

**Mickaelle Maria de Almeida Pereira**

**Aplicação da análise de dados funcionais nas emissões do CH<sub>4</sub>, CO<sub>2</sub> e N<sub>2</sub>O  
provenientes de diferentes usos do solo**

**Recife – PE  
26 de fevereiro de 2024**

Dados Internacionais de Catalogação na Publicação  
Universidade Federal Rural de Pernambuco  
Sistema Integrado de Bibliotecas  
Gerada automaticamente, mediante os dados fornecidos pelo(a) autor(a)

---

- P436a      Pereira, Mickaelle Maria de Almeida  
              Aplicação da análise de dados funcionais nas emissões do CH<sub>4</sub>, CO<sub>2</sub> e N<sub>2</sub>O provenientes de diferentes usos do solo  
              / Mickaelle Maria de Almeida Pereira. - 2024.  
              90 f. : il.
- Orientador: Paulo Jose Duarte Neto.  
              Inclui referências, apêndice(s) e anexo(s).
- Tese (Doutorado) - Universidade Federal Rural de Pernambuco, Programa de Pós-Graduação em Biometria e  
              Estatística Aplicada, Recife, 2027.
1. Curvas funcionais. 2. Derivativos. 3. Fluxos. 4. GEEs. 5. Mudança no uso da terra. I. Neto, Paulo Jose Duarte,  
              orient. II. Título

CDD 519.5

---



**UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO**  
**PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM BIOMETRIA E ESTATÍSTICA**  
**APLICADA**

**Aplicação da análise de dados funcionais nas emissões do CH<sub>4</sub>, CO<sub>2</sub> e N<sub>2</sub>O  
provenientes de diferentes usos do solo**

Tese apresentada ao Programa de Pós-Graduação em Biometria e Estatística Aplicada do Departamento de Estatística e Informática da Universidade Federal Rural de Pernambuco como requisito parcial para obtenção do grau de Doutorado em Biometria e Estatística Aplicada

Área de Concentração: Biometria e Estatística Aplicada.

**Orientador: Prof. Dr. Paulo José Duarte Neto**

**Recife – PE**  
**26 de fevereiro de 2024**

**Mickaelle Maria de Almeida Pereira**

**Aplicação da análise de dados funcionais nas emissões do CH<sub>4</sub>, CO<sub>2</sub> e N<sub>2</sub>O  
provenientes de diferentes usos do solo**

Tese julgada adequada para obtenção do título de Doutor em Biometria e Estatística Aplicada, defendida e aprovada por unanimidade em 26/02/2024 pela Banca Examinadora.

Orientador:

---

Prof. Dr. Paulo José Duarte Neto  
Universidade Federal Rural de Pernambuco

Banca Examinadora:

---

Prof. Dr. Tiago Alessandro Espínola Ferreira  
Universidade Federal Rural de Pernambuco

---

Prof. Luiz Antonio Martinelli  
Universidade de São Paulo

---

Prof.<sup>a</sup> Janaina Braga do Carmo  
Universidade Federal de São Carlos

---

Prof. José Romualdo de Sousa Lima  
Universidade Federal do Agreste de Pernambuco

# AGRADECIMENTOS

Agradeço a Deus e a nossa senhora do Perpétuo Socorro.

Ao meu esposo Lucimário Gois, a minha mãe Maria Marcia e aos meus avós Maria de Lourdes e José Vicente por todo amor, carinho e dedicação. As minhas tias e tios, em particular, Maria Alexandrina, Maria Mariza, Maria Mirian, Maria José, Mauricio Almeida e Manoel Murilo e aos meus sogros Luci Gois e Cláudio Mario. Aos meus primos e primas, em especial, Maurício Junior, Aleska Maria, Adones Melo e Bruna Costa. Aos meus afilhados Miguel, Gustavo e Davi.

Ao meu orientador Professor Doutor Paulo José Duarte Neto pela disponibilidade, apoio e amizade durante o desenvolvimento da pesquisa e por toda paciência durante esses quatro anos, em me explicar o assunto quantas vezes necessário.

As minhas amigas Leika Irabele, Quercia Elói, Denise Almeida e Jane Diniz e ao meu amigo Jucarlos Rufino, pelo apoio e incentivo todos esses anos. Agradeço a todos os professores, equipe escolar, colegas e amigos que passaram pelo meu caminho, desde a alfabetização até o doutorado.

Agradeço a todo o corpo docente e demais funcionais do DEINFO pelo compartilhamento dos seus conhecimentos.

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

*“Os que semeiam entre lágrimas,  
com alegria colherão.”  
(Salmo 126:5-6).*

# RESUMO

O manejo não adequado dos sistemas agrícolas, bem como mudanças na cobertura do solo são fatores significativos para aumento dos gases de efeito estufa (GEEs). Existem metodologias estatísticas para mensurar o fluxo dos GEEs, baseadas na variação temporal da concentração, em função do tempo. No entanto, essas análises podem não abranger características provenientes da variação e aleatoriedade presentes no fenômeno. Sendo assim, a análise de dados funcionais (*Functional data analysis*, FDA) consolida uma nova perspectiva para derivar modelos e otimizar técnicas na análise exploratória de dados, expressando notável potencial no estudo das variações de uma determinada variável em relação ao tempo, levando em consideração tanto a variação contínua (linear ou não linear) quanto sua aleatoriedade. O objetivo do estudo é representar e analisar o comportamento das emissões do dióxido de carbono (CO<sub>2</sub>), metano (CH<sub>4</sub>) e óxido nitroso (N<sub>2</sub>O) em quatro tipos de uso do solo (pasto extensivo, pasto intensivo, linha e entrelinha da cana-de-açúcar) utilizando a FDA sobre as perspectivas univariada e multivariada. Inicialmente, as observações discretas das concentrações foram reestruturadas como funções, utilizando a suavização *B-splines*. Consequentemente, derivadas das funções foram aplicadas no cálculo da variação de concentrações em função do tempo (fluxos), constatando grande variabilidade para os usos da terra, alternando-se entre efluxos e influxos. O modelo funcional usa a própria variação do gás em todo intervalo de tempo, conseguindo ter uma boa representação para o fluxo, tanto para variação na estrutura linear como não linear dos gases. Considerando a análise dos gases de forma conjunta, os três primeiros componentes principais funcional multivariado (MFPCA), capturaram, cumulativamente, mais de 90% da variação total presente nos dados. Com os *scores* dos componentes principais multivariado, foram utilizados para análise de cluster e de classificação. Portanto, a aplicação da FDA demonstra ser capaz de captar o comportamento do fenômeno estudado, englobando a natureza contínua do sistema. Sendo capaz de representar o processo das trocas dos gases entre solo e ar atmosférico, para gases que tenham um comportamento tanto linear como não linear.

**Palavras-chave:** curvas funcionais, derivativos, fluxo, GEEs, mudança no uso da terra.

# ABSTRACT

Inadequate management of agricultural systems and changes in soil cover are significant factors in the increase of greenhouse gases (GHGs). There are statistical methodologies for measuring the flow of GHGs, based on the temporal variation in concentration as a function of time. However, these analyses may not cover characteristics arising from the variation and randomness present in the phenomenon. Functional data analysis (FDA) therefore consolidates a new perspective for deriving models and optimizing techniques in exploratory data analysis, expressing remarkable potential in the study of variations of a given variable in relation to time, taking into account both continuous variation (linear or non-linear) and randomness. The aim of the study is to represent and analyze the behavior of carbon dioxide (CO<sub>2</sub>), methane (CH<sub>4</sub>) and nitrous oxide (N<sub>2</sub>O) emissions in four types of land use (extensive pasture, intensive pasture, sugarcane row and inter-row) using FDA from the univariate and multivariate perspectives. Initially, the discrete concentration observations were restructured as functions using B-splines smoothing. Consequently, derivatives of the functions were applied to calculate the variation in concentrations as a function of time (fluxes), finding great variability for land uses, alternating between effluxes and influxes. The functional model uses the variation of the gas itself over the entire time interval, providing a good representation of the flow, both for variations in the linear and non-linear structure of the gases. Considering the analysis of the gases as a whole, the first three multivariate functional principal components (MFPCA) cumulatively captured more than 90% of the total variation present in the data. The scores of the multivariate principal components were used for cluster and classification analysis. Therefore, the application of the FDA proves capable of capturing the behavior of the phenomenon studied, encompassing the continuous nature of the system. It is able to represent the process of gas exchange between soil and atmospheric air, for gases that have both linear and non-linear behavior.

