

**UNIVERSIDADE ESTADUAL PAULISTA “JÚLIO DE MESQUITA FILHO”
FACULDADE DE ENGENHARIA
CAMPUS DE ILHA SOLTEIRA**

ELTON GEAN ARAÚJO

**MODELAGEM DE DADOS DE CANA-DE-AÇÚCAR: UMA ABORDAGEM
MULTIVARIADA**

Ilha Solteira
2016

PROGRAMA DE PÓS-GRADUAÇÃO EM AGRONOMIA

ELTON GEAN ARAÚJO

**MODELAGEM DE DADOS DE CANA-DE-AÇÚCAR: UMA
ABORDAGEM MULTIVARIADA**

Tese apresentada à Faculdade de Engenharia de
Ilha Solteira – UNESP como parte dos requisitos
para obtenção do título de doutor em agronomia.
Especialidade: Sistema de produção.

Alan Rodrigo Panosso
Orientador

Morel de Passos e Carvalho
Co-orientador

Ilha Solteira
2016

FICHA CATALOGRÁFICA

Desenvolvido pelo Serviço Técnico de Biblioteca e Documentação

Araújo, Elton Gean.
A663m Modelagem de dados de cana-de-açúcar: uma abordagem multivariada /
Elton Gean Araújo. -- Ilha Solteira: [s.n.], 2016
116 f. : il.

Tese (doutorado) - Universidade Estadual Paulista. Faculdade de Engenharia
de Ilha Solteira. Área de conhecimento: Sistema de Produção, 2016

Orientador: Alan Rodrigo Panosso
Co-orientador: Morel De Passos E Carvalho
Inclui bibliografia

1. Ciência do solo. 2. Modelagem . 3. Produção da cana-de-açúcar e
dinâmica de carbono do solo.

CERTIFICADO DE APROVAÇÃO

TÍTULO DA TESE: Análise Estatística de dados de cana-de-açúcar: uma abordagem multivariada

AUTOR: ELTON GEAN ARAUJO

ORIENTADOR: ALAN RODRIGO PANOSSO

COORDINADOR: MOREL DE PASSOS E CARVALHO

Aprovado como parte das exigências para obtenção do Título de Doutor em AGRONOMIA,
especialidade: SISTEMAS DE PRODUÇÃO pela Comissão Examinadora:


Prof. Dr. ALAN RODRIGO PANOSSO
Departamento de Matemática / Faculdade de Engenharia de Ilha Solteira


Prof. Dr. RAFAEL MONTANARI
Departamento de Fitossanidade, Engenharia Rural e Solos / Faculdade de Engenharia de Ilha Solteira


Prof. Dr. EDSON DONIZETE DE CARVALHO
Departamento de Matemática / Faculdade de Engenharia de Ilha Solteira


Prof. Dr. OSMAR JESUS MACEDO
Departamento de Ciências Exatas / Universidade Federal de Mato Grosso do Sul


Prof. Dr. NEWTON LA SCALA JUNIOR
Departamento de Ciências Exatas / Faculdade de Ciências Agrárias e Veterinárias de Jaboticabal

Ilha Solteira, 29 de novembro de 2016

DEDICATÓRIA

À Deus, pois jamais teria conseguido realizar este trabalho sem a sua graça.

À minha esposa Vânia Gomes Coelho Araújo.

Aos meus pais Luiz Araújo Sobrinho e Tereza Gropo Araújo.

AGRADECIMENTOS

À minha esposa Vânia Gomes Coelho Araújo pelo companheirismo, apoio e paciência.

Aos meus pais Luiz Araújo Sobrinho e Tereza Gropo Araújo pela base de tudo; aos meus irmãos pelo incentivo e amizade e aos demais familiares pelo apoio e confiança.

Ao Professor Alan Rodrigo Panosso, grande responsável pelo meu aprendizado profissional e moral, pelos valiosos ensinamentos, pela paciência, pelo companheirismo, pela amizade, ajuda nas publicações. Muito obrigado por este tempo edificante de convívio.

Ao Professor Morel de Passos e Carvalho, pelos ensinamentos, apoio e parceria na condução deste trabalho. Agradeço imensamente por fazer parte deste trabalho, no qual não mediu esforços para contribuir.

Aos Professores membros da banca de Qualificação, Marcelo Andreotti e Rafael Montanari, cujas críticas e sugestões, foram substanciais para o desenvolver da pesquisa, dada a melhoria proporcionada.

Aos Professores membros da banca de Defesa, Osmar Jesus Macedo, Edson Donizete Carvalho, Rafael Montanari e Newton La Scala Júnior, grandes mestres, que dispenderam seu precioso tempo para contribuir com este estudo, às críticas e sugestões, às quais foram importantes para fortalecer o trabalho desenvolvido e para o avanço profissional.

Ao Professor Mário Luiz Teixeira Moraes, pela confiança, conselhos, amizade e apoio. O meu sincero agradecimento.

Aos demais professores e amigos do Programa de Pós-Graduação em Agronomia, pelo tempo de convívio, experiência compartilhada e ensinamentos.

Aos colegas de Pós-Graduação, em especial Wanderley Cardoso e Adrianly Rodrigues Corrêa, pelas parcerias nas pesquisas desenvolvidas, bem como amizade e integridade.

Ao Programa de Pós-Graduação em Agronomia, por acreditar em meu potencial.

“Talvez não tenhamos conseguido fazer o melhor,
mas lutamos para que o melhor fosse feito. Não somos o
que deveríamos ser, não somos o que iremos ser... Mas
graças a Deus, não somos o que éramos.”

Martin Luther King

RESUMO

Neste trabalho, foram estudadas as interações entre os componentes de produção e tecnológicos da cana-de-açúcar, bem como da dinâmica de carbono no solo e os fatores latentes formados pelos atributos físicos e químicos do solo, por meio de modelos de análise de equações estruturais, com auxílio das técnicas estatísticas multivariadas análise dos componentes principais e análise fatorial. Ensaio experimentais foram conduzidos em diferentes localidades dos estados de São Paulo (Áreas com Argissolo Vermelho Distrófico abrupto textura média/argilosa, Latossolo Vermelho Eutroférrico textura muito argilosa e Latossolo Vermelho Eutroférrico textura argilosa) e Mato Grosso do Sul (Área com Latossolo Vermelho-Amarelo distrófico textura franco argilo arenosa). Na camada de 0 a 0,20 m do solo, os componentes de produção e tecnológicos da cana-de-açúcar foram influenciados de forma negativa e significativa apenas pelo fator latente atributos físicos (densidade do solo, porosidade total e macroporos do solo) na ordem de 2,6 (em módulo) mais do que o fator latente atributos químicos associados a acidez do solo (pH_{Ca}, pH_K e pH_{ag}), evidenciando compactação causada pelo tráfego de máquinas (cana crua em 3º corte). No entanto, na camada de 0,20 a 0,40 m, apenas o fator latente atributos químicos influenciou de maneira positiva e significativa os componentes de produção e tecnológicos da cultura, na ordem de aproximadamente 3 vezes mais que os atributos físicos, evidenciando o efeito benéfico da aplicação de gesso após o primeiro corte. Em relação à dinâmica de carbono, expressa neste trabalho pelas variações da emissão de CO₂ do solo e do estoque de carbono no solo (calculado para a camada de 0-0,10 m), estas foram principalmente influenciadas pelo fator latente atributos físicos (porosidade livre de água, densidade do solo e macroporos), na ordem de pelo menos 4 vezes (em módulo) mais do que o fator latente atributos químicos (teor de cálcio, teor de magnésio e capacidade de troca de cátions do solo), para as áreas de estudos localizadas no estado de São Paulo (municípios de Pradópolis e Guariba). Na área de Aparecida do Taboado-MS, constatou-se que as variações de emissão de CO₂ não sofreram influência significativa dos fatores latentes atributos químicos e físicos; porém, as variações dos valores de estoque carbono, foram explicadas de forma estatisticamente significativas por esses dois fatores, sendo os coeficientes estruturais em relação aos fatores latentes atributos físicos e atributos químicos, iguais a -0,34 e 0,28, respectivamente.

Palavras-chave: Biomassa. Modelagem. Rendimento. Sustentabilidade Agrícola.

ABSTRACT

In this work, were studied the interactions between production components and of technology of sugarcane as well as the carbon dynamics in soil and latent factors formed by the physical and chemical properties of the soil, through of structural equations analysis models, with the help of the statistical multivariate principal components analysis and factor analysis. Experimental trials were conducted at different locations in the states of São Paulo (Areas with Dystrophic Red Argisol, with medium-clay texture , Eutroferic Red Latosol very clayey texture and Eutroferic Red Latosol with clay texture) and Mato Grosso do Sul (Area with Dystrophic Red-Yellow Latosol, with sandy clay franc texture). In the layer 0 to 0.20 m above the ground, the components of production and technology of sugarcane were influenced negatively and significantly only by the latent factor physical attributes (bulk density, porosity and macropores), in the order of 2.6 (in module), more than latent factor attributes chemical associated with soil acidity (PHCA, PHK and pHag), showing compaction caused by machinery traffic (crude cane in 3rd cut). However, at the layer 0.20 to 0.40 m, only the latent factor chemical attributes it influenced significantly and positively the production components and technological of the culture, on the order of about 3 times more than the physical attributes, highlighting the beneficial effect of plaster application after the first cut. Regarding the carbon dynamics expressed in this work by the variations of soil CO₂ emission and the stock of carbon in the soil (calculated for the layer of 0-0.10 m), these were mainly influenced by the latent factor physical attributes (porosity water-free, bulk density and macropores) in the order of at least 4 times (in module) more than the latent factor chemical attributes (calcium, magnesium content and exchange capacity of soil cations), for areas of localized studies in São Paulo (municipalities of Pradópolis and Guariba). In area Aparecida of Taboado-MS, it was found that variations in CO₂ emission suffered no significant influence of the latent factors chemical and physical attributes; however, variations in carbon stock values were explained in statistically significant way by these two factors, being the structural coefficients in relation to the latent factors physical attributes and attributes chemical, equal to -0.34 and 0.28, respectively.

Keywords: Biomass. Modeling. Yield. Agricultural Sustainability.

LISTA DE FIGURAS

Figura 1 - Representação Gráfica do Modelo Completo.	40
Figura 2 - Etapas da Análise de Equações Estruturais.....	42
Figura 3 - Mapa esquemático da área experimental de 10,5 ha e distribuição de 118 pontos amostrais na área de cana-de-açúcar no município de Suzanópolis, SP, Brasil.	66
Figura 4 - Etapas da Modelagem de Equações Estruturais.....	69
Figura 5- Gráfico para verificação de valores discrepantes (outliers multivariados) para as profundidades: a) 0,00-0,20 m e b) 0,20-0,40 m.....	73
Figura 6 - Screeplots da análise de componentes principais para todos os atributos estudados nas profundidades: a) 0,00-0,20 m e b) 0,20-0,40 m.	73
Figura 7 - Análise de fatores apresentado as cargas fatoriais entre a variável e o fator. Os valores referência à porcentagem de variação do conjunto de dados originais retida pelo fator nas profundidades: a) 0,00-0, 20 m e b) 0,20-0,40 m.	74
Figura 8 - Diagrama de Caminhos da análise de equações estruturais para as profundidades: a) 0,00-0, 20 m e b) 0,20-0,40 m.	77
Figura 9. Etapas da Análise de Equações Estruturais.....	95
Figura 10. Screeplots da análise de componentes principais para todos os atributos estudados da área de Pradópolis – SP.....	99
Figura 11. Diagrama Estrutural dos Modelos Conceituais para FCO ₂ e estoque de Carbono, área Pradópolis – SP.	102
Figura 12. Diagrama de Caminhos da análise de equações estruturais, área Pradópolis – SP.....	102
Figura 13. Diagrama de Caminhos da análise de equações estruturais, área Guariba–SP, Latossolo Vermelho Eutroférico.....	105
Figura 14. Diagrama de Caminhos da análise de equações estruturais, área Ap. do Taboado, Latossolo Vermelho-Amarelo Distrófico.	106

LISTA DE TABELAS

Tabela 1 - Índices de qualidade de ajustamento e seus valores de referência	50
Tabela 2. Descrição morfológica e dados analíticos do solo estudado.....	65
Tabela 3 - Estatística descritiva para os atributos de planta e de solo avaliados no experimento, nas duas profundidades.....	71
Tabela 4 - Resumo do resultado de ajuste dos modelos Análises de Equações Estruturais.	78
Tabela 5. Estatísticas e índices de qualidade de ajustamento dos modelos da análise de componentes de produção e tecnológicos da cana-de-açúcar equações estruturais.	79
Tabela 6. Médias das emissões de CO ₂ , temperatura, umidade do solo e demais atributos do solo e seus respectivos coeficientes de variação, profundidade de 0-0,10 m, área de Pradópolis-SP.	99
Tabela 7. Coeficientes das três primeiras componentes principais e a variância explicada por cada uma delas.	100
Tabela 8. Resumo do resultado de ajuste dos modelos Análises de Equações Estruturais, área Pradópolis – SP, Latossolo Vermelho Eutroférico.	103
Tabela 9. Estatísticas e índices de qualidade de ajustamento dos modelos de AEE FCO ₂ e estoque de carbono, área de Pradópolis – SP.....	104
Tabela 10. Estatísticas e índices de qualidade de ajustamento dos modelos de AEE FCO ₂ e estoque de carbono, áreas de Guariba – SP e Aparecida do Taboado – MS.	107

SUMÁRIO

CAPÍTULO 1 – CONSIDERAÇÕES GERAIS	11
1.1 OBJETIVOS.....	13
1.1.1 OBJETIVOS GERAIS	13
1.2 REVISÃO DE LITERATURA.....	13
1.2.1 ESTATÍSTICA MULTIVARIADA	13
1.2.1.1 CONCEITOS IMPORTANTES DA ESTATÍSTICA MULTIVARIADA	14
1.2.2 ANÁLISE DOS COMPONENTES PRINCIPAIS	18
1.2.3 ANÁLISE FATORIAL.....	23
1.2.4 ANÁLISE DE EQUAÇÕES ESTRUTURAIS.....	35
CAPÍTULO 2 – MODELAGEM DAS RELAÇÕES ENTRE OS ATRIBUTOS DO SOLO E COMPONENTES DE PRODUÇÃO DA CANA-DE-AÇÚCAR USANDO ANÁLISE DE EQUAÇÕES ESTRUTURAIS.	60
2.1 - INTRODUÇÃO.....	62
2.2 - MATERIAL E MÉTODOS.....	64
2.3 - RESULTADOS E DISCUSSÃO.....	70
2.4 - CONCLUSÕES	80
REFERÊNCIAS.....	81
CAPÍTULO 3 - MODELAGEM DA DINÂMICA E ESTOQUE DO CARBONO E DE OUTROS ATRIBUTOS DO SOLO EM ÁREAS DE CANA-DE-AÇÚCAR	86
3.1 - INTRODUÇÃO.....	89
3.2 - MATERIAL E MÉTODOS.....	92
3.3 - RESULTADOS E DISCUSSÃO.....	97
3.4 - CONCLUSÕES	108
REFERÊNCIAS.....	109
CAPÍTULO 4 – CONSIDERAÇÕES FINAIS	115

CAPÍTULO 1 – CONSIDERAÇÕES GERAIS

A estatística é uma ciência constituída de um conjunto de métodos que visam auxiliar na exploração, bem como na confirmação de hipóteses de pesquisas em diversas áreas do conhecimento. A estatística multivariada consiste em um conjunto de métodos estatísticos utilizados em situações nas quais várias variáveis são oriundas do mesmo elemento amostral (acessos). Em geral, as variáveis são correlacionadas entre si e quanto maior o número de variáveis, mais complexa torna-se a análise por métodos comuns de estatística univariada (MINGOTI, 2005).

As técnicas multivariadas são classificadas como técnicas de dependência e interdependência. Em adição, recomenda-se que, se as variáveis dependentes e independentes fizerem parte da hipótese da pesquisa, recomenda-se a utilização de uma das técnicas de dependência, como por exemplo: regressão múltipla, análise discriminante ou regressão logística. Caso contrário, pode-se utilizar uma das técnicas de interdependência, como a análise fatorial, análise de conglomerados ou escalonamento multidimensional (COOPER; SCHINDLER, 2003; CORRAR et al. 2009).

Estudos sobre a qualidade do solo apresentam inúmeras variáveis, as quais são, geralmente, descritas por meio de análises estatísticas univariadas. Contudo, a interação destes atributos pode ser estudada pelos métodos estatísticos multivariados, que permitem o agrupamento dos acessos com características similares permitindo um melhor entendimento das variações dos processos que ocorrem no solo e possibilitando a delimitação de áreas de manejo específico para a cultura e atributos em questão (FIDALSKI et al., 2007; JOSÉ et al., 2012; ALMEIDA JUNIOR, 2012; ARAÚJO et al., 2013).

Estudos internacionais na área de ciência do solo por meio de técnicas multivariadas tem sido frequentes, como as análises de componentes principais e correlação canônica aliadas à krigagem ordinária, entre os atributos da produtividade vegetal com os do solo; análise fatorial e análise de equações estruturais, como poderosas ferramentas para estudar relações e inter-relações entre os atributos do solo (JIANG; THELEN, 2004; MARTÍN et al., 2005; BRAHIM et al., 2011).

Contudo, no Brasil, são poucos os trabalhos que empregam essas técnicas, uma vez que a grande maioria dos estudos nacionais tem empregado as análises de correlação linear e a técnica da cokrigagem ordinária (DALCHIAVON et al., 2011; MONTANARI et al., 2012). Segundo Hair et al. (2005), qualquer pesquisador que examine apenas as relações entre duas variáveis e evite análise multivariada estará ignorando poderosas ferramentas que podem

fornecer informações úteis para a melhor compreensão dos fenômenos complexos do ambiente das organizações.

Outro fator que dificulta a utilização de tais técnicas é a escolha do programa computacional para processar as análises, na sua maioria pagos, com código-fonte fechado. A reprodução de resultados, em diferentes configurações de softwares e hardwares, torna-se impossível sem a utilização de programas de código aberto (INCE et al., 2012; HANSON et al., 2011). Atualmente, o ambiente R, ou simplesmente R, é um software de domínio público utilizado para análise de dados em geral que, além de ser gratuito, apresenta código-fonte aberto, ou seja, pode ser modificado ou implementado com novos procedimentos (pacotes), desenvolvidos pelos usuários (PETERNELLI; MELLO, 2007), garantindo a reprodutibilidade dos resultados.

O Brasil já tem um dos programas de maior sucesso de biocombustíveis atualmente, incluindo co-geração de energia elétrica a partir da biomassa (BORDONAL, 2016). Além disso, é o principal produtor mundial de cana-de-açúcar, cuja área cultivada na safra 2016/2017 foi de 9.073,7 mil hectares e produção estimada em 690,98 milhões de toneladas (CONAB, 2016). A produção de cana é mais concentrada na região centro-sul do Brasil, respondendo por cerca de 90% do total da área cultivada em 2015. São Paulo é o estado com a maior expansão de plantação de cana-de-açúcar durante os últimos anos, segurando cerca de 60% do cultivo no centro-sul do Brasil (UNICA, 2015).

Diferentemente dos países industrializados, no Brasil a maior parcela das emissões líquidas estimadas de CO₂ é proveniente do uso da terra (BORDONAL, 2012). Sendo assim, e considerando o contexto das mudanças climáticas, o Brasil é um potencial mitigador do efeito estufa adicional pela adoção de práticas agrícolas sustentáveis. A colheita da cana-de-açúcar no sistema cana crua (mecanizada) favorece o acúmulo de matéria orgânica do solo e redução na emissão de gás, quando comparado com o sistema de cana queimada (RAZAFIMBELO et al., 2006). Logo, entender a relação entre os componentes da dinâmica do carbono do solo em áreas de cana crua e os atributos do solo é fundamental para implementação de práticas agrícolas que proporcionem o acúmulo de carbono no solo, sem contudo aumentar as emissões de CO₂. Nesse sentido, estudos têm sido realizados com este propósito, tais como Panosso et al. (2012) e Bicalho (2016).

A despeito de todos os esforços, são poucos os estudos realizados com essa cultura, a fim de correlacionar sua produtividade agrícola e dinâmica do carbono com atributos do solo, por meio de modelos mais complexos. Assim, o objetivo do trabalho foi identificar por meio de técnicas multivariadas os principais atributos físico-químicos do solo que melhor se

relacionam com os atributos produtivos e tecnológicos da cultura da cana-de-açúcar. Em adição, como sistemas de cultivo associados à produção de energia, como a cultura da cana-de-açúcar, podem capturar o CO₂ atmosférico na forma de biomassa (sequestro), retornando para o solo quantidades adicionais de carbono orgânico; e consequentemente atuar na mitigação das emissões de CO₂ do solo para atmosfera, em tais sistemas.

1.1 OBJETIVOS

1.1.1 OBJETIVOS GERAIS

O objetivo do trabalho foi a aplicação de métodos estatísticos multivariados, principalmente a análise de equações estruturais, em dados oriundos de áreas de cultivo de cana-de-açúcar para a melhor caracterização das variáveis de desempenho da cultura e à dinâmica do carbono expressa pela emissão de CO₂ e pelo estoque de Carbono do solo associados aos atributos do solo, por meio do programa estatístico R. Adicionalmente, viabilizar a técnica generalizada análise de equações estruturais, que tem sido pouco utilizada no meio agrônomo e tão poderosa nas áreas da ciências sociais, com intuito de fornecer uma poderosa ferramenta no auxílio da modelagem em ciência do solo.

1.2 REVISÃO DE LITERATURA

1.2.1 ESTATÍSTICA MULTIVARIADA

De acordo com Mingoti (2005), a Estatística Multivariada se divide em dois grupos: um primeiro, consistindo em técnicas exploratórias de sintetização (ou simplificação) da estrutura da variabilidade dos dados, e um segundo, consistindo em técnicas de inferência estatística. No primeiro grupo estão: análise de componentes principais, análise fatorial, análise das correlações canônicas, análise de agrupamentos, análise discriminante e análise de correspondência. No segundo, encontram-se: os métodos de estimação de parâmetros, os testes de hipóteses, as análises de variância e covariância e de regressão multivariadas.

A Estatística Multivariada pode ser aplicada com diversas finalidades, mesmo quando não se dispõe de antemão de um modelo teórico rigorosamente estruturado das relações entre as variáveis (BAKKE; LEITE; SILVA, 2008). A finalidade de sua aplicação pode ser de reduzir os dados ou de simplificar sua estrutura, de classificar ou agrupar acessos e variáveis, de

investigar a dependência entre tais variáveis e de predizer e elaborar hipóteses e testá-las (JOHNSON; WICHERN, 1992).

Apenas na Estatística Multivariada é possível explorar o desempenho conjunto das variáveis, bem como determinar a influência de cada uma delas, dado que todas as outras são consideradas. Isso mostra a relevância da análise multivariada nas diversas áreas do conhecimento, inclusive em experimentos agrônômicos, dado a importância e complexidade das ciências agrárias.

1.2.1.1 CONCEITOS IMPORTANTES DA ESTATÍSTICA MULTIVARIADA

Alguns conceitos importantes utilizados no contexto da estatística multivariada serão descritos a seguir.

- a. Vetor de Médias. Considerando o vetor X contendo p atributos, onde cada atributo é uma variável aleatória, isto é, $X^T = (X_1, X_2, X_3, \dots, X_p)$, denota-se o vetor μ como vetor de médias de X , representado da seguinte forma:

$$\mu = \begin{bmatrix} E(X_1) \\ E(X_2) \\ \vdots \\ E(X_p) \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix} \quad (1)$$

- b. Variância, Covariância e Correlação. A variância da i -ésima variável do vetor X é denotada por $Var(X) = \sigma_i^2 = \sigma_{ii}$; indica se os valores X_i estão próximos ou distantes da média μ_i , sendo que valores grandes indicam grande dispersão. A covariância entre os valores da i -ésima e j -ésima variáveis do vetor X é dada por $Cov(X_i, X_j) = \sigma_{ij} = E[(X_i - \mu_i)(X_j - \mu_j)]$, medindo o grau de relacionamento linear entre duas variáveis aleatórias. O coeficiente de correlação entre as i -ésima e j -ésima variáveis do vetor X é definido por:

$$\rho_{ij} = \frac{\sigma_{ij}}{\sigma_i \sigma_j} \quad (2)$$

$-1 \leq \sigma_{ij} \leq 1$, $i, j = 1, 2, \dots, p$; σ_i, σ_j são os desvios padrões das variáveis aleatórias X_i e X_j , respectivamente.

- c. Matrizes de Covariâncias e de Correlação. A matriz de Variâncias e Covariâncias do vetor aleatório X é denotada por:

$$\text{Cov}(X) = V(X) = \text{Var}(X) = \Sigma_{p \times p} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{bmatrix} \quad (3)$$

Os elementos diagonais da matriz contêm as variâncias das variáveis e os elementos fora da diagonal contêm as covariâncias entre todos os pares de variáveis. A matriz de covariâncias é uma matriz quadrada, simétrica e positiva semi-definida, sendo invertível se $a^T \Sigma a > 0$, $a_{p \times 1} \in \mathbb{R}^p$ e não nulo, e sua inversa é denotada por $\Sigma^{-1}_{p \times p}$.

A matriz de correlação do vetor aleatório é denotada por:

$$P_{p \times p} = \begin{bmatrix} 1 & \rho_{12} & \rho_{13} & \cdots & \rho_{1p} \\ \rho_{21} & 1 & \rho_{23} & \cdots & \rho_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{p1} & \rho_{p2} & \rho_{p3} & \cdots & 1 \end{bmatrix} \quad (4)$$

- d. Combinações Lineares. Seja z uma combinação linear de um vetor aleatório X , com vetor de médias μ e matriz de covariâncias $\Sigma_{p \times p}$, então:

$$z = a_1 X_1 + a_2 X_2 + \cdots + a_p X_p, \text{ em que,}$$

$$a_{p \times 1} \in \mathbb{R}^p \text{ um vetor de constantes conhecidas, } \mu_z = \sum_{i=1}^p a_i \mu_i \text{ e}$$

$$\sigma_z^2 = a^T \Sigma_{p \times p} a.$$

- e. Autovalores e Autovetores. Dado uma matriz de covariância $\Sigma_{p \times p}$, é possível encontrar os autovalores desta matriz resolvendo a equação característica:

$$|\Sigma_{p \times p} - \lambda I_{p \times p}| = 0, \text{ sendo } I_{p \times p} \text{ a matriz identidade de dimensão } p.$$

Para cada autovalor, encontra-se o autovetor v_{pxl} correspondente a equação característica anterior (MINGOTI, 2005):

$$\sum_{p \times p} v_{pxl} = \lambda v_{pxl} \quad (5)$$

- f. Forma Quadrática: Considere uma matriz de covariâncias $\sum_{p \times p}$ e x um vetor de comprimento p .

Então a quantidade:

$$Q = x^T \sum_{p \times p} x \quad (6)$$

é um escalar que é chamado de uma forma quadrática. Tal escalar também pode ser expresso por:

$$Q = \sum_{i=1}^p \sum_{j=1}^p x_i \sigma_{ij} x_j \quad (7)$$

Em que x_i é o elemento na i -ésima linha de x e σ_{ij} é a covariância entre a i -ésima variável e a j -ésima variável de $\sum_{p \times p}$.

- g. Teorema da decomposição espectral: Seja uma matriz de covariâncias $\sum_{p \times p}$. Então $\sum_{p \times p}$ pode ser expressa em termos de seus p -pares de autovalores-autovetores (λ_i, e_i) , como (MINGOTI, 2005):

$$\sum_{p \times p} = O \Lambda O^T = \sum_{i=1}^p \lambda_i e_i e_i^T, \text{ em que } e_i, e_i^T \text{ são os autovetores}$$

normalizados (divididos pelos seus comprimentos) e os vetores normalizados transpostos da matriz de covariância, respectivamente; Λ e O são as matrizes de autovalores e autovetores da matriz de covariância \sum , respectivamente.

- h. Distribuição Normal P-Variada: Dado o vetor X com p variáveis aleatórias, se X tem distribuição de probabilidade normal P-variada, denota-se $X \sim N(\mu, \sum_{p \times p})$, e a função densidade de X é dada por:

$$F(x_1, x_2, \dots, x_p) = \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} \exp \left\{ -\frac{1}{2} (x - \mu)^T \Sigma^{-1} (x - \mu) \right\} \quad (8)$$

Para todo vetor $X \in \mathfrak{R}^p$, onde $X = (X_1, X_2, \dots, X_p)^T$, $-\infty < x_i < \infty$, $i = 1, 2, \dots, p$, $\mu \in \mathfrak{R}^p$, $\Sigma_{p \times p}$ invertível.

A identificação da distribuição do vetor aleatório X é muito útil para se utilizar de certos testes estatísticos multivariados, bem como estimar parâmetros destas distribuições, como será mostrado em seções futuras deste trabalho.

- i. Testes para verificação de Normalidade Multivariada e detecção de *Outliers* Multivariados.

Um teste bastante utilizado para verificação de uma distribuição multivariada quanto a sua normalidade ou não é o de Shapiro-Wilk, que é uma generalização do teste de Shapiro-Wilk no caso univariado, realizada por Royston (1983). Considerando X_1, X_2, \dots, X_n uma amostra aleatória e k variáveis dessa amostra, Royston propõe a seguinte estatística W_k :

$$W_k = \frac{\left[\sum_{j=1}^n a_j X_{(j)k} \right]^2}{\sum_{j=1}^n (X_{jk} - \bar{X}_{.k})^2} \quad (9)$$

em que a_j são valores tabelados por Shapiro-Wilk.

O procedimento para aplicação do teste prevê a estimação da estatística W de Shapiro-Wilk para cada uma das variáveis, e a estatística final do teste é baseada na soma dos seus valores (CANTELMO; FERREIRA, 2007).

De acordo com Ferreira (2011), deve ser utilizada uma transformação na estatística do teste acima citado, sendo que esta nova estatística possui distribuição aproximada de qui-quadrado com ν graus de liberdade. Os graus de liberdade ν são denominados graus de liberdades equivalentes, sendo obtidos em função do número de variáveis e das correlações entre as variáveis (ROYSTON, 1983). O teste é realizado sob a hipótese nula de normalidade multivariada.

Vários outros testes são encontrados na literatura, tais como o de Mardia (1970), que é baseado nos coeficientes de curtose e assimetria da distribuição multivariada e como o teste proposto por Henze e Zirkley (1990), sendo mais consistente e invariante para grandes amostras. Dos métodos estatísticos multivariados a serem abordados neste trabalho, a análise

fatorial e a análise de equações estruturais são mais restritivas com relação à distribuição normal multivariada das variáveis observadas, sendo que caso o método de estimação escolhido for o da máxima verossimilhança, esta distribuição se faz necessária. Com relação à análise dos componentes principais, esta técnica não requer uma suposição de normalidade multivariada (SARTÓRIO, 2008; ANDRIOTTI, 2010), entretanto, caso o intuito não seja apenas exploratório e sim fazer inferências, Mingoti (2005) afirma ser necessária a normalidade multivariada. Uma alternativa a não-normalidade multivariada e *outliers* é a utilização da análise dos componentes principais robusta (NAGA; ANTILLE, 1990; HUBERT; ENGELEN, 2004).

Para a verificação de *outliers* multivariados, pode-se citar o gráfico *Q-Q Plot*, o qual consta em um gráfico bi-dimensional, sendo a abscissa correspondente aos valores ordenados da distância de Mahalanobis ao quadrado e a ordenada os valores dos percentis da distribuição qui-quadrado. Tal fato se justifica, pois, para uma distribuição normal multivariada, os valores da distância de Mahalanobis ao quadrado tem distribuição qui-quadrado central com p graus de liberdade, sendo p a dimensionalidade dos dados (VALADARES et al., 2012). Ainda, segundo o mesmo autor, pode-se definir um *outlier* multivariado, como sendo aquelas medidas que ultrapassam determinado quantil da distribuição qui-quadrado. Entretanto, apesar de ser usada para detecção de *outliers*, esta técnica é sensível aos próprios *outliers*. A razão é a sensibilidade da média aritmética e da matriz de covariância amostral para *outliers* (HAMPEL et al., 1986). Uma alternativa para este problema é utilizar a distância de Mahalanobis Robusta, que pode ser encontrada com mais detalhes em Filzmoser; Garrett; Reimann (2005).

Este trabalho é base para análise de dados na área agronômica, especificamente, na cultura da cana-de-açúcar, em que métodos de análise multivariada de interdependência serão aplicados para a melhor caracterização das variáveis de desempenho da cultura associados aos atributos do solo e à dinâmica do carbono expressa pela emissão de CO₂ do solo e estoque de Carbono do solo. A descrição de tais métodos multivariados será abordada nas seções seguintes, sendo eles: Análise dos Componentes Principais, Análise Fatorial e Modelagem de Equações Estruturais.

1.2.2 ANÁLISE DOS COMPONENTES PRINCIPAIS

Na análise de componentes principais (ACP), a dimensionalidade dos dados é reduzida com intuito de facilitar a interpretação dos dados, bem como as relações que existem entre as

variáveis por meio dos componentes principais. Tais componentes são combinações lineares das variáveis originais, em que não são correlacionadas entre si.

Para definir os componentes principais, seja $X = [X_1, X_2, \dots, X_p]^T$ um vetor aleatório com matriz de covariância Σ com autovalores $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$.

Dado um conjunto de dados com p variáveis, define-se os componentes principais como:

$$\begin{aligned} Y_1 &= e_1^T X = e_{11}X_1 + e_{12}X_2 + \dots + e_{1p}X_p \\ Y_2 &= e_2^T X = e_{21}X_1 + e_{22}X_2 + \dots + e_{2p}X_p \\ &\vdots \\ Y_p &= e_p^T X = e_{p1}X_1 + e_{p2}X_2 + \dots + e_{pp}X_p, \end{aligned} \quad (10)$$

Em que $e_i^T = (e_{i1}, e_{i2}, \dots, e_{ip})$ é um vetor de dimensão $1 \times p$ com $e_i^T \cdot e_i = 1$, e $e_i^T \cdot e_j = 0$ sendo $i \neq j$ para a última condição e i e j variando de 1 a p para as duas condições. Tais restrições serão explicadas mais adiante.

Reescrevendo o sistema (10) na forma matricial, tem-se:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_p \end{bmatrix} = \begin{bmatrix} e_{11} & e_{12} & \dots & e_{1p} \\ e_{21} & e_{22} & \dots & e_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ e_{p1} & e_{p2} & \dots & e_{pp} \end{bmatrix}^T \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix} \Rightarrow Y = O^T \cdot X \quad (11)$$

A ideia geométrica dessa técnica é realizar uma rotação rígida no sistema de eixos coordenados, em que os novos eixos estejam na direção de maior variabilidade (FERREIRA, 2011). Este procedimento funciona bem quando se tem grande número de variáveis, em que poucas componentes poderão explicar a variabilidade de todas as variáveis, facilitando assim o poder interpretativo das mesmas. Portanto, o objetivo do método é encontrar os vetores e_i que maximizam as variâncias das componentes, levando-se em consideração as restrições anteriores sobre estes vetores. Porém, quanto maior os valores de e_i , maior a variabilidade da componente, tendendo ao infinito, portanto, considera-se a normalização deste vetor. Então, considerando a primeira componente principal, o problema a ser resolvido é o seguinte:

$$\max_{e_1} [\text{var}(Y_1)] \quad (12)$$

Sujeito à $e_1^T \cdot e_1 = 1$.

