, p.77–83, 2009.

CAIRES, E.; KUSMAN, M.; BARTH, G.; GARBUIO, F.; PADILHA, J. Alterações químicas do solo e resposta do milho à calagem e aplicação de gesso. **Revista Brasileira de Ciência do Solo**, v.28, n.1, p.125–136, 2004.

CALONEGO, J. C.; DOS SANTOS, C. H.; TIRITAN, C. S.; CUNHA JR, J. R. Estoques de carbono e propriedades físicas de solos submetidos a diferentes sistemas de manejo. **Revista Caatinga**, v.25, n.2, p.128–135, 2012.

COOLEY, W. W.; LOHNES, P. R. **Multivariate Procedure for the Behavioral Sciences**. New York: John Wiley and Sons, Inc, 1962. 211p.

D'ANDREA, A.; SILVA, M. L. N.; FREITAS, D.; CURI, N.; SILVA, C. A. Variações de curto prazo no fluxo e variabilidade espacial do CO<sub>2</sub> do solo em floresta nativa. **Pesquisa Florestal Brasileira, Colombo**, v.30, n.62, 2010.

EDEN, M.; SCHJØNNING, P.; MOLDRUP, P.; DE JONGE, L. W. Compaction and rotovation effects on soil pore characteristics of a loamy sand soil with contrasting organic matter content. **Soil Use and Management**, v.27, n.3, p.340–349, 2011.

EMBRAPA. **Manual de métodos de análise de solo**. Brasília: Ministério da Agricultura e do Abastecimento, 1997. 212p.

EMBRAPA. **Sistema brasileiro de classificação de solos**. Rio de Janeiro: Ministério da Agricultura e do Abastecimento, 2006. 412p.

FERNANDES, F.; DA SILVA, S. M. C. P. Manual prático para a compostagem de bi-ossólidos. In: **Manual prático para a compostagem de biossólidos**, PROSAB, 1999.

FERNANDES, L. A.; SIQUEIRA, J.; GUEDES, G. A. A.; CURI, N. Propriedades químicas e bioquímicas de solos sob vegetação de mata e campo cerrado adjacentes. **Ciência e Agrotecnologia**, v.21, n.1, p.58–70, 1997.

FERREIRA, A. D. S.; OLIVEIRA, R. S. D.; SANTOS, M. A. D.; BORGES, E. N. Respiratory activity of soil microbiota and glucose content in response to phosphorus addition in Cerrado soil-Brazil. **Revista Brasileira de Ciência do Solo**, v.32, n.5, p.1891–1897, 2008.

FERREIRA, D. F. **Estatística Multivariada**. Lavras, MG: UFLA, 2008. 661p.

FOLLETT, R. Soil management concepts and carbon sequestration in cropland soils. **Soil and Tillage Research**, v.61, n.1, p.77–92, 2001.

FRANZLUEBBERS, A. Soil organic matter stratification ratio as an indicator of soil quality. **Soil and Tillage Research**, v.66, n.2, p.95–106, 2002.

FREDDI, O. D. S.; CENTURON, J. F.; BEUTLER, A. N.; ARATANI, R. G.; LEONEL, C. L.; SILVA, Á. P. D. Compactação do solo e intervalo hídrico ótimo no crescimento e na produtividade da cultura do milho. **Bragantia**, p.477–486, 2007.

GODOI, D. R.; HONGYU, K.; DIAS, C. T. S. Os métodos Biplot de Gabriel na análise multivariada. In: 57° REUNIÃO ANUAL DA REGIÃO BRASILEIRA DA SOCIEDADE INTERNACIONAL DE BIOMETRIA, 2012. ; resumos. Piracicaba: , 2012.

GRAF, A.; HERBST, M.; WEIHERMÜLLER, L.; HUISMAN, J. A.; PRO-LINGHEUER, N.; BORNEMANN, L.; VEREECKEN, H. Analyzing spatiotemporal variability of heterotrophic soil respiration at the field scale using orthogonal functions. **Geoderma**, v.181, p.91–101, 2012.

HAIR JR., J. F.; BLACK, W. C.; BABIN, B. J.; ANDERSON, R. E.; TAHAM, R. L. **Análise multivariada de dados**. Porto Alegre - RS: Bookman, 2009. 688p.