**Keywords:** functional curves, derivatives, flow, GHGs, land use change.

# LISTA DE FIGURAS

- Figura 1 - (a) Delimitação da área experimental: de vegetação florestal (verde); parcelas do pasto extensivo (preto); parcelas do pasto intensificado (amarelo); parcelas da cana-de-açúcar (vermelho). (b) Cana-de-açúcar dois meses após o plantio. (c) Pastagem intensificada (verde escuro) dois meses após o plantio. Fonte: adaptado do segundo relatório científico do projeto: Consequências ambientais da conversão pastagem-cana-de-açúcar e intensificação de pastagens. Instituição Sede do projeto: Centro de Energia Nuclear na Agricultura (CENA). ..31
- Figura 2 - Fluxograma do modelo de classificação. Fonte: o autor. ....37
- Figura 3 - Exemplos de situações em que existe mudança de linearidade de maneira gradativa nas emissões das concentrações (pontos) do CO<sub>2</sub> para o pasto extensivo, coletadas através da câmara estática fechada inserida no solo. O fluxo ( $f_0$ ) é medido a partir da reta de regressão para o modelo linear (LM) (linha vermelha). O modelo HM é utilizado para mensurar o valor do fluxo para a curva exponencial (EXP) (linha azul) e o valor do fluxo para o modelo funcional (FDA) (linha preta) é mensurado a partir da derivada da função. Fonte: autor. ....40
- Figura 4 - Curvas funcionais (painel superior) representando as concentrações dos gases CH<sub>4</sub>, CO<sub>2</sub> e N<sub>2</sub>O para quatro usos do solo. Linha pontinha azul, representa quantis 2,5; 25; 75; 97,5. A curva em vermelho corresponde a mediana, e em preto a média. Painel intermediário representa o desvio padrão funcional referente as concentrações dos gases CH<sub>4</sub>, CO<sub>2</sub> e N<sub>2</sub>O para quatro usos do solo. Curvas derivativas (painel inferior) das concentrações suavizadas dos gases (fluxo). Fonte: o autor. ....45
- Figura 5 - Curvas funcionais representando as concentrações do gás N<sub>2</sub>O para quatro usos do solo: (a) Pasto extensivo; (b) pasto intensivo; (c) linha da cana-de-açúcar; (d) entrelinha da cana-de-açúcar. A curva em preto a média funcional. Fonte: o autor. ....46
- Figura 6 - Fluxo médio diário do CH<sub>4</sub>, CO<sub>2</sub> e N<sub>2</sub>O, em relação aos 42 dias de coletas para: pasto extensivo (a) - (c); pasto intensivo (d) - (f); linha da cana-de-açúcar (g) - (i); entrelinha da cana-de-açúcar (j) - (l). Fonte: o autor. ....49
- Figura 7 - Primeiro componente principal (painéis da esquerda) e segundo componente principal (painéis da direita) das emissões de: (a) CH<sub>4</sub>, (b) CO<sub>2</sub> e (c) N<sub>2</sub>O, em relação aos quatro usos do solo. As curvas mostram o perfil médio (linha sólida) e os efeitos da adição (+) e da subtração (-) de autofunções. Para os três PC, cada autofunção é multiplicada por um fator de 1,5 para aumentar a deformação das curvas (+) e (-). Fonte: o autor. ....52

Figura 8 - Biplot dos <i>scores</i> dos componentes principais funcional das emissões de: (a) CH <sub>4</sub> , (b) CO <sub>2</sub> e (c) N <sub>2</sub> O, em relação aos usos da terra: pasto extensivo (amarelo), pasto intensivo (vermelho), linha da cana-de-açúcar (verde) e entrelinha da cana-de-açúcar (azul). Fonte: o autor. ....	53
Figura 9 - Representação das emissões dos três componentes dos perfis médios do CH <sub>4</sub> , CO <sub>2</sub> e N <sub>2</sub> O. As curvas mostram o perfil médio (sólido) e os efeitos da adição (+) e da subtração (-) de autofunções. As porcentagens exibidas aos títulos dos cabeçalhos são a quantidade de variação explicada por cada componente. As porcentagens no rótulo do eixo vertical são as quantidades de variação explicada por cada variável. Para os três PCs, cada autofunção é multiplicada por um fator de 1,5 para aumentar a deformação das curvas (+) e (-). Fonte: o autor. ....	55
Figura 10 - Análise de <i>Cluster</i> baseado nos três componentes principais para o caso multivariado. Esquemas de cores: amarelo grupo 1 (pastagem); vermelho grupo 2 (intermediário); verde grupo 3 (cana-de-açúcar); azul grupo 4 (picos das emissões do CO <sub>2</sub> e N <sub>2</sub> O provocados pelo manejo no solo). Fonte: o autor. ....	57
Figura 11 - Temperatura média diária do Ar (painel esquerdo) e do Solo (painel direito) (C°) para cada uso da terra: pasto extensivo (PE), pasto intensivo (PI), linha da cana-de-açúcar (LC) e entrelinha da cana-de-açúcar (EC). Fonte: o autor. ....	60
Figura 12 - Matrizes de confusão dos modelos de classificação a respeito da base de teste referentes ao cenário 1 (a), cenário 2 (b), cenário 3 (c), cenário 4 (d), cenário 5 (e), cenário 6 (f), cenário 7 (g) e cenário 8 (h). Em relação aos usos do solo pasto extensivo (PE), pasto intensivo (PI), linha da cana-de-açúcar (LC), entrelinha da cana-de-açúcar (EC) e a união das observações da linha e entrelinha da cana-de-açúcar (C). Fonte: o autor. ....	63

# LISTA DE TABELAS

Tabela 1 - Adubação e variedades utilizados durante a intensificação dos agroecossistemas, com a conversão de pastagem extensiva para pastagem intensiva e canavial, nas áreas experimentais.....	32
Tabela 2 - Etapas de manejo do solo realizadas nos tratamentos pastagem extensiva (PE), pastagem intensiva (PI) e cana-de-açúcar (C) durante o primeiro ciclo da cana-de-açúcar. ....	33
Tabela 3 - Medidas de ajuste em relação aos modelos linear, exponencial e funcional para os CH <sub>4</sub> para cada uso da terra. ....	41
Tabela 4 - Medidas de ajuste em relação aos modelos linear, exponencial e funcional para os CO <sub>2</sub> para cada uso da terra. ....	42
Tabela 5 - Medidas de ajuste em relação aos modelos linear, exponencial e funcional para os N <sub>2</sub> O para cada uso da terra. ....	43
Tabela 6 - Emissões acumuladas dos fluxos do CH <sub>4</sub> , CO <sub>2</sub> e N <sub>2</sub> O em relação ao pasto extensivo (PE), pasto intensivo (PI), linha da cana-de-açúcar (LC) e entrelinha da cana-de-açúcar (EC).	51
Tabela 7 - Acurácia média obtida na validação cruzada durante o treinamento e acurácia obtida pelo teste referente ao modelo de classificação. ....	59
Tabela 8 - Métricas de avaliação do modelo de classificação para cada uso do solo. A descrição dos cenários foi escrita anteriormente na Tabela 7. ....	61