Em adiç o, considerando as principais defini  es anteriores, especificamente a defini  o (d), permite-se escrever como vari ncia da primeira componente principal, j  que   uma combina  o linear, a seguinte equa  o:

$$\text{var}(Y_1) = e_1^T \Sigma e_1 \quad (13)$$

Desta forma, a express o (12) pode ser reescrito da seguinte forma:

$$\max_{e_1} \left[e_1^T \Sigma e_1 \right] \quad (14)$$

sujeito a $e_1^T \cdot e_1 = 1$

A resolu  o deste problema n o seria t o simples desta forma, pois deve-se maximizar uma fun  o sujeito a uma restri  o. Para facilitar este procedimento, utiliza-se neste caso a t cnica de multiplicadores de Lagrange, e a nova express o a ser maximizada  :

$$\max_{e_1} \left[e_1^T \Sigma e_1 + \lambda (1 - e_1^T e_1) \right] \quad (15)$$

em que λ   o multiplicador de Lagrange.

A solu  o da express o (15)   encontrada derivando-se esta fun  o em rela  o a e_1 e igualando a zero. Para isso, calcula-se o gradiente da respectiva fun  o com respeito ao vetor e_1 , que indicar  a dire  o de maior varia  o, e iguala-se a zero. O gradiente   encontrado baseando-se em defini  es de derivadas de matrizes de acordo com Petersen e Pedersen (2012). A resolu  o deste problema   dada como:

$$\begin{aligned} \nabla \left(e_1^T \Sigma e_1 + \lambda (1 - e_1^T e_1) \right) &= 0 \\ \Rightarrow \nabla \left(e_1^T \Sigma e_1 \right) + \nabla \left(\lambda (1 - e_1^T e_1) \right) &= 0 \\ \Rightarrow \nabla \left(e_1^T \Sigma e_1 \right) - \nabla \left(\lambda e_1^T e_1 \right) &= 0 \\ \Rightarrow \Sigma e_1 + \Sigma^T e_1 - (\lambda I e_1 + \lambda I^T e_1) &= 0 \\ \Rightarrow 2 \Sigma e_1 - 2 \lambda e_1 &= 0 \\ \Rightarrow \Sigma e_1 &= \lambda e_1 \end{aligned} \quad (16)$$

O que mostra que a varia  o da primeira componente est  na dire  o do autovetor normalizado e_1 da matriz de covari ncia dos dados originais, aumentado do respectivo autovalor λ .   poss vel mostrar por meio dos resultados anteriores, que a vari ncia desta componente   o pr prio autovalor λ :

$$\text{var}(Z_1) = e_1^T \Sigma e_1 = e_1^T \lambda e_1 = \lambda e_1^T e_1 = \lambda \quad (\text{Pode-se considerar } \lambda = \lambda_1). \quad (17)$$

Portanto, a direção de máxima variabilidade da primeira componente é obtida por meio do primeiro autovetor normalizado da matriz de covariância dos dados e, sua variância é o autovalor correspondente, que é o maior entre todos os autovalores, isto é, a primeira componente principal está na direção de maior variabilidade dos dados.

Com os cálculos das demais componentes, é possível mostrar que a direção de máxima variabilidade da p -ésima componente principal é obtida pelo p -ésimo autovetor da matriz de covariância dos dados e, sua variância será o p -ésimo autovalor λ_p , com valor menor que os $p-1$ autovalores anteriores. A covariância entre duas quaisquer componentes é nula, pois os autovetores são ortogonais entre si, como mostrado a seguir:

$$\text{cov}(Y_i, Y_j) = e_i^T \Sigma e_j = e_i^T \lambda_j e_j = \lambda_j e_i^T e_j = 0, \quad \forall i, j = 1, \dots, p \text{ e } i \neq j \quad (18)$$

Podemos concluir que a definição de componentes principais, é na verdade, a obtenção dos autovalores e autovetores da matriz de covariância dos dados. Os autovetores representam o sentido da rotação dos eixos coordenados definidos pelas variáveis originais e os autovalores, definem as variâncias desses novos eixos coordenados (FERREIRA, 2011). Considerando-se uma amostra com n observações, mensurada através de p variáveis, utiliza-se a matriz de covariância amostral $S_{p \times p}$ como estimador não viesado de $\Sigma_{p \times p}$, para obtenção dos autovalores e autovetores, para fins de estimação do modelo de análise dos componentes principais amostrais.

Desta maneira, no final do processo, tem-se p componentes ortogonais representando grande parte da variabilidade dos dados originais com p variáveis. O intuito é gerar um modelo mais parcimonioso, com $m < p$ componentes principais. Para se ter uma ideia de quantas componentes escolher, pode-se calcular a proporção de variabilidade explicada pela componente amostral i (PVE_i) ou proporção acumulada da variabilidade explicada pelas m primeiras componentes amostrais ($PAVE$), como segue:

$$PVE_i = \frac{\text{var}(\hat{Y}_i)}{\text{Variância Total}} = \frac{\hat{\lambda}_i}{\text{tr}(S_{p \times p})} = \frac{\hat{\lambda}_i}{\sum_{i=1}^p \hat{\lambda}_i} \quad (19)$$

$$PAVE = \frac{\sum_{j=1}^m \hat{\lambda}_j}{\sum_{i=1}^p \hat{\lambda}_i} \quad (20)$$

A variância total é representada pela soma dos autovalores da matriz de covariância amostral $S_{p \times p}$, que é estimador não viesado da matriz de covariância populacional $\Sigma_{p \times p}$, em que:

$$tr(\Sigma) = tr(O\Lambda O^T) = tr(\Lambda O^T O) = tr(\Lambda I) = tr(\Lambda) = \sum_{i=1}^p \lambda_i \quad (21)$$

Onde Λ e O são as matrizes de autovalores e autovetores da matriz de covariância populacional Σ , respectivamente. De onde, tem-se que:

$$\sum_{i=1}^p \sigma_{ii} = \sum_{i=1}^p \lambda_i, \text{ pois } tr(\Sigma) = \sum_{i=1}^p \sigma_{ii},$$

E, portanto,

$$tr(S_{p \times p}) = \sum_{i=1}^p \hat{\lambda}_i \quad (22)$$

De acordo com Mingoti (2005), a correlação estimada entre a j -ésima componente principal amostral e a variável aleatória X_i , $i = 1, 2, \dots, p$ é dada por:

$$r_{Y_j, X_i}^{\wedge} = \frac{e_{ji}^{\wedge} \sqrt{\lambda_j}}{\sqrt{s_{ii}}} \quad (23)$$

Em que s_{ii} é a variância amostral da variável X_i .

Um critério para a escolha do número de componentes principais considerado na interpretação dos resultados foi proposto por Kaiser (1958), e consiste em levar em conta no sistema apenas as componentes relacionadas àqueles autovalores maiores que a unidade (1), isso quando as variáveis originais foram padronizadas. Pois o intuito é manter as combinações lineares que expliquem pelo menos a variância de uma variável padronizada. Uma outra ferramenta auxiliar na escolha do número de componentes principais é o gráfico scree-plot, que é plotado considerando-se a ordem do autovalor na abscissa e seu respectivo valor na ordenada (Cattell, 1966). O número de componentes principais é escolhido de acordo com o ponto de inflexão da curva gerada, ou seja, onde os valores dos autovalores tendem a se estabilizar.

Escolhidas as m primeiras componentes principais mais relevantes, é necessário estimar seus valores para cada elemento amostral, se o objetivo da pesquisa é fazer uso dos componentes principais ao invés das variáveis originais. Caso as variáveis originais forem padronizadas, a matriz de covariância amostral se reduzirá à matriz de correlação amostral e os procedimentos serão semelhantes.

Para detecção de possíveis *outliers* multivariados, utilizam-se os escores das últimas componentes principais para confecção de gráficos de dispersão bi e tridimensional e *Q-Q Plots* (SARTÓRIO, 2008). Os possíveis outliers serão aqueles pontos que destoarem da nuvem de pontos no gráfico de dispersão. Por comparação, é recomendado refazer a análise sem os possíveis pontos e verificar se houve melhoria no ajuste.

Por fim, de acordo com Mingoti (2005), se o vetor aleatório X tem distribuição normal multivariada, é possível construir intervalos de confiança e testes de hipóteses para os autovalores populacionais, bem como fazer inferência sobre a matriz de covariâncias ou correlações populacionais.

1.2.3 ANÁLISE FATORIAL

A análise fatorial (AF) deriva um conjunto de dimensões (componentes principais) que, são capazes de descrever os dados em uma estrutura menor de dimensões que as variáveis originais individuais (HAIR, 2005); e foi desenvolvida inicialmente por Spearman (1904). Neste sentido, a análise fatorial tem como objetivo principal explicar a correlação ou covariância, entre um conjunto de variáveis, em termos limitados de variáveis não observáveis, denominados fatores. Tal proposta é semelhante à ACP, contudo a explicação total das covariâncias ou de correlações na ACP é feita utilizando todos os componentes. Na análise fatorial, quando for possível modelar, pode-se explicar todas as covariâncias ou correlações, utilizando-se algumas poucas variáveis não observáveis ou fatores comuns (FERREIRA, 2011). As técnicas de ACP e AF podem ser usadas para explicação de estruturas de variância e covariância de dados de diversas áreas, tais como biologia, zootecnia, agronomia, etc. As técnicas podem ser usadas para o agrupamento de variáveis chamados fatores; adicionalmente, estes fatores são de extrema importância para sintetizar os dados originais e estudos posteriores.

Então, a análise dos componentes principais busca obter combinações lineares das variáveis originais que expliquem o máximo da variabilidade dos dados, por outro lado, a AF preocupa-se com a explicação da estrutura de covariância dos dados, em função de fatores comuns.

Seja $X_{p \times 1}$ um vetor aleatório com vetor de médias μ , matriz de variância e covariância $\Sigma_{p \times p}$ e um vetor de fatores comuns latentes não observáveis $F_{m \times 1}$; de acordo com Mingoti (2005) e Hair et al. (2005), o modelo de análise fatorial pode ser descrito desta forma:

e,

$$\begin{aligned}
 \text{cov}(X) &= E[(X - \mu)(X - \mu)^T] = E[(LF + \varepsilon)(LF + \varepsilon)^T] \\
 &= E[(LFF^T L^T + LF\varepsilon^T + \varepsilon F^T L^T + \varepsilon\varepsilon^T)] \\
 &= E[LFF^T L^T] + E[LF\varepsilon^T] + E[\varepsilon F^T L^T] + E[\varepsilon\varepsilon^T] \\
 &= LE[FF^T]L^T + LE[F\varepsilon^T] + E[\varepsilon F^T]L^T + E[\varepsilon\varepsilon^T] \\
 &= LIL^T + L0 + 0L^T + \psi
 \end{aligned} \tag{27}$$

Portanto, a covariância do vetor aleatório X é:

$$\text{cov}(X) = LL^T + \psi \tag{28}$$

O objetivo da análise fatorial é obter as matrizes $L_{p \times m}$ e $\psi_{p \times p}$ que possa representar a matriz de covariância $\Sigma_{p \times p}$ para um dado valor de m , menor que o número de variáveis originais p .

Desenvolvendo a equação (28) na forma matricial, tem-se:

$$\begin{aligned}
 \text{cov}(X) &= \begin{bmatrix} l_{11} & l_{12} & \cdots & l_{1m} \\ l_{21} & l_{22} & \cdots & l_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ l_{p1} & l_{p2} & \cdots & l_{pm} \end{bmatrix} \begin{bmatrix} l_{11} & l_{21} & \cdots & l_{p1} \\ l_{12} & l_{22} & \cdots & l_{p2} \\ \vdots & \vdots & \ddots & \vdots \\ l_{1m} & l_{2m} & \cdots & l_{pm} \end{bmatrix}^T + \begin{bmatrix} \psi_{11} & 0 & \cdots & 0 \\ 0 & \psi_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \psi_{pp} \end{bmatrix} \\
 \Rightarrow \text{cov}(X) &= \begin{bmatrix} \sum_{k=1}^m l_{1k}^2 & \sum_{k=1}^m l_{1k}l_{2k} & \cdots & \sum_{k=1}^m l_{1k}l_{pk} \\ \sum_{k=1}^m l_{2k}l_{1k} & \sum_{k=1}^m l_{2k}^2 & \cdots & \sum_{k=1}^m l_{2k}l_{pk} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{k=1}^m l_{pk}l_{1k} & \sum_{k=1}^m l_{pk}l_{2k} & \cdots & \sum_{k=1}^m l_{pk}^2 \end{bmatrix} + \begin{bmatrix} \psi_{11} & 0 & \cdots & 0 \\ 0 & \psi_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \psi_{pp} \end{bmatrix}
 \end{aligned} \tag{29}$$

Logo, pode-se escrever:

$$\begin{aligned}
 \text{var}(X_1) &= \sigma_{11} = \sum_{k=1}^m l_{1k}^2 + \psi_{11} \\
 \text{var}(X_2) &= \sigma_{22} = \sum_{k=1}^m l_{2k}^2 + \psi_{22} \\
 &\vdots \\
 \text{var}(X_j) &= \sigma_{jj} = \sum_{k=1}^m l_{jk}^2 + \psi_{jj}
 \end{aligned} \tag{30}$$

Caso as variáveis tiverem sido padronizadas antes da análise, então suas respectivas variâncias serão iguais a unidade, e, desta forma:

$$\text{var}(X_j) = \sigma_{jj} = \sum_{k=1}^m l_{jk}^2 + \psi_{jj} = 1 \tag{31}$$

O que mostra que a variância de X_j pode ser decomposta em duas partes. A primeira denotada por $\sum_{k=1}^m l_{jk}^2$, é a variância de X_j explicada pelos m fatores incluídos no modelo fatorial, chamado de comunalidade. A segunda parte, denotada por ψ_{jj} é a parte da variabilidade de X_j devido ao erro aleatório ε_j , o qual é específica de cada variável X_j , essa parte é chamada de unicidade ou variância específica.

Decorre ainda que:

$$\begin{aligned} \text{cov}(X, F) &= E[(X - E(X))(F - E(F))^T] = E[(X - \mu)F^T] \\ &= E[(LF + \varepsilon)F^T] = E[LFF^T + \varepsilon F^T] = LE[FF^T] + E[\varepsilon F^T] \\ &= LI + 0 \end{aligned} \quad (32)$$

E, portanto,

$$\text{cov}(X, F) = L \quad (33)$$

Assim, pode-se utilizar a matriz $L_{p \times m}$ na busca do entendimento e interpretação dos fatores F_j , $j = 1, 2, \dots, m$.

Deve-se, primeiramente, com vias de estimar o modelo fatorial, estimar a matriz de covariância populacional pela matriz de covariância amostral $S_{p \times p}$. Após isso, encontra-se os autovalores da matriz $S_{p \times p}$ que auxiliaram na seleção do número de fatores m . Os critérios para selecionar m são semelhantes aos critérios de seleção do número de componentes principais: análise da proporção da variância total, sendo m o número de autovalores retidos que representam maior proporção de variância explicada; critério de Kaiser (1958), sendo m o número de autovalores maiores que uma unidade e pelo gráfico scree plot de acordo com Cattell (1966), sendo m o valor da abcissa correspondente ao ponto de inflexão da curva plotada.

Pode-se citar ainda o método da análise paralela, inicialmente proposto por Horn (1965), que segundo Laros (2004) consiste em um procedimento estatístico de simulação de Monte-Carlo, em que são geradas matrizes de correlações de mesma dimensão da matriz original dos dados amostrais; os valores médios dos autovalores gerados são comparados de forma pareada com os autovalores reais para extração dos fatores.

Serão descritos a seguir os métodos de estimação do modelo de análise fatorial: método dos componentes principais, dos fatores principais e iterativo dos fatores principais e o método da máxima verossimilhança, respectivamente.

O método de estimação dos componentes principais consisti em estimar as matrizes L e ψ por meio das matrizes de autovalores e autovetores da matriz de covariância populacional $\Sigma_{p \times p}$. Do modelo de análise fatorial, tem-se a expressão (26) já descrita anteriormente:

$$\Sigma = LL^T + \psi$$

A partir de uma amostra aleatória de tamanho n estima-se Σ utilizando-se a matriz de covariância amostral $S_{p \times p}$, que é um estimador não viesado de Σ , como já dito anteriormente. Logo, tem-se o seguinte modelo estimado:

$$S = \hat{L}\hat{L}^T + \hat{\psi} \quad (34)$$

A matriz S pode ser decomposta com base no teorema da decomposição espectral, visto na subseção 1.2.1.1 deste trabalho, da seguinte forma:

$$S = \sum_{i=1}^p \lambda_i \hat{e}_i \hat{e}_i^T = \sum_{i=1}^m \lambda_i \hat{e}_i \hat{e}_i^T + \sum_{i=m+1}^p \lambda_i \hat{e}_i \hat{e}_i^T \quad (35)$$

Desta forma, considerando m fatores ($m < p$), a primeira parte da soma é uma aproximação para a matriz LL^T (MINGOTI, 2005). Logo:

$$\begin{aligned} \hat{L}\hat{L}^T &= \sum_{i=1}^m \lambda_i \hat{e}_i \hat{e}_i^T = \begin{bmatrix} \sqrt{\hat{\lambda}_1} \hat{e}_1 & \sqrt{\hat{\lambda}_2} \hat{e}_2 & \cdots & \sqrt{\hat{\lambda}_m} \hat{e}_m \end{bmatrix} \begin{bmatrix} \sqrt{\hat{\lambda}_1} \hat{e}_1 & \sqrt{\hat{\lambda}_2} \hat{e}_2 & \cdots & \sqrt{\hat{\lambda}_m} \hat{e}_m \end{bmatrix}^T \\ &= \hat{P} \hat{\Lambda}^{\frac{1}{2}} \left(\hat{P} \hat{\Lambda}^{\frac{1}{2}} \right)^T \end{aligned}$$

Sendo \hat{P} uma matriz $p \times m$ formada pelos primeiros m autovetores de S em suas colunas e $\hat{\Lambda}^{\frac{1}{2}}$ uma matriz diagonal $m \times m$ composta pela raiz quadrada dos autovalores de S .

A matriz $\hat{\psi}$ é uma matriz diagonal $p \times p$, como descrito anteriormente, e, de acordo com a equação (34), pode ser obtida da seguinte forma:

$$\hat{\psi} = \text{diag} \left(S - \hat{L}\hat{L}^T \right) = \begin{bmatrix} S_{11} - \sum_{j=1}^m \hat{l}_{1j}^2 & 0 & \cdots & 0 \\ 0 & S_{22} - \sum_{j=1}^m \hat{l}_{2j}^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & S_{pp} - \sum_{j=1}^m \hat{l}_{pj}^2 \end{bmatrix} \quad (36)$$

Assim, por meio da decomposição espectral de S foi possível obter os estimadores de L e ψ .

A matriz residual que mede a qualidade do ajuste do modelo fatorial é:

$$MRES = S - \hat{L}\hat{L}^T - \hat{\psi} \quad (37)$$

A matriz residual possui diagonal nula e o ideal, portanto, é que os valores fora da diagonal também fossem próximos de zero, o que ocorreria só se $m = p$, o que na prática não é viável, pois espera-se que o modelo fatorial seja parcimonioso, isto é, a estrutura de covariância dos dados estimadas com um número menor de fatores ($m < p$).

Há outros critérios para se avaliar a qualidade de ajuste do modelo fatorial, que, de acordo com Mingoti (2005), são definidos da seguinte forma:

EQM = média dos valores que estão acima da diagonal principal de $MRES$;

$REQM$ = raiz quadrada da média dos valores, ao quadrado, que estão acima da diagonal principal de $MRES$.

Um outro modo de avaliar a qualidade de ajuste tomando como base a matriz $MRES$, é descrito em Ferreira (2011) como:

$$tr(MRES^2) \leq \sum_{i=m+1}^p \hat{\lambda}_i^2 \quad (38)$$

em que $tr(MRES^2)$ é a soma de quadrado de resíduos, isto é, a soma de quadrados dos valores de $MRES$, e, $\sum_{i=m+1}^p \hat{\lambda}_i^2$ é a soma de quadrado dos autovalores que não foram retidos para análise.

A i -ésima comunalidade é estimada da seguinte forma:

$$\hat{h}_i^2 = \sum_{j=1}^m \hat{l}_{ij}^2 \quad (39)$$

A variância da i -ésima variável é o valor da i -ésima comunalidade acrescida do termo de erro ou variância específica, estimada por:

$$S_{ii} = \hat{h}_i^2 + \hat{\psi}_i = \sum_{j=1}^m \hat{l}_{ij}^2 + \hat{\psi}_i \quad (40)$$

E, por fim, a variância explicada pelo j -ésimo fator VF_j , em relação a variância total, é estimada por:

$$\begin{aligned} VF_j &= \sum_{i=1}^p \hat{l}_{ij}^2 = \left(\sqrt{\hat{\lambda}_j} \hat{e}_{1j} \right)^2 + \left(\sqrt{\hat{\lambda}_j} \hat{e}_{2j} \right)^2 + \cdots + \left(\sqrt{\hat{\lambda}_j} \hat{e}_{pj} \right)^2 = \\ &\hat{\lambda}_j \left(\hat{e}_{1j}^2 + \hat{e}_{2j}^2 + \cdots + \hat{e}_{pj}^2 \right) = \hat{\lambda}_j, \end{aligned} \quad (41)$$

pois $\hat{e}_{ij}^T \hat{e}_{ij} = 1$, para $i = 1, 2, \dots, p$.

Assim, a proporção da variância total explicada pelo j -ésimo fator PVF_j é:

$$PVF_j = \frac{\hat{\lambda}_j}{tr(S)} \quad (42)$$

No método de estimação dos componentes principais, a matriz de covariância é reproduzida de forma satisfatória na diagonal principal, ou seja, as variâncias das variáveis em estudo, o que não ocorre com as covariâncias, pois a matriz de erro amostral é diagonal. No método visto a seguir, este problema é minimizado.

Finalmente, é conveniente salientar que ao se utilizar as variáveis padronizadas, deve-se modelar a matriz de correlações amostrais R . Isso é útil para o método de estimação, principalmente quando as escalas das variáveis são muito diferentes. As estimativas das cargas fatoriais e das variâncias específicas podem ser obtidas pelos autovalores e autovetores de R (FERREIRA, 2011).

O métodos de estimação dos fatores principais e, iterativo dos fatores principais, consideram uma estimativa inicial da matriz ψ , fato que não acontece no método dos componentes principais. Nestes métodos, o princípio é a minimização da soma de quadrado dos elementos de $S - \Sigma$, isto é, minimizar a quantidade Q , dada por:

$$Q = tr[(S - \Sigma)^2] = tr[(S - LL^T - \psi)^2] \quad (43)$$

Derivando a função objetivo (43) em relação a L e ψ e, igualando a expressão resultante a zero, obtêm-se o seguinte sistema de equações normais:

$$\left(S - \hat{L}\hat{L}^T - \hat{\psi} \right) \hat{L} = 0 \quad (44)$$

$$diag\left(S - \hat{L}\hat{L}^T \right) - \hat{\psi} = 0 \quad (45)$$

Seja $\hat{\psi}_0$ uma estimativa de ψ e $S_r = S - \hat{\psi}_0$, então a equação (44) pode ser reescrita e simplificada da seguinte forma:

$$\begin{aligned} \left(S - \hat{\psi}_0 \right) \hat{L} - \hat{L}\hat{L}^T \hat{L} &= 0 \\ \left(S - \hat{\psi}_0 \right) \hat{L} &= \hat{L}\hat{L}^T \hat{L} \\ S_r &= \hat{L}\hat{L}^T \end{aligned} \quad (46)$$

Utilizando a decomposição espectral para a matriz S_r pode-se estimar a matriz \hat{L} da seguinte forma:

$$\hat{L} = \hat{P}_{p \times m} \hat{\Lambda}^{\frac{1}{2}}_{p \times m} \quad (47)$$

O i -ésimo valor da diagonal de S_r é a comunalidade associada a i -ésima variável, $i=1,2,\dots,p$. De acordo com Johnson e Wichern (2007), uma forma bastante utilizada para estimativa inicial da matriz $\hat{\psi}_0$ é:

$$\hat{\psi}_0 = [\text{diag}(S^{-1})]^{-1} \quad (48)$$

Portanto, para aplicação do método dos fatores principais deve-se obter primeiramente a matriz de covariâncias amostrais dos dados originais $S_{p \times p}$, logo após a matriz diagonal $\hat{\psi}_0$ de ordem $p \times p$, em seguida calcular a matriz $S_r = S - \hat{\psi}_0$. Por fim, estima-se a matriz \hat{L}_0 , com base na decomposição espectral de S_r . Uma nova estimativa para a matriz de variâncias específicas $\hat{\psi}_1$ pode ser obtida por:

$$\hat{\psi}_1 = \text{diag}\left(S - \hat{L}_0 \hat{L}_0^T\right) \quad (49)$$

Para aplicação do método iterativo dos fatores principais, recalcula-se a matriz S_r utilizando-se $\hat{\psi}_1$ e obtêm-se uma nova matriz de cargas fatoriais \hat{L}_1 por decomposição espectral, e, uma nova estimativa para $\hat{\psi}$, no caso, $\hat{\psi}_2$. O processo é repetido até que em um dado estágio as estimativas das comunalidades e das variâncias específicas converjam, considerando-se uma precisão pré-fixada (FERREIRA, 2011). Os procedimentos de estimação são semelhantes caso a matriz de correlação amostral R seja utilizada ao invés da matriz de covariância amostral S .

Existem outros procedimentos de estimação do modelo de análise fatorial, como o da Máxima Verossimilhança, que exige que a matriz de dados amostrais tenha distribuição normal multivariada. O método da máxima verossimilhança para estimação das matrizes L e ψ , introduzido por Lawley (1940), exige algumas suposições além das admitidas anteriormente, tais como que os vetores aleatórios F e ε sejam normalmente distribuídos. Com base no modelo (26) e todas essas suposições, tem-se que o vetor aleatório das variáveis observáveis X também tem distribuição normal com média 0 e matriz de variância e covariância Σ .

Considerando uma amostra aleatória de tamanho n , supondo que o vetor aleatório das variáveis observáveis X tem distribuição normal e a equação (26), o intuito deste método é maximizar a função de verossimilhança de X , dada por:

$$LV(X; \mu, L, \Psi) = \frac{1}{(2\pi)^{\frac{np}{2}} |LL^T + \Psi|^{\frac{n}{2}}} \exp \left\{ -\frac{1}{2} \text{tr} \left[(n-1)(LL^T + \Psi)^{-1} S \right] - \frac{1}{2} \text{tr} \left[n(LL^T + \Psi)^{-1} (\bar{X} - \mu)(\bar{X} - \mu)^T \right] \right\} \quad (50)$$

Por comodidade, calcula-se o logaritmo natural da função de verossimilhança, deriva-se a função resultante em relação a L e Ψ , obtendo-se o seguinte sistemas de equações normais:

$$\hat{L} \left(I + \hat{L}^T \hat{\Psi}^{-1} \hat{L} \right) = S \hat{\Psi}^{-1} \hat{L} \quad (51)$$

$$\hat{\Psi} = \text{diag} \left(S - \hat{L} \hat{L}^T \right) \quad (52)$$

De acordo com Ferreira (2011), a maximização da função de verossimilhança deve atender a condição de unicidade em que

$$\hat{L}^T \hat{\Psi}^{-1} \hat{L} = \Delta \quad (53)$$

tem de ser matriz diagonal. O sistema de equações normais requer solução numérica, de acordo com Lawley (1958).

No entanto, métodos iterativos tem sido propostos e melhorados para resolução do sistema de equações normais anterior; Ferreira (2011) cita procedimentos desenvolvidos por Maxwell (1964) e, recentemente, Hayashi e Bentler (2000). Caso as variáveis observadas estejam padronizadas, utiliza-se a matriz de correlação amostral R ao invés da matriz de covariância amostral S , sendo o procedimento de estimação realizado de maneira semelhante.

O método da máxima verossimilhança é um método mais sofisticado que os métodos de componentes e fatores principais e produz estimativas mais precisas, no entanto é fundamentado a suposição normalidade de X , F e ε , e, também a especificação do número de fatores m . Para qualidade de ajuste da estimação pode-se recorrer a matriz residual, como definida nos métodos de estimação anteriores, ou ainda, como vantagem, utilizar o teste de razão de máxima verossimilhança, sob a hipótese nula de que $\Sigma = (LL^T + \Psi)$, conforme Lawley (1940).

A estimação em análise fatorial é realizada sobre as matrizes de covariância e correlação amostral, no entanto, se os dados sofrem desvios de normalidade multivariada e possuem *outliers*, a matriz de correlação amostral estimada pode ser severamente tendenciosa (PISON et al., 2003; TREIBLMAIER; FILZMOSE, 2010). Uma solução para este problema, é a utilização de métodos robustos para estimar a matriz de covariância ou correlação amostral. Um procedimento (algoritmo) comumente utilizado é o Método do Determinante Mínimo – MCD (*Minimum Covariance Determinant*).

De acordo com Valadares (2012), o objetivo deste algoritmo é encontrar um subconjunto h de observações que torne o determinante da matriz de covariância clássica mínimo; sendo o estimador de localização a média destes h pontos, enquanto o estimador de escala sua matriz de covariância. Ainda segundo o mesmo autor, o ponto de ruptura aproximado do MCD é $(n - h)/n$, que considerando o subconjunto de observações $h \approx 0,75n$, seu valor correspondente é de 25%, sendo n o número de observações. De acordo com Treiblmaier e Filzmoser (2010), este procedimento torna a estimação em análise fatorial mais robusto e estável que utilizando a matriz de correlação clássica na violação dos pressupostos desta técnica.

De acordo com Ferreira (2011), uma das características da análise fatorial é a possibilidade de se ter diversas soluções capazes de obter a mesma matriz de covariâncias $\Sigma_{p \times p}$. Quando o número de fatores $m > 1$, há sempre alguma ambiguidade com o modelo fatorial (JOHNSON; WICHERN, 2007). Ainda, segundo os mesmos autores, esta característica é verificada tomando uma matriz ortogonal T , e, dado que $T^{-1} = T^T$ tem-se $TT^T = T^T T = I$. Considerando a expressão do modelo fatorial pode-se verificar que:

$$X - \mu = LF + \varepsilon = LTT^T F + \varepsilon = L^* F^* + \varepsilon \quad (54)$$

em que,

$$L^* = LT \text{ e } F^* = T^T F$$

Assim, mostra-se que:

$$E(F^*) = T^T E(F) = 0 \text{ e } \text{cov}(F^*) = T^T \text{cov}(F)T = T^T T = I_{m \times m}$$

Além disso:

$$\Sigma = L^* L^{*T} + \psi = LTT^T L^T + \psi = LL^T + \psi \quad (55)$$

Portanto, as novas cargas fatoriais reproduzem a matriz de Σ do mesmo modo que a matriz das cargas fatoriais L . As propriedades estatísticas do modelo não são alteradas com o novo fator F^* . A escolha de T é direcionada por critérios que levam em conta a maior

facilidade de interpretação dos novos fatores (FERREIRA, 2011); pode-se citar o método de Kaiser (1958), conhecido como critério *varimax*.

De acordo com Ferreira (2011), uma outra categoria de rotação refere-se às rotações oblíquas, que são de simples interpretação em função de possibilitar que os eixos fatoriais movam de forma diferencial uns dos outros, neste caso permitindo que os fatores sejam correlacionados. Uma interessante transformação oblíqua, resultante de uma modificação do método *varimax*, tem seu uso recomendado. Sendo tal transformação denominada *oblimin* e corresponde a minimização da covariância dos quadrados das cargas entre as colunas da matriz de cargas fatoriais. De acordo com Malhotra (2011), deve-se usar rotação oblíqua quando os fatores na população tendem a estar bem correlacionados.

Por fim, há alguns procedimentos para verificar se a estrutura de variâncias e covariâncias ou correlações foram reproduzidas adequadamente pelo modelo fatorial, principalmente no que diz respeito a matriz estimada $\hat{L}_{p \times m}$. Jonhson e Wichern (2002), sugerem, entre outras coisas, verificar as mudanças na interpretação dos fatores por meio do método dos componentes principais sem e com rotação por *varimax*. Os mesmos autores, aconselham ainda, a utilização de mais de um método de estimação do modelo, como por exemplo, a dos componentes principais e fatores principais para comparação das soluções obtidas. Além disso, sugerem utilizar valores diferentes para m e avaliar os resultados. Existem outros procedimentos, citados a seguir.

A medida ou coeficiente desenvolvido por Kaiser (1970), é um critério de ajuste do modelo fatorial definido por:

$$KMO = \frac{\sum_{i \neq j} R_{ij}^2}{\sum_{i \neq j} R_{ij}^2 + \sum_{i \neq j} Q_{ij}^2} \quad (56)$$

Os termos R_{ij} e Q_{ij} representam as matrizes de correlação amostral e correlação parcial amostral entre X_i e X_j , respectivamente. De acordo com Johnson e Wichern (2002), a correlação parcial é a correlação que existe entre duas variáveis, quando todas as outras variáveis são consideradas constantes. Rencher (2002) sugere que para um modelo de análise fatorial seja adequadamente ajustado aos dados amostrais, a inversa da matriz de correlações amostrais seja uma matriz diagonal. O critério KMO leva em consideração tal princípio e, quanto mais próximo de 1, maior a adequação do modelo de análise fatorial aos dados amostrais. De acordo com Corrar et al. (2009), se o KMO for menor que 0,50, os fatores

encontrados na análise fatorial não são capazes de descrever, satisfatoriamente, as variações dos dados amostrais.

Quando os dados são provenientes de uma distribuição normal p -variada, é possível fazer o teste de hipótese esfericidade de Bartlett para verificar se a matriz de correlação populacional é próxima da matriz identidade. Considerando a matriz identidade $I_{p \times p}$ e a matriz de correlação teórica das p -variáveis $P_{p \times p}$, as hipóteses a serem testadas no teste de esfericidade de Bartlett são, de acordo com Mingoti (2005):

$$\begin{aligned} H_0 : P_{p \times p} &= I_{p \times p} \\ H_1 : P_{p \times p} &\neq I_{p \times p} \end{aligned} \quad (57)$$

A estatística T do teste é definida como:

$$T = - \left[n - \frac{1}{6}(2p + 11) \right] \left[\sum_{j=1}^p \ln(\hat{\lambda}_j) \right] \quad (58)$$

em que $\hat{\lambda}_i$, $i = 1, 2, \dots, p$ são os autovalores da matriz de correlação amostral $R_{p \times p}$. Para que o modelo de análise fatorial possa ser ajustado, o teste de esfericidade de Bartlett deve rejeitar a hipótese nula.