HALL, J.; PATERSON, E.; KILLHAM, K. The effect of elevated CO<sub>2</sub> concentration and soil pH on the relationship between plant growth and rhizosphere denitrification potential. **Global Change Biology**, v.4, n.2, p.209–216, 1998.

HAYNES, R. Lime and phosphate in the soil-plant system. **Advances in Agronomy**, v.37, p.249–315, 1984.

HOTELLING, H. Relations between two sets of variates. **Biometrika**, v.28, n.3/4, p.321–377, 1936.

IPCC. **Climate Change 2007: The Physical Science Basis: Summary for Policymakers**. Intergovernmental Panel on Climate Change Secretariat, 2007.

IPCC. **Climate change 2014: mitigation of climate change**. Cambridge University Press, 2014. 3v.

JOBSON, J. D. **Applied Multivariate Data Analysis; Volume II: Categorical and Multivariate Methods**. New York: Springer-Verlag, 1992. 731p.

JOHNSON, R. A.; WICHERN, D. W. **Applied Multivariate Statistical Analysis**. New Jersey: Prentice Hall, 2007. 773p.

JOLLIFFE, I. T. **Principal component analysis**. New York: Springer, 2002. 487p.

JORGE, R. F.; DE ALMEIDA, C. X.; BORGES, E. N.; PASSOS, R. R. Distribuição de poros e densidade de latossolos submetidos a diferentes sistemas de uso e manejo. **Bioscience Journal**, v.28, n.1, p.159–169, 2012.

KAISER, H. F. The varimax criterion for analytic rotation in factor analysis. **Psychometrika**, v.23, n.3, p.187–200, 1958.

- KLEFENS, P. C. O. O Biplot na análise fatorial multivariada. Piracicaba, 2009. 102p. Tese (Doutorado) - Universidade de São Paulo.
- LAL, R. Forest soils and carbon sequestration. **Forest ecology and management**, v.220, p.242–258, 2005.
- LATTIN, J. M.; CARROLL, J. D.; GREEN, P. E. **Análise de dados multivariados**. São Paulo: CENGAGE Learning, 2011. 455p.
- LEAL, F. T.; FRANÇA, A. B. C.; SIQUEIRA, D. S.; TEIXEIRA, D. B.; MARQUES JR, J.; LA SCALA JR, N. Characterization of potential CO<sub>2</sub> emissions in agricultural areas using magnetic susceptibility. **Scientia Agricola**, v.72, n.6, p.535–539, 2015.
- MAESTRE, M. R.; SARTORIO, S. D.; MARCIANO, C. R.; OLIVEIRA, F. C.; PIEDADE, S. M. S.; LIMA, C. G. Aplicação da análise de componentes principais na seleção de características de solo. In: 19º SIMPÓSIO NACIONAL DE PROBABILIDADE E ESTATÍSTICA, 2010. ; resumos. Campinas: , 2010.
- MANLY, B. J. F. **Multivariate Statistical Methods: Primer**. USA: CHAPMAN and HALL/CRC, 2004. 207p.
- MARTINS, A. Dinâmica do carbono do solo. In: FIEDLER, N. C.; LOUSADA, J. L. P. C.; CARMO, F. C. A. (Ed.). **Emissão de gases e sequestro de carbono em sistemas florestais**. Alegre: CAUFES, 2016. p.193–211.
- MINGOTI, S. A. **Análise de Dados Através de Métodos de Estatística Multivariada: uma Abordagem Aplicada**. Belo Horizonte - MG: Editora UFMG, 2005. 295p.
- MOREIRA, F. M. S.; SIQUEIRA, J. O. **Microbiologia e bioquímica do solo**. Ufla, 2006. 744p.
- MORRISON, D. F. **Multivariate statistical methods**. EUA: McGraw-Hill, 1967. 338p.

NOGUEIRA, S. F.; GREGO, C. R.; RODRIGUES, C. A. G.; GIMENES, F. M. A.; FURTADO, A. L. S.; BRANCALIAO, S. R.; DEMARCHI, J. J. A. A. Análise espacial de atributos do solo e correlações com fluxos de gases em diferentes sistemas de produção agrícola. In: , 2013. **Anais XVI Simpósio Brasileiro de Sensoriamento Remoto - SBS**; resumos. Foz do Iguaçu: INPE, 2013.