# LISTA DE APÊNDICES

- Apêndice A1 - Análise do fator de correlação de tendência, uma medida de quão longe uma linha de melhor ajuste está de um ângulo de  $45^\circ$  através da origem para a comparação do modelo linear vs. funcional do  $\text{CH}_4$ , em relação ao pasto intensivo (a), pasto intensivo (b) linha da cana-de-açúcar (c) e entrelinha da cana-de-açúcar (d). A linha tracejada vermelha é a linha de concordância e representa a concordância perfeita entre o modelo linear e o modelo funcional (declive = 1 e intercepto = 0). A linha preta representa a regressão linear entre os modelos. Fonte: o autor..... 77
- Apêndice A2 - Análise do fator de correlação de tendência, uma medida de quão longe uma linha de melhor ajuste está de um ângulo de  $45^\circ$  através da origem para a comparação do modelo exponencial vs. funcional do  $\text{CH}_4$ , em relação ao pasto intensivo (a), pasto intensivo (b) linha da cana-de-açúcar (c) e entrelinha da cana-de-açúcar (d). A linha tracejada vermelha é a linha de concordância e representa a concordância perfeita entre o modelo linear e o modelo funcional (declive = 1 e intercepto = 0). A linha preta representa a regressão linear entre os modelos. Fonte: o autor..... 78
- Apêndice A3 - Análise do fator de correlação de tendência, uma medida de quão longe uma linha de melhor ajuste está de um ângulo de  $45^\circ$  através da origem para a comparação do modelo linear vs. funcional do  $\text{CO}_2$ , em relação ao pasto intensivo (a), pasto intensivo (b) linha da cana-de-açúcar (c) e entrelinha da cana-de-açúcar (d). A linha tracejada vermelha é a linha de concordância e representa a concordância perfeita entre o modelo linear e o modelo funcional (declive = 1 e intercepto = 0). A linha preta representa a regressão linear entre os modelos. Fonte: o autor..... 79
- Apêndice A4 - Análise do fator de correlação de tendência, uma medida de quão longe uma linha de melhor ajuste está de um ângulo de  $45^\circ$  através da origem para a comparação do modelo exponencial vs. funcional do  $\text{CO}_2$ , em relação ao pasto intensivo (a), pasto intensivo (b) linha da cana-de-açúcar (c) e entrelinha da cana-de-açúcar (d). A linha tracejada vermelha é a linha de concordância e representa a concordância perfeita entre o modelo linear e o modelo funcional (declive = 1 e intercepto = 0). A linha preta representa a regressão linear entre os modelos. Fonte: o autor..... 80
- Apêndice A5 - Análise do fator de correlação de tendência, uma medida de quão longe uma linha de melhor ajuste está de um ângulo de  $45^\circ$  através da origem para a comparação do modelo

linear vs. funcional do N<sub>2</sub>O, em relação ao pasto intensivo (a), pasto intensivo (b) linha da cana-de-açúcar (c) e entrelinha da cana-de-açúcar (d). A linha tracejada vermelha é a linha de concordância e representa a concordância perfeita entre o modelo linear e o modelo funcional (declive = 1 e intercepto = 0). A linha preta representa a regressão linear entre os modelos.

Fonte: o autor..... 81

Apêndice A6 - Análise do fator de correlação de tendência, uma medida de quão longe uma linha de melhor ajuste está de um ângulo de 45° através da origem para a comparação do modelo exponencial vs. funcional do N<sub>2</sub>O, em relação ao pasto intensivo (a), pasto intensivo (b) linha da cana-de-açúcar (c) e entrelinha da cana-de-açúcar (d). A linha tracejada vermelha é a linha de concordância e representa a concordância perfeita entre o modelo linear e o modelo funcional (declive = 1 e intercepto = 0). A linha preta representa a regressão linear entre os modelos.

Fonte: o autor..... 82

Apêndice B - Parâmetros e distribuições consideradas no *RandomizedSearchCV*. ..... 83

Apêndice C - Valor dos intervalos de confiança (95%) a partir da distribuição t de Student em relação ao ponto de tempo que estabiliza o fluxo do N<sub>2</sub>O para os usos da terra. Valor de *p* significativo ( $p < 0.05$ ). ..... 84

Apêndice D - Acurácia obtida na validação cruzada para as cinco rodadas durante o treinamento. .... 85

# LISTA DE SIGLAS E ABREVIATURAS

APTA	Agência Paulista de Tecnologia dos Agronegócios
C	Carbono
C.b	Fator de correlação de tendência
CCC / $\rho_c$	Coefficiente de Correlação de Concordância
CH <sub>4</sub>	Metano
CO <sub>2</sub>	Dióxido de Carbono
EC	Entrelinha da cana-de-açúcar
FDA	Análise de Dados Funcionais
FPCA	Análise de Componentes Principais Funcional
MFPCA	Análise de Componentes Principais Funcional Multivariado
GEEs	Gases de Efeito Estufa
GCV	Validação Cruzada Generalizada
PAG	Potencial de Aquecimento Global
IPCC	Painel Intergovernamental de Mudanças Climáticas
LC	Linha da cana-de-açúcar
$\lambda$	Parâmetro de Suavização
$\rho$	Coefficiente de Correlação de Pearson
N <sub>2</sub>	Diazoto / Dinitrogênio
N <sub>2</sub> O	Óxido Nitroso
NO	Monóxido Azoto / Óxido Nítrico
NO <sub>2</sub> <sup>-</sup>	Nitrito
NO <sub>3</sub> <sup>-</sup>	Nitrato
NO <sub>4</sub> <sup>+</sup>	Amônio
<i>PENNSE</i> <sub><math>\lambda</math></sub>	Medida de Penalidade de Rugosidade
PE	Pastagem extensiva
PI	Pastagem intensiva
R <sup>2</sup>	Coefficiente de Determinação

# SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	<b>12</b>
1.1	Introdução.....	12
1.2	Objetivos.....	14
1.2.1	Geral .....	14
1.2.2	Específicos .....	14
1.3	Estrutura da tese .....	14
<b>2</b>	<b>REVISÃO DA LITERATURA</b>	<b>15</b>
2.1	Usos do solo: pastagens intensificadas e cana-de-açúcar .....	15
2.2	Principais Gases de Efeito Estufa (GEEs).....	16
2.2.1	Dióxido de Carbono (CO <sub>2</sub> ).....	17
2.2.2	Metano (CH <sub>4</sub> ) .....	17
2.2.3	Óxido Nitroso (N <sub>2</sub> O) .....	18
2.3	Modelo Exponencial para mensuração dos fluxos dos GEEs .....	19
2.4	Coefficiente de Correlação de Concordância (CCC).....	19
2.5	Análise de Dados Funcionais ( <i>Functional Data Analysis</i> , FDA).....	20
2.6	Suavização das Observações .....	21
2.6.1	Suavização <i>B-splines</i> .....	21
2.6.2	Medida de rugosidade da suavização .....	23
2.6.3	Validação Cruzada Generalizada (GCV) .....	23
2.7	Derivadas das Funções .....	24
2.8	Análise de Componentes Principais Funcional (FPCA) para os casos univariado e multivariado.....	24
2.9	Análise de agrupamento: <i>Cluster</i> hierárquico .....	27
<b>3</b>	<b>METODOLOGIA</b>	<b>30</b>
3.1	Descrição da área de estudo.....	30
3.2	Desenho experimental .....	30
3.3	Manejo do solo .....	31
3.4	Amostragem para análises das emissões de CO <sub>2</sub> , CH <sub>4</sub> e N <sub>2</sub> O.....	34
3.5	Procedimentos de análises .....	35
<b>4</b>	<b>RESULTADOS E DISCUSSÕES</b>	<b>39</b>

4.1	Comparação entre os modelos funcional, linear e exponencial sobre os diferentes usos do solo .....	39
4.2	Representação dos dados como funcional e sua derivação .....	44
4.3	Emissões das médias diárias do CH <sub>4</sub> , CO <sub>2</sub> e N <sub>2</sub> O .....	48
4.4	Análise de componentes principais funcional (FPCA) para os casos univariado e multivariado.....	51
4.5	Análise de agrupamento .....	55
4.6	Análise de classificação do uso dos solos .....	57
<b>5</b>	<b>CONCLUSÕES</b>	<b>64</b>
	<b>REFERÊNCIAS BIBLIOGRÁFICAS</b>	<b>66</b>
	<b>APÊNDICE A</b>	<b>77</b>
	<b>APÊNDICE B</b>	<b>83</b>
	<b>APÊNDICE C</b>	<b>85</b>
	<b>APÊNDICE D</b>	<b>85</b>
	<b>ANEXO A</b>	<b>86</b>

# 1 INTRODUÇÃO

## 1.1 Introdução

A emissão de gases de efeito estufa (GEEs) de atividades agrícolas é de grande preocupação mundialmente, porque diferentes práticas de manejo do uso do solo têm diversos impactos na composição da comunidade microbiana, influenciando a atuação do solo como fonte ou sumidouro dos principais GEEs tais como: o dióxido de carbono ( $\text{CO}_2$ ), metano ( $\text{CH}_4$ ) e óxido nitroso ( $\text{N}_2\text{O}$ ) (Bandyopadhyay; Lal, 2014; Begun *et al.*, 2019; Dowhower *et al.*, 2020). Obter um conhecimento sobre as emissões dos GEEs através dos sistemas dos usos da terra é de fundamental importância para o planejamento de estratégias eficazes de adaptação e mitigação em cenários futuros dos sistemas globais climatológicos (Meena *et al.*, 2018).