O método apresentado anteriormente, foi o método da análise fatorial exploratória. Existe ainda, a análise fatorial confirmatória, onde o pesquisador deve partir de uma hipótese pré-estabelecida de relacionamento entre as variáveis e os fatores. Portanto, a análise fatorial confirmatória tem o intuito de confirmar ou não a hipótese do pesquisador. Este método será visto posteriormente, como um caso particular do modelo de equações estruturais.

Na análise fatorial exploratória, o intuito é explicar a variação das p variáveis por meio de $m < p$ fatores não observáveis. No entanto, esses fatores são não observáveis e não se pode utilizar as colunas da matriz das cargas fatoriais de forma direta para estimar os fatores, como ocorre de forma semelhante na análise dos componentes principais. Sendo assim, deve-se prever os valores dos fatores para cada unidade amostral e, a esses valores, dá-se o nome de escores fatoriais. Portanto, deve-se encontrar os escores fatoriais, no caso do pesquisador necessitar utilizar essas novas variáveis em análises posteriores. O método geralmente utilizado para obtenção dos escores fatoriais é o de Mínimos Quadrados Ponderados, que tem como princípio minimizar a soma de quadrado do resíduo, ponderada pelas variâncias específicas, proposto por Bartlett (1937).

Portanto, considerando desconhecidos os parâmetros μ , L e ψ do modelo fatorial, e, tomando seus estimadores amostrais, pode-se calcular o vetor de escores fatoriais para a j -ésima unidade amostral por:

$$\hat{F}_j = \left(\hat{L}^T \hat{\psi}^{-1} \hat{L} \right)^{-1} \hat{L}^T \hat{\psi}^{-1} (X_j - \bar{X}) \quad (59)$$

De acordo com Ferreira (2011), se a matriz de correlações amostrais for utilizada ao invés da matriz de covariâncias amostrais, deve-se considerar as variáveis padronizadas Z_j em substituição as variáveis originais, sendo o preditor de escores fatoriais:

$$\hat{F} = \left(\hat{L}^{*T} \hat{\psi}^{*-1} \hat{L}^* \right)^{-1} \hat{L}^{*T} \hat{\psi}^{*-1} Z_j \quad (60)$$

1.2.4 ANÁLISE DE EQUAÇÕES ESTRUTURAIS

A análise de modelos de equações estruturais (AEE) é uma extensão de outras técnicas multivariadas, sendo possível avaliar, de forma conjunta, as relações de dependência e independência entre uma ou mais variáveis. Tal procedimento, combina as técnicas de regressão e análise fatorial.

De acordo com Silva (2006), o conceito de modelagem de equações estruturais também teve início nos trabalhos de Wright (1921,1934). Ele trabalhou com padrões de covariância entre várias características de suínos, transformando as correlações observadas em um sistema de equações, as quais descreviam matematicamente suas hipóteses concernentes as relações causais, por meio dos parâmetros do modelo.

Segundo Bistaffa (2010), foi somente nos trabalhos de Jöreskog (1973), Keesling (1972) e Wiley (1973) que o modelo geral foi proposto, o qual era composto de dois submodelos, sendo um submodelo relacionando variáveis latentes ou não observáveis (fatores) e outro submodelo relacionando as variáveis observadas e variáveis latentes; em que desenvolveram métodos de estimação para a estrutura de variância e covariância do modelo geral. De acordo com Marôco (2010), análise de equações estruturais é uma técnica de modelagem generalizada, com objetivo de testar a validade de modelos teóricos, os quais definem relações causais, hipotéticas, entre variáveis. Tais relações são medidas pelos parâmetros do modelo, os quais representam o tamanho do efeito das variáveis independentes sobre as variáveis dependentes.

Algumas características do modelo de equações estruturais o diferencia da modelagem linear clássica, as quais: os modelos são construídos com variáveis não mensuráveis,

denominadas variáveis latentes ou construtos; os modelos levam em conta erros de medição para cada variável observada, os quais são parâmetros que podem ser estimados e testados; os modelos são, geralmente ajustados por matrizes de índices de inter-relacionamento, que são as matrizes de covariâncias ou correlações, entre todos os pares de variáveis observadas (RAIKOV; MARCOULIDES, 2006). Outro fato que torna o modelo de equações estruturais diferenciado, é que o modelo teórico é estabelecido a priori pelo pesquisador, depois coleta-se os dados que irão confirmar ou não o modelo formulado. Portanto, a teoria é o motor da análise, enquanto que na estatística clássica, os dados estão no centro do processo de análise (MARÔCO, 2010).

Ainda de acordo com Marôco (2010), a análise de equações estruturais tem se mostrado útil em áreas como ciências sociais e da saúde, em que o pesquisador se depara com variáveis latentes, sendo variáveis não diretamente observáveis, tais como "Depressão" medida por indicadores cujos valores indiquem tal estado. Assim, o pesquisador deve definir operacionalmente a variável latente de interesse em termos do construto acreditado para representá-lo (BOLLEN, 1989). Brahim et al. (2011), com o objetivo de desenvolver dois modelos para o teor de carbono orgânico em solos argilosos e arenosos na região semiárida do Mediterrâneo (Tunísia), utilizando a AEE, concluíram que, para solos argilosos, o acúmulo de carbono orgânico foi controlado pelos atributos químicos do solo (pH, matéria orgânica, teor de Nitrogênio) e a densidade do solo, quando comparados aos demais atributos físicos do solo, por outro lado, em solos arenosos, o acúmulo de carbono orgânico foi controlado, principalmente pelos atributos químicos (pH e teor de matéria orgânica do solo) quando comparados aos atributos físicos do solo. Os autores concluem que tais resultados indicam que atributos físicos e químicos do solo apresentam efeitos independentes no teor de carbono orgânico do solo nessas áreas.

A utilização da AEE como alternativa aos métodos tradicionais se justifica por fornecer ao pesquisador a habilidade de acomodar múltiplas relações de dependência inter-relacionadas em um só modelo (HAIR et al. 2009). A AEE permite purgar os erros nas variáveis por intermédio de modelos de medidas e de modelo estrutural, que descontaminam as variáveis (BOLLEN, 1989), o que não ocorre nos modelos clássicos, causando prejuízos nas estimativas dos parâmetros.

Os resultados da AEE não podem, geralmente, garantir a existência de uma relação de causalidade, pois uma utilização frequentemente incorreta, é a da inferência de causalidade, isto é, de uma relação de causa-e-efeito em que uma variável dita independente ou preditora, é a causa de outra variável, dita dependente ou critério (KLINE, 2010). A hipótese de causalidade

é suposta pelo pesquisador na construção do modelo teórico, e não propriamente pela significância do teste estatístico. Entretanto, se a hipótese for confirmada pelo teste estatístico, a suposição de causalidade do pesquisador não pode ser constatada como verdadeira. A AEE "raramente produz quaisquer resultados que tenham qualquer interpretação com efeitos causais" (WILKINSON; TFSI, 1999). Segundo Bistaffa (2010), para que uma relação de causa e efeito possa ser verificada, existe a necessidade de se verificar três condições: isolamento (as variáveis devem estar isoladas de qualquer outro tipo de influência externa); associação (as variáveis estão correlacionadas, respeitando o isolamento); direção (a relação de causa-e-efeito vai da causa para o efeito).

Na análise de equações estruturais, as variáveis em geral assumem escala quantitativa (intervalar ou razão). No entanto, em diversos estudos, é muito comum o uso de variáveis qualitativas (nominal e ordinal). De acordo com Bistaffa (2010), vários autores, entre eles, cita Olsson (1979), Boosma (1983), Johnson e Creech (1983), Muthén e Kaplan (1985), Babakus et al. (1987), vem realizando estudos simulados com o objetivo de testar a robustez dos ajustes de modelos de equações estruturais quando se considera uma variável ordinal como contínua.

Os tipos de modelos de equações estruturais comumente utilizados na literatura são: Análise de Caminhos (Path Analysis), Análise Fatorial Confirmatória (AFC) e os Modelos de Regressão Estrutural. A Análise de Caminhos é uma extensão do modelo de regressão linear múltipla, composto por um sistema de equações lineares que busca estudar as relações causais entre variáveis observadas e foi inicialmente desenvolvida por Wright (1921,1934), na área de genética quantitativa, análise fatorial confirmatória e modelo estrutural;

Os modelos de (AFC), considerados submodelos de medida do modelo de AEE, são frequentemente utilizados para examinar padrões de inter-relações entre vários construtos (fatores). Cada construto incluído no modelo é normalmente medido por um conjunto de indicadores observados. Assim, em um modelo de análise fatorial confirmatória, relações direcionais não específicas são assumidas entre os construtos, só que eles são potencialmente correlacionados entre si (RAIKOV; MARCOULIDES, 2006). De acordo com Marôco (2010), a variabilidade dos indicadores não explicada pelo construto ou fator comum é explicada por outros fatores latentes, específicos de cada indicador, não considerado no modelo (erros ou resíduos) e que tais fatores específicos podem estar correlacionados entre si, refletindo o efeito ou a existência de outros fatores comuns aos itens diferentes daqueles presentes no estudo.

Finalmente, os Modelos de Regressão Estrutural, considerados também submodelos do modelo de AEE, se assemelham aos modelos de análise fatorial confirmatória, exceto que eles também postulam relações particulares entre os construtos (regressões latentes), em vez de

considerá-los apenas interligados entre si. Os modelos podem ser usados para testar ou desconfirmar teorias sobre relações explicativas entre várias variáveis latentes sob investigação (RAIKOV; MARCOULIDES, 2006). Tais modelos definem relações causais ou de associação entre as variáveis latentes, em que se considera fatores específicos ou resíduos, também denominados *disturbances* (MORÔCO, 2010)

O modelo é constituído de uma série de parâmetros estruturais e variáveis, sendo que os parâmetros estruturais medem a associação entre as variáveis incluídas no modelo, que são as variáveis observadas ou manifestas e as variáveis latentes ou construtos (não observadas diretamente). Relativamente, à sua função no modelo de equações estruturais, as variáveis (quer latentes quer manifestas) podem ser classificadas em variáveis independentes ou exógenas, quando não são influenciadas por outra variável no modelo; e variáveis dependentes ou endógenas, quando a variação dessas variáveis é explicada por variáveis presentes no modelo (MARÔCO, 2010).

O modelo geral da AEE pode ser descrito da seguinte forma (BOLLEN, 1989):

$$y = \Lambda_y \eta + \varepsilon \quad (61)$$

$$x = \Lambda_x \xi + \delta \quad (62)$$

$$\eta = B\eta + \Gamma\xi + \zeta \quad (63)$$

em que,

$y^T = (y_1, y_2, \dots, y_p)$ são as variáveis manifestas no modelo de medida (61);

Λ_y é a matriz $p \times r$ dos efeitos de η sobre y no modelo de medida (61);

$x^T = (x_1, x_2, \dots, x_q)$ são as variáveis manifestas no modelo de medida (62);

Λ_x é a matriz $q \times s$ dos efeitos de ξ sobre x no modelo de medida (62);

$\eta^T = (\eta_1, \eta_2, \dots, \eta_r)$ são as variáveis latentes dependentes;

$\xi^T = (\xi_1, \xi_2, \dots, \xi_s)$ são as variáveis latentes independentes;

$\varepsilon^T = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_p)$ e $\delta^T = (\delta_1, \delta_2, \dots, \delta_q)$ são os vetores de erros de medida ou

mensuração de y e de x , respectivamente, modelos de medida (61) e (62);

$B_{r \times r}$ é a matriz dos coeficientes de η no modelo estrutural (63);

$\Gamma_{r \times s}$ é a matriz dos coeficientes de ξ no modelo estrutural (63);

$\zeta^T = (\zeta_1, \zeta_2, \dots, \zeta_r)$ são os resíduos ou erros do modelo estrutural (63);

O modelo geral assume as seguintes pressuposições:

- ε é não correlacionado com η ;
- δ é não correlacionado com ξ ;
- ζ é não correlacionado com ξ ;
- ζ , ε e δ são mutuamente não correlacionadas;
- As esperanças matemáticas dos erros são nulas;

Como uma variável não é causa e efeito dela mesma (modelo estrutural), decorre:

- $B_{ii} = 0$ e,
- $(I - B)$ é uma matriz singular (possui inversa), sendo I a matriz identidade.

E, com as seguintes matrizes de covariâncias:

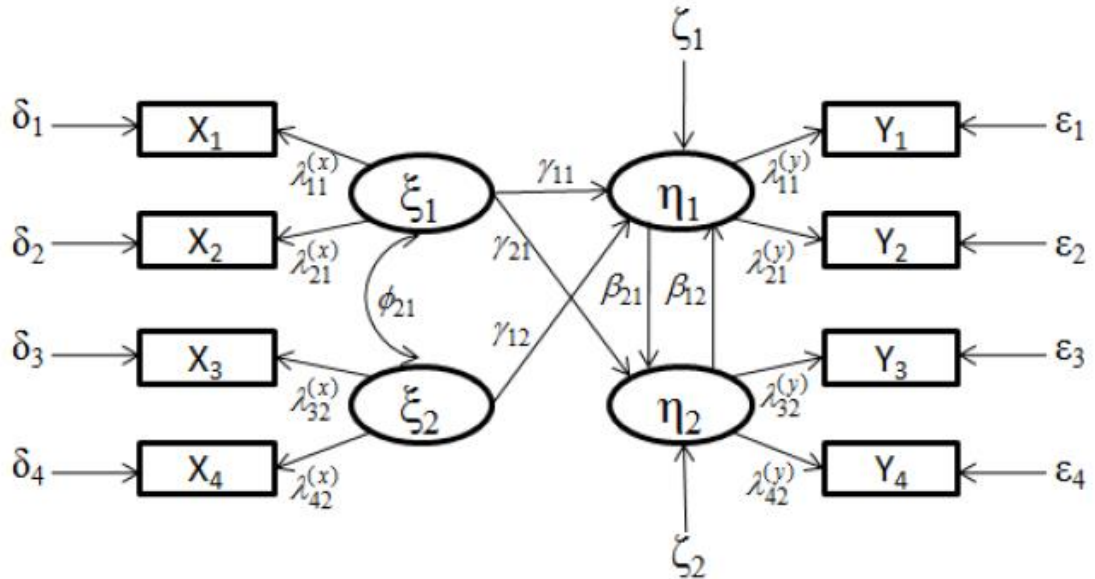
- $\Theta^{\varepsilon}_{(pxp)} = E[\varepsilon\varepsilon^T]$;
- $\Theta^{\delta}_{(qxq)} = E[\delta\delta^T]$;
- $\Psi_{(rxr)} = E[\zeta\zeta^T]$;
- $\Phi_{(sxs)} = E[\xi\xi^T]$;

As equações (61) e (62) referem-se aos modelos de medidas ou mensuração de y e de x , respectivamente, e a equação (63) refere-se ao modelo estrutural, relacionando as variáveis latentes do modelo. O modelo matemático de equações estruturais pode ser representado de forma gráfica, utilizando-se símbolos e figuras geométricas na composição do chamado diagrama de caminhos. Os símbolos que poderão ser utilizados na representação gráfica de um modelo de equações estruturais, bem como seus respectivos significados são descritos de acordo com Silva (2006), Bistaffa (2010) e Marôco (2010):

- Quadrados ou retângulos representam variáveis observadas ou manifestas;
- Círculos ou elipses representam variáveis latentes (fatores ou erros);
- Flechas retilíneas únicas significam que a variável da base da flecha "causa" a variável do outro extremo (de causa para efeito);
- Flechas retilíneas duplas significam que a variável "causa" e é causada pela outra (relação recursiva ou de feedback);
- Flechas curvilíneas de duplo sentido significam correlação entre as duas variáveis;

Como exemplo de um modelo matemático representado de forma gráfica, considera-se a Figura 1, conforme Bistaffa (2010).

Figura 1 - Representação Gráfica do Modelo Completo.



Fonte: Bistaffa (2010)

A Figura 1 representa graficamente um modelo de equações estruturais completo, sendo que na parte central tem-se o modelo estrutural relacionando as variáveis latentes e nos extremos tem-se os modelos de medidas relacionando as variáveis manifestas x e y e as variáveis latentes. Pode-se observar duas variáveis exógenas (ξ_1 e ξ_2), operacionalizadas por duas variáveis manifestas (x) cada uma; duas variáveis latentes endógenas (η_1 e η_2), operacionalizadas por duas variáveis manifestas (y) cada uma.

Nota-se ainda que as variáveis latentes exógenas estão correlacionadas, medida representada por (ϕ_{21}); considera-se ainda que os índices dos pesos fatoriais e coeficientes estruturais λ e γ , respectivamente, são representados por ordem de efeito para causa. Por exemplo, o peso fatorial do fator ξ_1 em x_2 é $\lambda_{21}^{(x)}$ e, de forma semelhante, o coeficiente estrutural de ξ_1 para η_2 é γ_{21} .

Dado que o modelo está representado, o próximo passo é escrevê-lo matematicamente, e, conseqüentemente, na forma matricial, para estimação dos parâmetros do modelo. O modelo de equações de mensuração que medem a relação entre as variáveis latentes e as manifestas, respeitando a Figura 1 e considerando as equações (61) e (62) são:

$$\begin{aligned}
y_1 &= \lambda_{11}^y \eta_1 + \varepsilon_1 & x_1 &= \lambda_{11}^x \xi_1 + \delta_1 \\
y_2 &= \lambda_{21}^y \eta_1 + \varepsilon_2 & x_2 &= \lambda_{21}^x \xi_1 + \delta_2 \\
y_3 &= \lambda_{32}^y \eta_2 + \varepsilon_3 & x_3 &= \lambda_{32}^x \xi_2 + \delta_3 \\
y_4 &= \lambda_{42}^y \eta_2 + \varepsilon_4 & x_4 &= \lambda_{42}^x \xi_2 + \delta_4
\end{aligned}$$

Que, reescritas na forma matricial, tem-se:

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} = \begin{bmatrix} \lambda_{11}^y & 0 \\ \lambda_{21}^y & 0 \\ 0 & \lambda_{32}^y \\ 0 & \lambda_{42}^y \end{bmatrix} \begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \end{bmatrix} \quad \text{e} \quad \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} \lambda_{11}^x & 0 \\ \lambda_{21}^x & 0 \\ 0 & \lambda_{32}^x \\ 0 & \lambda_{42}^x \end{bmatrix} \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix} + \begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \end{bmatrix}$$

Com matrizes de covariância:

$$\Theta^\varepsilon = \begin{bmatrix} \theta_{11}^\varepsilon & 0 & 0 & 0 \\ 0 & \theta_{22}^\varepsilon & 0 & 0 \\ 0 & 0 & \theta_{33}^\varepsilon & 0 \\ 0 & 0 & 0 & \theta_{44}^\varepsilon \end{bmatrix} \quad \text{e} \quad \Theta^\delta = \begin{bmatrix} \theta_{11}^\delta & 0 & 0 & 0 \\ 0 & \theta_{22}^\delta & 0 & 0 \\ 0 & 0 & \theta_{33}^\delta & 0 \\ 0 & 0 & 0 & \theta_{44}^\delta \end{bmatrix}$$

O modelo estrutural de acordo com a Figura 1 e a equação (63) representa as relações entre as variáveis latentes, e, suas equações são:

$$\begin{aligned}
\eta_1 &= \quad + \beta_{12}\eta_2 + \gamma_{11}\xi_1 + \gamma_{12}\xi_2 + \varsigma_1 \\
\eta_2 &= \beta_{21}\eta_1 + \quad + \gamma_{21}\xi_1 + \quad + \varsigma_2
\end{aligned}$$

E, na forma matricial, pode-se reescrever o modelo estrutural da seguinte forma:

$$\begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} = \begin{bmatrix} 0 & \beta_{12} \\ \beta_{21} & 0 \end{bmatrix} \begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} + \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & 0 \end{bmatrix} \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix} + \begin{bmatrix} \varsigma_1 \\ \varsigma_2 \end{bmatrix}$$

Com matrizes de covariâncias descritas abaixo:

$$\Phi = \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix} \quad \text{e} \quad \Psi = \begin{bmatrix} \psi_{11} & \psi_{12} \\ \psi_{21} & \psi_{22} \end{bmatrix}$$

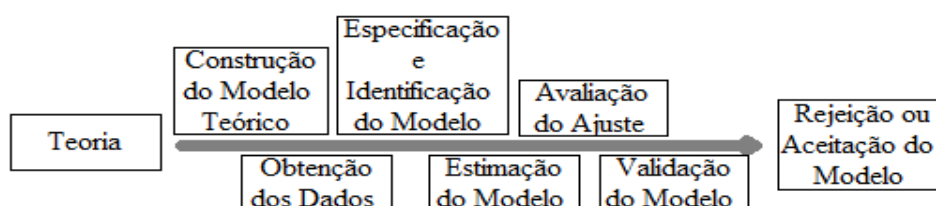
De acordo com Bistaffa (2010), os parâmetros que são fixados em 0 ou qualquer outro valor, são ditos parâmetros fixos, e, quando algum parâmetro é restringido ao valor de outro, este é dito parâmetro restrito, os demais parâmetros sem restrições são denominados parâmetros livres. Ainda segundo o mesmo autor, quando o modelo de equações estruturais possui uma relação recíproca, tal como a de η_1 e η_2 , este modelo é denominado modelo não-recursivo, sendo mais complicado de ser analisado. Por fim, se as relações de um modelo são bem definidas, este modelo é dito estável, caso contrário, instável.

De acordo com Raikov e Marcoulides (2006), além de erros de medidas, o modelo de equações estruturais permite obter estimativas, testar modelos multivariados complexos, bem como estudar os efeitos diretos e indiretos das variáveis envolvidas em determinado modelo. Na definição de Bollen (1987), os efeitos diretos são aquelas influências não mediadas por qualquer outra variável no modelo; os efeitos indiretos são aqueles mediados por pelo menos uma variável interveniente, sendo determinado pela subtração entre os efeitos diretos e efeitos totais (efeitos diretos mais indiretos).

Considerando a Figura 1, pode-se tomar um exemplo para entendimento dos efeitos, conforme Bistaffa (2010). O efeito direto de ξ_1 em η_1 é γ_{11} ; o efeito indireto de ξ_1 em η_1 é $\phi_{21}\gamma_{12}$; e o efeito total é a soma $\gamma_{11} + \phi_{21}\gamma_{12}$. Portanto, para entender as relações entre as variáveis do modelo, é importante a compreensão de cada tipo de efeito. Os procedimentos da aplicação da AEE (Figura 2), bem como descrição da relevância de cada um deles, serão apresentados a seguir.

Como dito anteriormente, na AEE o modelo teórico é o motor da análise, diferentemente da análise estatística clássica, em que os dados é a fonte principal. Para tanto, a primeira etapa neste caso, é a construção do modelo teórico, exigindo do pesquisador vasta revisão na literatura, afim de compreender o fenômeno em estudo, bem como seus relacionamentos, por meio do modelo operacionalizado.

Figura 2 - Etapas da Análise de Equações Estruturais



Fonte: Marôco (2010).

O próximo passo da AEE é a obtenção dos dados e uma preocupação que surge é no que tange ao tamanho da amostra. Segundo Marôco (2010), em modelos de AEE, pode-se ter 2 ou 3 parâmetros associados a uma variável manifesta, como exemplo; cita um erro, um peso fatorial e uma associação entre erros. Com isso, a dimensão da amostra necessária para produzir variabilidade suficiente para estimar cada um dos parâmetros do modelo, passa-se a ser maior que nos modelos clássicos, como na regressão linear. Com isso Marôco (2010) sugere entre 10 e 15 observações por cada variável manifesta, ou seja, pelo menos 5 observações por cada parâmetro. No trabalho de Hu, Bentler e Kano (1992), quando avaliaram sete índices de qualidade de ajuste de um modelo de análise fatorial confirmatória considerando-se seis tamanhos amostrais, sugeriram um tamanho de amostra superior a 10 vezes o número de parâmetros livres do modelo. Contudo, Kline (2010), afirma que um tamanho "típico" de amostra onde AEE é utilizada é cerca de 200 casos, em que este número corresponde ao tamanho de amostra mediano aproximado de um levantamento de artigos publicados em que se utilizaram a AEE nas áreas de psicologia e gestão, de Breckler (1990) e Goldstein, Bonnet e Rocher (2007), respectivamente.

Definido o modelo teórico e obtidos os dados, deve-se especificar e identificar o modelo teórico. Marôco (2010) define a especificação modelo como: decidir que variáveis manifestas operacionalizam que variáveis latentes, que relações causais entre variáveis latentes e/ou manifestas devem ser incluídas/excluídas, que associações (não causais) devem ser incluídas/omitidas do modelo, que erros ou resíduos devem ser correlacionados. De acordo com Hox e Bechger (1998), a especificação do modelo é direcionada por uma combinação de resultados teóricos e empíricos de pesquisas anteriores. Nesta etapa, pode-se elaborar a representação gráfica, como exemplificado na Figura 1.

Na fase de identificação do modelo, deve-se incluir as informações sobre os parâmetros a serem estimados e verificar se o modelo é identificável ou não, sendo, portanto, uma das etapas mais importantes da AEE. Como os parâmetros do modelo são estimados em função dos elementos da matriz de variâncias e covariâncias dos dados, tais informações devem ser consideradas no que tange a identificação do modelo.

Considerando $p + q$ variáveis manifestas e t o número de parâmetros a estimar, então os graus de liberdade são calculados da seguinte forma:

$$gl = \frac{(p + q)(p + q + 1)}{2} - t \quad (64)$$

A primeira parte da equação refere-se ao número de elementos não redundantes da matriz de covariância dos dados. As classificações de identificação do modelo são similares às de um sistema linear, mais detalhes em Raikov e Marcoulides (2006) e Kline (2010).

Após o passo de identificação do modelo, a próxima etapa é o processo de estimação dos parâmetros do modelo, com base na matriz de variância e covariâncias dos dados. Considerando que os "dados" em AEE corresponde a matriz de variâncias e covariâncias, deseja-se então obter estimativas dos parâmetros que reproduzam de forma satisfatória a estrutura de variância e covariâncias ou correlações dos dados.

Segundo Marôco (2010), se o modelo de equações estruturais estiver correto, então a diferença entre as covariâncias observadas nas amostras e as covariâncias previstas pelo modelo será mínima, sendo o objetivo da estimação encontrar um vetor θ dos parâmetros do modelo que reproduza melhor possível a matriz de covariâncias observadas ou correlações observadas.

De acordo com Bistaffa (2010), além da escolha do método de estimação a ser utilizado, o pesquisador deve escolher a matriz de entrada para efetuar os cálculos, sendo a matriz de covariâncias mais vantajosa quanto a fornecer comparações válidas entre diferentes populações ou amostras e a matriz de correlações mais vantajosa quando o intuito é compreender as relações entre os construtos do modelo, mais não explicar a variância total de uma variável latente. Com base na validade dos pressupostos do modelo, Jöreskog e Sörbom (1996) definem a matriz de covariância das variáveis manifestas na população da seguinte forma:

$$\Sigma = \begin{bmatrix} \Sigma_{yy} & \Sigma_{yx} \\ \Sigma_{xy} & \Sigma_{xx} \end{bmatrix} = \begin{bmatrix} \Lambda_y (I - B)^{-1} (\Gamma \Phi \Gamma^T + \Psi) [(I - B)^{-1}]^T \Lambda_y^T + \Theta^\epsilon & \Lambda_y (I - B)^{-1} \Gamma \Phi \Lambda_x^T \\ \Lambda_x \Phi \Gamma^T [(I - B)^{-1}]^T \Lambda_y & \Lambda_x \Phi \Lambda_x^T + \Theta^\delta \end{bmatrix}$$

Sendo a matriz de variâncias e covariâncias dos dados, Σ , e a matriz de parâmetros do modelo, $\Sigma(\theta)$, espera-se que após obtidas as estimativas dos parâmetros, seja possível escrever:

$$\Sigma = \hat{\Sigma}(\hat{\theta}) \quad (65)$$

Na maioria das vezes, o censo é impraticável, de onde se utiliza a matriz S ao invés da matriz Σ , e, se o modelo estiver correto, objetiva-se:

$$S = \hat{\Sigma}(\hat{\theta}) \quad (66)$$

De acordo com Marôco (2010), o modelo pode ser estimado por métodos exatos ou quase exatos de equações diferenciais que minimizem a diferença entre S e $\hat{\Sigma}(\hat{\theta})$, sendo que os softwares utilizam, em sua maioria, métodos iterativos (exatos), tornando a tarefa mais

demorada e complexa. Portanto, o objetivo é minimizar a função da diferença da matriz de covariância dos dados S e da matriz de covariância gerada pelo modelo teórico $\hat{\Sigma}(\theta)$, sendo então esta função definida como:

$$f = F\left(S - \hat{\Sigma}(\theta)\right) \quad (67)$$

Segundo Bistaffa (2010), a função definida em (67) baseada em S e $\hat{\Sigma}(\theta)$, é denominada função de discrepância e possui as seguintes propriedades:

- f é um escalar para cada $\hat{\theta}$ fixado;
- $f \geq 0$;
- $f = 0$ se e somente se $\hat{\Sigma}(\theta) = S$;
- f é contínua em $\hat{\Sigma}(\theta)$ e em S .

Marôco (2010) afirma que se o ajustamento for perfeito, então $f = 0$; portanto, quanto menor for f , melhor será o ajustamento do modelo teórico aos dados (covariâncias). Ainda segundo o mesmo autor, o vetor θ é estimado por um algoritmo iterativo ou uma transformação monótona, que depende do método de estimação usado. A seguir será apresentado três métodos de estimação muito utilizados em AEE, sendo o Método da Máxima Verossimilhança (ML), o método dos mínimos quadrados generalizados (GLS) e o método Distribuição Assimptótica Livre (ADF) ou Mínimos Quadrados Ponderados (WLS).

Dado que um vetor de variáveis manifestas z (com p e q variáveis manifestas) possui distribuição normal multivariada, com base numa amostra de N observações, com vetor de médias de z igual a μ e matriz de variâncias e covariâncias de z igual a Σ , pode-se escrever a seguinte função densidade de probabilidade de z (AMORIM et al., 2012):

$$f(z) = (2\pi)^{-\left(\frac{p+q}{2}\right)} |\Sigma|^{-\frac{1}{2}} \exp\left[-\frac{1}{2} z^T \Sigma^{-1} z\right] \quad (68)$$

em $p + q$ é o número total de variáveis manifestas.

A ideia do método ML é estimar a matriz de parâmetros $\Sigma(\theta)$ que maximiza a verossimilhança de se observar a matriz S . A função de discrepância de máxima verossimilhança a ser minimizada ao invés de maximizada, devido a necessidade de se ter

adicionado termos que não dependessem dos parâmetros do modelo é (JÖRESKOG; SÖRBOM, 1996; AMORIM et al., 2012):

$$f_{ML} = \log|\hat{\Sigma}(\hat{\theta})| + tr\left(S\hat{\Sigma}(\hat{\theta})^{-1}\right) - \log|S| - (p + q) \quad (69)$$

em que $\log|\hat{\Sigma}(\hat{\theta})|$ é o logaritmo do determinante de $\hat{\Sigma}(\hat{\theta})$ e $tr\left(S\hat{\Sigma}(\hat{\theta})^{-1}\right)$ é o traço de $S\hat{\Sigma}(\hat{\theta})^{-1}$.

O método de estimação da máxima verossimilhança é um dos mais tradicionais em AEE e produz estimativas dos parâmetros centradas e consistentes, em que quanto maior a amostra, maior a precisão das estimativas dos parâmetros. O método ML é robusto a violação do pressuposto de normalidade se a assimetria e achatamento das distribuições das variáveis manifestas não forem muito grandes (MARÔCO, 2010).

Ainda segundo o mesmo autor, o método GLS é um método iterativo que estima os parâmetros ponderando os erros de estimação da matriz residual $E = S - \hat{\Sigma}(\hat{\theta})$ com pesos correspondentes ao inverso da matriz de covariância amostral, isto é, elementos da matriz E com maior variância amostral, tem menor peso no modelo. Uma vantagem potencial do método GLS sobre a estimativa ML é que ele requer menos tempo de computação e menos memória do computador. No entanto, essa vantagem potencial não é tão significativa atualmente, devido a processadores rápidos e memória abundante em computadores pessoais e relativamente de baixo custo (KLINE, 2010).

Com base em Jöreskog e Sorbom (1996), a função de discrepância a ser minimizada no método GLS é:

$$f_{GLS} = \frac{1}{2} tr\left[S^{-1}\left(S - \hat{\Sigma}(\hat{\theta})\right)\right]^2 = \frac{1}{2} tr\left[\left(I - S^{-1}\hat{\Sigma}(\hat{\theta})\right)^2\right] \quad (70)$$

O método GLS é equivalente a minimizar a soma de quadrado dos erros (SQE), ponderada pelo inverso da matriz de covariância amostral (MAROCÔ, 2010). De acordo com Kline (2010), sob a suposição de normalidade multivariada, os métodos ML e GLS são assintóticos, porém, o método ML tem menores probabilidades de erro tipo I no teste χ^2 da qualidade de ajustamento global e produz estimativas mais corretas. Por isso e de acordo com Kline (2010), o método ML é mais utilizado que o método GLS.

O método ADF é menos restritivo, pois não exige que as variáveis manifestas tenham distribuição de probabilidades normal multivariada. De acordo com Raikov e Marcoulides (2006), o método da ADF minimiza a seguinte função de discrepância:

$$f_{ADF} = \left[\left(s - \sigma(\hat{\theta}) \right)^T W^{-1} \left(s - \sigma(\hat{\theta}) \right) \right] \sigma(\hat{\theta}) \quad (71)$$

em que s é o vetor de elementos da matriz triangular inferior S incluindo a diagonal; $\sigma(\hat{\theta})$ é o vetor de elementos da matriz triangular inferior $\Sigma(\hat{\theta})$ incluindo a diagonal e W é uma matriz de distâncias de todas as observações às médias de todas as variáveis. A função de discrepância ADF permite obter estatísticas de teste da qualidade global do modelo com distribuição assintótica χ^2 e erros-padrão dos parâmetros não enviesados, contudo o cálculo da matriz W pode ser complexo, uma vez que sua dimensão depende do número de variáveis manifestas, que pode ser bem grande, caso se tenha muitas variáveis manifestas, além disso, a matriz W deve ser invertível (MARÔCO, 2010).