NORDGREN, B. A method for determining microbially available N and P in an organic soil. **Biology and Fertility of Soils**, v.13, n.4, 1992.

NOVAIS, R. F. Fósforo, 2016. Notas de aula da disciplina Fertilidade do Solo e Nutrição de Plantas. Disponível em: <<http://www.passeidireto.com/arquivo/5751171/fosforo>> Acesso em: 20 de dezembro de 2016.

OLIVEIRA, G. C. D.; DIAS JR, M. D. S.; RESCK, D. V. S.; CURTI, N. Alterações estruturais e comportamento compressivo de um Latossolo Vermelho distrófico argiloso sob diferentes sistemas de uso e manejo. **Pesquisa Agropecuária Brasileira**, v.38, n.2, p.291–299, 2003.

OMETTO, J. C. **Bioclimatologia Vegetal**. Piracicaba: Agronomica CERES Ltda, 1981. 425p.

OURIVES, O. E. A.; SOUZA, G. M.; TIRITAN, C. S.; SANTOS, D. H. Organic fertilizer as phosphorus supplier for Brachiaria Brizantha cv. Marandú. **Pesquisa Agropecuária Tropical**, v.40, n.2, 2010.

PANOSSO, A. R. Variabilidade espacial da emissão de CO<sub>2</sub> e sua relação com propriedades do solo em área de cana-de-açúcar no Sudeste do Brasil. Jaboticabal, 2011. 107p. Tese (Doutorado) - Faculdade de Ciências Agrárias e Veterinárias, Universidade Estadual Paulista "Júlio de Mesquita Filho".

PANOSSO, A. R.; MARQUES, J.; MILORI, D. M. B. P.; FERRAUDO, A. S.; BARBIERI, D. M.; PEREIRA, G. T.; N LA SCALA JR. Soil CO<sub>2</sub> emission and its

relation to soil properties in sugarcane areas under slash-and-burn and green harvest. **Soil and Tillage Research**, v.111, n.2, p.190–196, 2011.

PARIENTE, S.; LAVÉE, H. Soil Organic Matter and degradation. **SCAPE – Soil conservation and Protection for Europe**, 2000.

PEREIRA, G. T.; TEIXEIRA, D. B. B.; BICALHO, E. S.; L, A.; LA SCALA JR, N. Incertezas na predição espacial da respiração do solo e propriedades relacionadas. In: 57º REUNIÃO ANUAL DA REGIÃO BRASILEIRA DA SOCIEDADE INTERNACIONAL DE BIOMETRIA, 2012. ; resumos. Piracicaba: , 2012.

PINTO JR, O. B.; SANCHES, L.; DALMOLIN, A. C.; NOGUEIRA, J. S. Efluxo de CO<sub>2</sub> do solo em floresta de transição Amazônia Cerrado e em área de pastagem. **Acta amazonica**, v.39, n.4, p.813–821, 2009.

PIRES, L. F.; ROSA, J. A.; TIMM, L. C. Comparação de métodos de média da densidade do solo. **Acta Scientiarum Agronomy**, v.33, n.1, p.161–170, 2011.

POSNER, A.; BOWDEN, J. Adsorption isotherms: should they be split? **Journal of Soil Science**, v.31, n.1, p.1–10, 1980.

R CORE TEAM. **R: A Language and Environment for Statistical Computing**. R Foundation for Statistical Computing, Vienna, Austria, 2016.

RAIJ, B. V. **Análise química para avaliação da fertilidade em solos tropicais**. Campinas - SP: Instituto Agronômico, 2001. 285p.

RAIJ, B. V. **Fertilidade do solo e manejo de nutrientes**. Piracicaba: International Plant Nutrition Institute, 2011. 420p.

REINERT, D. J.; ALBUQUERQUE, J. A.; REICHERT, J. M.; AITA, C.; ANDRADA, M. M. C. Bulk density critical limits for normal root growth of cover crops. **Revista Brasileira de Ciência do Solo**, v.32, n.5, p.1805–1816, 2008.

REIS, E. **Estatística Multivariada Aplicada**. Lisboa: Edições Sílabo, 2001. 343p.