Uma ferramenta que tem sido amplamente aplicada para medir emissões de gases de efeito estufa são as câmaras estáticas fechadas (Clough *et al.*, 2020). Elas consistem em restringir um volume de ar dentro de um recipiente sem fundo com tampa fechada colocado na superfície do solo para monitorar as mudanças na concentração do gás ao longo do tempo de amostragem. Essas câmaras podem ser aplicadas em ambientes naturais ou agrícolas, sendo ideal para o estudo de pequenas parcelas experimentais, locais diversos com variados terrenos ou em situações em que múltiplos locais distintos devem ser estudados sem investimentos significativos em infraestrutura (Collier *et al.*, 2014; Galdino; Signor; de Moraes, 2023).

As câmaras têm um alto grau de adaptabilidade e sensibilidade, e são eficazes para medir simultaneamente os fluxos de  $\text{CH}_4$ ,  $\text{CO}_2$  e  $\text{N}_2\text{O}$ , além de seu baixo custo financeiro (Silva *et al.*, 2015). As concentrações são medidas entre 3 e 4 pontos de tempo e o fluxo desses gases é mensurado baseando-se na variação temporal da concentração do gás no interior das câmaras (Keller; Kaplan; Wofsy, 1986; Lambert; Fréchette, 2005; Denmead, 2008).

Dos métodos mais utilizados nessas mensurações, destaca-se o modelo linear. Esse modelo calcula o fluxo a partir da inclinação  $\left(\frac{dc(t)}{dt}\right)$  dos dados de concentração em relação aos dados no tempo, ou seja, o cálculo representa a inclinação da reta de regressão linear (Parkin; Venterea, 2010; Christiansen *et al.*, 2011). Esse modelo correlaciona as concentrações medidas na câmara estática em relação ao tempo decorrido desde o começo das medidas. Entretanto, ao aplicar um modelo linear nos dados de concentração pode haver um erro sistemático por meio

da não linearidade, além disso uma série de fatores podem influenciar nas mudanças das concentrações dos gases de um modo sistemático, levando ao desenvolvimento de mudança curvilínea na concentração (Pirk *et al.*, 2016).

Para minimizar esse efeito, vários autores têm proposto alternativas de mensurar o fluxo, por exemplo, o modelo de Hutchinson-Mosier (HM), baseado no modelo exponencial, o estimador de fluxo difusivo não estacionário (NDFE), o método de HM modificado, o método Quad, entre outros (Christiansen *et al.*, 2011; Parkin; Venterea; Hargreaves, 2012). Porém, modelos não lineares podem não levar em consideração a variação aleatória nos dados de concentração (Pedersen; Petersen; Schelde, 2010). Nesse sentido, necessita-se de um método que leve em consideração tanto a variação contínua quanto a aleatoriedade dos dados de concentração dos gases.

Nessa perspectiva, uma família estatística que expressa um grande potencial para estudar as variações de uma determinada variável em relação ao contínuo, seja tempo, espaço ou frequência, é a análise de dados funcionais (*Functional data analysis*, FDA). Ela é uma metodologia bem aplicada em modelagem de dados, principalmente aos que são construídos a partir de equações diferenciais. Apesar da FDA ser relativamente nova, a mesma traz diversos fundamentos estatísticos (testes de hipóteses, métodos de exploração aos dados, entre outros), que podem ser usados diretamente nas observações sem perder sua funcionalidade.

As principais características da FDA são: i) representar os dados através de funções a fim de simplificar as manipulações e evidenciar as principais características que esses dados apresentam, permitindo a análise de padrões e variações nessas representações; ii) trabalhar com observações que não precisam ter espaçamentos iguais; iii) lidar com conjunto de dados tanto lineares como não lineares; iv) utilizar derivativos e outros valores funcionais para as análises (Ramsay; Silverman, 2005).

Sua aplicabilidade está em diversas áreas de estudo como, por exemplo: economia (Shang; Kearney, 2022; Sum *et al.*, 2022), educação (Fortuna; Maturo, 2019), medicina (Ramsay; Bock; Gasser, 1995; Ramsay, 2000; Coffey *et al.*, 2011; Ngo *et al.*, 2015; Barua *et al.*, 2018; Happ; Greven, 2018), meio ambiente (Gao; Niemeir, 2008; Sum; Genton, 2012; Di Salvo; Ruggieri; Plaia, 2015; Schmutz *et al.*, 2020), oceanografia (Pauthenet *et al.*, 2017; Pauthenet *et al.*, 2018), previsão demográfica (Hyndman; Ullah, 2007; Shang *et al.*, 2016), entre outros. Todavia, ainda não existem estudos que visem a aplicação da FDA em emissões dos principais gases do efeito estufa. Neste contexto, tem-se como objetivo representar e analisar o comportamento das emissões do dióxido de carbono (CO<sub>2</sub>), metano (CH<sub>4</sub>) e óxido nitroso

(N<sub>2</sub>O) em quatro tipos de uso do solo (pasto extensivo, pasto intensivo, linha e entrelinha da cana-de-açúcar) utilizando a FDA sobre as perspectivas univariada e multivariada.

## **1.2 Objetivos**

### **1.2.1 Geral**

Representar e analisar o comportamento das emissões do dióxido de carbono (CO<sub>2</sub>), metano (CH<sub>4</sub>) e óxido nitroso (N<sub>2</sub>O) em diferentes usos do solo utilizando a FDA sobre as perspectivas univariada e multivariada.

### **1.2.2 Específicos**

- Representar as concentrações dos gases de maneira contínua no tempo, utilizando métodos de suavização existentes na FDA condizentes com o comportamento das observações;
- Avaliar as taxas de mudança em relação ao tempo das curvas funcionais das concentrações através do estudo das derivadas funcionais para extrair informações intrínsecas das funções;
- Utilizar métodos estatísticos funcionais univariados e multivariados com intuito de analisar os comportamentos dos gases em relação aos usos da terra.

## **1.3 Estrutura da tese**

A discussão deste trabalho está estruturada em cinco capítulos contando com esta introdução. No capítulo (2), é feita uma revisão de literatura onde são discutidos os processos de formação dos principais gases de efeito estufa e como são mensurados seus fluxos através dos modelos tradicionais. Além disso, apresentam-se os conceitos básicos da FDA e as metodologias estatísticas funcionais utilizadas neste trabalho. Em relação ao capítulo (3), são delimitadas a área de estudo, experimento e a metodologia proposta. Para o capítulo (4), são apresentados os resultados obtidos e as discussões. Por fim, tem-se o capítulo (5) onde são apresentadas as conclusões, exibindo as principais contribuições obtidas no decorrer do estudo.

## 2 REVISÃO DA LITERATURA

### 2.1 Usos do solo: pastagens intensificadas e cana-de-açúcar

Áreas de pastagens de gado e o cultivo de cana-de-açúcar são reconhecidas como duas das principais bases econômicas do Brasil proveniente da agropecuária, em que a mesma expandiu sua área ocupada em 50% no país nos últimos 38 anos (Freitas, 2023; MapBiomias, 2023). Áreas de cultivos agrícolas (lavouras de grãos e cana-de-açúcar) triplicaram durante esse período, destacando-se também o aumento nas áreas de desmatamento para pastagens (32%) (MapBiomias, 2023).