Um dos pressupostos da Análise de equações estruturais é que, caso o método de estimação dos parâmetros seja o da máxima verossimilhança, as variáveis manifestas sejam normalmente distribuídas, como mencionado anteriormente. Estimar os parâmetros do modelo por máxima verossimilhança quando as variáveis são contínuas, porém, não apresentam distribuição normal multivariada, podem produzir estimativas menores de erro padrão (25 – 50% menores), levando a rejeitar a hipótese nula de que o parâmetro populacional é zero mais frequentemente do que o esperado, o que significa, incorrer no erro tipo I inflacionado (MATTOS, 2009).

De acordo com Newsom (2015), caso os dados não sejam normalmente distribuídos, pode ocorrer ainda inflação da estatística χ^2 de ajustamento, provocando a rejeição de muitos modelos, quando na verdade descreveriam de forma satisfatória a estrutura de variância e covariância dos dados. Para solucionar este problema, Satorra e Bentler (1988; 2001), propuseram o método de estimação da máxima verossimilhança com correção ou ajustamento da estatística χ^2 , bem como erros padrão robustos, calculados com base na curtose multivariada.

Por fim, são realizados testes, calculados índices e realizada a análise de resíduos, com intuito de avaliar a adequabilidade do modelo teórico da AEE, isto é, se por meio da matriz de estimativas dos parâmetros conseguiu-se reproduzir de forma satisfatória a estrutura de variâncias e covariâncias ou correlações das variáveis observadas. Marôco (2010) aponta uma série de estudos simulados e observação empírica em que dezenas de estatísticas foram desenvolvidas para avaliação da qualidade de ajuste do modelo.

O primeiro teste comumente utilizado para avaliar a adequabilidade do modelo é o teste χ^2 de Ajustamento, realizado sob a hipótese nula de que o modelo ajustado reproduz perfeitamente a estrutura de variâncias e covariâncias populacional. A estatística do teste é descrita da seguinte forma (RAIKOV; MARCOULIDES, 2006):

$$T = (n-1)f_{\min} \quad (72)$$

em que f_{\min} é o valor mínimo das funções de discrepância para os métodos de estimação apresentados anteriormente e n é o tamanho da amostra ou número de observações. A estatística T tem distribuição χ^2 .

Como o teste mede a perfeição do ajustamento e, em sua maioria, os modelos sob teste, quando muito, são uma simplificação do fenômeno real, Marôco (2010) afirma que, como qualquer modelo apresenta um grau de erro, a hipótese testada pelo teste pode ser facilmente considerada falsa. Entretanto, ainda segundo o mesmo autor, esses testes são sensíveis a dimensão da amostra, por isso, para amostras pequenas se tem probabilidades acrescidas do erro Tipo II, e, para amostras grandes, o teste tem probabilidades acrescidas do erro Tipo I. Estes testes ainda são sensíveis a violação do pressuposto de normalidade multivariada e, nestes casos, a estatística do teste qui-quadrado de ajustamento pode ser corrigida, de acordo com Satorra e Bentler (2001).

De acordo com Raikov e Marcoulides (2006), as limitações da estatística do teste qui-quadrado de ajustamento T , indicam a importância da disponibilidade de outros índices para auxiliar no processo de avaliação do modelo. Existem diversos índices que podem ser usados em alternativa ao teste χ^2 , geralmente enquadrados em cinco grupos: Índices Absolutos, Índices Relativos, Índices de Parcimônia, Índice de discrepância populacional e Índices baseados na teoria da informação.

Os principais índices Absolutos são: T/ gl (Estatística do Teste χ^2 dividida pelos graus de liberdade); Raiz do Quadrado Médio Residual (RMR, do inglês Root Mean-square Residual), sendo a raiz quadrada da matriz dos erros dividida pelos graus de liberdade e o Índice de Qualidade do Ajuste (GFI, do inglês Goodness of Fit Index); os Índices Relativos avaliam a qualidade do modelo sob análise em relação ao modelo de pior ajuste e ao modelo de melhor ajuste, os principais índices desta família são: Índice de Ajuste Normalizado (NFI, do inglês Normed Fit Index); Índice de Ajuste Comparativo (CFI, do inglês Comparative Fit Index), Índice de Ajuste Relativo (RFI, do inglês Relative Fit Index) e o Índice de Tucker-Lewis (TLI, do inglês Tucker-Lewis Index); os Índices de Parcimônia são uma correção ou penalização dos

índices relativos por um fator de parcimônia estimado em função dos graus de liberdade do modelo em questão e os graus de liberdade do modelo basal, sendo gl / gl_b . Os principais índices de parcimônia são: Índice de Ajuste Comparativo de Parcimônia (PCFI, do inglês Parsimony Comparative Fit Index), Índice de Qualidade de Ajuste de Parcimônia (PGFI, do inglês Parsimony Goodness of Fit Index), e o Índice de Ajuste Normalizado de Parcimônia (PNFI, do inglês Parsimony Normed Fit Index).

Os índices de discrepância populacional medem o desajustamento de modelos aninhados, pela função de discrepância, a principal medida é a Raiz do Erro Quadrático Médio de Aproximação (RMSEA, do inglês Root Mean Square Error of Aproximation), estatística que mede o grau de desajustamento do modelo teórico a estrutura de variâncias e covariâncias observadas; os índices baseados na teoria da informação penalizam modelos mais complexos, sendo os mais utilizados: Critério de Informação de Akaike (AIC, do inglês Akaike Information Criterion), Critério de Browne-Cudeck (BCC, do inglês Browne-Cudeck Criterion) e o Índice Esperado de Validação Cruzada (ECVI, do inglês Expected Cross-Validation Index).

Tabela 1 - Índices de qualidade de ajustamento e seus valores de referência

Estatística	Valores de Referências
T	$p - valor > 0,05$
T / gl	> 5 - Ajustamento mau $]2;5]$ - Ajustamento sofrível $]1;2]$ - Ajustamento bom ~ 1 - Ajustamento muito bom
CFI	$< 0,8$ - Ajustamento mau
GFI	$[0,8;0,9[$ - Ajustamento sofrível
TLI	$[0,9;0,95[$ - Ajustamento bom $\geq 0,95$ - Ajustamento muito bom
$PGFI$	$< 0,6$ - Ajustamento mau
$PCFI$	$[0,6;0,8[$ - Ajustamento bom $\geq 0,8$ - Ajustamento muito bom
$RMSEA$	$> 0,10$ - Ajustamento inaceitável $]0,05;0,10]$ - Ajustamento bom $\leq 0,05$ - Ajustamento muito bom
AIC	Apenas para comparar modelos, de preferência modelos não aninhados. Não tem valor de referência, quanto menor, melhor.
BCC	
$ECVI$	
$MECVI$	

Fonte: Marôco (2010).

Os índices de ajuste vistos anteriormente (Tabela 1) são com respeito ao ajuste global do modelo, por isso não fornecem informações individuais do modelo, referente a cada parâmetro. De acordo com Raikov e Marcoulides (2006), um modelo pode apresentar fraca relação entre duas variáveis e no restante excelente, neste caso, o teste qui-quadrado pode não acusar e julgar o modelo plausível, ou ainda, a estatística qui-quadrado pode ser não significativa, porém o modelo explica de forma satisfatória todas as variâncias e covariâncias restantes da matriz S .

Uma alternativa para este tipo de problema é o exame dos resíduos do modelo proposto; Marôco (2010) destaca o exame dos resíduos estandarizados, os quais são os resíduos dividido pelos seus erros padrões (o quociente mantém a unidade de medida das variáveis manifestas), sendo que resíduos estandarizados superior a 2 indicam observações muito dispare das demais (*Outliers*). Latif (2000) indica a realização do gráfico quantil-quantil gaussiano dos resíduos

estandardizados, sendo o ajuste satisfatório quando se forma uma nuvem de pontos acima da primeira bissetriz (BULHÕES, 2013).

Caso o modelo teórico proposto não apresente um bom ajuste, isso pode evidenciar que o modelo não é apropriado para descrever a estrutura correlacional dos dados; não necessariamente pode-se dizer que o modelo está totalmente errado. Portanto, neste caso, o modelo pode ser reespecificado utilizando-se Índices de Modificação, os quais de acordo com Marôco (2010) tem o intuito de indicar a redução da estatística do teste qui-quadrado quando um parâmetro fixo é libertado, novas trajetórias estruturais forem adicionadas e erros de medidas forem dados como correlacionados. De uma forma geral, busca-se excluir parâmetros não significativos e incluir parâmetros a serem estimados no modelo, com base nestes índices.

Alguns métodos de estimação presumem que as variáveis manifestas tenham distribuição normal multivariada, sendo esse um dos principais requisitos da AEE (PILATI; LAROS, 2007). No entanto, a violação desta pressuposição pode estar atrelada ao fato da presença de *outliers* na matriz residual. Tal fato pode ser constatado pela análise de resíduo, citada anteriormente, bem como pela *Distância de Mahalanobis* (D_i) (MARÔCO, 2010):

$$D_i = \sqrt{(x_i - \bar{x})^T S^{-1} (x_i - \bar{x})} \quad (73)$$

em que D_i mede a distância de uma observação à média de todas as observações de todas as variáveis (\bar{x}), denotada centroide.

Pode-se utilizar o teste estatístico de Arbuckle (2009), cuja estatística tem distribuição normal padrão e é calculada em função da distância D_i e da curtose multivariada. Para amostras grandes calcula-se a probabilidade de uma dada observação ter um valor de D_i^2 superior a d_i^2 calculada para essa observação (p_1), e também a probabilidade de a maior distância de Mahalanobis ser superior a d_i^2 calculada (p_2) (MARÔCO, 2010). De acordo com Arbuckle (2009), valores das duas probabilidades inferiores a 0,05 para dada observação, a credencia como *outlier* multivariado.

REFERÊNCIAS

- ANDRIOTTI, J. L. S. **Interpretação dos dados de prospecção geoquímica com auxílio de estatística**. Porto Alegre: Companhia de Pesquisa de Recursos Minerais – CRPM, 2010. Disponível em: <www.cprm.gov.br/publique/media/rli_Andriotti.pdf>. Acesso em: 10 fev. 2016.
- ARBUCKLE, J. L. **Amos 18 reference guide**. Chicago: SPSS, 2009.
- ALMEIDA JÚNIOR, J. Aplicações da análise multivariada em solos que podem ser utilizados em melhorias de processos produtivos agrícolas. In: SIMPÓSIO DA ENGENHARIA DE PRODUÇÃO DA REGIÃO NORDESTE, 7., 2012, Mossoró. **Anais...** Mossoró: SEPRONe, 2012. p. 1-8
- AMORIM, L. D. A. F.; FIACCONE, R. L.; SANTOS, C. A. S. T.; MORAES, L. T. L. P.; OLIVEIRA, N. F.; OLIVEIRA, S. B.; SANTOS, T. N. L. **Modelagem com equações estruturais: princípios básicos e aplicações**. Salvador: UFBA, 2012. Disponível em: <https://repositorio.ufba.br/ri/bitstream/ri/17684/1/ebook_SEM_2012.pdf>. Acesso em: 06 out. 2015.
- ARAÚJO, A. O.; MENDONÇA, L. A. R.; LIMA, M. G. S.; FEITOSA, J. V.; SILVA, F. J. A.; NESS, R. L. L.; FRISCHKORN, H.; SIMPLÍCIO, A. A. F.; REGINA, M. Modificações nas propriedades do solo de uma área de manejo florestal na chapada do Araripe. **R. Bras. Ci. Solo**, Viçosa, MG, v. 37, n. 3, p. 754-762, maio/jun. 2013.
- BABAKUS, E.; FERGUSON JUNIOR, C. E.; JÖRESKOG, K. G. The sensitivity of confirmatory maximum likelihood factor analysis to violations of measurement scale and distributional assumptions. **Journal of Marketing Research**, Chicago, v. 24, n. 2, p. 222-228, maio 1987.
- BARTLETT, M. S. The statistical conception of mental factors. **British Journal of Psychology**, London, v. 28, p. 97-104, jul. 1937.
- BAKKE, H. A.; LEITE, A. S. M.; SILVA, L. B. Estatística multivariada: aplicação da análise fatorial na engenharia de produção. **Revista Gestão Industrial**, Ponta Grossa, v. 4, n. 4, p. 1-14, dez. 2008.
- BISTAFFA, B. C. **Incorporação de indicadores categóricos ordinais em modelos de equações estruturais**. 2010. 127 f. Dissertação (Mestrado em Estatística) - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2010.

- BOLLEN, K. A. Total, direct, and indirect effects in structural equation models. In: BOHRNSTEDT, G. W. **Sociological methodology**. Washington: C. C. Clogg, 1987. v. 37, p. 37-69.
- BOLLEN, K. A. **Structural equations with latent variables**. New York: Wiley, 1989. 528 p.
- BOOMSMA, A. **On the robustness of LISREL (maximum likelihood estimation) against small sample size and non-normality**. 1983. 228 f. Tese (Doutorado em Filosofia) - University Of Groningen, Groningen, 1983.
- BORDONAL, R. O. **Balço de gases de efeito estufa em áreas de cana-de-açúcar sob diferentes cenários de manejo, considerando-se mitigações**. 2012. 67 f. Dissertação (Mestrado em Agronomia) – Faculdade de Ciências Agrárias e Veterinária, Universidade Estadual Paulista, Jaboticabal, 2012.
- BORDONAL, R. O. **Greenhouse gas balance associated with sugarcane production in south-central Brazil, considering the management and expansion**. 2016. 81 f. Tese (Doutorado em Agronomia) - Faculdade de Ciências Agrárias e Veterinária, Universidade Estadual Paulista, Jaboticabal, 2016.
- BRAHIM, N.; BLAVET, D.; GALLALI, T.; BERNOUX, M. Application of structural equation modeling for assessing relationships between organic carbon and soil properties in semiarid Mediterranean region. **International Journal of Environmental Science and Technology**, Berlin, v. 8, n. 2, p. 305-320, 2011.
- BRECKLER, S. J. Applications of covariance structure modeling in psychology: cause for concern?. **Psychological Bulletin**, Washington, v. 107, n. 2, p. 260–273, mar. 1990.
- BULHÕES, R. S. **Contribuições a análise de outliers em modelos de equações estruturais**. 2013. 79 f. Dissertação (Mestrado em Estatística) - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2013.
- CANTELMO, N. F.; FERREIRA, D. F. Desempenho de testes de normalidade multivariados avaliado por simulação Monte Carlo. **Ciência e Agrotecnologia**, Lavras, v. 31, n. 6, p. 1630-1636, 2007.
- CATTELL, R. B. The screen test for the number of factors. **Multivariate Behavioral Research**, Abingdon, v. 1, n. 5, p. 140-161, jan. 1966.
- COMPANHIA NACIONAL DE ABASTECIMENTO – CONAB. **Acompanhamento de safra brasileira: cana-de-açúcar**. Brasília: Conab, 2016. (Primeiro Levantamento, v. 1).

Disponível

em:

<http://www.conab.gov.br/OlalaCMS/uploads/arquivos/16_04_18_14_27_15_boletim_cana_portugues_-_1o_lev_-_16.pdf>. Acesso em: 20 abr. 2016.

COOPER, D. R.; SCHINDLER, P. S. **Métodos de pesquisa em administração**. 7. ed. Porto Alegre: Bookman, 2003. 640 p.

CORRAR, L. J. et al. **Análise multivariada**: para os cursos de administração, ciências contábeis e economia. São Paulo: Atlas, 2009.

DALCHIAVON, F. C.; CARVALHO, M. P.; FREDDI, O. S.; ANDREOTTI, M.; MONTANARI, R. Variabilidade espacial da produtividade do feijoeiro correlacionada com atributos químicos de um Latossolo Vermelho distroférrico sob sistema de semeadura direta. **Bragantia**, Campinas, v. 70, n. 4, p. 908-916, 2011.

FERREIRA, D. F. **Estatística multivariada**. 2. ed. Lavras: UFLA, 2011. 676 p.

FIDALSKI, J.; TORMENA, C. A.; SCAPIM, C. A. Espacialização vertical e horizontal dos indicadores de qualidade para um Latossolo Vermelho. . **R. Bras. Ci. Solo**, Viçosa, MG, v. 31, p. 9-19, 2007.

FILZMOSE, P.; GARRETT, R. G.; REIMANN, C.. Multivariate outlier detection in exploration geochemistry. **Computers & Geosciences**, Amsterdam, v. 31, n. 5, p. 579-587, jun. 2005.

GOLDSTEIN, H.; BONNET, G.; ROCHER, T.. Multilevel structural equation models for the analysis of comparative data on educational performance. **Journal of Educational And Behavioral Statistics**, Stanford, v. 32, n. 3, p. 252-286, set. 2007. DOI: 10.3102/1076998606298042.

HAIR, J. F. J. et al. **Análise multivariada de dados**. 5. ed. Porto Alegre: Bookman, 2005.

HAIR, Joseph F. et al. **Análise Multivariada de Dados**. 6. ed. Porto Alegre: Bookman, 2009. 688 p.

HAMPEL, F. R.; RONCHETTI, E. M.; ROUSSEEUW, P. J.; STAHEL, W. A. **Robust statistics**: the approach based on influence functions. New York: John Wiley & Sons, 1986. 502 p.

HANSON, B.; SUGDEN, A.; ALBERTS, B. Making data maximally available. **Science**, Washington, v. 331, p. 649, 2011.

- HAYASHI, K.; BENTLER, P. M. The asymptotic covariance matrix of maximum-likelihood estimates in factor analysis: the case of nearly singular matrix of estimates of unique variances. **Linear Algebra And its Applications**, Amsterdam, v. 321, n. 1-3, p. 153-173, dez. 2000.
- HENZE, N.; ZIRKLER, B. A class of invariant consistent tests for multivariate normality. **Communications In Statistics: theory and methods**, New York, v. 19, n. 10, p. 3595-3617, jan. 1990.
- HORN, J. L. A rationale and test for the number of factors in factor analysis. **Psychometrika**, New York, v. 30, n. 2, p. 179-185, jun. 1965.
- HOX, J. J.; BECHGER, T. M. An introduction to structural equation modeling. **Family Science Review**, Weber, v. 11, p. 354-373, 1998.
- HOYSTON, J. P. Some techniques for assessing multivariate normality based on the Shapiro-Wilk W. **Applied Statistics - Journal of The Royal Statistical Society - Série C.**, London, v. 32, n. 2, p. 121-133, 1983.
- HU, L.; BENTLER, P. M.; KANO, Y. Can test statistics in covariance structure analysis be trusted?. **Psychological Bulletin**, Washington, v. 112, n. 2, p. 351-362, 1992. DOI: 10.1037/0033-2909.112.2.351.
- HUBERT, M.; ENGELN, S. Robust PCA and classification in biosciences. **Bioinformatics**, Oxford, v. 20, n. 11, p. 1728-1736, 2004.
- INCE, D. C.; HATTON, L.; GRAHAM-CUMMING, J. The case for open computer programs. **Nature**, New York, v. 482, n. 7386, p. 485-488, 2012.
- JIANG, P.; THELEN, K. D. Effect of soil and topographic properties on crop yield in a north-central corn-soybean cropping system. **Agron. J.**, Madison, v. 96, n. 1, p. 252-258, 2004.
- JOHNSON, D. R.; CREECH, J. C. Ordinal Measures in Multiple Indicator Models: A Simulation Study of Categorization Error. **American Sociological Review**, Washington, v. 48, n. 3, p. 398-407, 1983.
- JOHNSON, R. A.; WICHERN, D. W. **Applied multivariate statistical analysis**. 3. ed. Englewood Cliffs: Prentice Hall, 1992.
- JOHNSON, R. A.; WICHERN, D. W. **Applied multivariate statistical analysis**. New Jersey: Pearson: Prentice Hall, 2002.

JOHNSON, R. A.; WICHERN, D. W. **Applied multivariate statistical analysis**. 6. ed. New Jersey: Pearson: Prentice Hall, 2007. 800 p.

JÖRESKOG, K. G. A general method for estimating as linear structural equation system. In: GOLDBERGER, A. S. D.; OTIS, D. (Ed.). **Structural equation models in the social sciences**. New York: Academic, 1973. p. 85-112.

JÖRESKOG, K. G.; SÖRBOM, D. **LISREL8: user's reference guide**. Chicago: Licolnwood: Scientific Software Internacional, 1996.

JOSÉ, J. V.; WOLFF, W.; DIOTTO, A.V.; ALVES, D. S.; MARQUES, P. A. A.; GAVA, R. Análise multivariada de variáveis físico-hídricas de diferentes solos. In: WORKSHOP DE INOVAÇÕES TECNOLÓGICAS NA IRRIGAÇÃO, 4., 2012, Fortaleza. **Anais...** Fortaleza: INOVAGRI, 2012. p. 1-5.

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

KAISER, H. F. A second generation little jiffy. **Psychometrika**, New York, v. 35, p. 401-415, 1970.

KEESLING, J. W. **Maximum likelihood approaches to causal analysis**. 1972. **n. f.?** Tese (Doutorado **em?**) - University Of Chicago, Chicago, 1972.

KLINE, R. B. **Principles and practice of structural equation modeling**. 3. ed. New York: Guilford, 2010. 427 p.

LATIF, S. A. **Modelagem de equações estruturais**. 2000. 98 f. Dissertação (Mestrado em Estatística) - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2000.

LAROS, J. A. O uso da análise fatorial: algumas diretrizes para pesquisadores. In: PASQUALI, L. (Org.) **Análise fatorial para pesquisadores**. Petrópolis: Vozes, 2004.

LAWLEY, D. N. The estimation of factor loadings by the method of maximum likelihood. **Proc. Roy. Soc. Edingburgh - Série A**, London, v. 60, n. 1, p. 64–82, 1940.

LAWLEY, D. N. Some new results in maximum likelihood factor analysis. **British Journal Of Statistical Psychology**, New Jersey, v. 11, p. 1-12, 1958.

MALHOTRA, N. K. **Pesquisa de marketing: uma orientação aplicada**. 6. ed. Porto Alegre: Bookman, 2011. 768 p.

MARDIA, K. V. Measures of multivariate skewness and kurtosis with applications. **Biometrika**, London, v. 57, n. 3, p. 519-530, dez. 1970.

MARÔCO, J. **Análise de equações estruturais: fundamentos teóricos, software e aplicações**. Pero Pinheiro: Report Number, 2010. 384 p.

MARTÍN, N. F.; BOLLERO, G.; BULLOCK D. G. Associations between field characteristics and soybean plant performance using canonical correlation analysis. **Plant and Soil**, Switzerland, v. 273, n. 1, p. 39–55, jun. 2005.

MATTOS, A. **Validade dimensional da escala de apoio social do Medical Outcomes Study adaptada para o português no Estudo Pró-Saúde**. 2009. 51 f. Dissertação (Mestrado em Saúde Pública) - Escola Nacional de Saúde Pública, Fundação Oswaldo Cruz, Rio de Janeiro, 2009.

MAXWELL, A. E. Calculating maximum likelihood factor loadings. **Journal Of The Royal Statistical Society - Série A**, London, v. 127, n. 2, p. 238-241, 1964.

MINGOTI, S. A. **Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada**. Belo Horizonte: Editora UFMG, 2005.

MONTANARI, R.; ZAMBIANCO, E. C.; CORRÊA, A. R.; PELLIN, D. M. P.; CARVALHO, M. P.; DALCHIAVON, F. C. Atributos físicos de um Latossolo Vermelho correlacionados linear e espacialmente com a consorciação de guandu com milheto. **Revista Ceres**, Viçosa, MG, v. 59, n. 1, p. 407-410, 2012.

MUTHÉN, B.; KAPLAN, D. A comparison of some methodologies for de factor anaysis of non-normal Likert variables. **British Journal of Mathematical and Statistical Psychology**, New Jersey, v. 35, n. 1, p. 171-189, maio 1985.

NAGA, R. A.; ANTILLE, G. Stability of robust and non-robust principal components analysis. **Computational Statistics & Data Analysis**, Amsterdam, v. 10, n. 2, p.169-174, 1990.

NEWSOM, J. T. **Practical approaches to dealing with nonnormal and categorical variables**. [S. l.: s. n.], 2015. Disponível em: <http://www.upa.pdx.edu/IOA/newsom/semclass/ho_estimate2.pdf>. Acesso em: 11 fev. 2016.

OLSSON, U. Maximum likelihood estimation of the polychoric correlation coefficient. **Psychometrika**, New York, v. 44, n. 4, p. 443-460, 1979.

PETERNELLI, L. A.; MELLO, M. P. **Conhecendo o R: uma visão estatística**. Viçosa, MG: Editora UFV, 2007.

PETERSEN, K. B.; PEDERSEN, M. S.. **The matrix Cook Book**. [S. l.: s. n.], 2012. Disponível em: <<http://www.math.uwaterloo.ca/~hwolkowi/matrixcookbook.pdf>>. Acesso em: 19 ago. 2015

PISON, G.; ROUSSEEUW, P. J.; FILZMOSER, P.; CROUX, C. Robust factor analysis. **Journal of Multivariate Analysis**, Amsterdam, v. 84, n. 1, p. 145-172, 2003.

R CORE TEAM. **R: a language and environment for statistical computing**. Vienna: R Foundation for Statistical Computing, 2014. Disponível em: <<http://www.R-project.org/>>. Acesso em: 19 ago. 2015.

RAZAFIMBELO, T.; BARTHÈS, B. G.; LARRÉ-LARROUY, M. C.; DE LUCA, E. F.; LAURENT, J. Y.; CERRI, C. C.; FELLER, C. Effect of sugarcane residue management (mulching versus burning) on organic matter in a clayey Oxisol from southern Brazil. **Agriculture, Ecosystems & Environment**, Amsterdam, v. 115, n. 1-4, p. 285-289, 2006.

RAYKOV, T.; MARCOULIDES, G. A. **A first course in structural equation modeling**. 2. ed. New Jersey: Lawrence Erlbaum Associates, 2006. 235 p.

RENCHE, A. C. **Methods of multivariate analysis**. 2. ed. New York: John Wiley, 2002. 708 p.

SARTORIO, S. D. **Aplicações de técnicas de análise multivariada em experimentos agropecuários usando o software R**. 2008. 130 f. Dissertação (Mestrado em Agronomia) - Escola Superior de Agricultura Luiz de Queiroz – ESALQ, Universidade de São Paulo - USP, Piracicaba, 2008.

SATORRA, A.; BENTLER, P. M. Scaling corrections for chi-square statistics in covariance structure analysis. In: BUSINESS AND ECONOMIC STATISTICS SECTION OF THE AMERICAN STATISTICAL ASSOCIATION, 1988, Alexandria. **Proceedings...** Washington: [s. n.], 1988. p. 308-313.

SATORRA, A.; BENTLER, P. M. A scaled difference chi-square test statistic for moment structure analysis. **Psychometrika**, New York, v. 66, n. 4, p.507-514, 2001.

SILVA, J. S. F. **Modelagem de equações estruturais: apresentação de uma metodologia**. 2006. 105 f. Dissertação (Mestrado em Engenharia de Produção) - Universidade Federal do Rio

Grande do Sul, Porto Alegre, 2006. Disponível em: <http://www.lume.ufrgs.br/handle/10183/8628?locale=pt_BR>. Acesso em: 14 abr. 2015.

SPEARMAN, C. General intelligence, objectively determined and measured. **The American Journal of Psychology**, Champaign, v. 15, p. 201-292, 1904.

TREIBLMAIER, H.; FILZMOSER, P. Exploratory factor analysis revisited: How robust methods support the detection of hidden multivariate data structures in is research. **Information & Management**, Amsterdam, v. 47, n. 4, p. 197-207, 2010.

UNIÃO DA INDÚSTRIA DE CANA-DE-AÇÚCAR – ÚNICA. **UNICADATA produção:** área cultivada com cana-de-açúcar em 2015. São Paulo, 2015. Disponível em: <<http://www.unicadata.com.br/>>. Acesso em: 15 mar. 2016.

VALADARES, F. G.; AQUINO, A. L. L.; PEREIRA JUNIOR, Á. R. Detecção de outliers multivariados em redes de sensores. In: CONGRESO LATINO-IBEROAMERICANO DE INVESTIGACIÓN OPERATIVA/ SIMPÓSIO BRASILEIRO DE PESQUISA OPERACIONAL, 16., 2012, Rio de Janeiro. **Anais...** . Rio de Janeiro: Claio/sbpo, 2012. p. 1351 - 1360.

WILEY, D. E. The identification problem for structural equation models with unmeasured variables. In: GOLDBERGER, A. S.; DUNCAN, O. D. (Ed.) **Structural equation models in the social sciences**. New York: Academic Press, 1973. p. 69 – 83,

WILKINSON, L.; TASK FORCE ON STATISTICAL INFERENCE (TFSI). Statistical methods in psychology journals: guidelines and explanations. **American Psychologist**, Washington, v. 54, n. 8, p. 594-604, 1999.

WRIGHT, S. Correlation and causation. **Journal of Agricultural Research**. Washington, v. 20, p. 557-585, 1921.

WRIGHT, S. The method of path coefficients. **Annals of Mathematical Statistics**, Beachwood, v. 5, n. 3, p. 161-215, 1934.

CAPÍTULO 2 – MODELAGEM DAS RELAÇÕES ENTRE OS ATRIBUTOS DO SOLO E COMPONENTES DE PRODUÇÃO DA CANA-DE-AÇÚCAR USANDO ANÁLISE DE EQUAÇÕES ESTRUTURAIS.

RESUMO

A cana-de-açúcar é a principal fonte de energia renovável no Brasil e com um futuro promissor em todo o mundo, tanto na questão econômica quanto na ambiental, o que justifica cada vez mais as pesquisas sobre o entendimento do seu comportamento no sistema solo-planta. Objetivou-se explicar a produtividade da cultura da cana-de-açúcar por meio de fatores latentes formados a partir de atributos químicos e físicos de um Argissolo Vermelho Distrófico abrupto textura média/argilosa, simultaneamente, utilizando a técnica de modelagem de equações estruturais. Para a coleta de dados da planta e do solo, foram definidos 118 pontos amostrais de uma grade regular, nas profundidades 0 a 0,20 m e 0,20 a 0,40 m. Os três primeiros fatores compuseram os atributos químicos, físicos e de produção com explicação em torno de 60 % da variabilidade dos dados. Os componentes de produção e tecnológicos da cana-de-açúcar foram influenciados significativamente e negativamente pelo fator latente atributos físicos na camada de 0 a 0,20. Entretanto, houve influência significativa e positiva do fator latente atributos químicos associados à acidez, na camada de 0,20 a 0,40 m.

Palavras chave: Fatores. Modelos. Biomassa. Argissolo.

ABSTRACT

The sugarcane is the main source of renewable energy in Brazil and with a promising future in the world, both in the economic issue as environmental, which justifies more and more research on the understanding of their behavior in the system soil-plant. It aimed to explain the crop yield of sugarcane by latent factors formed from chemical and physical attributes of one Dystrophic Red Argisol, with medium/clay texture, simultaneously, using the technique of structural equation modeling. To collect data from plant and soil, were defined 118 sampling points the a regular grid defined at the depths from 0 to 0.20 m and 0.20 to 0.40m. The first three factors it composed the attributes chemical, physical and production with explanation 60% of the variability of the data. The components of production and technology of sugarcane were significantly and negatively influenced by latent factor physical attributes in layer 0 to 0.20 m. However, there was a significant and positive influence of latent factor chemical attributes associated with acidity, at the layer from 0.20 to 0.40 m.

Keywords: Factors. Models. Biomass. Argisol.

2.1 - INTRODUÇÃO

A crescente preocupação mundial com a proteção do meio ambiente vem gerando pressões sobre o uso de combustíveis fósseis, que são os grandes responsáveis pela emissão de gases poluentes para a atmosfera. Vários países têm buscado reduzir o uso desses combustíveis para diminuir a carga poluidora, sendo uma fonte alternativa os combustíveis não fósseis, como os biocombustíveis, derivados de biomassa renovável (MAULE; MAZZA; MARTHA JUNIOR, 2001; VAKKILAINEN; KUPARINEN; HEINIMÖ, 2013).

Nesse cenário, a cana-de-açúcar é a cultura mais importante como fonte primária de energia do Brasil, estando entre as melhores opções dentre as fontes de energia renováveis como a biomassa, com grande importância no cenário agrícola brasileiro, bem como um futuro promissor no cenário mundial (EPE, 2008; MARTINS et al., 2015). No entanto, de acordo com Ololade et al. (2010), a produção agrícola é uma complexa interação entre as variáveis ambientais, os atributos do solo e a dinâmica de nutrientes no sistema solo-planta, assim, novos estudos sobre o potencial produtivo da cultura da cana-de-açúcar em solos brasileiros são necessários.

Avaliar o enfraquecimento da fertilidade do solo, incluindo esgotamento de nutrientes (diminuição do pH e/ou um aumento de alumínio trocável), perda do teor de matéria orgânica do solo e aumento dos teores de elementos tóxicos é difícil, pois a maioria dos atributos químicos do solo quer mudam muito lentamente ou têm grandes flutuações sazonais. Neste sentido, o pH é talvez o atributo mais importante ligado ao uso eficiente de fertilizantes, sendo que a disponibilidade de macronutrientes pode mudar de baixa (solo ácido) para alta (pH 6,0 a 7,0); entretanto, a disponibilidade de ferro, cobre, manganês e zinco é maior sob condições ácidas, diminuindo com a elevação do pH (LOPES; GUILHERME, 2000; HARTEMINK, 2006).

Atributos que também podem influenciar no crescimento e desenvolvimento da planta são aeração, teor de água no solo, temperatura do solo, resistência do solo à penetração de raízes, sendo que estes, por sua vez, são alterados por mudanças na porosidade total, estabilidade de agregados e densidade do solo (SILVA; TORMENA; IMHOFF, 2002; BRAUNACK, 1991).

Segundo Maule; Mazza e Martha Júnior (2001), estudar a cultura no seu ambiente de desenvolvimento pode gerar uma enorme quantidade de informações para adequar o melhor manejo e cultivar para os específicos ambientes (solo e clima), sendo então, possível explorar ao máximo o local de produção para promover o melhor rendimento da cultura e

consequentemente maior produtividade ou maior competitividade para as agroindústrias da cana-de-açúcar. Nesse contexto, o solo é um dos componentes de um conjunto complexo de fatores que influenciam a produção da cultura, destacando-se pelo seu importante papel de fornecer às plantas o suporte físico, a água e os nutrientes necessários para o bom desenvolvimento da cultura. Portanto, o conhecimento dos atributos inerentes a cada solo, os chamados fatores edáficos, é importante para julgar o potencial de produção agrícola (LEPSCH, 1987).

Vários estudos têm utilizado a ferramenta geoestatística, correlações e regressões, com o objetivo de estudar a variabilidade espacial entre atributos físicos, químicos e biológicos do solo e de produtividade das culturas agrícolas, bem como avaliar as relações entre eles, pode-se citar Dias (1999), Gioia (2011) e Sória (2014).