RENCHER, A. C. **Methods of multivariate analysis**. New York: John Wiley, 2002. 708p.

RODRIGUES, V. A.; FENNER, P. T.; SANSÍGOLO, C. A.; MORAES, M. H. Estimativa Da Água No Solo Em Floresta de Eucalyptus Grandis. **Irriga**, v.17, n.4, p.523, 2012.

RONQUIM, C. C. Conceitos de fertilidade do solo e manejo adequado para as regiões tropicais. **Embrapa Monitoramento por Satélite. Boletim de Pesquisa e Desenvolvimento**, 2010.

DOS SANTOS, D. R.; GATIBONI, L. C.; KAMINSKI, J. Fatores que afetam a disponibilidade do fósforo e o manejo da adubação fosfatada em solos sob sistema plantio direto. **Ciência Rural**, v.38, n.2, p.576–586, 2008.

SARTORI, F.; LAL, R.; EBINGER, M. H.; PARRISH, D. J. Potential soil carbon sequestration and CO<sub>2</sub> offset by dedicated energy crops in the USA. **Critical Reviews in Plant Sciences**, v.25, n.5, p.441–472, 2006.

SCHARLEMANN, J. P. W.; TANNER, E. V. J.; HIEDERER, R.; KAPOS, V. Global soil carbon: understanding and managing the largest terrestrial carbon pool. **Carbon Management**, v.5, n.1, p.81–91, 2014.

SCHWENDENMANN, L.; VELDKAMP, E.; BRENES, T.; O'BRIEN, J. J.; MACKENSEN, J. Spatial and temporal variation in soil CO<sub>2</sub> efflux in an old-growth neotropical rain forest, La Selva, Costa Rica. **Biogeochemistry**, v.64, n.1, p.111–128, 2003.

SILVA, E. F. Variabilidade espaço-temporal da emissão de CO<sub>2</sub> do solo em curto período sob influência de eventos de precipitação. Jaboticabal, 2016. 58p. Dissertação (Mestrado) - Universidade Estadual Paulista (UNESP).

SILVA, E. F.; MOITINHO, M. R.; TEIXEIRA, D. B.; PEREIRA, G. T.; LA SCALA JR, N. Soil CO<sub>2</sub> emission associated with liming in area of conversion of orange for sugar cane. **Engenharia Agrícola**, v.34, n.5, p.885–898, 2014.

SILVA, N. R.; PADOVANI, C. R. Utilização de componentes principais em experimentação agrônômica. **Revista de Energia na Agricultura**, v.21, n.4, p.98–113, 2006.

SIQUEIRA NETO, M.; PICCOLO, M. D. C.; FEIGL, B. J.; VENZKE FILHO, S. D. P.; CERRI, C. E. P.; CERRI, C. C. Crop rotation under no-tillage in Tibagi (Paraná State, Brazil): II-CO<sub>2</sub> and N<sub>2</sub>O emissions. **Revista Brasileira de Ciência do Solo**, v.33, n.4, p.1023–1029, 2009.

SIX, J.; MERCKX, R.; KIMPE, K.; PAUSTIAN, K.; ELLIOTT, E. A re-evaluation of the enriched labile soil organic matter fraction. **European Journal of Soil Science**, v.51, n.2, p.283–293, 2000.

SMART, D. R.; PENUELAS, J. Short-term CO<sub>2</sub> emissions from planted soil subject to elevated CO<sub>2</sub> and simulated precipitation. **Applied Soil Ecology**, v.28, n.3, p.247–257, 2005.

SOUZA, E. C. Os métodos biplot e escalonamento multidimensional nos delineamentos experimentais. Piracicaba, 2010. 134p. Tese (Doutorado) - Universidade de São Paulo.

SOUZA, J. L.; PREZOTTI, L. C.; GUARÇONI, M. Potencial de sequestro de carbono em solos agrícolas sob manejo orgânico para redução da emissão de gases de efeito estufa. **Idesia (Arica)**, v.30, n.1, p.7–15, 2012.

SPEARMAN, C. "General Intelligence", objectively determined and measured. **The American Journal of Psychology**, v.15, n.2, p.201–292, 1904.