O Brasil é o maior produtor mundial de etanol e açúcar, produzidos a partir da cana-de-açúcar (Voora *et al.*, 2023), com safra de 2023/2024 de 667.602,1 mil toneladas de cana, 46.880,7 mil toneladas de açúcar e 27,98 bilhões de litros de etanol, em uma área de 8.352,1 mil hectares (Conab, 2023). A região Sudeste do Brasil lidera o setor canavieiro, com 64,19% da produção nacional para a safra 2023/2024, seguida pelas regiões Centro-Oeste (21,2%), Nordeste (8,8%), Sul (5,21%) e Norte (0,6%). Notavelmente, o estado de São Paulo é o maior produtor nacional de cana-de-açúcar (51,75% do total), seguido por Minas Gerais (11,83%) e Goiás (11,24%) (Conab, 2023). Além disso, São Paulo possui o segundo maior rebanho bovino comercial do mundo, com, aproximadamente, 232 milhões de cabeças, sendo o maior exportador de carne bovina (Zia *et al.*, 2019). Sua produção é baseada em grande parte na pastagem, que são pastagens cultivadas com gramíneas exóticas tropicais do gênero *Brachiaria* spp. (Gracindo *et al.*, 2014; Zia *et al.*, 2019).

A transformação de pastagens extensas e degradadas para a produção da cana-de-açúcar é o cenário central de mudança do uso da terra, empregado para impulsionar a ampliação da produção da cana-de-açúcar no Brasil (Adami *et al.*, 2012; Canisares *et al.*, 2019). Se por um lado, a utilização destas áreas para a expansão agrícola reduziria a desflorestação e a competição pelo uso da terra para a produção de alimentos versus biocombustíveis, além de oferecer muitas vantagens em relação à dívida de carbono associada às mudanças diretas e indiretas no uso da terra (Alkimim; Clarke, 2018), por outro lado, a mecanização intensiva utilizada nos canaviais, incluindo o preparo do solo por aração e gradagem e o tráfego de máquinas pesadas durante a colheita mecanizada, degrada a estrutura do solo, afetando seus múltiplos processos e funções (Cherubin *et al.*, 2016; Rabot *et al.*, 2018; Canisares *et al.*, 2019).

Os processos de mudança da terra, tanto a intensificação de pastagens como a conversão pastagem-cana-de-açúcar, necessitam procedimentos de manejo que envolvem a aplicação de fertilizantes e pesticidas, o que pode resultar em danos nos ecossistemas, como a degradação do solo, perda de biodiversidade, lixiviação de agroquímicos aos ecossistemas aquáticos e aumento das emissões dos GEEs (Silva *et al.* 2007; Girotto *et al.*, 2023; Freitas, 2023).

Observando a situação atual e futura de intensificação de pastagens e ampliação das plantações de cana-de-açúcar no Brasil (Oliveira *et al.*, 2019), é fundamental aprimorar a compreensão atual a respeito das consequências ambientais desses dois sistemas de mudança de uso do solo que acontecem simultaneamente, com o intuito de melhor prever resultados e reduzir possíveis impactos negativos (Freitas *et al.*, 2023), bem como orientar as políticas de expansão da produção de cana e as decisões de uso da terra em direção a uma maior sustentabilidade (Cherubin *et al.*, 2021; Garofalo *et al.*, 2022).

## 2.2 Principais Gases de Efeito Estufa (GEEs)

As emissões de gases mais prevalentes na atmosfera são na forma de dióxido de carbono (CO<sub>2</sub>), metano (CH<sub>4</sub>) e óxido nitroso (N<sub>2</sub>O) (Walling; Vaneckhaute, 2020; Canadell *et al.*, 2021). O CO<sub>2</sub> é o principal contribuinte antropogênico do aquecimento global. O CH<sub>4</sub> é o segundo gás de efeito estufa mais abundante na atmosfera. E o N<sub>2</sub>O é um gás de longa vida na atmosfera (tempo de residência de cerca de 120 anos) e desempenha um papel importante na destruição da camada de ozônio. Entre os principais gases que contribuem para o aumento do efeito estufa o CO<sub>2</sub> é gás mais abundante na atmosfera. Dessa forma, o potencial de aquecimento global (PAG) de cada GEE é calculado em função desse gás (Vasconcelos *et al.*, 2019; Xing *et al.*, 2011).

O PAG é um método acordado internacionalmente e publicado pelo Painel Intergovernamental sobre Alterações Climáticas (IPCC) para converter emissões de gases de efeito estufa em emissões equivalentes de CO<sub>2</sub> (IPCC, 2014). O PAG é definido como o forçamento radiativo cumulativo entre o presente e algum horizonte de tempo causado por uma unidade de massa de gás emitido atualmente expressado com relação a um gás de referência, tal como o CO<sub>2</sub> (Dalcin; Mangini; De Godoi, 2021; Wang *et al.*, 2021). Os valores padrão de PAG por unidade de massa de CH<sub>4</sub> e N<sub>2</sub>O medidos ao longo de um período de 100 anos são, aproximadamente, de 28 e 273 vezes maiores que as emissões de CO<sub>2</sub>, respectivamente

(Kamran *et al.*, 2023). Desta forma, esses gases são considerados os principais gases de efeito estufa que resulta em mudanças climáticas (Anderson; Hawkins; Jones, 2016).

### **2.2.1 Dióxido de Carbono (CO<sub>2</sub>)**

O dióxido de carbono (CO<sub>2</sub>) é gerado de forma natural, produzido por meio da decomposição de material orgânico, respiração ou liberação no oceano, e por ações humanas, através das queimas de combustíveis fósseis (carvão, petróleo e gás), biomassa e as mudanças na cobertura e no uso da terra (Kram *et al.*, 2000; Leung *et al.*, 2014; Huntingford *et al.*, 2017; Bhatia *et al.*, 2019). Sobre as emissões do CO<sub>2</sub> proveniente do solo, elas são controladas por vários fatores, como temperatura do solo e do ar, conteúdo de água do solo, preparo do solo, vegetação e disponibilidade de substrato do solo. Além disso, as propriedades físicas, químicas e biológicas do solo também contribuem na alterações dessas emissões (Yerli; Sahin; Oztas, 2022). Walsh *et al.* (2017) relata que a parcela de CO<sub>2</sub> gerado por atividades humanas é pequena quando comparadas com as fontes naturais, contudo interfere no equilíbrio natural, porque a natureza consegue retirar aproximadamente a mesma quantidade de CO<sub>2</sub> que é produzida por fontes naturais.

Entender os controles sobre as emissões de CO<sub>2</sub> do solo é fundamental, pois alterações nas suas emissões podem modificar a taxa de sequestro de carbono no solo (Gong *et al.*, 2021). Visto que um terço das terras cultiváveis do planeta é usada para a agricultura, pesquisas realçam a necessidade de mensurar os fluxos de CO<sub>2</sub> no solo oriundos do uso de terras agrícolas (Zomer *et al.*, 2017; Beegum *et al.*, 2023). Desse modo, a estimacão e caracterização das emissões do gás carbono consegue ajudar a aferir com exatidão o efeito das práticas de manejo da terra, no aquecimento global, na ciclagem do carbono e propiciar a previsão dos padrões de armazenamento de carbono orgânico no solo (Gong *et al.*, 2021).

### **2.2.2 Metano (CH<sub>4</sub>)**

As emissões do metano (CH<sub>4</sub>) do solo são conduzidas pelo equilíbrio líquido de dois processos microbianos neutralizantes, a produção (metanogênese) e a oxidação (metanotrofia), determinando assim se os solos são fontes ou sumidouros desse potente GEE. A produção de

metano por micro-organismos é responsável por mais de dois terços das emissões globais desse gás (Chen *et al.*, 2013; Wang *et al.*, 2023).

Os micróbios produtores de CH<sub>4</sub> (i.e. metanogênicos) são as metanogênicas acetoclásticas e hidrogenotróficas, considerando-se como as principais fontes de CH<sub>4</sub> nos solos (Conrad, 2009; Bridgham *et al.*, 2013; Lyu *et al.*, 2018; Täumer *et al.*, 2022). Em relação a oxidação, as metanotróficas aeróbicas são consideradas as principais oxidantes em solos úmidos, além disso, as metanotróficas fornecem o único sumidouro biológico conhecido para o CH<sub>4</sub> atmosférico (Ciais *et al.*, 2013). No entanto, não é totalmente compreendido quais microrganismos oxidam CH<sub>4</sub> em concentrações atmosféricas nos solos (Täumer *et al.*, 2022).