Deste modo, a necessidade de modelos matemáticos mais complexos relacionando a produtividade da cultura com os atributos do solo são necessários. Nazmi (2013) utilizou modelos de regressão linear múltipla e análise de equações estruturais com o objetivo de relacionar fatores latentes (fatores não observáveis no campo e associados a atributos observados no campo, que estejam relacionados), e o tamanho da influência desses nos atributos físicos, químicos e de produção do trigo; observou que as mensurações dos atributos físicos e químicos foram estatisticamente significativas na predição e no entendimento dos componentes de produção de trigo por meio do modelo de regressão e do modelo estrutural.

Freitas et al. (2014) se beneficiaram das técnicas multivariadas para avaliar de forma comparativa os atributos físicos e químicos, em função de classes texturais nas profundidades 0 a 0,10 m e 0,10 a 0,20 m, sob cultivo contínuo de cana-de-açúcar. O estudo mostrou que áreas com menor teor de argila apresentaram comportamento diferenciado de áreas com maior teor de argila.

Diante desse contexto, o objetivo do trabalho foi explicar a produtividade da cultura da cana-de-açúcar por meio de fatores latentes, formados a partir de atributos químicos e físicos do solo, em diferentes profundidades de avaliação. Para tanto, foram utilizados os métodos multivariados: análise dos componentes principais e análise fatorial para reduzir a dimensionalidade dos dados, bem como para auxiliar na formação de cada fator; e aplicação da análise de equações estruturais, visando obter uma melhor caracterização dos componentes de desempenho da cultura da cana-de-açúcar associados aos atributos do solo.

2.2 - MATERIAL E MÉTODOS

Área experimental

O experimento foi conduzido em área de plantio comercial de cana-de-açúcar de 10,5 ha, localizada no município de Suzanópolis, SP, Brasil, com coordenadas 20°29'54" de latitude Sul, 51°01'38" de longitude Oeste, com elevação de 350 m acima do nível do mar (Figura 3). A classificação climática, de acordo com Köppen, é Aw, caracterizado como tropical úmido com verão chuvoso e inverno seco, típico da região do cerrado brasileiro, com temperatura média anual de 24,5 °C e precipitação total média anual de 1400 mm. A classificação climática da região foi C_{2r}A'a' de acordo com o sistema de Thornthwaite (ROLIM et al., 2007). O solo da área foi classificado como Argissolo Vermelho Distrófico abrupto textura média/argilosa, moderado, hiperdistrófico, meso-epieutrófico, endoálico, endodêmico, epicompactado, muito profundo, fortemente ácido (ALVES et al. 2009). Na Tabela 2 estão apresentados os dados da caracterização morfológica e analíticos do solo estudado, obtidos por ocasião do Levantamento Taxonômico Semidetalhado dos Solos da Usina Vale do Paraná S/A Álcool e Açúcar (ALVES et al. 2009).

Para a instalação da cultura foi utilizada a variedade RB 85 5453 de cana-de-açúcar. A operação de preparo do solo foi realizada por meio de uma gradagem pesada com duas passadas da grade média. Foi realizada a distribuição a lanço de 2,0 t ha⁻¹ de calcário dolomítico incorporado com um arado de aiveca seguido de uma gradagem leve. A operação de plantio foi realizada no mês de junho de 2009, por meio de sulcador de duas linhas, no espaçamento de 1,50 m e com adubação de plantio na dose de 500 kg ha⁻¹ da fórmula 6-30-24 e para cobertura dos toletes no sulco de plantio. Após o primeiro corte, realizou-se aplicação de 1 t.ha⁻¹ de gesso. Para a coleta de dados, a cultura encontrava-se no terceiro corte (24/05/2014). A operação de colheita ocorreu na forma mecanizada.

Para a coleta de dados da planta e do solo, foram definidos 118 pontos amostrais, distribuídos na área de maneira homogênea, caracterizando uma malha georreferenciada com distância de 34 m entre os pontos amostrais da grande malha e 17 m na malha de refinamento, de forma a cobrir a área do talhão escolhido (Figura 3). Os atributos da planta (componentes de produção e tecnológicos) estudados foram: população de plantas (POP), expresso em pl.m⁻²;

Tabela 2. Descrição morfológica e dados analíticos do solo estudado

Horizonte	Profundidade	Descrição
Ap	0-33 cm	bruno-avermelhado escuro (2,5YR 3/4, úmida); média; moderada média blocos angulares; dura, friável, ligeiramente plástica e ligeiramente pegajosa; transição ondulada e abrupta.
BA	33-63 cm	vermelho escuro (2,5YR 3/6, úmida); média; moderada grande laminar; dura, friável, plástica e pegajosa; transição ondulada e abrupta.
Bt1	63-124 cm	vermelho (2,5YR 4/8, úmida); argilosa; moderada grande blocos angulares; ligeiramente dura, firme; plástica pegajosa; transição plana e gradual.
Bt2	124-164 cm	vermelho (2,5YR 4/8, úmida); argilosa; moderada grande blocos angulares; ligeiramente dura, friável, plástica e pegajosa; transição plana e gradual.
Bt3	164- 200 cm+	vermelho (2,5YR 4/8, úmida); argilosa; moderada média blocos angulares; ligeiramente dura, friável, plástica e pegajosa.

Horizonte		Composição granulométrica da TFSA			Relação
Símbolo	Profundidade	areia	silte	Argila	silte/argila
	(cm)g kg ⁻¹			
Ap	0-33	725	115	160	0,72
BA	33-63	537	120	343	0,35
Bt1	63-124	569	78	353	0,22
Bt2	124-164	557	74	369	0,20
Bt3	164- 200 +	541	74	385	0,19

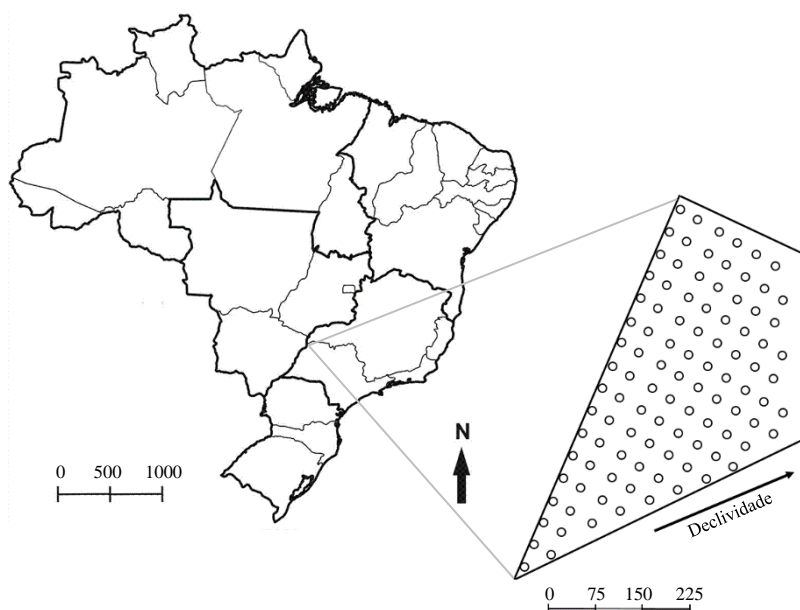
Horizonte		Atributos físicos				
Símbolo	Profundidade	Densidade do solo	Densidade da partícula	Macroporosidade	Microporosidade	Porosidade total
	(cm)kg dm ⁻³m ³ m ⁻³			
Ap	0-33	1,56	2,59	0,07	0,33	0,40
BA	33-63	1,58	2,55	0,06	0,29	0,35
Bt1	63-124	1,50	2,58	0,06	0,30	0,37
Bt2	124-164	1,44	2,62	0,09	0,34	0,43
Bt3	164- 200 +	1,49	2,59	0,08	0,33	0,41

Horizonte		Atributos químicos											
Símbolo	Profundidade	P	MO	pH	K ⁺	Ca ²⁺	Mg ²⁺	H ⁺	Al ³⁺	SB	T	V%	m%
	(cm)	mg dm ⁻³	g.dm ⁻³	CaCl ₂mmolc dm ⁻³								
Ap	0-33	5,0	17	4,5	2,6	17	10	21	3	30	54	56	10
BA	33-63	2,8	10	4,5	1,6	13	10	13	13	24	50	48	36
Bt1	63-124	1,4	5	4,2	1,2	6	6	15	18	13	46	29	57
Bt2	124-164	1,4	3	4,2	0,6	2	7	14	16	10	40	26	60
Bt3	164- 200 +	1,4	5	4,2	0,6	2	5	4	20	7	31	23	74

Fonte: Alves et al. (2009).

tonelada de cana-de-açúcar por hectare (TCH), expresso em $t\ ha^{-1}$; açúcares totais recuperáveis (ATR), expresso em $kg\ t^{-1}$; sólidos solúveis totais (BRI), sacarose no caldo (POL), pureza aparente (PUR) e fibra (FIB), expressos em % (BIDOIA, 2008; CONSECAN, 2006).

Figura 3 - Mapa esquemático da área experimental de 10,5 ha e distribuição de 118 pontos amostrais na área de cana-de-açúcar no município de Suzanápolis, SP, Brasil.



Fonte: Dados do próprio autor.

Para a obtenção dos atributos físicos do solo (29/05/2014), foram determinados: densidade do solo (DS) ($kg\ dm^{-3}$), resistência à penetração do solo (RP) (MPa), umidade gravimétrica do solo (UG) ($kg\ kg^{-1}$), umidade volumétrica do solo (UV) ($m^3.m^{-3}$), determinada pelo fórmula apresentada em Kiehl (1979): $UV = UG.DS$, macroporosidade do solo (MA) ($m^3\ m^{-3}$), microporosidade do solo (MI) ($m^3\ m^{-3}$) e porosidade total do solo (PT) ($m^3\ m^{-3}$) (KIEHL, 1979; STOLF, 1991; EMBRAPA, 1997; MONTANARI, 2009). Os atributos químicos do solo avaliados foram carbono orgânico (CO) ($t\ ha^{-1}$), o pH em água (pHag), pH em cloreto de cálcio (pHCa), pH em cloreto de potássio (pHK) (RAIJ et al., 2001; EMBRAPA, 2009). Todos os atributos foram coletados nas profundidades 1 (0-0,20 m) e 2 (0,20-0,40 m). As determinações dos atributos do solo foram realizadas nos Laboratórios de Fertilidade do solo e Análises Físicas do Solo da Faculdade de Engenharia de Ilha Solteira – UNESP.

Análise dos dados

Na primeira parte da análise de dados, foi verificado o comportamento de cada atributo, por meio da análise descritiva dos dados, bem como a existência de valores discrepantes (outliers multivariados), além da avaliação da hipótese de normalidade multivariada e,

posteriormente, a aplicação dos métodos estatísticos multivariados: análise dos componentes principais (ACP) e análise fatorial robusta (AFR) nas duas profundidades estudadas. Para verificação da existência de outliers multivariados, utilizou-se o gráfico bidimensional *Q-Q Plot*, com base na distância de mahalanobis robusta (FILZMOSE; GARRETT; REIMANN, 2005).

Em seguida foram aplicados os métodos estatísticos multivariados ACP e AFR. O método ACP (74) consistiu na realização de uma rotação rígida no sistema de eixos coordenados originais de modo que os novos eixos estivessem na direção de maior variabilidade dos dados, em que os coeficientes dos novos eixos são os autovetores da matriz de covariância amostral dos dados. Na análise fatorial (76), explicou-se a covariância entre o conjunto de atributos do estudo, em termos limitados de fatores não observáveis; e, para maior robustez, foi utilizado o Método do Determinante Mínimo – MCD (*Minimum Covariance Determinant*) para estimar a matriz de covariância amostral, cujo objetivo foi encontrar um subconjunto h de observações que torne o determinante da matriz de covariância clássica mínimo, sendo o estimador de locação a média destes h pontos, enquanto o estimador de escala sua matriz de covariância; o método de estimação foi o fator principal (PISON et al., 2003).

Para uma melhor interpretação dos resultados da AFR, foi utilizado o método de rotação oblíqua (*oblimin*) uma vez que foi suposta a existência de associação entre os fatores. Os números de componentes e fatores da ACP e AFR, respectivamente, foram determinados pelos critérios de Kaiser (1958): fatores com autovalor maiores ou iguais a 1 (um), da análise paralela de Horn (1965): o valores médios dos autovalores gerados por simulação comparados aos autovalores reais; além disso, para os componentes e fatores selecionados na ACP e AFR, consideraram-se apenas os atributos associados a estes componentes com coeficientes maiores que 0,35 e fatores com cargas fatoriais maiores que 0,4; sendo a adequação aos dados avaliada por meio do critério de Kaiser – Meyer – Olkin (KMO): leva em conta a inversa da matriz de correlação próxima de uma matriz diagonal; quando mais próximo de 1 (um), maior adequabilidade.

Sendo $X = [X_1, X_2, \dots, X_p]^T$ um vetor aleatório com p atributos, matriz de covariância amostral $S_{p \times p}$ com autovalores $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$. A equação (74) descreve o modelo ACP, em sua forma matricial:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_p \end{bmatrix} = \begin{bmatrix} e_{11} & e_{12} & \cdots & e_{1p} \\ e_{21} & e_{22} & \cdots & e_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ e_{p1} & e_{p2} & \cdots & e_{pp} \end{bmatrix}^T \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix} \Rightarrow Y = O^T \cdot X, \quad (74)$$

em que $e_i^T = (e_{i1}, e_{i2}, \dots, e_{ip})$ é um autovetor de dimensão $1 \times p$ com $e_i^T \cdot e_i = 1$, e $e_i^T \cdot e_j = 0$ sendo $i \neq j$ para a última condição e i e j variando de 1 a p para as duas condições.

Seja μ o vetor de médias de X_{px1} , matriz de variância e covariância amostral S_{pxp} e um vetor de fatores comuns latentes não observáveis F_{mx1} ; de acordo com Mingoti (2005) e Hair et al. (2009), o modelo de análise fatorial pode ser descrito pela equação (75):

$$\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix}_{px1} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix}_{px1} + \begin{bmatrix} l_{11} & l_{12} & \cdots & l_{1m} \\ l_{21} & l_{22} & \cdots & l_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ l_{p1} & l_{p2} & \cdots & l_{pm} \end{bmatrix}_{pxm} \begin{bmatrix} F_1 \\ F_2 \\ \vdots \\ F_m \end{bmatrix}_{mx1} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_p \end{bmatrix}_{px1}, \quad (75)$$

em que l_{ij} são as cargas fatoriais, F_j são os fatores comuns e ε_i são os erros aleatórios ou fatores específicos, em que $j = 1, \dots, m$ e $i = 1, \dots, p$.

E, então, o modelo (75) pode ser reescrito como:

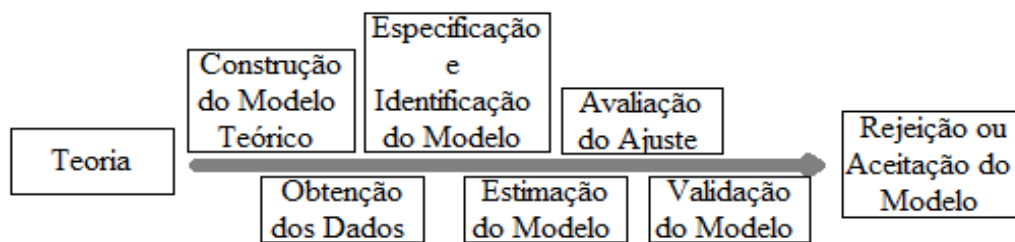
$$X = \mu + LF + \varepsilon \Rightarrow X - \mu = LF + \varepsilon, \quad (76)$$

em que a matriz L_{pxm} é a matriz das cargas fatoriais associadas i -ésima variável X_i e ao j -ésimo fator F_j .

Na segunda parte de análise, cada grupo de atributos selecionados de seus respectivos fatores ou componentes e dado o conhecimento de estudos anteriores, foi associado a um fator latente e o modelo conceitual da análise de equações estruturais (AEE) foi construído (BRAHIN et al., 2011).

A AEE é uma técnica de modelagem generalizada, cujo objetivo foi testar e validar modelos teóricos, que definem relações causais e hipotéticas entre as variáveis estudadas. Tais relações foram medidas pelos parâmetros do modelo que representam o tamanho do efeito dos atributos independentes sobre os atributos dependentes (MARÔCO, 2010). Na Figura 4 estão os procedimentos da aplicação da AEE.

Figura 4 - Etapas da Modelagem de Equações Estruturais.



Fonte: Marôco (2010)

O modelo geral da análise de equações estruturais é composto dos modelos de medidas, os quais relacionam os atributos observados no campo a seus respectivos fatores latentes, e, um modelo de regressão estrutural, que representa as relações entre os fatores latentes.

Para evitar problemas de desvios de normalidade multivariada, foi utilizado o método de estimação da máxima verossimilhança com correção ou ajustamento da estatística χ^2 , bem como erros padrão robustos, calculados com base na curtose multivariada (SATORRA; BENTLER, 1988; 2001). As medidas utilizadas para avaliar a qualidade de ajustamento do modelo foram: χ^2 = valor qui-quadrado do teste de ajustamento global, χ^2/gl = razão entre valor qui-quadrado do teste de ajustamento global e os graus de liberdade do modelo, CFI = Índice de ajuste comparativo, GFI = Índice de qualidade do ajuste, TLI = Índice de Tucker-Lewis, PGFI = Índice de qualidade de ajuste de parcimônia, RMSEA = raiz do erro quadrático médio de aproximação.

As análises de dados foram conduzidas para as camadas de 0,00 – 0,20 m e de 0,20 – 0,40 m, separadamente, com intuito de comparar os resultados entre as camadas do solo. Todas as análises foram realizadas por meio do software estatístico R, o qual é gratuito e apresenta código-fonte-aberto.

2.3 - RESULTADOS E DISCUSSÃO

Análise preliminares

Inicialmente, os dados foram analisados em termos de sua estatística descritiva (Tabela 3). Pode-se observar que os valores de assimetria e curtose dos atributos na profundidade 0 – 0,20 m ficaram próximos de 0,00 e 3,00, respectivamente. Isto é, as distribuições de probabilidade dos atributos se aproximaram ou possuem distribuição normal. A única exceção é para o atributo CO, que apresentou assimetria 0,29 e curtose 2,37, o que indica uma distribuição de probabilidade com assimetria a direita ($0,29 > 0,00$) e forma da distribuição mais achatada ou platicúrtica ($2,37 < 3,00$), o que mostra um desvio de normalidade.

No que tange a profundidade 0,20 – 0,40 m (Tabela 3), notaram-se valores atípicos de assimetria e curtose conjuntamente para as variáveis pHCa e CO, com assimetria 0,32 e 0,62, respectivamente, e curtose 2,58 e 3,89, respectivamente, o que indica uma distribuição com leve assimetria a direita ($0,32 > 0$) e achatada ou platicúrtica ($2,58 < 3,00$) para pHCa e distribuição com assimetria a direita ($0,62 > 0$) e leptocúrtica ($3,89 > 3,00$) para CO, permitindo-se dizer que estes dois atributos possuem algum desvio de normalidade.

Cruz et al. (2010) obtiveram comportamento semelhante na distribuição das mensurações de carbono orgânico quando estudaram análise espacial de atributos físicos e carbono orgânico em um argissolo vermelho-amarelo cultivado com cana-de-açúcar na profundidade 0 a 0,20 m, com curva assimétrica a direita (0,12) e forma achatada ou platicúrtica.

Os atributos químicos (Tabela 3), apresentaram mensurações maiores na camada de 0,20 a 0,40m do que na camada de 0 a 0,20 m, para pHCa, pHK e pHag. Tais resultados corroboram com os de Freitas et al. (2010), quando avaliaram propriedades físicas e químicas de um argissolo vermelho na reforma de canavial para a produção de oleaginosas, cujos valores de pH em água foram 6,3 na camada de 0 a 0,10 m; 6,2 na camada de 0,10 a 0,26 e 6,7 de 0,26 a 0,42 m, mostrando um aumento do pH das camadas superficiais para a camada subsuperficial.

Em Carvalho et al (2013), estudou-se a produtividade da cana soca sem queima em função do uso do gesso e vinhaça de uma latossolo vermelho-amarelo distrófico, textura média; observaram incremento de 0,4 unidades do pH em cloreto de cálcio (passou de 4,3 para 4,7) na camada de 0,20 a 0,40, comparando-se sem e com a aplicação de gesso; valor considerado ideal para a cultura da cana-de-açúcar, que segundo os autores estaria no intervalo de 4,5 a 5,0. O fato constata a importância da gessagem na redução da acidez do solo nas camadas subsuperficiais.

Tabela 3 - Estatística descritiva para os atributos de planta e de solo avaliados no experimento, nas duas profundidades.

Variável	Profundidade de 0,00-0,20 m						Profundidade de 0,20-0,40 m					
	Média	Mediana	Desvio padrão	CV	Assimetria	Curtose	Média	Mediana	Desvio padrão	CV	Assimetria	Curtose
TCH (t ha ⁻¹)	89,51	90,34	20,62	23,04	0,14	3,02	89,51	90,34	20,62	23,04	0,14	3,02
POP (pl m ⁻²)	6,58	6,56	1,37	20,89	0,05	3,68	6,58	6,56	1,37	20,88	0,05	3,68
ATR (kg t ⁻¹)	136,00	136,65	16,18	11,90	- 0,46	3,14	136,00	136,65	16,18	11,90	- 0,46	3,15
BRI (%)	19,59	19,38	2,21	11,26	0,17	3, 11	19,59	19,38	2,21	11,26	0,17	3,11
PUR (%)	85,48	85,68	3,22	3,77	- 0,44	2,71	85,48	85,69	3,22	3,77	-0,44	2,71
DS (kg dm ⁻³)	1,56	1,58	0,09	5, 93	-0,24	2,35	1,63	1,62	0,07	4,54	0,24	2,68
UV (m ³ .m ⁻³)	0,14	0,14	0,03	23,23	0,01	2,67	0,09	0,09	0,02	23,91	0,21	3,01
MA (m ³ m ⁻³)	0,09	0,08	0,04	43,18	0,49	2,71	0,06	0,06	0,02	34,95	0,46	2,83
PT (m ³ m ⁻³)	0,36	0,36	0,34	8,65	-0,26	3,49	0,35	0,35	0,03	9,73	0,08	3,63
pHag	5,87	5,90	0,25	4,21	0,24	3,04	6,01	6,00	0,24	3,99	0,08	2,85
pHK	4,82	4,85	0,26	5,32	0,05	2,49	4,97	4,90	0,28	5,67	0,17	2,61
pHCa	4,64	4,70	0,23	5,01	-0,02	2,64	4,75	4,70	0,25	5,25	0,32	2,58
CO (t ha ⁻¹)	16,77	16,24	1,88	11,21	0,29	2,37	15,25	15,0	1,96	12, 86	0,62	3,89

Fonte: Dados do próprio autor.

TCH = tonelada de colmo por hectare, POP = população de plantas, ATR = açúcar total recuperável, BRI = brix, PUR = pureza, DS = densidade do solo, UV = umidade volumétrica, MA = macroporosidade, PT = porosidade total, pHag = potencial hidrogeniônico em água, pHK = potencial hidrogeniônico em potássio, pHCa = potencial hidrogeniônico em cálcio, CO = carbono orgânico.

No tocante aos atributos físicos (Tabela 3), observou valores médios de densidade do solo iguais a $1,56 \text{ kg dm}^{-3}$ de 0 a 0,20 m e $1,63 \text{ kg dm}^{-3}$ para a camada 0,20 a 0,40, com maior desvio padrão para a camada superficial, valores considerados altos e que evidenciam possível compactação do solo na área estudada. A macroporosidade foi menor que 0,10 nas duas profundidades, valor mínimo necessário para o desenvolvimento do sistema radicular. Os coeficientes de variação para macroporosidade nas duas profundidades foram considerados altos, com valores acima de 30%; resultados semelhantes aos encontrados em Carvalho et al. (2003) e Souza et al. (2006).

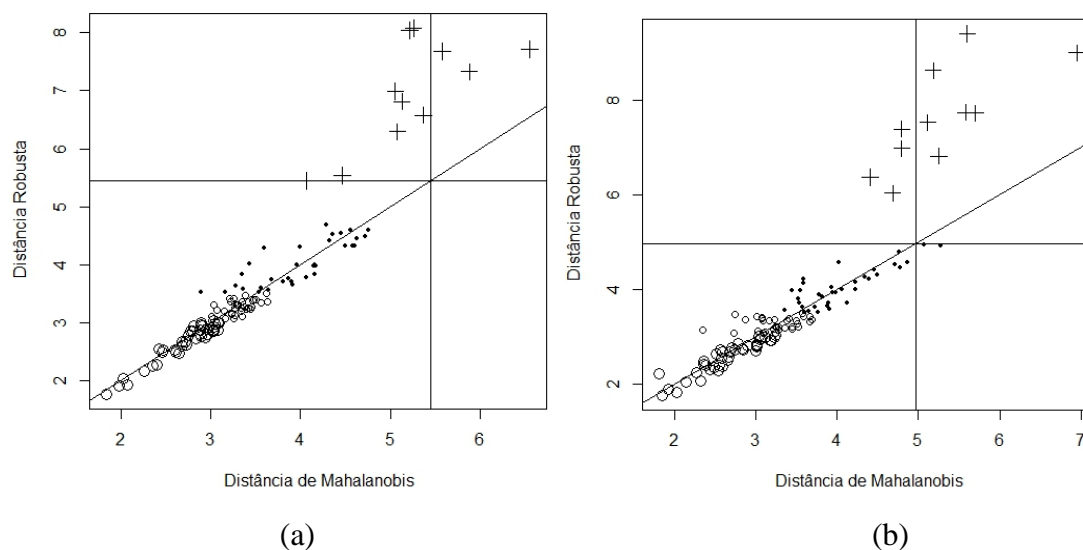
Oliveira Filho (2013), encontraram diferença significativa na avaliação dos atributos físicos do solo em diferentes anos de cultivo da cana-de-açúcar, com valores de densidade do solo acima de $1,5 \text{ kg dm}^{-3}$, bem como macroporosidades abaixo de $0,10 \text{ m}^3.\text{m}^{-3}$, em que detectou-se maior compactação com o passar dos anos. A colheita mecanizada da cana-de-açúcar, mesmo realizada na zona de friabilidade, pode provocar compactação adicional do solo (SEVERIANO et al., 2010).

A umidade volumétrica na profundidade 0 a 0,20 m foi $0,14 \text{ m}^3.\text{m}^{-3}$, enquanto que de 0,20 a 0,40 m o valor correspondente foi igual a $0,09 \text{ m}^3.\text{m}^{-3}$. Nota-se um decréscimo na umidade volumétrica da camada superficial para camada subsuperficial, concordando com Rodrigues (2014), quando estudou a variabilidade espacial de atributos físicos, químicos e biológicos de um latossolo vermelho eutroférico, textura argilosa e a produtividade de cana-de-açúcar.

Por fim, com relação médias dos atributos da planta (Tabela 3), observou-se uma produção de $89,51 \text{ t.ha}^{-1}$, ATR igual a 136 kg.t^{-1} , BRI (19,59 %) e pureza correspondente a 85,48%, valores inferiores aos encontrados por Alves et al. (2014), com produtividade média de $112,12 \text{ t.ha}^{-1}$, ATR igual a $160,68 \text{ kg.t}^{-1}$, BRI (22,52 %) e pureza mensurada em 86,60 %.

Observando a Figura 5, foram detectadas 11 observações candidatas a outliers multivariados pela distância de Mahalanobis Robusta (*Q-Q Plot*), porém, considerando o tamanho da amostra (118 observações), optou-se por não retirar tais observações e trabalhar com métodos robustos de estimação (PISON et al., 2003).

Figura 5- Gráfico para verificação de valores discrepantes (outliers multivariados) para as profundidades: a) 0,00-0,20 m e b) 0,20-0,40 m.

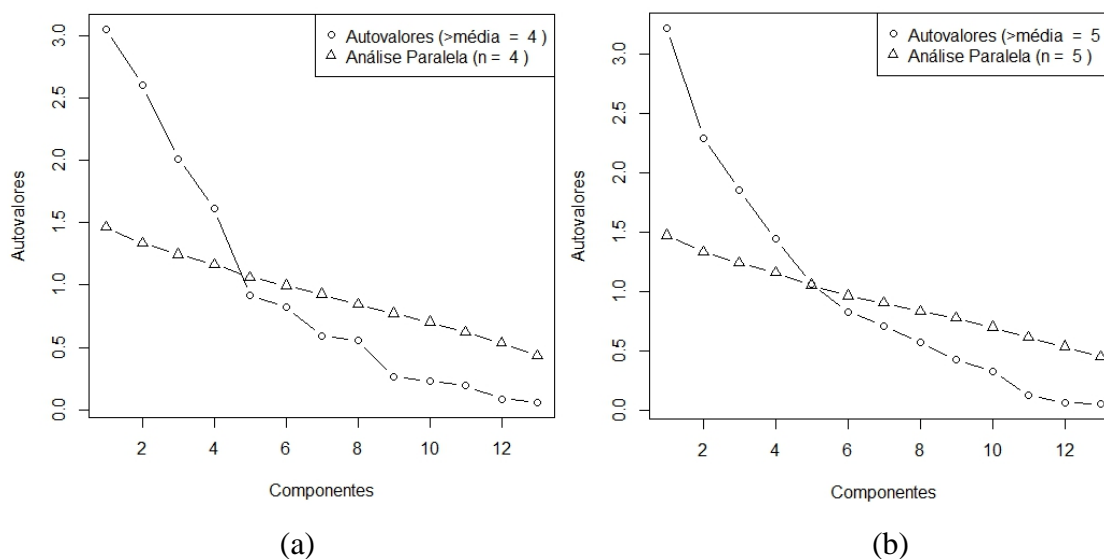


Fonte: Dados do próprio autor.

Análise fatorial e Componentes Principais

Os números de fatores e de componentes principais nas duas profundidades foram determinados pela análise paralela e pelo método de Kaiser, sobrepostos ao *screeplot* (Figura 6). Os resultados sugerem a utilização de 4 componentes e fatores para a profundidade 0 a 0,20 m e, para um modelo mais parcimonioso, optou-se por trabalhar com 4 componentes e fatores ao invés de 5 na profundidade 0,20 a 0,40 m (Figura 6).

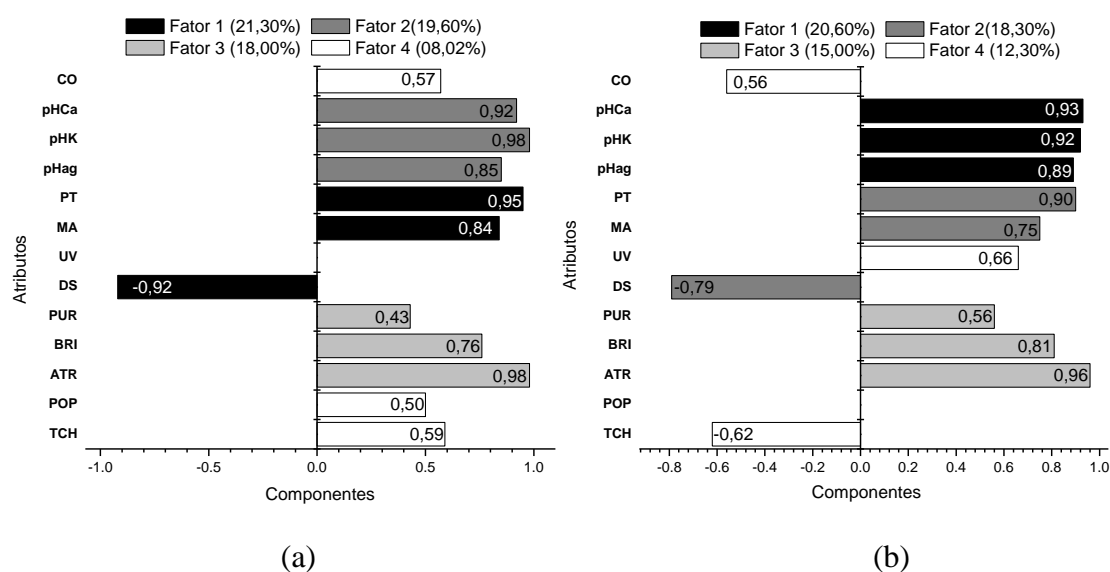
Figura 6 - Screeplots da análise de componentes principais para todos os atributos estudados nas profundidades: a) 0,00-0,20 m e b) 0,20-0,40 m.



Fonte: Dados do próprio autor.

Foram conduzidas as análises de componentes principais e análises fatoriais robustas, para auxiliar na formação dos fatores latentes na análise de equações estruturais (BRAHIN et al., 2011). Os resultados das duas técnicas foram similares para as profundidades 1 e 2, considerando os coeficientes das componentes principais maiores que 0,35 e das cargas fatoriais dos fatores maiores que 0,40, portanto, na Figura 7 são apresentados os resultados das análises fatoriais. Pode-se observar que para a profundidade 0 a 0,20 m, o Fator 1 reteve os atributos físicos DS, MA e PT, o Fator 2 referiu-se aos atributos químicos pHag, pHK, pHCa, para o Fator 3 foram os componentes de produção e tecnológicos da cana-de-açúcar, ATR, BRI e PUR e finalmente, o Fator 4 referiu-se aos atributos de produção TCH e POP, bem como o atributo químico CO. Com relação a profundidade 0,20 a 0,40 m, o Fator 1 referiu-se aos atributos químicos (pHag, pHK, pHCa), o Fator 2 referiu-se aos atributos físicos (DS, MA e PT), o Fator 3 compôs os componentes de produção e tecnológicos da cana-de-açúcar (ATR, BRI e PUR), o Fator 4 referiu-se ao atributo de produção TCH, bem como os atributos UV e CO.

Figura 7 - Análise de fatores apresentado as cargas fatoriais entre a variável e o fator. Os valores referência à porcentagem de variação do conjunto de dados originais retida pelo fator nas profundidades: a) 0,00-0, 20 m e b) 0,20-0,40 m.



Fonte: Dados do próprio autor.

TCH = tonelada de colmo por hectare, POP = população de plantas, ATR = açúcar total recuperável, BRI = brix, PUR = pureza, DS = densidade do solo, UV = umidade volumétrica, MA = macroporosidade, PT = porosidade total, pHag = potencial hidrogeniônico em água, pHK = potencial hidrogeniônico cloreto de potássio, pHCa = potencial hidrogeniônico cloreto de cálcio, CO = carbono orgânico do solo.

Nas duas profundidades, os três primeiros fatores explicaram a maior parte da variação dos dados originais (58,9% e 53,9%, respectivamente), sendo que, os quatro fatores explicaram 66,92% da variabilidade dos dados na profundidade 0 a 0,20 m e 66,20% na profundidade 0,20 a 0,40. Deve-se ressaltar que, na profundidade 0 a 0,20 m, o Fator 1, correspondente aos atributos físicos do solo, foi o responsável pela maior proporção de variância explicada (21,3%), enquanto que, na profundidade 0,20 a 0,40 m a maior variância explicada (20,6%) foi correspondente ao Fator 1, representado pelos atributos químicos do solo. As medidas de adequabilidade do modelo KMO para as profundidades 1 e 2 foram 0,62 e 0,60, respectivamente, podendo ser consideradas suficientes para que os fatores conseguissem explicar a variação dos dados amostrais (CORRAR et al., 2009).