TABACHNICK, B. G.; FIDELL, L. S. **Using Multivariate Statistics**. USA: Pearson Education, Inc, 2007. 980p.

TAVARES, R. L. M. Emissão de CO<sub>2</sub> e atributos físicos, químicos e microbiológicos do solo em sistemas de manejo de cana-de-açúcar. Campinas, 2014. 100p. Tese (Doutorado) - Engenharia Agrícola, Universidade Estadual de Campinas.

TAVARES, R. L. M.; SIQUEIRA, D. S.; PANOSSO, A. R.; CASTIONI, G. A. F.; SOUZA, Z. M. D.; SCALA JR, N. L. Soil management of sugarcane fields affecting CO<sub>2</sub> fluxes. **Scientia Agricola**, v.73, n.6, p.543–551, 2016.

TAVARES, R. L. M.; DE SOUZA, Z. M.; SIQUEIRA, D. S.; LA SCALA JR, N.; PANOSSO, A. R.; CAMPOS, M. C. C. Soil CO<sub>2</sub> emission in sugarcane management systems. **Acta Agriculturae Scandinavica, Section B - Soil & Plant Science**, v.65, n.8, p.755–762, 2015.

TEBRÜGGE, F.; DÜRING, R.-A. Reducing tillage intensity – a review of results from a long-term study in Germany. **Soil and tillage research**, v.53, n.1, p.15–28, 1999.

TEIXEIRA, D. B.; BICALHO, E. S.; PANOSSO, A. R.; CERRI, C. E. P.; PEREIRA, G. T.; LA SCALA JR, N. Spatial variability of soil CO<sub>2</sub> emission in a sugarcane area characterized by secondary information. **Scientia Agricola**, v.70, n.3, p.195–203, 2013.

TEIXEIRA, D. B.; BICALHO, E. S.; PANOSSO, A. R.; PERILLO, L. I.; IAMAGUTI, J. L.; PEREIRA, G. T.; LA SCALA JR, N. Uncertainties in the prediction of spatial variability of soil CO<sub>2</sub> emissions and related properties. **Revista Brasileira de Ciência do Solo**, v.36, n.5, p.1466–1475, 2012.

TEIXEIRA, D. B.; PANOSSO, A. R.; CERRI, C. E. P.; PEREIRA, G. T.; LA SCALA JR, N. Soil CO<sub>2</sub> emission estimated by different interpolation techniques. **Plant and Soil**, v.345, n.1-2, p.187–194, 2011.

TOEBE, M.; CARGNELUTTI FILHO, A. Não normalidade multivariada e multicolinearidade na análise de trilha em milho. **Pesquisa Agropecuária Brasileira**, v.48, n.5, p.466–477, 2013.

TREVISAN, R.; MATTOS, M.; HERTER, F. Atividade microbiana em argissolo vermelho-amarelo distrófico típico coberto com aveia preta (*Avena sp.*) no outono, em um pomar de pessegueiro. **Científica Rural, Bagé**, v.7, n.2, p.83–89, 2002.

VAN RAIJ, B. **Fertilidade do solo e adubação**. Editora Agronômica Ceres, 1991.

WARRICK, A. W.; NIELSEN, D. R. Spatial variability of soil physical properties in the field. **Applications of soil physics**, p.319–344, 1980.

WENDLING, B.; VINHAL-FREITAS, I. C.; DE OLIVEIRA, R. C.; BABATA, M. M.; BORGES, E. N. Densidade, agregação e porosidade do solo em áreas de conversão do cerrado em floresta de pinus, pastagem e plantio direto = Density, porosity and soil aggregation in areas of cerrado conversion in pine forest, pasture and no-tillage. **Bioscience Journal**, v.28, n.1, 2012.

XU, M.; QI, Y. Soil-surface CO<sub>2</sub> efflux and its spatial and temporal variations in a young ponderosa pine plantation in northern California. **Global Change Biology**, v.7, n.6, p.667–677, 2001.

ZANINI, J. R.; BARRETO, A. K.; LA SCALA JÚNIOR, N. Taxa de emissão de CO<sub>2</sub> de um latossolo fertirrigado com ácido fosfórico por gotejamento. **Engenharia Agrícola**, v.25, p.179–188, 2005.