A abundância relativa de metanogênese e metanotróficos variam ao longo do perfil do solo e ao longo das estações, levando à flutuação das emissões de CH<sub>4</sub> em relação a esses dois perfis (Wang *et al.*, 2023). Também, suas emissões alteram conforme o conjunto de técnicas agrícolas empregados e com os sistemas de manejo, que podem originar a diminuição na capacidade de oxidação do CH<sub>4</sub> pelo solo, devido à mudança de uso da terra e devido a perturbação nos organismos metanotróficos (Smith *et al.*, 2003; Mosier *et al.*, 2004; Bento, 2020).

### 2.2.3 Óxido Nitroso (N<sub>2</sub>O)

A agricultura é responsável por 60-80% das emissões antropogênicas globais de óxido nitroso (N<sub>2</sub>O) (Coskun *et al.*, 2017) e está a crescer em uma velocidade mais acelerado do que a projeção do IPCC (Thompson *et al.*, 2019). A previsão de Davidson e Kanter (2014) é que essas emissões de N<sub>2</sub>O das terras agrícolas dupliquem até 2050. Em relação aos setores de cana-de-açúcar, a fertilização com N é a principal fonte de emissões de N<sub>2</sub>O, podendo ser ainda mais intensificada pela aplicação de vinhaça (resíduos líquidos da produção de etanol), assim como pelo depósito de palha na camada superficial da terra (Pitombo *et al.*, 2016; Borges *et al.*, 2019).

A geração de N<sub>2</sub>O no solo ocorre por meio de processos biogênicos, tais como nitrificação e desnitrificação. A nitrificação é a oxidação microbiana de amônio (NH<sub>4</sub><sup>+</sup>) a nitrato (NO<sub>3</sub><sup>-</sup>), com N<sub>2</sub>O emitido como subproduto. A desnitrificação do nitrificante é a diminuição do nitrito (NO<sub>2</sub><sup>-</sup>) a monóxido de azoto (NO), em seguida a N<sub>2</sub>O e, por fim, a dinitrogênio (N<sub>2</sub>). A desnitrificação é um processo de duas etapas em que NO<sub>3</sub><sup>-</sup> é convertido em N<sub>2</sub>O e depois em

$N_2$  inerte sob condições anaeróbicas (Butterbach-Bahl *et al.*, 2013; Fowler *et al.*, 2015; Wang *et al.*, 2021).

Múltiplos fatores, tanto ambientais quanto de manejo do solo, são capazes de regular as emissões de  $N_2O$ , afetando direta ou indiretamente a disponibilidade de substrato e o estado de aeração do solo. Os ambientais são: população microbiana, relação carbono (C) para nitrogênio (N), concentrações de  $NH_4^+$  e  $NO_3^-$ , pH, umidade, temperatura, textura e salinidade do solo. Os de manejo estão relacionados a aplicação de fertilizante, sistema de lavoura, resíduos de colheita e irrigação (Bouwman; Boumans; Batjes, 2002a; Wang *et al.*, 2021).

## 2.3 Modelo Exponencial para mensuração dos fluxos dos GEEs

Hutchinson e Mosier (1981) apresentaram um procedimento para estimar o fluxo de gás traço solo-atmosfera com câmaras estáticas, quando a taxa de variação da concentração dos gases no espaço livre da câmara não for constante. Essa metodologia é baseada no método exponencial e é conhecido como o método Hutchinson e Mosier (HM) (Parkin; Venterea, 2010). Pedersen, Pedersen e Schelde (2010) fizeram uma parametrização da equação do método HM, com o intuito de mensurar qualquer fluxo não linear, definida como:

$$C_t = \varphi + f_0 \frac{\exp(-kt)}{-kh}, \quad (2.1)$$

em que,  $C_t$  é a concentração de gás no tempo,  $\varphi$  é a concentração constante da fonte em uma certa profundidade abaixo da câmara,  $f_0$  é o fluxo inicial no tempo zero quando a câmara é fechada, portanto, o fluxo sem efeito de câmara,  $k$  (kappa) é o parâmetro de forma não linear que deve ser maior que zero, estimado pelo método dos mínimos quadrados ordinários para  $-\infty < f_0 < +\infty$ ,  $t$  é o tempo (após o fechamento da câmara) e  $h$  é a altura da câmara (Jiang *et al.*, 2010).

## 2.4 Coeficiente de Correlação de Concordância (CCC)

Um meio de tratar o problema de concordância em relação as medições contínuas é desenvolver um índice de resumo graduado que pode variar de  $-1$  a  $1$ , similar a um coeficiente de correlação de Pearson ( $\rho$ ). Com essa perspectiva, Lin (1989) propôs o coeficiente de

correlação de concordância (CCC/  $\rho_c$ ) para escalas contínuas. O CCC analisa a concordância entre duas variáveis contínuas mensurando a variação da relação linear entre os pares observados que estão na reta de 45° passando pela origem (Vallejos *et al.*, 2020). Assim, seja  $X$  e  $Y$  duas variáveis, o CCC é determinado da seguinte forma:

$$\rho_c = 1 - \frac{E[(X-Y)^2]}{E_I[(X-Y)^2]} = \frac{2\sigma_{XY}}{\sigma_X^2 + \sigma_Y^2 + (\mu_X - \mu_Y)^2}, \quad (2.2)$$

em que  $\mu_X = E(X)$ ,  $\mu_Y = E(Y)$ ,  $\sigma_X^2 = Var(X)$ ,  $\sigma_Y^2 = Var(Y)$  e  $\sigma_{XY} = Cov(X, Y)$ , e  $E_I[\ ]$  é a esperança, assumindo a independência entre dois observadores (Lin, 1989; Tsai; Lin, 2023).

O coeficiente de correlação de concordância tem as seguintes características (Lin, 1989; Pandit; Schuller, 2019):

- i.  $-1 \leq -|\rho| \leq \rho_c \leq |\rho| \leq 1$ ;
- ii.  $\rho_c = 0$ , se e só se:  $\rho = 0$ , isto é  $\sigma_{XY} = 0$ ;
- iii.  $\rho_c = \rho$ , se e só se:  $\sigma_X = \sigma_Y$ , e  $\mu_X = \mu_Y$ ;
- iv.  $\rho_c = 1$ , se e só se:  $\rho = 1$ ,  $\sigma_X = \sigma_Y$ , e  $\mu_X = \mu_Y$ , i. e. se e só se:  $x_i = y_i \forall i : i \in [1, N], i \in \mathbb{N}$ ;
- v.  $\rho_c = -1$ , se e só se:  $\rho = -1$ ,  $\sigma_X = \sigma_Y$ , e  $\mu_X = \mu_Y$ , i. e. se e só se:  $x_i = -y_i \forall i : i \in [1, N], i \in \mathbb{N}$ .

Portanto, o CCC de Lin mede tanto a precisão quanto a exatidão. De acordo com Altman (1990), o  $\rho_c$  deveria ser interpretado próximo a outros coeficientes de correlação, como o de Pearson, com  $\rho_c < 0,2$  como ruim e  $\rho_c > 0,8$  como excelente. De outro modo, McBride (2005) foi mais rígido em sua interpretação, indicando valor de  $\rho_c > 0,99$  como quase perfeito,  $\rho_c$  entre 0,95 a 0,99 substancial,  $\rho_c$  entre 0,90 a 0,95 moderado e  $\rho_c < 0,9$  ruim (Akoglu, 2018).