Na prática, tem-se atributos físicos relacionados na maior direção de variabilidade na camada superficial, cujas mensurações indicam incidência de compactação, devido ao tráfego de máquinas e sua textura mais arenosa (Argissolo); e atributos químicos relacionados na direção de maior variabilidade na camada subsuperficial, cujos valores de pH estão no intervalo considerado adequado para cultura da cana-de-açúcar, o que pode estar associado ao efeito da gessagem realizada após o primeiro corte (CARVALHO et al, 2013).

Em Brahini et. al (2011), avaliaram o efeito dos fatores latentes na dinâmica do carbono de um argissolo. Os fatores latentes foram construídos tomando-se como base a revisão bibliográfica, bem como a técnica de análise fatorial, com rotação oblíqua. Os atributos químicos pH, matéria orgânica, Nitrogênio e o atributo físico densidade do solo compuseram o primeiro fator latente, que explicou 38,422 % da variação dos dados, enquanto que os atributos físicos argila, silte e areia compuseram o segundo fator latente, o qual explicou 28,955 % da variação dos dados; cuja variância explicada acumulada foi 67,377 %.

Análise de equações estruturais

Tomando como base os resultados apresentados na Figura 7, bem como a importância dos atributos físicos e químicos para os componentes de produção da cultura, foram formados três fatores latentes para cada profundidade, sendo que na profundidade 0 a 0,20 m o primeiro fator latente (F1) foi formado pelos atributos físicos densidade (DS), porosidade total (PT) e macroporosidade (MA), o segundo fator latente (F2) foi formado pelos atributos químicos pHK, pHCa e pHag, e o terceiro fator latente (F3) foi formado pelos componentes de produção e tecnológicos da cana-de-açúcar, açúcares totais recuperáveis (ATR), pureza (PUR) e população de plantas (POP), configurando-se assim, os modelos de medidas.

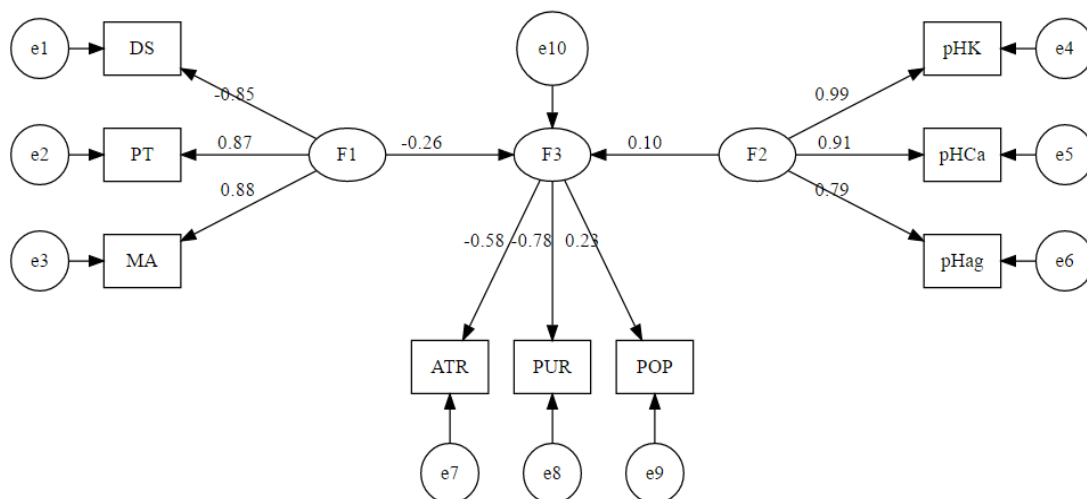
Para a profundidade 0,20 a 0,40 m, o primeiro fator latente (F1) foi formado pelos atributos químicos pHK, pHCa e pHag, o segundo fator latente (F2) foi baseado nos atributos físicos densidade (DS), porosidade total (PT) e macroporosidade (MA), e o terceiro fator latente (F3) foi formado pelos componentes de produção e tecnológicos da cultura, açúcares totais recuperáveis (ATR), pureza (PUR) e população de plantas (POP), configurando-se assim, os modelos de medidas para essa profundidade.

Nas duas profundidades, o modelo estrutural foi construído de forma a equacionar os três fatores latentes, sendo que em ambas foram testadas a influência dos fatores atributos físicos e químicos sobre o fator latente resposta componentes de produção e tecnológicos da cana-de-açúcar.

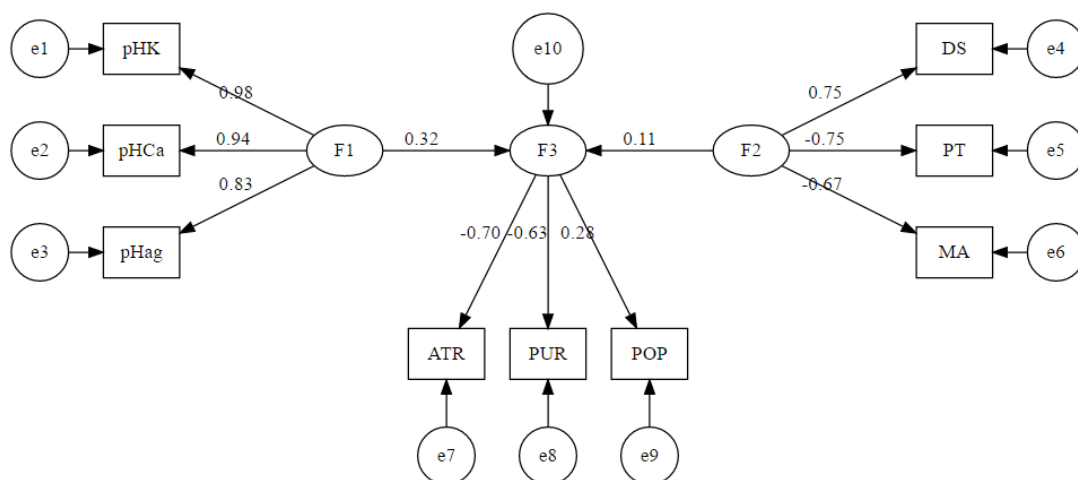
Na Figura 8 são apresentados os diagramas de caminhos com as estimativas dos parâmetros padronizados. Os parâmetros do modelo foram estimados pelo método da máxima verossimilhança robusto (MLM). Os dois modelos foram bem ajustados aos dados, foi o que indicou o teste χ^2 qui-quadrado de ajustamento (p-valores $> 0,05$) e os índices de qualidade de ajuste. A avaliação de ajuste, assim como os índices de ajustamento foram apresentados na Tabela 5. De acordo com os valores de referências das estatísticas e índices de qualidade de ajustamento dos modelos de AEE citados em Marôco (2010), concluiu-se que os modelos ajustados para profundidades 0 a 0,20 m e 0,20 a 0,40 m tiveram bons ajustes.

Figura 8 - Diagrama de Caminhos da análise de equações estruturais para as profundidades: a) 0,00-0, 20 m e b) 0,20-0,40 m.

a)



b)



Fonte: Dados do próprio autor.

POP = população de plantas, ATR = açúcar total recuperável, PUR = pureza, DS = densidade do solo, MA = macroporosidade, PT = porosidade total, pHag = potencial hidrogeniônico em água, pHK = potencial hidrogeniônico cloreto de potássio, pHCa = potencial hidrogeniônico cloreto de cálcio, F1 = Fator latente 1, F2 = Fator latente 2, F3 = Fator latente 3, e_i , $i = 1, 2, \dots, 10$ = erros do modelo.

Nos modelos de medida para ambas as profundidades, todos os parâmetros apresentaram significância estatística, ao nível de 5% de probabilidade (Tabela 4). Para a profundidade 0 a 0,20 m, os atributos que mais foram influenciados pelo fator latente correspondente aos atributos físicos do solo foram a porosidade total e densidade do solo,

enquanto o fator latente atributos químicos se manifestou fortemente sobre todos os seus atributos: pHCa, pHK e pHag (Figura 8). Com relação ao fator latente componentes de produção e tecnológicos da cana-de-açúcar, este se manifestou fortemente nos atributos açúcares totais redutores e pureza (Figura 8).

A respeito da profundidade 0,20 a 0,40 m, o fator latente atributos químicos do solo se manifestou fortemente em todos seus atributos: pHCa, pHK e pHag; e, da mesma forma, todos os atributos físicos (porosidade, densidade e macroporosidade) sofreram forte influência do fator latente atributos físicos (Figura 8). Por fim, o fator latente componentes de produção e tecnológicos da cana-de-açúcar se manifestou com maior efeito nos componentes açúcares totais redutores e pureza (Figura 8).

Tabela 4 - Resumo do resultado de ajuste dos modelos Análises de Equações Estruturais.

Variáveis Latentes	Profundidade 0,00 - 0,20 m				Variáveis Latentes	Profundidade 0,20 - 0,40 m			
	Estimativa	Erro Padrão	Valor Z	P - Valor		Estimativa	Erro Padrão	Valor Z	P - Valor
F1					F1				
DS	-0,85	0,06	-13,92	<0,01	pHK	0,88	0,02	61,39	<0,01
PT	0,87	0,04	19,68	<0,01	pHCa	0,94	0,02	60,96	<0,01
MA	0,88	0,03	27,62	<0,01	pHag	0,83	0,03	32,33	<0,01
F2					F2				
pHK	0,99	0,01	69,88	<0,01	DS	0,75	0,08	8,82	<0,01
pHCa	0,91	0,03	30,90	<0,01	PT	-0,75	0,10	-7,95	<0,01
pHag	0,79	0,04	21,53	<0,01	MA	-0,67	0,08	-8,52	<0,01
F3					F3				
ATR	-0,58	0,14	-4,08	<0,01	ATR	-0,70	0,14	-5,00	<0,01
PUR	-0,78	0,16	-4,99	<0,01	PUR	-0,63	0,12	-5,32	<0,01
POP	0,23	0,11	2,04	0,04	POP	0,28	0,11	2,53	0,01
Regressão					Regressão				
F3					F3				
F1	-0,26	0,12	-2,15	0,03	F1	0,32	0,13	2,36	0,02
F2	0,10	0,12	0,87	0,39	F2	0,11	0,12	0,85	0,39

Fonte: Dados do próprio autor.

POP = população de plantas, ATR = açúcar total recuperável, PUR = pureza, DS = densidade do solo, MA = macroporosidade, PT = porosidade total, pHag = potencial hidrogeniônico em água, pHK = potencial hidrogeniônico em potássio, pHCa = potencial hidrogeniônico em cálcio, F1 = Fator Latente 1, F2 = Fator Latente 2, F3 = Fator Latente 3, Valor Z = Valor da estatística Z, P – Valor = nível descritivo.

Em relação aos modelos estruturais, para a primeira profundidade, o fator latente componentes de produção e tecnológicos da cultura, sofreu influência negativa e significativa apenas do fator latente atributos físicos (Tabela 4), sendo os mais importantes: porosidade total e densidade, na ordem de 2,6 (em módulo) vezes mais que o fator latente atributos químicos

(Figura 8), o qual não apresentou significância estatística. Por outro lado, para a segunda profundidade, apenas o fator latente atributos químicos (pH_{Ca}, pH_K e pH_{ag}) influenciou, de forma positiva e significativa o fator latente componentes de produção e tecnológicos da cana-de-açúcar (Tabela 4), na ordem de aproximadamente 3 vezes mais do que os atributos físicos, que não apresentaram significância estatística (Figura 8).

Nazmi (2013) aplicou a análise de equações estruturais com o objetivo de relacionar fatores latentes e o tamanho da influência desses nos atributos físicos, químicos e de produção do trigo; e seus resultados mostraram que o fator atributos químicos do solo influenciou mais os componentes de produção de trigo do que os fatores atributos físicos do solo, com destaque para os atributos carbono orgânico do solo, pH e carbonato de cálcio equivalente.

A razão do fator atributos físicos influenciar significativamente e de forma negativa o fator componentes de produção e tecnológicos da cana-de-açúcar, na camada de 0 a 0,20 m, está associada a compactação do solo devido ao tráfego de máquinas (SEVERIANO et al., 2010). Por outro lado, na camada de 0,20 a 0,40 m, o fator atributos químicos que influenciou significativamente e de forma positiva o fator componentes de produção e tecnológicos da cana-de-açúcar, efeito da gessagem realizada após o primeiro corte, corrigindo a acidez do solo nesta subsuperfície (CARVALHO; RAIJ, 1997; CAIRES et al., 1999; FOLTRAN, 2008).

Tabela 5. Estatísticas e índices de qualidade de ajustamento dos modelos da análise de componentes de produção e tecnológicos da cana-de-açúcar equações estruturais.

Estatística	Profundidade 0,00 - 0,20 m	Profundidade 0,20 - 0,40 m
χ^2	20,264 (p = 0,682)	33,005 (p = 0,104)
χ^2/gl	0,84	1,375
CFI	0,998	0,979
GFI	1,000	1,000
TLI	1,000	0,969
PGFI	0,61	0,60
RMSEA	0,01	0,056

Fonte: Dados do próprio autor.

χ^2 = valor qui-quadrado do teste de ajustamento, χ^2/gl = razão entre valor qui-quadrado do teste de ajustamento e os graus de liberdade do modelo, CFI = índice de ajuste comparativo, GFI = índice de qualidade de ajuste, TLI = índice Tucker-Lewis, PGFI = índice de qualidade de ajuste de parcimônia, RMSEA = raiz quadrada do erro médio de aproximação.

2.4 - CONCLUSÕES

Os modelos de análise fatorial e componentes principais foram adequados para reduzir a dimensionalidade dos dados das profundidades estudadas, bem como para descrever a estrutura de variância e covariância destes, por meio de componentes e fatores associados aos atributos estudados.

Foram ajustados os modelos teóricos das análises de equações estruturais para as profundidades 0 a 0,20 m e 0,20 a 0,40 m. Os modelos de equações estruturais se ajustaram bem aos dados, podendo reproduzir por meio da matriz de seus parâmetros a matriz de covariância dos dados amostrais, podendo-se avaliar todas as interações entre atributos e fatores e entre fatores, de forma simultânea.

Constatou-se que na profundidade 0 a 0,20 m, o fator componentes de produção e tecnológicos da cana-de-açúcar apresentou influência negativa e significativa do fator atributos físicos do solo, evidenciando o efeito da compactação devido ao tráfego de máquinas (3º corte).

Por fim, na profundidade de 0,20 a 0,40 m, o fator componentes de produção e tecnológicos da cana-de-açúcar foi influenciado de forma positiva e significativa pelo fator atributos químicos do solo associado a acidez do solo; efeito benéfico da aplicação de gesso após o primeiro corte.

Os modelos conceituais de análise de equações estruturais construídos neste estudo, podem ser considerados como modelos base em estudos na área agrícola, inclusive para outras classes de solo, quando o intuito seja a modelagem de dados de desempenho de culturas, para o melhor entendimento de suas relações com o grupo de atributos físicos e químicos do solo.

REFERÊNCIAS

- ALVES, M. C.; CARVALHO, M. P.; VANZELA, S. L. **Levantamento taxonômico semidetalhado dos solos da Usina Vale do Paraná S/A Álcool e Açúcar – Suzanápolis: estudo agrotécnico da Usina Vale do Paraná S/A Álcool e Açúcar.** Ilha Solteira: Unesp, 2009. 324 p.
- ALVES, V.; MONTANARI, R.; CARVALHO, M. P.; CORREA, A. R.; ROQUE, C. G. Atributos tecnológicos, stand e produtividade da cana-planta correlacionados com aspectos da fertilidade do solo. **Revista de Agricultura Neotropical**, Cassilândia, v. 1, n. 1, p. 75-91, 2014.
- BIDOIA, M. A. P. Instalação, condução e colheita de experimentos. In: DINARDO-MIRANDA, L. L.; VASCONCELOS, A. C. M.; LANDELL, M. G. A. (Ed.). **Cana-de-açúcar.** Campinas: Instituto Agrônômico e Fundação- IAC, 2008. p. 821-838.
- BRAUNACK, M. V. **The effect of soil physical properties on growth and yield of sugarcane.** Australia: Bureau of Sugar Experiment Station Queensland, 1991. Disponível em: <<http://elibrary.sugarresearch.com.au/bitstream/handle/11079/514/528.pdf?sequence=1&isAllowed=y>>. Acesso em: 15 mar. 2016.
- CAIRES, E. F.; FONSECA, A. F.; MENDES, J.; CHUEIRI, W. A.; MADRUGA, E. F. Produção de milho, trigo e soja em função das alterações das características químicas do solo pela aplicação de calcário e gesso na superfície, em sistema de plantio direto. **Revista Brasileira Ciência do Solo**, Viçosa, v. 23, p. 315-27, 1999.
- CARVALHO, M. C. S.; RAIJ, B. V. Calcium sulphate, phosphogypsum and calcium carbonate in the amelioration of acid subsoils for root growth. **Plant and Soil**, Switzerland, v. 192, n. 1, p. 37-48, 1997.
- CARVALHO, M. P.; TAKEDA, E. Y.; FREDDI, O. S. Variabilidade espacial de atributos de um solo sob videira em Vitória Brasil (SP). **Revista Brasileira de Ciência do Solo**, Viçosa, v.27, n.4, p.695-703, 2003.
- CARVALHO, J. M.; ANDREOTTI, M.; BUZETTI, S.; CARVALHO, M. P. Produtividade de cana soca sem queima em função do uso de gesso e vinhaça. **Pesq. Agropec. Trop.**, Goiânia, v. 43, n. 1, p. 1-9, 2013.

CESAR, M. A. A.; DELGADO, A. A.; CAMARGO, A. P.; BISSOLI, B. M. A.; SILVA, F. C. Capacidade de fosfatos naturais e artificiais em elevar o teor de fósforo no caldo de cana-de-açúcar (cana-planta), visando o processo industrial. **STAB**, Piracicaba, v. 6, p. 32-38, 1987.

CONSELHO DOS PRODUTORES DE CANA DE AÇÚCAR, AÇÚCAR E ÁLCOOL DO ESTADO DE SÃO PAULO- CONSECANA. **Manual de instruções**. 5. ed. Piracicaba: Consecana, 2006. 54 p.

CORRAR, L. J.; PAULO, E.; DIAS FILHO, J. M.. **Análise multivariada:** para os cursos de administração, ciências contábeis e economia. São Paulo: Atlas, 2009.

CRUZ, J. S.; ASSIS JÚNIOR, R. N.; MATIAS, S. S. R.; CAMACHO-TAMAYO, J. H.; TAVARES, R. C. Análise espacial de atributos físicos e carbono em um argissolo vermelho-amarelo cultivado com cana-de-açúcar. **Ciênc. Agrotec.**, Lavras, v. 34, n. 2, p. 271-278, 2010.

DIAS, F. L. F.; MAZZA, J. A.; MATSUOKA, S.; PERECIN D.; MAULE, R. F. Relação entre produtividade, clima, solos e variedades de cana-de-açúcar na região noroeste do estado de São Paulo. **R. Bras. Ci. Solo**, Viçosa, MG, v. 23, p. 627-634, 1999.

EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA - EMBRAPA. Centro Nacional de Pesquisa de Solos. **Manual de análises químicas de solos, plantas e fertilizantes**. 2. ed. Brasília, DF: Embrapa Informação Tecnológica, 2009.

EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA - EMBRAPA. **Manual de métodos de análise de solos**. 2. ed. Rio de Janeiro: Embrapa/Solos, 1997.

EMPRESA DE PESQUISA ENERGÉTICA – EPE. **Cana-de-açúcar confirmada como a segunda fonte primária de energia no Brasil**. Brasília, DF, 2008. Disponível em: <[http://www.epe.gov.br/Estudos/Documents/Estudos_13/Informativo BEN - Resultado final 2008.pdf](http://www.epe.gov.br/Estudos/Documents/Estudos_13/Informativo_BEN_-_Resultado_final_2008.pdf)>. Acesso em: 20 mar. 2016.

FILZMOSE, P.; GARRETT, R. G.; REIMANN, C. Multivariate outlier detection in exploration geochemistry. **Computers & Geosciences**, Amsterdam, v. 31, p. 579-587, 2005.

FOLTRAN, R. **Aplicação de calcário, silicato e gesso em soqueira de cana-de-açúcar sem despalha a fogo**. 2008. 100 f. Dissertação (Mestrado em Agronomia) – Faculdade de ciências Agrônômicas, Universidade Estadual Paulista, Botucatu, 2008.

FREITAS, L.; CASAGRANDE, J. C.; OLIVEIRA, I. A.; CAMPOS, M. C. C. Análise multivariada na avaliação de atributos de solos com diferentes texturas cultivados com cana-de-açúcar. **Rev. Cienc. Agrar.**, Recife, v. 57, n. 3, p. 224-233, 2014.

FREITAS, P. L.; DONAGEMMA, G. K.; LUMBRERAS, J. F.; SILVA, F. C.; VIEIRA JÚNIOR, P. A.; FINOTO, E. L. Propriedades físicas e químicas de um argissolo vermelho na reforma de canavial para produção de oleaginosas. In: CONGRESSO BRASILEIRO DE CIÊNCIA DO SOLO, 23., 2010, Uberlândia. **Anais...** . Uberlândia: CBCS, 2010. p. 1-4.

GALDIANO, L. C.. **Qualidade da cana-de-açúcar (*saccharum spp*) submetida à aplicação de maturadores químicos em final de safra**. 2008. 45 f. Dissertação (Mestrado em Agronomia) - Faculdade de Ciências Agrárias e Veterinária, Universidade Estadual Paulista, Jaboticabal, 2008.

GIOIA, M. T.. **Produtividade e qualidade da cana soca correlacionadas com atributos físicoquímicos de um argissolo vermelho distrófico do noroeste paulista**. 2011. 127 f.

Dissertação (Mestrado) - Curso de Agronomia, Universidade Estadual Paulista, Ilha Solteira, 2011.

HAIR, J. F; BLACK, W. C.; BABIN, B. J.; ANDERSON, R. E.; TATHAM, R. L. **Análise multivariada de dados**. 6. ed. Porto Alegre: Bookman, 2009.

HARTEMINK, A. E. Soil fertility decline: definitions and assessment. In: LAL, R. (Ed.) **Encyclopedia of soil science**. 2. ed. New York: Dekker, 2006. p. 1618 – 1621.

HORN, J. L. A rationale and test for the number of factors in factor analysis. **Psychometrika**, New York, v. 30, n. 2, p. 179-185, 1965.

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

KIEHL, E. J. **Manual de edafologia: relações solo-água**. São Paulo: Agronômica Ceres, 1979.

LEPSCH, I. F. Influência dos fatores edáficos na produção. In: CASTRO, P. R. C.; FERREIRA, S. O.; YAMADA, T. (Ed.) **Ecofisiologia da produção**. Piracicaba: Associação Brasileira para Pesquisa da Potassa e do Fosfato, 1987. p. 83-98.

LOPES, A. S.; GUILHERME, L. R. G. **Uso eficiente de fertilizantes e corretivos agrícolas: aspectos agronômicos**. 3. ed. São Paulo: Anda, 2000.

MARÔCO, J. **Análise de equações estruturais: fundamentos teóricos, software e aplicações**. Pero Pinheiro: Report Number, 2010.

MARTINS, J. A.; FIORIO, P. R.; BARROS, P. P. S.; MUNIZ, R. A.; SEIXAS, H. T. Alterações na resposta espectral da cana-de-açúcar cultivada em diferentes condições edáficas. In:

SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 17., 2015, João Pessoa. **Anais....** João Pessoa: SBSR, 2015. p. 7389-7396.

MAULE, R. F.; MAZZA, J. A.; MARTHA JÚNIOR, G. B. Produtividade agrícola de cultivares de cana-de-açúcar em diferentes solos e épocas de colheita. **Sci. Agric.**, Piracicaba, v. 58, p. 295-301, 2001.

MINGOTI, S. A. **Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada.** Belo Horizonte: Editora UFMG, 2005.

MONTANARI, R.. **Aspectos da produtividade do feijoeiro correlacionados com atributos do solo sob sistemas de manejo de elevado nível tecnológico.** 2009. 172 f. Tese (Doutorado em Agronomia) – Faculdade de Engenharia, Universidade Estadual Paulista, Ilha Solteira, 2009.

NAZMI, L. Modeling for relationships between soil properties and yield components of wheat using multiple linear regression and structural equation modeling. **Advances In Environmental Biology**, Amman-Jordan, v. 7, n. 2, p. 235-242, 2013.

OLIVEIRA FILHO, F. X.; MIRANDA, N. O.; MEDEIROS, J. F.; SILVA, P. C. M.; MESQUITA, F. O.; COSTA, T. K. G. Zona de manejo para preparo do solo na cultura da cana-de-açúcar. **Revista Brasileira de Engenharia Agrícola e Ambiental**, Campina Grande, v. 19, n. 2, p. 186-193, 2015.

LOLADE, I. A.; AJAYI, I. R.; GBADAMOSI, A. E.; MOHAMMED, O. Z.; SUNDAY, A. G. A study on effects of soil physico-chemical properties on cocoa production in ondo state. **Modern Applied Science**, Finchurst Plaza, v. 4, n. 5, p. 35-43, 2010.

PISON, G.; ROUSSEUW, P. J.; FILZMOSER, P.; CROUX, C. Robust factor analysis. **Journal Of Multivariate Analysis**, Amsterdam, v. 84, p. 145-172, 2003.

RAIJ, B.V.; ANDRADE, J. C.; CANTARELLA, H.; QUAGGIO, J. A. **Análise química para avaliação da fertilidade de solos tropicais.** Campinas: Instituto Agronômico, 2001. 285 p.

RODRIGUES, K. M. **Variabilidade espacial de atributos físicos, químicos e biológicos do solo e a produtividade da cana-de-açúcar.** 2014. 176 f. Tese (Doutorado em Agronomia) - Instituto Agronômico de Campinas, Campinas, 2014.

ROLIM, G. S.; CAMARGO, M. B. P.; LANIA, D. G.; MORAES, J. F. L. Classificação climática de Köppen e de Thornthwaite e sua aplicabilidade na determinação de zonas agroclimáticas para o estado de São Paulo. **Bragantia**, Campinas, v. 66, n. 4, p. 711-720, 2007.

SATORRA, A.; BENTLER, P. M. Scaling corrections for chi-square statistics in covariance structure analysis. In American Statistical Association, editor. In: BUSINESS AND ECONOMIC STATISTICS SECTION, 1988, Alexandria. **Proceedings...** Alexandria: American Statistical Association, 1988. p. 308-313.

SATORRA, A.; BENTLER, P. M. A scaled difference chi-square test statistic for moment structure analysis. **Psychometrika**, New York, v. 66, n. 4, p. 507-514, 2001.

SEVERIANO, E. C.; OLIVEIRA, G. C.; DIAS JÚNIOR, M. S.; CASTRO, M. B.; OLIVEIRA, L. F. C.; COSTA, K. A. P. Compactação de solos cultivados com cana-de-açúcar: i - modelagem e quantificação da compactação adicional após as operações de colheita. **Eng. Agríc.**, Jaboticabal, v. 30, n. 3, p. 404-413, 2010.

SILVA, A. P.; TORMENA, C. A.; IMHOFF, S. Intervalo hídrico ótimo. In: MORAES, M. H.; MÜLLER, M. M. L.; FOLONI, J. S. S. (Ed.) **Qualidade física do solo: métodos de estudo sistemas de preparo e manejo do solo**. Jaboticabal: FUNEP, 2002. p. 1-18.

SORIA, J. E. **Correlações lineares e espaciais entre atributos físico-químicos de solos de diferentes ambientes de produção de cana-de-açúcar no noroeste do estado de São Paulo**. 2014. 97 f. Dissertação (Mestrado em Agronomia) – Faculdade de Engenharia, Universidade Estadual Paulista, Ilha Solteira, 2014.

SOUZA, Z. M.; MARQUES JÚNIOR, J.; PEREIRA, G. T.; MONTANARI, R.; CAMPOS, M. C. C. Mostragem de solo para determinação de atributos químicos e físicos em área com variação nas formas do relevo. **Científica**, Jaboticabal, v. 34, n. 2, p.249-256, 2006.

STOLF, R. Teoria e teste experimental de fórmulas de transformação dos dados de penetrômetro de impacto em resistências do solo. **R. Bras. Ci. Solo**, Viçosa, MG, v. 15, p. 229-235, 1991.

VAKKILAINEN, E.; KUPARINEN, K.; HEINIMÖ, J. **Large industrial users of energy biomass**. [S. l.: s. n.], 2013. Disponível em: <<http://www.bioenergytrade.org/downloads/t40-large-industrial-biomass-users.pdf>>. Acesso em: 12 fev. 2016.

CAPÍTULO 3 - MODELAGEM DA DINÂMICA E ESTOQUE DO CARBONO E DE OUTROS ATRIBUTOS DO SOLO EM ÁREAS DE CANA-DE-AÇÚCAR

RESUMO

A cana-de-açúcar tem sido uma das mais viáveis fontes de combustível renovável a partir da biomassa. A maioria dos canaviais apresentam o manejo de cana crua, sem queima, onde a colheita é realizada de forma mecanizada. A palhada acumulada sobre o solo após a colheita da cana-de-açúcar tem papel importante na dinâmica da água, dos nutrientes e, não somente da matéria orgânica do solo, mas também na perda de carbono do solo via respirações do solo. Assim, novas pesquisas sobre o entendimento destas interações entre os atributos do solo e a perda ou acúmulo de carbono no solo são necessárias. O objetivo do trabalho foi propor um modelo de análise de equações estruturais para o entendimento das variações de emissão de CO₂ e estoque de carbono em solos de cultivo de cana-de-açúcar por meio de fatores latentes formados a partir de atributos químicos e físicos do solo, com o auxílio da análise dos componentes principais. Os ensaios experimentais foram conduzidos em áreas de cultivo comercial de cana-de-açúcar, em sistema de cana crua, localizadas nos municípios de Guariba e Pradópolis (SP) (Latossolo Vermelho Eutroférico textura muito argilosa e Latossolo Vermelho Eutroférico textura argilosa, respectivamente) e Aparecida do Taboado (MS) (Latossolo Vermelho Amarelo Distrófico textura franco argilo arenosa). A emissão de CO₂ do solo e temperatura do solo foram determinadas por meio do sistema LI-8100, foram determinados os atributos físicos e químicos do solo e posteriormente foi realizado o cálculo do estoque de carbono na camada de 0-0,10 m de profundidade. Os dois primeiros fatores latentes compuseram os atributos químicos e físicos com explicação em torno de 50 % da variabilidade dos dados, sendo que, as variações de emissão de CO₂, bem como estoque de carbono foram principalmente influenciadas pelo fator latente do grupo atributos físicos (porosidade livre de água, densidade do solo e macroporos), na ordem de pelo menos 4 vezes (em módulo) mais do que o fator latente do grupo atributos químicos (teores de cálcio, magnésio e capacidade de troca de cátions do solo), para as áreas do estado de São Paulo. Na área de Aparecida do Taboado (MS), constatou-se que as variações de emissão de CO₂ não sofreram influência significativa dos fatores latentes atributos químicos e físicos; porém, as variações dos valores de estoque carbono, foram explicadas de forma estatisticamente significativas por

esses dois fatores, sendo os coeficientes estruturais em relação aos fatores latentes atributos físicos e atributos químicos, iguais a -0,34 e 0,28, respectivamente.

Palavras chave: Fatores Latentes. Modelos, Biomassa, Latossolo.

ABSTRACT

The sugarcane has been one of the most viable renewable fuel from biomass. Most sugarcane field presents the management crude cane, without burning, where the harvest is carried out mechanically. The accumulated straw on the soil after harvest of sugarcane has an important role in the dynamics of water, nutrients and, not only of soil organic matter, but also in the loss of soil carbon via soil respiration. Thus, new research on the understanding of the interactions between soil attributes and loss or accumulation soil carbon are needed. The objective was to propose a model of analysis of structural equations to explain the variations in CO₂ emissions and carbon stock in soils of farming sugarcane through latent factors formed from chemical and physical attributes of soil with the help of principal components analysis. Experimental trials were conducted in areas of commercial cultivation of sugarcane in crude cane system, located in the municipalities of Guariba and Pradópolis (SP) (Areas with Eutroferic Red Latosol very clayey texture and Eutroferic Red Latosol with clay texture, respectively) and Aparecida do Taboado (MS) (Area with Dystrophic Red-Yellow Latosol, with sandy clay franc texture), soil CO₂ emissions and temperature soil were determined by the LI-8100 system, were determined the attributes physical and chemical of soil and was later carried out carbon stock calculation in the layer of 0-0.10 m deep. The first two latent factors composed the chemical and physical attributes with explanation of around 50% of the variability of the data, and the variations of CO₂ emissions and carbon stocks were mainly influenced by latent factor called physical attributes (free porosity water, soil density and macropores) in the order of at least 4 times more than the latent factor called chemical attributes (calcium, magnesium and exchange capacity of the soil cation) for the areas of São Paulo state. In area of Aparecida of Taboado (MS), it was found that variations in CO₂ emission not they suffered influence significant of the latent factors attributes chemical and physical. However, variations in stock carbon were explained in a statistically significant manner by the two factors, with structural coefficients in relation to the latent factors physical attributes and chemical attributes, equals to -0.34 and 0.28, respectively

Keywords: Latent Factors. Models. Biomass. Latossol.

3.1 - INTRODUÇÃO

Nos ecossistemas terrestres a quantidade de carbono armazenado no solo é geralmente maior que a quantidade estocada na massa seca das vegetações vivas. Por sua vez, os solos estocam em seus primeiros 30 cm de profundidade cerca de 1.550 Pg de carbono orgânico, tal valor é mais que duas vezes a quantidade de carbono presente na atmosfera (720 Pg) e cerca de três vezes mais que o carbono da biota terrestre (POST; KWON, 2000; LAL, 2001; FOLLETT, 2001). A perda de carbono no solo ocorre, principalmente, via emissão de CO₂ (FCO₂), também conhecida como respiração do solo, resultado da mineralização da matéria orgânica do solo devido à atividade microbiana. Nesse contexto, os níveis de carbono orgânico nos solos são os resultados das complexas interações entre variáveis relacionadas aos processos de produção e transporte do gás do solo para a atmosfera (PANOSSO, 2011).