## 2.5 Análise de Dados Funcionais (*Functional Data Analysis, FDA*)

Dados que podem ser representados como uma função suave e contínua, seja em relação ao tempo, espaço, frequência, entre outros, são chamados de dados funcionais. A FDA representa dados cuja  $i$ -ésima observação é uma função real,  $r_i(t), i = 1, 2, \dots, n; t \in T$ , em que  $T$  é um intervalo real definido em um espaço de função integráveis ao quadrado no conjunto compacto  $I$ , com produto interno  $(f, g) = \int_I fg$  e norma  $\|f\|_2 = (\int_I f^2)^{1/2}$  (Ramsay;

Dalzell, 1991). Desta forma, ao invés de pontos no tempo, cada função, produzida a partir de uma série de observações sucessivas, é analisada como uma só observação, possibilitando a retirada de informações de um procedimento temporal como um todo.

## 2.6 Suavização das Observações

A primeira etapa a ser usada no contexto da FDA é a geração da amostra funcional a partir da amostra vetorial, ou seja, a construção de curvas de melhor ajuste com os pontos correspondentes aos valores discretos das medições experimentais (Lasheras *et al.*, 2018; Curceac *et al.*, 2019). Para preparar os dados observados como funções contínuas e suaves,  $r(\cdot)$ , em relação ao tempo ( $t$ ), precisa-se utilizar técnicas de interpolação ou suavização, descrito como:

$$r_i(t) = \sum_{k=1}^K c_{ik}(t) \phi_k(t), \quad (2.3)$$

em que  $c_{ik}(t)$  são números reais chamados de coeficientes de expansão e os  $\phi_k(t)$ ,  $i = 1, 2, \dots, K$  são o conjunto de funções bases combinados linearmente (Ramsay; Silverman, 1997).

Através da suavização é possível avaliar a função em qualquer ponto de tempo, bem como, analisar as  $m$ -ésimas derivadas ( $D^m r(t)$ ) existentes. A escolha do método de suavização vai depender do subjacente comportamento dos dados em análise (Ramsay; Silverman, 2005). Segundo Ullah e Finch (2013), a suavização *B-spline* é mais utilizada por causa da sua simplicidade e flexibilidade para lidar com um vasto conjunto de situações de modelagem semi-paramétrica e não paramétrica. Ela é utilizada quando as observações têm um comportamento não periódico. Os alisadores de Fourier são tradicionalmente usados quando os dados são cíclicos ou periódicos. As bases *Wavelets* são escolhidos para representar dados exibindo descontinuidades e/ou mudanças rápidas no comportamento (Ramsay; Hooker; Graves, 2009; Ullah; Finch, 2013). A suavização de Kernel vizinho mais próximo (kNN) é utilizada para a FDA não paramétrica. O kNN inclui um parâmetro de suavização localmente adaptável permitindo o controle da heterogeneidade local nos dados (Kara *et al.*, 2017; Aneiros *et al.*, 2019).

### 2.6.1 Suavização *B-splines*

Os polinômios são bastantes usados em estudos para aproximar funções, devido a sua formação simples e obtenção de boas propriedades de aproximação. Schumaker (1981) propôs a definição para as funções *Splines*. Seja um intervalo  $T = [a, b]$ , em que a função está definida. Divide-se esse intervalo em subintervalos menores da forma de  $[x_0, x_1], [x_1, x_2], \dots, [x_k, x_{k+1}]$  e então um polinômio ( $p_i$ ), com grau reduzido em relação a função que será suavizada, é utilizado para aproximação em cada subintervalo. Esse método gera uma função de aproximação polinomial por partes  $s(\cdot)$ , i.e.,  $s(x) = p_i(x)$  em  $[x_i, x_{i+1}]$ ,  $i = 1, \dots, k$ . Os pontos  $x_0, x_1, \dots, x_k, x_{k+1}$  são conhecidos como nós, os quais  $x_0$  e  $x_{k+1}$  são os nós externos e  $x_1, \dots, x_k$  os nós internos.

Como as funções *splines* têm suas propriedades definidas pelos polinômios, então suas somas, subtrações e multiplicações geram outra função *spline*, assim qualquer combinação linear dessas funções bases forma uma nova função *spline*. Desta forma, uma extensão dos *splines*, conhecido como *B-splines*, também foi proposto por Schumaker (1981), e compõem uma base de espaços *splines*.

*B-splines* são polinômios por partes construídos dividindo o intervalo  $T = [a, b]$  em  $k + 1$  subintervalos menores  $[x_0, x_1], \dots, [x_k, x_{k+1}]$  com  $k + 2$  nós. Cada intervalo de ordem  $m$  dos *B-splines* são não nulos, então a quantidade total final de nós para construção da função é  $k + 2m$ , por isso necessita-se acrescentar alguns nós ( $\tau$ ), com valores arbitrários, no começo e no final da sequência, de maneira que  $\tau_1 \leq \tau_2 \leq \dots \leq \tau_{m-1} \leq x_0$  e  $x_{k+1} \leq \tau_{m+k+2} \leq \dots \leq \tau_{k+2m}$  (Schumaker, 1981). Por consequência, Boor (2001) desenvolveu um algoritmo que possui uma relação de recorrência para calcular um *B-spline* de ordem  $m$ , a partir de um *B-spline* de ordem  $m - 1$  (i.e utilizando uma ordem menor). Sendo assim, uma sequência crescente com o  $i$ -ésimo *B-splines* de ordem  $m$  e nós  $\tau_i = \{\tau_i\}$  é dado por:

$$B_{i,m}(t) = \frac{t-\tau_i}{\tau_{i+m-1}-\tau_i} B_{i,m-1}(t) + \frac{\tau_{i+m}-t}{\tau_{i+m}-\tau_{i+1}} B_{i+1,m-1}(t), \quad (2.4)$$

em que:

$$B_{i,m}(t) = \begin{cases} 1 & \text{se } \tau_i \leq t \leq \tau_{i+m}, \\ 0, & \text{caso contrário.} \end{cases} \quad (2.5)$$

Portanto, considerando um *B-spline* de ordem  $m$  com  $k$  nós interiores, é possível escrever a função  $s(t)$  como:

$$s(t) = \sum_{i=1}^{m+k} c_i B_{i,m}(t), \quad (2.6)$$

em que  $c_i$  são os coeficientes de expansão e os  $B_{i,m}(t)$  refere-se à função base na ordem  $m$  com  $k$  nós em relação ao tempo. Assim, a estimativa suave de  $s(t)$  é calculada por:

$$\hat{f}(t) = \sum_{i=1}^K \hat{c}_i B_i(t), \quad (2.7)$$

correspondente à equação 2.3.

Até certo ponto, a representação básica é um compromisso entre a interpolação e a suavização dos dados originais (Gao; Niemeier, 2008). Geralmente, as medições contêm algum erro ou ruído de medição ( $\varepsilon_{ij}$ ) (Coffey *et al.*, 2011). Então, dado o erro aleatório nos dados brutos, algum grau de suavização é fundamental para minimizá-lo. Uma técnica de filtrar esse erro é utilizar uma penalidade de aspereza ou medida de rugosidade (Ramsay; Hooker; Graves, 2009).

### 2.6.2 Medida de rugosidade da suavização

Estima-se a medida de rugosidade usando critério de soma dos quadrados residuais penalizados que negocia entre a rugosidade da curva e a falta de ajuste dos dados, definido por:

$$PENNSE_{\lambda} = \sum_i (y_i - r(t_i))^2 + \lambda \int (D^2 x)(t)^2 dt, \quad (2.8)$$

em que  $\lambda$  é o parâmetro de suavização que especifica o destaque na curvatura penalizadora do segundo termo em relação à qualidade do ajuste quantificada na soma dos resíduos quadrados no primeiro termo. Portanto, aplica-se a suavização de  $\lambda$  junto com a  $\int (D^2 x)(t)^2 dt$  para reduzir os resíduos. Para estimar o melhor valor do parâmetro de suavização e o tamanho de  $K$ , que é a quantidade de funções bases, utiliza o critério de validação cruzada generalizada (Ramsay; Hooker; Graves, 2009).