O uso e o manejo dos solos contribuem de maneira significativa para o efeito estufa adicional, isto é, o aquecimento elevado da superfície da terra e da baixa atmosfera, como resultado do aumento das emissões de gases estufas para atmosfera terrestre, também conhecidos como gases do efeito estufa (GEE), são eles: dióxido de carbono (CO₂), metano (CH₄) e óxido nitroso (N₂O). O aumento da concentração de GEE na atmosfera se deve principalmente às atividades antrópicas, principalmente após o período da revolução industrial, e tais emissões podem ser consideradas superiores a quaisquer outras oriundas de processos naturais (SCHOLLES; ANDREAE, 2000; SMITH et al., 2003). No Brasil, uma das principais fontes de gases de efeito estufa (GEE) e aerossóis antrópicos é a queima de biomassa, utilizada como prática agrícola ou como auxiliar na mudança da cobertura do solo.

De acordo com o IPCC (2007), 60% do efeito estufa adicional pode ser atribuído ao aumento de CO₂ na atmosfera. Smith et al. (2003), afirmam que estes gases são produzidos (ou consumidos) como resultado de processos microbianos do solo, sendo que a intensidade do fluxo entre o solo e a atmosfera depende fortemente dos fatores físicos e químicos do solo. Os atributos físicos e químicos do solo tais como a umidade, temperatura, conteúdo de matéria orgânica, acidez, conteúdo de argila, capacidade de troca catiônica, densidade e a porosidade do solo são os grandes responsáveis por controlar os níveis de acúmulo ou decréscimo de carbono orgânico do solo (PAUSTIAN, 1997; XU; QI, 2001; EPRON et al., 2006; FONTANA et al., 2006; BORTOLUZZI et al., 2009; RYU et al., 2009; KEMMITT et al., 2008).

Assim, o conhecimento da distribuição das fontes e sumidouros de carbono e suas variações ao longo do tempo é fundamental para a recomendação de práticas que influenciem no ciclo do carbono terrestre global, aumentando assim a sustentabilidade dos sumidouros de

carbonos atuais (CHEN et al., 2003; GOYAL et al., 2008; BRAHIM et al., 2011). A importância da agricultura neste cenário não está relacionada apenas aos processos de perda de carbono do solo, mas também está relacionada ao seu significativo potencial de mitigação de carbono atmosférico. Dependendo do uso e manejo, os solos agrícolas podem atuar como fontes ou sumidouros de carbono da atmosfera (USSIRI; LAL, 2009, BRITO, 2009).

A predição dos efeitos do clima, composição atmosférica e mudanças no uso da terra é essencial na formulação de políticas ambientais, agrícolas, sociais e econômicas. Modelos matemáticos otimizam o entendimento da matéria orgânica do solo e são ferramentas essenciais na predição dos efeitos das mudanças ambientais, para testar cenários específicos e desenvolver estratégias que mitiguem os efeitos destas mudanças (LEITE; MENDONÇA, 2003). A caracterização do fluxo de CO₂ do solo é um dos grandes desafios da comunidade científica em função das incertezas relacionadas a essas medidas (MOITINHO et al., 2013). No estudo de Zeraatpishe; Khormali (2012), investigaram o efeito de materiais orgânicos sobre os atributos físicos e químicos e o efeito dos fatores minerais na estabilização e estoque de carbono orgânico de solos na província de Golestan, Iran, e observaram que o estoque de carbono foi controlado principalmente pelo fator clima (temperatura e precipitação) e minerais de argila. Brahim et al. (2011), utilizando-se de dados construídos a partir de resultados de análise analítica de solos escavados por equipes de pedólogos, aplicaram a modelagem de equações estruturais para estudar o efeito dos atributos físicos *versus* a função dos atributos químicos do solo sobre o teor de carbono orgânico do solo, e concluíram que em solos argilosos, o teor de carbono orgânico foi controlado de maneira significativa, principalmente pelos atributos físicos, por outro lado, em solos arenosos os atributos químicos do solo controlaram de maneira significativa as alterações do teor de carbono orgânico do solo.

Sistemas de cultivo associados à produção de energia, como exemplo a cultura da cana-de-açúcar, podem capturar o CO₂ atmosférico na forma de biomassa (sequestro), retornando para o solo quantidades adicionais de carbono orgânico do solo (PANOSSO, 2011). Pitombo et al. (2015) estudaram os estoques de carbono e as emissões de CO₂, a partir de solo com diferentes históricos de aplicação de lodo de esgoto no município de Campinas – SP. Os resultados demonstraram que a aplicação de lodo de esgoto propiciou um incremento de carbono no solo, enquanto que o fluxo de CO₂ foi mais influenciado pela temperatura do solo nos tratamentos com aplicação de lodo de esgoto.

Por meio de um estudo geoestatístico e com auxílio da análise dos componentes principais, Graf et al. (2012), avaliaram a possibilidade da identificação de padrões espaço-temporais do fluxo de CO₂ do solo com relações estatisticamente significativas para os atributos

do solo. Os autores concluíram que a análise dos componentes principais foi importante neste aspecto, sendo que o padrão espacial mais importante no fluxo de CO₂ foi claramente relacionado com a umidade do solo e as propriedades físicas do solo de condução. Neste sentido, Allaire et al. (2012), estudaram a variabilidade espacial da emissão de CO₂ e a correlação dessas emissões com as propriedades físico-químicas do solo em duas escalas espaciais, por meio da análise espacial multiescala; observando-se correlações significativas entre o fluxo de CO₂ e as propriedades do solo, sendo estas maiores em grande escala espacial.

Em Moitinho et al. (2013), estudaram o efeito do preparo do solo e da deposição de resíduos da cultura da cana-de-açúcar na emissão de CO₂ e seus efeitos na temperatura e umidade durante a reforma do canavial; obtendo nos resultados, menor média na emissão de CO₂ (2,16 $\mu\text{mol m}^2 \text{s}^{-1}$) para manejo sem preparo do solo e com resíduos da colheita sobre a superfície do solo. Diante deste cenário e da relevância do entendimento da dinâmica do carbono no solo, neste estudo objetivou-se propor um modelo conceitual para melhor entender de que forma os fatores formados pelos atributos físicos e químicos do solo estão influenciando, de forma simultânea, nas emissões de CO₂ e no estoque de carbono do solo em áreas com cultivo de cana-de-açúcar, por meio da análise de equações estruturais.

3.2 - MATERIAL E MÉTODOS

Área experimental

Os experimentos foram conduzidos em três áreas de plantio de cana-de-açúcar (*Saccharum* spp.), sendo duas áreas localizadas no estado de São Paulo e uma área localizada no estado do Mato Grosso do Sul. As áreas do estado de São Paulo apresentam mais de 35 anos de cultivo comercial de cana-de-açúcar, localizadas na região nordeste do estado. O clima da região foi classificado como B2rB'4a' pelo sistema de Thornthwaite (ROLIM et al., 2007), indicando uma região mesotermal de verões chuvosos e invernos secos, com temperatura média anual de 22,2 °C. A precipitação média anual registrada é de 1.438 mm, com período de maior concentração entre os meses de outubro e março, e precipitações mais espaçadas, e de menores intensidades, no período de abril a setembro. A vegetação natural era constituída por floresta tropical subcaducifolia. *Área 1 (experimento de 2008)*: O estudo foi conduzido no ano de 2008, em área com 38 anos de cultivo de cana-de-açúcar, na Fazenda São Bento, localizada na cidade de Guariba – SP, pertencente à Usina São Martinho. As coordenadas geográficas são 21° 24' de latitude Sul e 48° 09' de longitude Oeste, com altitude ao redor de 550 m acima do nível do mar, com relevo suave ondulado. O histórico de manejo da área foi no sistema cana crua, com 8 anos de colheita mecanizada e 12 t ha⁻¹ de resíduos da cultura (palha) sobre a superfície do solo. A variedade plantada na área foi a CTC-6. O solo da área foi classificado como Latossolo Vermelho eutroférico, textura muito argilosa (SANTOS et al., 2013). Para o estudo foi instalada uma grade de 50 × 50 m, com distâncias mínimas de separação entre pontos de 0,50 m, contendo 89 pontos amostrais. *Área 2 (experimento de 2012)*: No ano de 2012, o estudo foi conduzido em área localizada na Fazenda Santa Cândida em área com 50 anos de cultivo de cana-de-açúcar, com coordenadas de 21° 20' de latitude Sul e 48° 08' de longitude Oeste, com altitude ao redor de 515 m acima do nível do mar, declividade baixa, município de Pradópolis, SP. O solo, semelhante aos demais, foi classificado como Latossolo Vermelho eutroférico, textura argilosa (SANTOS et al., 2013); há 15 anos em sistema de manejo cana crua, sobre o qual é depositado anualmente 12 t ha⁻¹ de resíduos da cultura. A variedade plantada na área foi a CTC14, que estava em oitavo corte. Nessa área, foi instalada uma malha regular de 50 × 50 m contendo 133 pontos, com distância mínima de separação de 1 m entre os pontos adensados do centro da malha amostral.

A terceira área experimental (*Área 3 – experimento de 2014*) localizou-se no Estado do Mato Grosso do Sul, com 20 anos de cultivo comercial de cana-de-açúcar. O clima da região

foi classificado como C1dAa' pelo sistema de Thornthwaite (ROLIM et al., 2007), indicando uma região sub-úmido seco, sem excedentes hídricos, megatérmico com evapotranspiração no verão menor que 48% da evapotranspiração anual, temperatura média anual de 23,5 °C e precipitação média anual de 1.300 mm, com período de maior concentração entre os meses de setembro a junho, invernos secos de junho a agosto. A vegetação original encontrada na área foi descrita como cerrado típico, com predomínio de formas arbustivas em solo profundo, pouco fértil. O estudo foi conduzido no ano de 2014, em área localizada na cidade de Aparecida do Taboado – MS, pertencente à Usina Alcoolvale. As coordenadas geográficas são 20° 19' de latitude Sul e 51° 13' de longitude Oeste, com altitude ao redor de 367 m acima do nível do mar, com relevo suave ondulado. O histórico de manejo da área foi no sistema cana crua, com 5 anos de colheita mecanizada e 16 t ha⁻¹ de resíduos da cultura (palha) sobre a superfície do solo. A variedade plantada na área foi a CTC-1. O solo da área foi classificado como Latossolo Vermelho-Amarelo distrófico, textura franco argilo arenosa (SANTOS et al., 2013). Para o estudo foi instalada uma grade de 100 × 100 m, com distâncias mínimas de separação entre pontos de 5 m, contendo 102 pontos amostrais.

Avaliação da emissão de CO₂, estoque de carbono e atributos do solo

A emissão de CO₂ do solo (FCO₂) foi registrada por meio de dois sistemas LI-COR (LI-8100), os quais monitoram as mudanças na concentração de CO₂ dentro de uma câmara fechada por meio de espectroscopia na região do infravermelho. A câmara para solos tem um volume interno de 854,2 cm³ com área de contato circular de 83,7 cm². Essa câmara foi colocada sobre colares de PVC previamente inseridos no solo em cada ponto do gradeado.

Um sensor portátil do sistema LI-8100 foi utilizado para medir a temperatura do solo (Ts), usando uma haste de 0,20 m, 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 é constituído por uma sonda, apresentando duas hastes de 12 cm, inseridas no interior do solo em região próxima aos colares de PVC.

Após todas as medidas de FCO₂, Ts e Us terem sido gravadas, amostras de solo foram coletadas de cada ponto do gradeado na profundidade de 0 a 0,10 m. As amostras foram secas ao ar e peneiradas através de uma malha de 2 mm antes de serem submetidas à análises subsequentes, as quais compreenderam em determinar o teor de carbono orgânico do solo (CO),

de P disponível, pH, K, Ca, Mg e H + Al (RAIJ et al., 2001), que permitiu o cálculo da capacidade de troca de cátions (CTC).

A densidade do solo (D_s) foi determinada em amostra indeformada coletada com amostrador adaptado a cilindros com dimensões médias de 5,0 cm de diâmetro interno e 4,0 cm de altura (EMBRAPA, 1997). A porosidade total do solo (PT) foi calculada com base no valor de densidade. A distribuição de poros por tamanho, macroporosidade (Macro) e microporosidade (Micro) foram determinadas utilizando-se funil de placa porosa sob a tensão de 60 cm 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 foram calculados por diferença (EMBRAPA, 1997). A porosidade livre de água (PLA, em % do volume), foi calculada por diferença entre a porosidade total e a umidade do solo.

O estoque de carbono do solo (EstC) foi calculado com base na massa de solo equivalente (BAYER et al., 2000), pela equação (77):

$$EstC = \frac{CO \cdot D_s \cdot E}{10} \quad (77)$$

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

Análise dos dados

Na primeira parte da análise de dados, foi verificado o comportamento de cada atributo, por meio da análise descritiva dos dados, bem como a existência de valores discrepantes (outliers multivariados), além da avaliação da hipótese de normalidade multivariada e, posteriormente, a aplicação do método estatístico multivariado: análise dos componentes principais (ACP). Para verificação da existência de outliers multivariados, utilizou-se o gráfico bidimensional *Q-Q Plot*, com base na distância de mahalanobis robusta (FILZMOSER; GARRETT; REIMANN, 2005).

Em seguida foi aplicado a ACP, que consistiu na realização de uma rotação rígida no sistema de eixos coordenados originais de modo que os novos eixos estivessem na direção de maior variabilidade dos dados, em que os coeficientes dos novos eixos são os autovetores da matriz de covariância amostral dos dados.

Para uma melhor interpretação dos resultados da ACP, foi utilizado o método de rotação ortogonal (*varimax*). O número de componentes da ACP foi determinado pelo critério de Kaiser

(1958): componentes com autovalores maiores ou iguais a 1 (um), da análise paralela de Horn (1965): os valores médios dos autovalores gerados por simulação comparados aos autovalores reais; além disso, para os componentes selecionados na ACP, consideraram-se apenas os atributos associados a estes componentes com coeficientes maiores que 0,70.

Sendo $X = [X_1, X_2, \dots, X_p]^T$ um vetor aleatório com p atributos, matriz de covariância amostral $S_{p \times p}$ com autovalores $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$. A Equação (78) descreve o modelo ACP, em sua forma matricial (MINGOTI, 2005):

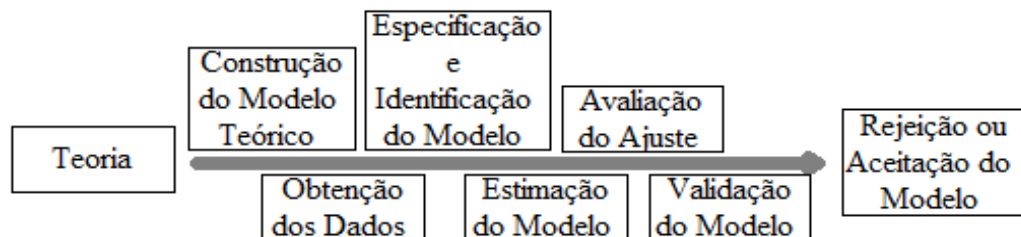
$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_p \end{bmatrix} = \begin{bmatrix} e_{11} & e_{12} & \cdots & e_{1p} \\ e_{21} & e_{22} & \cdots & e_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ e_{p1} & e_{p2} & \cdots & e_{pp} \end{bmatrix}^T \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix} \Rightarrow Y = O^T \cdot X, \quad (78)$$

em que $e_i^T = (e_{i1}, e_{i2}, \dots, e_{ip})$ é um autovetor de dimensão $1 \times p$ com $e_i^T \cdot e_i = 1$, e $e_i^T \cdot e_j = 0$ sendo $i \neq j$ para a última condição e i e j variando de 1 a p para as duas condições.

Na segunda parte de análise, cada grupo de atributos selecionados de seus componentes e dado o conhecimento de estudos anteriores, foi associado a um fator latente e o modelo conceitual da análise de equações estruturais (AEE) foi construído (BRAHIN et al., 2011).

A AEE é uma técnica de modelagem generalizada, cujo objetivo foi testar e validar modelos teóricos, que definem relações causais e hipotéticas entre as variáveis estudadas. Tais relações foram medidas pelos parâmetros do modelo que representam o tamanho do efeito dos atributos independentes sobre os atributos dependentes (MARÔCO, 2010). Na Figura 9 estão os procedimentos da aplicação da AEE.

Figura 9. Etapas da Análise de Equações Estruturais



Fonte: Marôco (2010).

O modelo geral da análise de equações estruturais é composto dos modelos de medidas, os quais relacionam os atributos observados no campo a seus respectivos fatores latentes, e, um

modelo de regressão estrutural, que representa as relações entre os fatores latentes. O modelo de regressão estrutural com r fatores latentes dependentes $\eta^T = (\eta_1, \eta_2, \dots, \eta_r)$, com matriz de coeficientes de η do modelo estrutural $B_{r \times r}$, s variáveis latentes independentes $\xi^T = (\xi_1, \xi_2, \dots, \xi_s)$, com matriz de coeficientes de ξ do modelo estrutural $\Gamma_{r \times s}$ e $\zeta^T = (\zeta_1, \zeta_2, \dots, \zeta_r)$ os erros de medida do modelos estrutural, é descrito de acordo com Bollen (1989) pela equação (79).

$$\eta = B\eta + \Gamma\xi + \zeta \quad (79)$$

Para evitar problemas de desvios de normalidade multivariada, foi utilizado o método de estimação da máxima verossimilhança com correção ou ajustamento da estatística χ^2 , bem como erros padrão robustos, calculados com base na curtose multivariada (SATORRA; BENTLER, 1988; 2001). As medidas utilizadas para avaliar a qualidade de ajustamento do modelo foram: χ^2 = valor qui-quadrado do teste de ajustamento global, χ^2/gl = razão entre valor qui-quadrado do teste de ajustamento global e os graus de liberdade do modelo, CFI = Índice de ajuste comparativo, GFI = Índice de qualidade do ajuste, TLI = Índice de Tucker-Lewis, PGFI = Índice de qualidade de ajuste de parcimônia, RMSEA = raiz do erro quadrático médio de aproximação.

Os dados coletados em Pradópolis - SP foram considerados dados base para construção e ajuste do modelo teórico de equações estruturais, por ter maior tamanho amostral que os outros dois bancos de dados. Posteriormente, ajustou-se o modelo construído anteriormente aos dados coletados em Guariba - SP e Aparecida do Taboado - MS, com intuito de avaliar a invariância do modelo. Todas as análises foram realizadas por meio do software estatístico R, o qual é gratuito e apresenta código-fonte-aberto.

3.3 - RESULTADOS E DISCUSSÃO

Análise preliminares

Inicialmente, os dados foram analisados em termos de suas médias e respectivos coeficientes de variação (Tabela 6). Os valores médios de FCO_2 para as três áreas estudadas, variaram entre 1,33 e 2,19 $\mu\text{mol m}^{-2}\text{s}^{-1}$, com coeficientes de variação entre 35,73% e 45,55%, valores considerados altos (Tabela 6). Os valores encontrados de FCO_2 e respectivos coeficientes de variação foram próximos aos de outros estudos de solos com cultivos de cana-de-açúcar (BRITO et al., 2010; PANOSSO, 2011; BICALHO, 2016), bem como de outros trabalhos realizados com intuito de caracterizar a variabilidade dos valores das emissões de CO_2 , cita-se Graf et al. (2012) e Allaire (2012). O fato da ocorrência do maior valor médio de FCO_2 na área de Guariba – SP, 2,19 $\mu\text{mol m}^{-2}\text{s}^{-1}$, pode ser justificada pelos maiores valores médios de temperatura e umidade do solo em comparação com as outras áreas, 25,98 °C e 27,51%, respectivamente, corroborando com Moitinho et al. (2013).

Pode-se observar ainda, que os atributos físicos densidade do solo, porosidade total e porosidade livre de água também contribuíram para a variação nos valores de FCO_2 (Tabela 6), sendo que para as áreas de Pradópolis ($\text{FCO}_2=1,33 \mu\text{mol m}^{-2}\text{s}^{-1}$), Guariba ($\text{FCO}_2=2,19 \mu\text{mol m}^{-2}\text{s}^{-1}$) e Aparecida do Taboado ($\text{FCO}_2=1,71 \mu\text{mol m}^{-2}\text{s}^{-1}$), obteve-se valores de D_s , PT e PLA iguais a 1,45 kg.dm^{-3} , 49,74 m.m^{-3} e 40,49 m.m^{-3} ; 1,17 kg.dm^{-3} e 42,78 m.m^{-3} e 15,27 m.m^{-3} ; 1,47 kg.dm^{-3} , 43,05 m.m^{-3} e 39,14 m.m^{-3} , respectivamente. Brito et al. (2009) encontraram maior FCO_2 em áreas com maior porosidade. Os principais atributos do solo que influenciam o FCO_2 incluem a temperatura e o teor de água no solo (LA SCALA et al., 2000; KOSUGI et al., 2007; LEON et al., 2014).

No que tange ao estoque de carbono, além da relação direta com o teor de carbono orgânico dos solos e a densidade, notou-se uma redução nos seus valores, com decréscimos de Ca, Mg e P (Tabela 5). Um dos grandes responsáveis por controlar os níveis de crescimento do conteúdo de carbono orgânico é o fator físico-químico do solo. A fixação de carbono no solo inclui estabilização química, processos bioquímicos e proteções físicas (PAUSTIAN, 1997; FONTANA et al., 2006; BORTOLUZZI et al., 2009; CHRISTENSEN, 1996).

Análise dos Componentes Principais e o Modelo Teórico da AEE

Procedeu-se com a análise dos dados da área de Pradópolis-SP. Foi detectada 1 observação (linha 81) candidata a outlier multivariado, pela distância de Mahalanobis Robusta

(*Q-Q Plot*), a qual foi retirada por dois motivos: destoar significativamente das demais observações e promover redução nos desvios de normalidade multivariada. A técnica estatística multivariada análise dos componentes principais foi aplicada a estes dados com intuito de reduzir a dimensionalidade destes e auxiliar na formação dos fatores latentes a serem utilizados posteriormente na AEE (BRAHIM et al., 2011). Preferiu-se não considerar os atributos FCO₂ e EstC na ACP, pois serão as variáveis dependentes no modelo estrutural da AEE, a ser construído. O número de componentes retidas foi verificado pelo gráfico Screeplot (Figura 10), considerando os critérios de Kaiser (1958) e análise paralela de Horn (1965).

Observando a Figura 10, tem-se 5 componentes retidos pelo método de Kaiser (1958) e 3 componentes retidos pelo critério da análise paralela de Horn (1965). Optou-se por utilizar apenas as três primeiras componentes principais (PANOSSO et al. 2012). Os coeficientes das três componentes principais, bem como a variação explicada por cada uma delas, é mostrado na Tabela 6. A variância acumulada foi de 60 %, ou seja, as componentes retidas explicaram 60% da variação total dos dados amostrais.

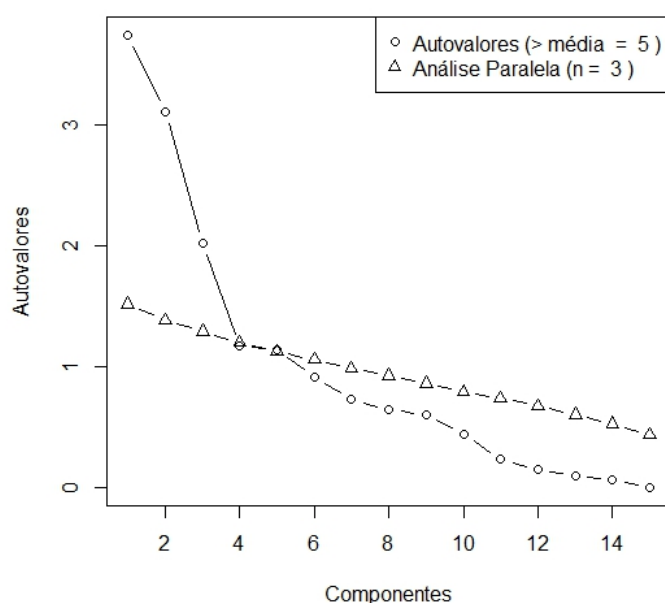
Considerando os coeficientes com valores acima 0,70 (Tabela 7), observou-se que, a primeira componente principal referiu-se aos atributos químicos do solo, cuja variância explicada foi 23%. A segunda componente principal se mostrou fortemente associada aos atributos físicos do solo, com variação explicada correspondente a 21%, resultados semelhantes aos de Brahim (2011) e Moitinho (2014).

Tabela 6. Médias das emissões de CO₂, temperatura, umidade do solo e demais atributos do solo e seus respectivos coeficientes de variação, profundidade de 0-0,10 m, área de Pradópolis-SP.

Atributo	Pradópolis		Guariba		Aparecida do Taboado	
	Média	CV	Média	CV	Média	CV
FCO ₂ (μmol m ⁻² s ⁻¹)	1.33	45.55	2.19	35.73	1.71	37.49
EstC (Mg ha ⁻¹)	8.38	17.6	8.03	9.62	6.47	18.82
Ts (°C)	20.55	2.04	25.98	2.23	21.79	6.07
Us (%)	9.25	9.31	27.51	13.97	9.46	12.23
PLA (m.m ⁻³)	40.49	10.13	15.27	33.36	39.14	2.95
Ds (kg dm ⁻³)	1.45	7.89	1.17	5.48	1.47	8.00
Macro (m.m ⁻³)	19.61	33.8	6.60	47.30	10.62	45.82
Micro (m.m ⁻³)	30.13	9.88	36.18	5.44	32.43	9.78
PT (m.m ⁻³)	49.74	8.77	42.78	5.49	43.05	10.58
pH	5.43	6.16	4.60	5.48	5.32	7.15
CO (g dm ⁻³)	16.68	15.47	14.04	9.37	11.37	17.60
P (mg dm ⁻³)	26.26	47	18.07	39.52	9.05	31.00
Ca (mmol _c dm ⁻³)	32.96	35.82	27.74	23.70	22.43	34.43
Mg (mmol _c dm ⁻³)	14.74	22.58	11.62	26.42	10.27	31.63
K (mmol _c dm ⁻³)	2.12	40.4	5.79	31.46	1.42	28.65
H+Al (mmol _c dm ⁻³)	33.42	23.81	58.60	20.18	24.64	21.01
CTC (mmol _c dm ⁻³)	83.28	14.18	103.97	9.09	58.73	14.08

CV – Coeficiente de Variação; FCO₂ – emissão de CO₂; EstC – Estoque de Carbono no solo; Ts – Temperatura; Us – Umidade; PLA – Porosidade livre de água; Ds – Densidade; Macro – Macroporosidade; Micro – Microporosidade; PT – Porosidade Total; pH – potencial hidrogeniônico; CO – Carbono orgânico; P – Fósforo; Ca – Cálcio; Mg – Magnésio; K – Potássio; H.Al – Acidez Potencial; CTC – Capacidade de troca de Cátions.

Figura 10. Screeplots da análise de componentes principais para todos os atributos estudados da área de Pradópolis – SP.



Fonte: Dados do próprio autor.

Análise de Equações Estruturais

Por fim, o modelo conceitual de análise de equações estruturais para os dados de Pradópolis-SP foi construído com base em pesquisas realizadas, tais como (CHRISTENSEN, 1996, BRITO et al. 2009; LA SCALA et al., 2000; LEON et al., 2014, TAVARES et al., 2016; BICALHO, 2016), as quais evidenciam os fatores mais importantes nas mudanças dos níveis de carbono no solo, bem como nos valores das emissões de CO₂, principalmente em solos com cultivos de cana-de-açúcar.

A ACP também foi muito importante para a construção do modelo de AEE. Os atributos com maiores coeficientes na sua respectiva componente principal, foram os selecionados para compor os fatores latentes da AEE. Para isto, foi considerado o ponto de corte 0,70, como mencionado anteriormente. Tal procedimento foi realizado por Brahim (2011).

Tabela 7. Coeficientes das três primeiras componentes principais e a variância explicada por cada uma delas.

Componentes Principais	CP1	CP2	CP3
Autovalores	3,48	3,10	2,28
Variância Explicada (%)	0,23	0,21	0,16
Variância Acumulada (%)	0,23	0,44	0,60
Coeficientes (Cargas)			
Ts	0.20	-0.28	-0.40
Us	-0.07	0.47	0.16
PLA	0.15	0.91	-0.07
Ds	-0.07	-0.93	-0.02
Macro	0.06	0.95	-0.07
Micro	0.67	-0.14	0.54
PT	-0.01	0.09	0.68
pH	-0.15	0.09	0.68
CO	0.47	0.16	0.08
P	-0.22	0.17	-0.54
Ca	0.93	-0.06	0.18
Mg	0.85	-0.09	0.04
K	0.29	0.15	-0.07
H.AI	-0.51	0.06	-0.69
CTC	0.84	-0.03	-0.27

Fonte: Dados do próprio autor.

Ts – Temperatura; Us – Umidade; PLA – Porosidade livre de água; Ds – Densidade; Macro – Macroporosidade; Micro – Microporosidade; PT – Porosidade Total; pH - potencial hidrogeniônico; CO – Carbono orgânico; P – Fósforo; Ca – Cálcio; Mg – Magnésio; K – Potássio; H.AI – Acidez Potencial; CTC – Capacidade de troca de Cátions.

O primeiro fator latente (F1), foi composto pelos atributos químicos Ca, Mg e CTC, correspondentes a primeira componente principal. Para o segundo fator latente (F2), foram retidos os atributos físicos PLA, Ds e Macro (Tabela 7). Na prática, tem-se neste argissolo, os

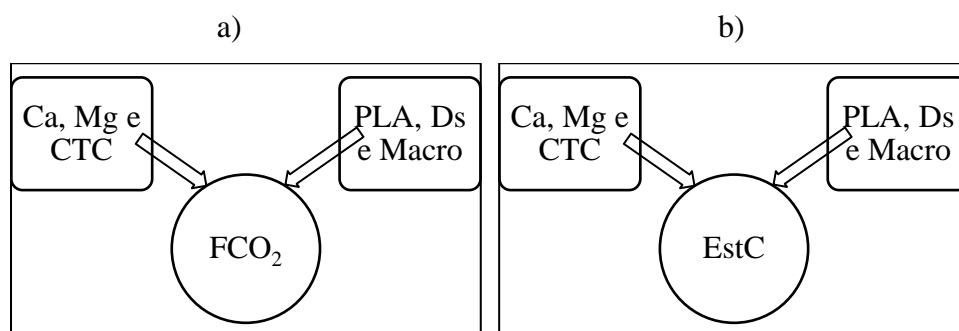
atributos químicos Ca, Mg e CTC correlacionados entre si (BEBÉ et al, 2009) e na direção de maior variabilidade dos dados, correspondentes aos maiores coeficientes da primeira componente principal. Por outro lado, os atributos físicos PLA, Ds e Macro estão associados entre si (BAHIA, 2012), e, estão na segunda direção de maior variabilidade dos dados amostrais.

Em Brahlin et. al (2011) avaliaram o efeito dos fatores latentes na dinâmica do carbono de um argissolo. Os fatores latentes foram construídos tomando-se como base a revisão bibliográfica, bem como a ACP, com rotação oblíqua. Os atributos químicos pH, matéria orgânica, Nitrogênio e o atributo físico densidade do solo compuseram o primeiro fator latente, que explicou 38,422 % da variação dos dados, enquanto que os atributos físicos argila, silte e areia compuseram o segundo fator latente, o qual explicou 28,955 % da variação dos dados; cuja variância explicada acumulada foi 67,377 %.

Portanto, os modelos estruturais para a área de Pradópolis - SP foram construídos de forma a equacionar os dois fatores latentes, sendo que foi testado a influência dos fatores atributos químicos e físicos sobre as emissões de CO₂ (modelo FCO₂) e estoque de carbono do solo (modelo EstC). Na Figura 11, tem-se a forma dos dois modelos de AEE a serem ajustados, considerando-se os dados de Pradópolis-SP.

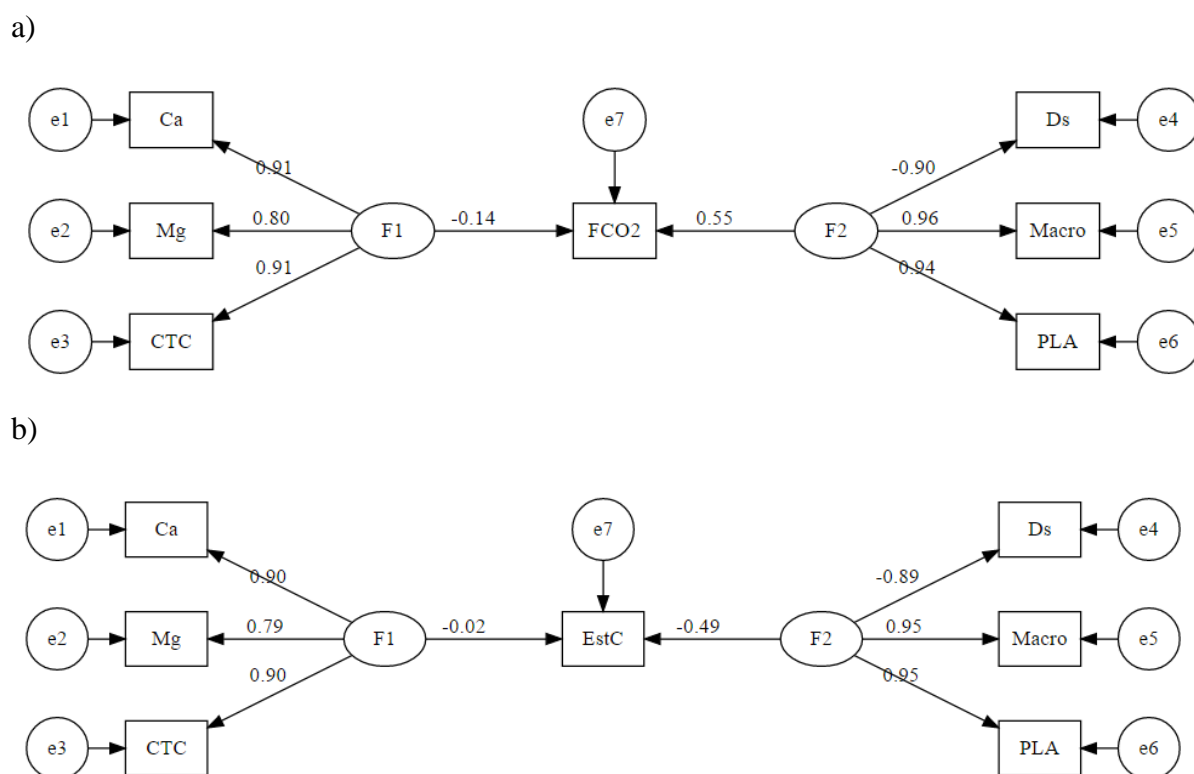
Na Figura 12, são apresentados os diagramas de caminhos com as estimativas dos parâmetros padronizados. Os parâmetros dos modelos foram estimados pelo método da máxima verossimilhança robusto (MLM). Os dois modelos foram bem ajustados aos dados, foi o que indicou o teste χ^2 qui-quadrado de ajustamento (p-valores > 0,05) e os índices de qualidade de ajuste. A avaliação de ajuste, assim como os índices de ajustamento foram apresentados na Tabela 9. De acordo com os valores de referências das estatísticas e índices de qualidade de ajustamento dos modelos de AEE citados em Marôco (2010), concluiu-se que os modelos ajustados para FCO₂ e estoque de carbono tiveram bons ajustes.

Figura 11. Diagrama Estrutural dos Modelos Conceituais para FCO₂ e estoque de Carbono, área Pradópolis – SP.



Fonte: Dados do próprio autor.

Figura 12. Diagrama de Caminhos da análise de equações estruturais, área Pradópolis – SP.



Fonte: Dados do próprio autor.

De acordo com o resumo do ajuste do modelo FCO₂ e do modelo EstC, todos os parâmetros dos seus respectivos submodelos (modelos de medidas) apresentaram significância estatística, ao nível de 5% de probabilidade (Tabela 8), mostrando que os fatores latentes conseguiram explicar de forma satisfatória a variabilidade amostral de seus atributos (Figura 12). Oliveira et al. (2016), estudou o potencial da AEE nas relações entre qualidade e diversidade da fauna no sistema solo-serapilheira em agroflorestas de cacau no sul da Bahia, e,

em seu modelo de AEE, todos os parâmetros do modelo de medida qualidade do solo foram estatisticamente significativos.

Em relação aos modelos estruturais, para o modelo FCO₂, o atributo emissão de CO₂, sofreu influência positiva e significativa apenas do fator latente atributos físicos, F2 (Tabela 8), na ordem de aproximadamente 4 vezes (em módulo) mais que o fator latente atributos químicos, F1 (Figura 12), o qual não apresentou significância estatística. No modelo EstC, o fator latente atributos físicos (F2) influenciou de forma negativa e significativa o estoque de carbono (Tabela 8), na ordem de 24,5 vezes mais do que o fator latente atributos químicos (F1), que não apresentaram significância estatística.

Ficou evidente que o fator latente atributos físicos (PLA, Ds e Macro) foi mais importante na explicação das variações de FCO₂ (efeito positivo) e, principalmente estoque de carbono (efeito negativo), no latossolo altamente argiloso de Pradópolis, cultivado com cana-de-açúcar. Estes resultados corroboram com estudos recentes desenvolvidos em áreas cultivadas com cana-de-açúcar (BRITO et al., 2009; PANOSSO et al., 2012; BICALHO, 2016).

Tabela 8. Resumo do resultado de ajuste dos modelos Análises de Equações Estruturais, área Pradópolis – SP, Latossolo Vermelho Eutroférico.

Variáveis Latentes	FCO ₂				Variáveis Latentes	Estoque de Carbono			
	Estimativa	Erro Padrão	Valor Z	P - Valor		Estimativa	Erro Padrão	Valor Z	P - Valor
F1					F1				
Ca	0,91	0,02	38,21	<0,01	Ca	0,90	0,03	34,00	<0,01
Mg	0,80	0,04	21,04	<0,01	Mg	0,79	0,04	19,54	<0,01
CTC	0,91	0,03	27,14	<0,01	CTC	0,91	0,04	24,02	<0,01
F2					F2				
Ds	-0,90	0,02	-51,54	<0,01	Ds	-0,89	0,02	-45,66	<0,01
Macro	0,96	0,01	69,41	<0,01	Macro	0,96	0,01	64,48	<0,01
PLA	0,94	0,02	59,65	<0,01	PLA	0,94	0,02	59,98	<0,01
Regressão					Regressão				
FCO₂					EstC				
F1	-0,14	0,08	-1,88	0,06	F1	-0,02	0,09	-0,26	0,79
F2	0,55	0,06	9,34	<0,01	F2	-0,49	0,07	-7,07	<0,01

Fonte: Dados do próprio autor.

PLA – Porosidade livre de água; Ds – Densidade do solo; Macro – Macroporosidade; Ca – Cálcio; Mg – Magnésio; CTC – Capacidade de troca de Cátions; FCO₂ – emissão de CO₂; EstC – Estoque de Carbono no solo; F1 – Fator Latente 1; F2 – Fator Latente 2.

Por fim, com intuito de avaliar a invariância dos modelos de AEE construídos e avaliados anteriormente, nos dados de Pradópolis – SP, testou-se os mesmos modelos teóricos nos dados da área de Guariba – SP (ano 2008) e Aparecida do Taboado – MS (ano 2014). O

mesmo processo de estimação foi adotado, e os diagramas de caminhos com as estimativas dos parâmetros padronizados para ambas as áreas, são apresentados nas Figuras 13 e 14.

A avaliação de ajuste, assim como os índices de ajustamento foram apresentados na Tabela 10. De acordo com os valores de referências das estatísticas e índices de qualidade de ajustamento dos modelos de AEE citados em Marôco (2010), os modelos de FCO₂ e estoque de carbono para as áreas de Guariba-SP e Aparecida do Taboado-SP, tiveram bom ajustamento, o que mostra que a estrutura dos modelos pode ser mantida.

Com relação aos modelos estruturais de Guariba-SP (área com Latossolo Vermelho Eutroférico), os resultados foram semelhantes aos de Pradópolis (área com Latossolo altamente argiloso), sendo que, o FCO₂ e o estoque de carbono tiveram efeito significativo positivo e significativo negativo, respectivamente, apenas do fator latente atributos físicos do solo (F2); o fato pode ser justificado por pesquisas anteriores (PANOSSO, 2011; BICALHO, 2016), que também mostram valores semelhantes nestas duas áreas, as quais são pertencentes à mesma região. A diferença foi no tamanho da influência (Área de Guariba-SP), sendo que o FCO₂, neste caso, sofreu 39 vezes mais efeito do fator latente atributos físicos do que do fator latente atributos químicos (F1), e, o estoque de carbono foi influenciado aproximadamente 5 vezes mais pelos atributos físicos do que pelos atributos químicos (Figura 13).

No modelo de AEE de Guariba-SP, referente ao FCO₂ (Figura 13), foi considerado a covariância entre as variáveis CTC e PLA, indicado pelo índice de modificação, o qual mostra o decréscimo na estatística Qui-Quadrado de avaliação de ajuste, caso haja liberdade ou fixação de algum parâmetro ou se erros de medida forem correlacionados (MARÔCO, 2010).

Tabela 9. Estatísticas e índices de qualidade de ajustamento dos modelos de AEE FCO₂ e estoque de carbono, área de Pradópolis – SP.

Estatística	FCO ₂	Estoque de Carbono
χ^2	22,527 (p = 0,068)	23,903 (p = 0,057)
χ^2/gl	1,880	1,991
CFI	0,986	0,979
GFI	1,000	1,000
TLI	0,960	0,964
PGFI	0,655	0,620
RMSEA	0,080	0,096

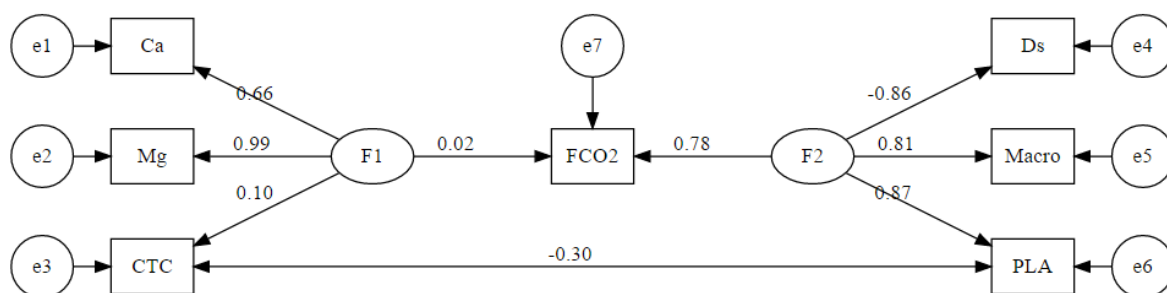
Fonte: Dados do próprio autor.

FCO₂ – emissão de CO₂, χ^2 = valor qui-quadrado do teste de ajustamento, χ^2/gl = razão entre valor qui-quadrado do teste de ajustamento e os graus de liberdade do modelo, CFI = comparative fit index, GFI = goodness fit index, TLI = Tucker-Lewis index, PGFI = parsimony GFI, RMSEA = root mean square error of approximation.

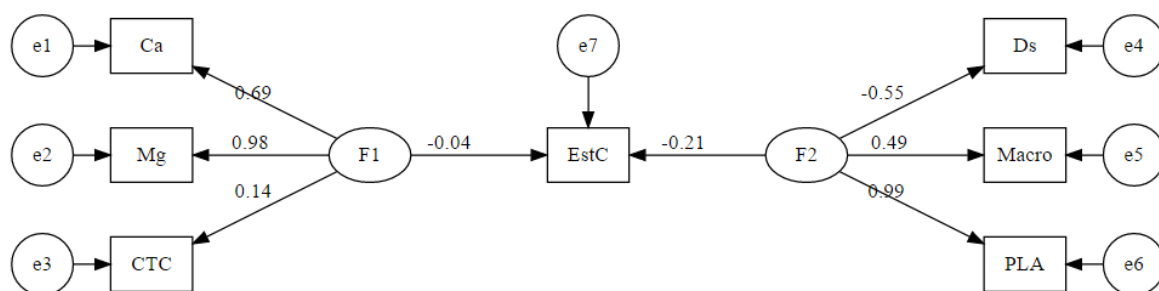
Estudos conduzidos por Panosso et al. (2012) e Moitinho (2014), em experimentos de cultivos de cana-de-açúcar, mostraram que as primeiras componentes principais tiveram maiores correlações com os atributos físicos, e, a emissão de CO₂ está diretamente relacionada ao aumento da PLA, Macro e CTC, como mostrou o modelo de AEE deste trabalho, para as áreas de Pradópolis-SP e Guariba-SP. Na área de Pradópolis-SP, os atributos físicos Macro e PLA foram mais importantes na explicação da emissão de CO₂ (efeito positivo) e no estoque de carbono (efeito negativo); enquanto que na área de Guariba-SP, os atributos físicos Ds e PLA foram mais importantes na explicação da emissão de CO₂ (efeito positivo) e no estoque de carbono (efeito negativo), sendo que PLA e Ds apresentaram relação inversa no fator latente atributos físicos.

Figura 13. Diagrama de Caminhos da análise de equações estruturais, área Guariba- SP, Latossolo Vermelho Eutroférico.

a)



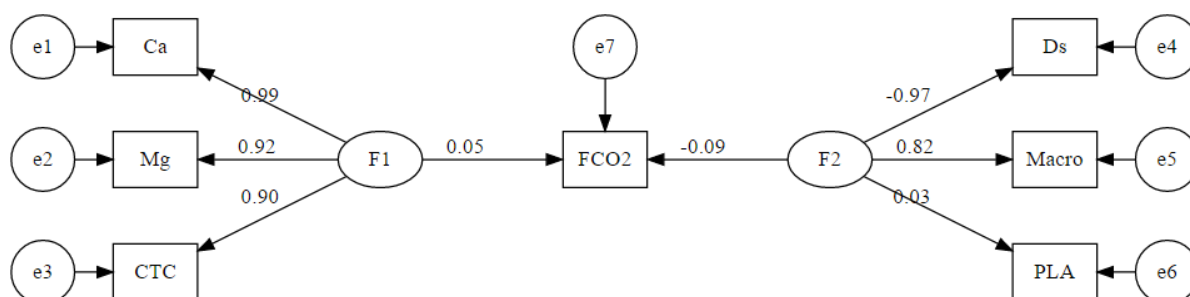
b)



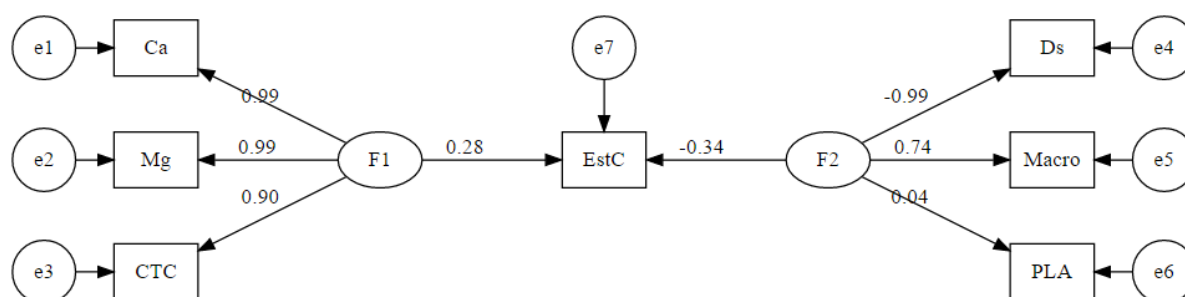
Fonte: Dados do próprio autor.

Figura 14. Diagrama de Caminhos da análise de equações estruturais, área Ap. do Taboado, Latossolo Vermelho-Amarelo Distrófico.

a)



b)



Fonte: Dados do próprio autor.

Nos modelos estruturais dos dados da área de Aparecida de Taboado – MS (Latossolo Vermelho Distroférrico), o atributo dependente emissão de FCO_2 não sofreu influência significativa dos fatores atributos químicos (F1) e atributos físicos (F2). Por outro lado, o atributo dependente estoque de carbono sofreu influência significativa tanto do fator latente atributos químicos (F1, influência positiva), como do fator latente atributos físicos (F2, influência negativa), ao nível de 5% de significância estatística (Figura 14). No entanto, o atributo PLA não apresentou significância estatística para nenhum dos dois modelos de medida.

Tabela 10. Estatísticas e índices de qualidade de ajustamento dos modelos de AEE FCO₂ e estoque de carbono, áreas de Guariba – SP e Aparecida do Taboado – MS.

Estatística	Dados - 2008 (Guariba - SP)		Dados – 2014 (Ap. do Taboado - MS)	
	FCO ₂	Estoque de Carbono	FCO ₂	Estoque de Carbono
χ^2	21.807(p=0,057)	30,388(p=0,051)	14,737(p=0,256)	20,670(0,061)
χ^2/gl	1,982	2,762	1,228	1,722
CFI	0,860	0,934	0,964	0,900
GFI	0,930	1,000	0,958	0,940
TLI	0,838	0,873	0,937	0,824
PGFI	0,665	0,625	0,680	0,640
RMSEA	0,09	0,10	0,04	0,08

Fonte: Dados do próprio autor.

FCO₂ – emissão de CO₂, χ^2 = valor qui-quadrado do teste de ajustamento, χ^2/gl = razão entre valor qui-quadrado do teste de ajustamento e os graus de liberdade do modelo, CFI = comparative fit index, GFI = goodness fit index, TLI = Tucker-Lewis index, PGFI = parsimony GFI, RMSEA = root mean square error of approximation.

Apesar dos modelos para área de Aparecida do Taboado-MS terem se ajustado bem aos dados (Tabela 10), a emissão de CO₂ não pode ser explicada de forma significativa nem pelo fator atributos físicos, nem pelo fator atributos químicos. Porém, para o estoque de carbono, os dois fatores o explicaram de forma significativa, sendo assim, importantes para entender as variações na estocagem de carbono no solo; sendo os atributos químicos Ca e Mg e os atributos físicos Ds e Macro, os mais importantes. Os atributos Ds e Macro apresentaram relação inversa dentro do fator atributos físicos.

Bahia (2012) escreve que a dinâmica de carbono tem sido condicionada principalmente à mudanças nos atributos físicos do solo, que são afetados por práticas de preparo e manejo do solo. Portanto, deve-se atentar às práticas de manejo para evitar variações negativas principalmente nos atributos físicos do solo, aqui destaca-se a PLA, DS e Macro; apontados pelo modelo de AEE, e, desta forma, admitir e realizar atividades agrícolas rentáveis e sustentáveis (LA SCALA JR, et al., 2000; CERRI et al., 2009; FIGUEIREDO et al., 2010).

Os modelos conceituais de análise de equações estruturais construídos neste estudo, podem ser considerados como modelos base em estudos na área agrícola, inclusive para outras classes de solo, quando o intuito seja a modelagem de dados da dinâmica e estoque de carbono, para o melhor entendimento de suas relações com os demais atributos físicos e químicos do solo.

3.4 - CONCLUSÕES

Foram construídos e ajustados os modelos conceituais de análises de equações estruturais para emissão de FCO_2 e estoque de carbono, considerando-se os dados da área de Pradópolis-SP. Em seguida, foi possível testar os modelos anteriores aos dados das áreas de Guariba-SP e Aparecida do Taboado-MS, com intuito de avaliar a invariância dos modelos.

Os modelos de equações estruturais se ajustaram bem aos dados, podendo reproduzir por meio da matriz de seus parâmetros a matriz de covariância dos dados amostrais, e, avaliar as interações de forma simultânea, entre a emissão de CO_2 , estoque de carbono e os fatores latentes formados pelos atributos químicos e físicos.

Os modelos estruturais para as áreas de Pradópolis-SP e Guariba-SP apresentaram resultados semelhantes, e, mostraram que tanto as variações de emissão CO_2 , quanto as de estoque de carbono, podem ser explicadas de forma significativa pelo fator latente atributos físicos, com destaque para Macro e PLA.

No modelo estrutural da área de Aparecida do Taboado – MS, constatou-se que as variações de emissão de CO_2 não sofreram influência significativa dos fatores latentes atributos químicos e físicos. Porém, as variações dos valores de estoque carbono, foram explicadas de forma estatisticamente significativas pelos dois fatores latentes.

Portanto, pode-se propor um modelo de equações estruturais, em que foi capaz de explicar as variações de emissão de CO_2 e estoque de carbono por meio de dois fatores latentes formados por atributos químicos e físicos; evidenciando-se assim um melhor entendimento da dinâmica de carbono associada aos atributos do solo, e possibilitando a obtenção de informações úteis como recomendação para as práticas agrícolas. Ressalta-se, que o modelo pode ser adaptado, bem como melhorado, devido a dinamicidade das mudanças no solo e utilizado como base para melhor entendimento das variações dos atributos relacionados a qualidade física e química do solo, bem como dos componentes de desempenho de culturas, em áreas com outras classes de solo.

REFERÊNCIAS

- ALLAIRE, S. E.; LANGE, S. F.; LAFOND, J. A.; PELLETIER, B.; CAMBOURIS, A. N.; DUTILLEUL, P. Multiscale spatial variability of CO₂ emissions and correlations with physico-chemical soil properties. **Geoderma**, Amsterdam, v. 170, p.251-260, 2012.
- BAHIA, A. S. R. S. **Espectroscopia de reflectância difusa como técnica auxiliar na caracterização da emissão de CO₂ e atributos do solo**. 2012. 74 f. Dissertação (Mestrado em Agronomia) – Faculdade de Ciências Agrárias e Veterinária, Universidade Estadual Paulista, Jaboticabal, 2012.
- 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 & Tillage Research**, Amsterdam, v. 54, n. 1-2, p. 101-109, 2000.
- BEBÉ, F. V.; ROLIM, M. M.; PEDROSA, E. M. R.; SILVA, G. B.; OLIVEIRA, V. S. Avaliação de solos sob diferentes períodos de aplicação com vinhaça. **Revista Brasileira de Engenharia Agrícola e Ambiental**, Campina Grande, v. 13, n. 6, p. 781-787, abr. 2009.
- BICALHO, E. S. **Soil greenhouse gas emissions and their relations to soil attributes in a sugarcane area**. 2016. 80 f. Tese (Doutorado em Agronomia) - Faculdade de Ciências Agrárias e Veterinária, Universidade Estadual Paulista, Jaboticabal, 2016.
- BORTOLUZZI, E. C.; RHEINHEIMER, D. dos S.; PETRY, C.; KAMINSKI, J. Contribuição de constituintes de solo à capacidade de troca de cátions obtida por diferentes métodos de extração. **Revista Brasileira de Ciência do Solo**, Viçosa, MG, v. 33, p. 507-515, 2009.
- BRAHIM, N.; BLAVET, D.; GALLALI, T.; BERNOUX, M. Application of structural equation modeling for assessing relationships between organic carbon and soil properties in semiarid Mediterranean region. **International Journal of Environmental Science and Technology**, Berlin, v. 8, n. 2, p. 305-320, 2011.
- BRITO, L. F.; MARQUES JÚNIOR, J.; PEREIRA, G. T.; LA SCALA JÚNIOR, N. Spatial variability of soil CO₂ emission in different topographic positions. **Bragantia**, Campinas, v. 69, n. Suplemento, p. 19-27, 2010.
- BRITO, L. F.; MARQUES JÚNIOR, J.; PEREIRA, J. T.; SOUZA, Z. M.; LA SCALA JÚNIOR, N. Soil CO₂ emission of sugarcane field as affected by topography. **Scientia Agricola**, Piracicaba, v. 66, n. 1, p. 77-83, 2009.

CERRI, C. C.; MAIA, S. M. F.; GALDOS, M. V.; CERRI, C. E. P.; FEIGL, B. J.; BERNOUX, M. Brazilian greenhouse gas emissions: the importance of agriculture and livestock. **Scientia Agricola**, Piracicaba, v. 66, n. 6, p. 831-843, 2009.

CHEN, Y.; MILLER, J. R.; FRANCIS, J. A. Observed and modeled relationships among Arctic climate variables. **Journal of Geophysical Research: Atmospheres**, Washington, v. 108, n. 24, p. 1-13, 2003.

CHRISTENSEN, B.T. **Carbon in primary and secondary organomineral complexes**. In: CARTER, M. R.; STEWART, B. A. (Ed.). Structure and organic matter storage in agricultural soils. Boca Raton: CRC, 1996. p. 97-165.

EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA - EMBRAPA. **Manual de métodos de análise de solo**. 2 ed. Brasília, DF, 1997. 212 p.

EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA - EMBRAPA. **Sistema brasileiro de classificação de solos**. Brasília, DF, 2006. 412 p.

EPRON, D.; NOUVELLON, Y.; ROUPSARD, O.; MOUVONDY, W.; MABIALA, A.; SAINT-ANDRE', L.; JOFFRE, R.; JOURDAN, C.; BONNEFOND, J.; BERBIGIER, P.; HAMEL, O. Spatial and temporal variations of soil respiration in a Eucalyptus plantation in Congo. **Forest Ecology and Management**, Amsterdam, v. 202, p. 149-160, 2004.

FIGUEIREDO, E. B.; PANOSSO, A. R. ; ROMÃO, R.; LA SCALA JÚNIOR, N. Greenhouse gas emission associated with sugar production in southern Brazil. **Carbon Balance and Management**, New York, v. 5, p. 3, 2010.

FILZMOSER, P.; GARRETT, R. G.; REIMANN, C. Multivariate outlier detection in exploration geochemistry. **Computers & Geosciences**, Amsterdam, v. 31, n. 5, p. 579-587, 2005.

FOLLETT, R. F. Soil management concepts and carbon sequestration zin cropland soils. **Soil & Tillage Research**, Amsterdam, v. 61, n. 1-2, p. 77-92, 2001.

FONTANA, A.; PEREIRA, M. G.; LOSS, A.; CUNHA, T. J. F.; SALTON, J. C. Atributos de fertilidade e frações húmicas de um Latossolo Vermelho no Cerrado. **Pesquisa Agropecuária Brasileira**, Brasília, DF, v. 41, p. 847-853, 2006.

GOYAL, P.; SHARMA, P.; SRIVASTAVA, S.; SRIVASTAVA, M. M. Saraca indica leaf powder for decontamination of Pb: removal, recovery, adsorbent characterization and

equilibrium modeling. **International Journal of Environmental Science and Technology**, Berlin, v. 5, n. 1, p. 27-34, 2008.

GRAF, A.; HERBST, M.; WEIHERMÜLLER, L.; HUISMAN, J. A.; PROLINGHEUER, N.; BORNEMANN, L.; VEREECKEN, H. Analyzing spatiotemporal variability of heterotrophic soil respiration at the field scale using orthogonal functions. **Geoderma**, Amsterdam, v. 181-182, p. 91-101, 2012.

HORN, J. L. A rationale and test for the number of factors in factor analysis. **Psychometrika**, New York, v. 30, n. 2, p. 179-185, 1965.

IPCC. Climate change 2007: the physical science basis. In: _____. **Contribution of working group i to the fourth assessment report of the intergovernmental panel on climate change**. Cambridge: Cambridge University Press, 2007. p. 220-269.

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

KEMMITT, S. J.; LANYON, C. V.; WAITE, I. S.; WEN, Q.; ADDISCOTT, T. M.; BIRD, N. R. A.; O'DONNELL, A. G.; BROOKES, P. C. Mineralization of native soil organic matter is not regulated by the size, activity or composition of the soil microbial biomass—a new perspective. **Soil Biology and Biochemistry**, Amsterdam, v. 40, n. 3, p. 61–73, 2008.

KOSUGI, Y.; MITANI, T.; ITOH, M.; NOGUCHI, S.; TANI, M.; MATSUO, N.; TAKANASHI, S.; OHKUBO, S.; NIK, A. R. Spatial and temporal variation in soil respiration in a Southeast Asian tropical rainforest. **Agric For Meteorol**, Amsterdam, v. 147, p.35-47, 2007.

LA SCALA JÚNIOR, N.; MARQUES JÚNIOR, J.; PEREIRA, G. T.; CORÁ, J. E. Shortterm temporal changes in the spatial variability model of CO₂ emissions from a Brazilian bare soil. **Soil Biology and Biochemistry**, Amsterdam, v. 32, n. 10, p. 1459-1462, 2000.

LAL, R. World cropland soils as a source or sink for atmospheric carbon. **Advances in Agronomy**, San Diego, v. 71, p. 145-191, 2001.

LEITE, L. F. C.; MENDONÇA, E. S. Modelo century de dinâmica da matéria orgânica do solo: Equações e pressupostos. **Ciência Rural**, Santa Maria, v. 33, n. 4, p. 679-686, 2003.

LEON, E.; VARGAS, R.; BULLOCK, S.; LOPEZ, E.; PANOSSO, A. R.; LA SCALA JÚNIOR, N. Hot spots, hot moments, and spatio-temporal controls on soil CO₂ efflux in a water-limited ecosystem. **Soil Biology And Biochemistry**, Amsterdam, v. 77, p. 12-21, 2014.

MARÔCO, J. **Análise de equações estruturais: fundamentos teóricos, software e aplicações.** Pero Pinheiro: Report Number, 2010.

MINGOTI, S. A. **Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada.** Belo Horizonte: Editora UFMG, 2005.

MOITINHO, M. R.; PADOVAN, M. P.; PANOSSO, A. R.; LA SCALA JÚNIOR, N. Efeito do preparo do solo e resíduo da colheita de cana-de-açúcar sobre a emissão de CO₂. **Rev. Bras. Ciênc. Solo**, Viçosa, MG, v. 37, n. 6, p. 1720-1728, 2013.

MOITINHO, M. R.; TEIXEIRA, D. B.; ALMEIDA, R. F.; PADOVAN, M. P.; LA SCALA, N. Emissão de CO₂ Associada ao Manejo e Atributos do Solo em Áreas de Cana-de-açúcar no Interior de São Paulo. **Cadernos de Agroecologia**, Recife, v. 9, n. 4, p. 1-12, 2014.

OLIVEIRA, P. H. G. **Modelagem de equações estruturais aplicada ao estudo das relações entre qualidade e diversidade da fauna no sistema solo-serapilheira.** 2016. 117 f. Tese (Doutorado em Agronomia) - Universidade Estadual do Norte Fluminense Darcy Ribeiro, Campos dos Goytacazes - RJ, 2016.

PANOSSO, A. R. **Variabilidade espacial da emissão de CO₂ e sua relação com propriedades do solo em área de cana-de-açúcar no sudeste do Brasil.** 2011. 107 f. Tese (Doutorado em Agronomia) – Faculdade de Ciências Agrárias e Veterinária, Universidade Estadual Paulista, Jaboticabal, 2011.

PANOSSO, A. R.; PERILLO, L. I.; FERRAUDO, A. S.; PEREIRA, G. T.; MIRANDA, J. G. V.; LA SCALA, N. Fractal dimension and anisotropy of soil CO₂ emission in a mechanically harvested sugarcane production area. **Soil and Tillage Research**, Amsterdam, v. 124, n. 1, p. 8–16, 2012.

PAUSTIAN, K.; LEVINE, E.; POST, W. M.; RYZHOVA, I. M. The use of models to integrate information and understanding of soil C at the regional scale. **Geoderma**, Amsterdam, v. 79, p. 227–260, 1997.

PITOMBO, L. M.; CARMO, J. B.; MARIA, I. C.; ANDRADE, C. A. Carbon sequestration and greenhouse gases emissions in soil under sewage sludge residual effects. **Scientia Agricola**, Piracicaba, v. 72, n. 2, p. 147-156, 2015.

POST, W. M.; KWON, K. C. Soil carbon sequestration and land-use change: processes and potential. **Global Change Biology**, Washington, v. 6, n. 3, p. 317-327, mar. 2000.

R CORE TEAM. **R**: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing, 2014. Disponível em: <<http://www.R-project.org/>>. Acesso em: 10 fev. 2016.

RAIJ, B. V. **Análise química para avaliação da fertilidade de solos tropicais**. Campinas: Instituto Agrônômico, 2001. 285 p.

RYU, S.; CONCILIO, A.; CHEN, J.; NORTH, M.; MA, S. Prescribed burning and mechanical thinning effects on belowground conditions and soil respiration in a mixed-conifer forest, California. **Forest Ecology and Management**, Amsterdam, v. 257, n. 4, p. 1324–1332, 2009.

SATORRA, A.; BENTLER, P. M. A scaled difference chi-square test statistic for moment structure analysis. **Psychometrika**, New York, v. 66, n. 4, p. 507-514, 2001.

SATORRA, A.; BENTLER, P. M. Scaling corrections for chi-square statistics in covariance structure analysis. In: BUSINESS AND ECONOMIC STATISTICS SECTION, 1988, Alexandria. **Proceedings...** Washington: American Statistical Association, 1988. p. 308-313, 1988.

SCHOLE, M.; ANDREAE, M. O biogenic and pyrogenic emissions from africa and their impact on the global atmosphere. **A Journal of the Human Environment**, Dordrecht, v. 29, n. 1, p. 23-29, fev. 2000.

SMITH, K. A.; BALL, T; CONEN, F.; DOBBIE, K.E.; MASSHEDER, J.; REY, A. Exchange of greenhouse gases between soil and atmosphere: interactions of soil physical factors and biological processes. **Eur J. Soil Science**, Washington, v. 54, n. 4, p. 779-791, dez. 2003.

TAVARES, R; SOUZA, Z.M.; LA SCALA JR, N.; CASTIONI, G.A.F.; SOUZA, G.S.; TORRES, J.L. Spatial and Temporal Variability of Soil CO₂ Flux in Sugarcane Green Harvest Systems. **Rev. Bras. Ciênc. Solo**, Viçosa, MG, v. 40, p.1-14, 2016.

USSIRI, A. N.; LAL, R., 2009. Long-term tillage effects on soil carbon storage and carbon dioxide emissions in continuous corn cropping system from an alfisol in Ohio. **Soil & Tillage Research**, Amsterdam, v. 104, n. 1, p. 39-47, 2009.

XU, M.; QI, Y. Soil-surface CO₂ efflux and its spatial and temporal variations in a Young ponderosa pine plantation in northern California. **Global Change Biology**, Washington, v. 7, p. 667–677, 2001.

ZERAATPISHE, M; KHORMALI, F. Carbon stock and mineral factors controlling soil organic carbon in a climatic gradient, Golestan province. **Journal of Soil Science And Plant Nutrition**, Temuco, v. 12, n. 4, p. 637-654, 2012.

CAPÍTULO 4 – CONSIDERAÇÕES FINAIS

O entendimento das interações no sistema solo-planta-atmosfera está cada vez mais necessário para realização de uma agricultura mais rentável e sustentável. No entanto, este sistema é bem complexo, devido à enorme quantidade de fatores e variações que o envolve. A literatura aponta para eficácia de modelos estatísticos mais avançados, os quais conseguem contribuir de forma satisfatória para interpretabilidade destes fenômenos.

Neste trabalho, aplicou-se e propôs-se modelos estatísticos multivariados, que proporcionaram entender as interações ocorridas entre os componentes de produção e tecnológicos da cana-de-açúcar, bem como explicar a dinâmica de carbono; por meio de fatores latentes formados pelos atributos físicos e químicos do solo.

As técnicas estatísticas multivariadas utilizadas neste trabalho foram a análise dos componentes principais (ACP), análise fatorial exploratória (AFE) e a análise de equações estruturais (AEE). Destas técnicas, a ACP e a AFE foram úteis para entender a variação dos dados amostrais em números menores de dimensões, bem como auxiliarem na composição dos fatores latentes a serem utilizados nos modelos de medida da AEE, que buscou entender e explicar as variações ocorridas entre os atributos do solo e os fatores latentes, de forma simultânea.

Especificamente, para as áreas estudadas, os componentes de produção e tecnológicos da cana-de-açúcar foram influenciados significativamente e de forma negativa pelo fator latente atributos físicos do solo DS, PT e Macro, evidenciando efeito de compactação na camada de 0 a 0,20 m devido ao tráfego de máquinas (3º corte). Porém, na camada de 0,20 a 0,40 m do solo, o fator latente atributos químicos associados a acidez do solo, pHCa, pHK e pHag, influenciou de forma significativa e positiva os componentes de produção e tecnológicos da cana-de-açúcar, efeito benéfico da aplicação de gesso após o primeiro corte.

No que tange a dinâmica de carbono, os modelos conceituais para as áreas de Pradópolis-SP e Guariba-SP apresentaram resultados semelhantes, e, mostraram que tanto as variações de emissão CO_2 , quanto as de estoque de carbono, podem ser explicadas de forma significativa pelo fator latente atributos físicos, com destaque para Macro e PLA. No modelo de AEE para a área de Aparecida do Taboado-SP, as variações de estoque de carbono foram explicadas de forma significativa tanto pelo fator latente atributos químicos (Ca, Mg e CTC), quanto pelo fator latente atributos físicos (Ds e Macro).

Os modelos conceituais de análise de equações estruturais propostos neste estudo podem ser adaptados, bem como melhorados, devido a dinamicidade das mudanças no solo e utilizados

como base para melhor entendimento das variações dos atributos relacionados a qualidade física e química dos solos, bem como dos componentes de desempenho de culturas, em áreas com outras classes de solo; pois a quantidade de informações extraídas e de forma simultânea é o diferencial deste modelo, tão pouco utilizado no meio agrícola.

Deve-se levar em conta alguns cuidados com a modelagem de equações estruturais, tais como a estrutura da matriz de covariância ou de correlação e os desvios de normalidade multivariada dos dados amostrais. Observou-se neste trabalho, problemas como pequena correlação entre certos pares de atributos, ocasionando prejuízos como a singularidade da matriz de correlação, optando-se assim pela não utilização destes atributos e, desvios de normalidade multivariada, levando à retirada de observações ou aplicação de métodos robustos de estimação dos parâmetros.