### 2.6.3 Validação Cruzada Generalizada (GCV)

A validação cruzada generalizada (GCV) foi desenvolvida como uma alternativa mais simples do procedimento de validação cruzada (CV), sendo considerada mais segura do que a CV, no sentido de possuir uma menor tendência na suavização (Craven; Wahba, 1979; Ramsay; Silverman, 2005). O método é definido como:

$$GCV(\lambda) = \left( \frac{n}{n-df(\lambda)} \right) \left( \frac{SSE}{n-df(\lambda)} \right), \quad (2.9)$$

em que  $n$  é o número de observações,  $df(\lambda) = \text{trace}[H(\lambda)]$  é uma medida equivalente aos graus de liberdade, onde  $H(\lambda)$  é uma matriz de suavização, quadrada, simétrica e de ordem  $n$ .  $SSE = \sum_i^n [y_i - r(t_i)]^2$  é a soma dos erros quadrados que deve ser minimizada.

O objetivo da GCV é otimizar tanto o valor de  $\lambda$ , quanto a quantidade de variáveis selecionadas em relação à previsão das observações, ou seja, a quantidade de funções bases necessária para a suavização (Ramsay; Hooker; Graves, 2009; Jansen, 2015).

## 2.7 Derivadas das Funções

A FDA tem duas aplicações principais na análise de dados: por um lado, fornece resultados tanto flexíveis, como robustos e, por outro lado, permite extrair mais informação dos dados, informações que geralmente não estão disponíveis em aplicações de técnicas estatísticas clássicas (Mas; Pumo, 2009; Lasheras *et al.*, 2018). Uma maneira de extrair informações das funções é através das suas derivadas, pois possibilita analisar a dinâmica de uma função.

Deste modo, através de um modelo funcional, pode-se obter uma ou mais derivadas associadas ao tempo. Usa-se a notação  $D^m r(t)$  para designar a  $m$ -ésima derivada da função  $r(t)$  em relação ao tempo. A maneira mais habitual do modelo funcional é uma equação que uni duas ou mais ordens de derivadas. Dessa forma, com base em  $r(t)$  podemos calcular suas derivadas (Ramsay; Silverman, 2002):

$$D^m r(t) = \sum_{i=1}^{m-2} c_i(t) D^i \phi_i(t). \quad (2.10)$$

Quando se suaviza as observações a partir da técnica *B-splines* de ordem  $m$ , têm-se derivadas contínuas até a ordem  $m - 2$ .

## 2.8 Análise de Componentes Principais Funcional (FPCA) para os casos univariado e multivariado

A ideia por trás da maioria dos métodos de redução de dimensão é transformar o conjunto original de variáveis de forma que apenas algumas das novas variáveis transformadas incorporem a maior parte das informações contidas nas originais. A técnica mais popular é a

análise de componentes principais (PCA). Na PCA, as novas variáveis são combinações lineares não correlacionadas das originais (Berrendero; Justel; Svarc, 2011).

A PCA funcional (FPCA) procede de maneira análoga à PCA convencional. O objetivo é determinar combinações lineares mutuamente ortogonais das variáveis originais que maximizam a variação explicada, mas em vez de usar vetores, a FPCA utiliza funções para conduzir o processo, sendo a mesma uma das técnicas de análise multivariada mais populares para a extração de informações da FDA (Sánchez-Sánchez *et al.*, 2014; Di Salvo; Ruggieri; Plaia, 2015).

Em relação ao contexto funcional univariado, cada componente principal é estabelecido por meio de uma função,  $\xi(t)$ , conhecida como função peso ou como autofunções, delimitada sobre o mesmo intervalo de tempo que foi estabelecido para as observações funcionais. Cada função peso tem a incumbência de determinar o modo mais marcante de variação nas curvas. Os *scores* dos componentes principais dos indivíduos da amostra são os valores  $z_i$  dados por:

$$z_i = \int \xi(t) r_i(t) dt. \quad (2.11)$$

A finalidade da FPCA é achar a função de ponderação  $\xi_1(t)$  que maximize a variação das pontuações do  $z_i$  sujeita à restrição  $\int \xi_1^2 dt = 1$ . A restrição unitária da soma dos quadrados nos pesos é fundamental para que o problema se torne preciso. Sem essa restrição, os valores dos quadrados médios da combinação linear poderiam ficar aleatoriamente grandes. Os próximos componentes principais (segunda, terceira e ordem superior) são resolvidos de maneira similar, porém com restrições adicionais (Ramsay; Silverman, 2002).

A extensão da FPCA univariado para dados funcionais multivariados é de alta relevância prática para revelar a variação conjunta das diferentes variáveis (Happ; Greven, 2018). Pauthenet *et al.* (2017) e Pauthenet *et al.* (2019) descrevem os cálculos da FPCA para o caso multivariado.

Sejam as curvas  $\theta$  e  $S$  das variáveis projetadas numa *B-spline* com  $K$  coeficientes, para cada posição temporal. Tem-se uma matriz de tamanho  $\mathbf{X}$  com  $N$  linhas e  $L = 2 * K$  colunas, formada pelos coeficiente dessa decomposição. Cada linha estabelece um vetor  $\mathbf{v}_n$ , dado por  $\mathbf{v}_n = (v_{n,1}^\theta, \dots, v_{n,K}^\theta; v_{n,1}^S, \dots, v_{n,K}^S)'$ , que unem ambos os coeficientes das curvas. A observação média do vetor pode ser mensurada por:

$$\bar{\mathbf{v}} = (\bar{v}_1^\theta, \dots, \bar{v}_K^\theta; \bar{v}_1^S, \dots, \bar{v}_K^S)', \quad (2.12)$$

em que  $\bar{v}_k^i, i \in \{\theta, S\}, k = 1, \dots, K$  é à média dos  $N$  coeficientes da coluna  $k$  em  $\mathbf{X}$ . O perfil funcional médio pode ser obtido para variáveis como:

$$\bar{r}^i(t) = \sum_{k=1}^K \bar{v}_k^i \phi_k(t), \quad i \in \{\theta, S\}, \quad (2.13)$$

A função do PC consiste em encontrar a decomposição única da matriz  $\mathbf{VWM}$  através da resolução do seguinte problema de valores próprios:

$$\mathbf{VWM}\mathbf{b}_l = h_l \mathbf{b}_l, \quad (2.14)$$

onde  $h_l$  valor próprio associado ao  $l$ -ésimo vector próprio  $\mathbf{b}_l$ . A matriz  $\mathbf{V}$  é a matriz de covariância cruzada de tamanho  $L \times L$  dos coeficientes em  $\mathbf{C}$  calculados com  $\mathbf{V} = \frac{1}{N} \mathbf{C}'\mathbf{C}$ . A organização da matriz de dados  $\mathbf{C}$  produz uma matriz de covariância formada em blocos  $\mathbf{V}$  que também pode ser escrita como  $\mathbf{V} = \begin{pmatrix} \mathbf{V}_{\theta\theta} & \mathbf{V}_{\theta S} \\ \mathbf{V}_{S\theta} & \mathbf{V}_{SS} \end{pmatrix}$ , em que a matriz  $\mathbf{V}_{\theta S}$  é a matriz de covariância de dimensão  $K \times K$  entre os coeficientes da variável  $S$  e da variável  $\theta$ .

Como as *B-splines* não formam uma base ortonormal, a matriz  $\mathbf{W}$  de dimensão  $L \times L$  garante a equidade métrica entre o problema funcional e a sua versão discreta. Sendo construída por blocos da seguinte forma:

$$\mathbf{W} = \begin{pmatrix} \mathbf{W}_\theta & 0 \\ 0 & \mathbf{W}_S \end{pmatrix}, \quad (2.15)$$

em que as matrizes  $\mathbf{W}_i$  são constituídas por produtos escalares das funções de base  $\{\phi_1, \dots, \phi_K\}$ , que são integrais nesse caso. A matriz de ponderação diagonal de bloco  $\mathbf{M}$ , de dimensão  $L \times L$ , utilizada para equilibrar os valores dos coeficientes durante a decomposição da variância dos dados.

Um valor próprio  $h$ , associado com o vector próprio  $\mathbf{b} = (b_1^\theta, \dots, b_K^\theta; b_1^S, \dots, b_1^S)'$ , fornecerá duas funções próprias  $(\xi^\theta, \xi^S)$ , expressas como:

$$\xi^i(z) = \sum_{k=1}^K \beta_k^i \phi_k(z), \quad i \in \{\theta, S\}, \quad (2.16)$$

com coeficientes  $\boldsymbol{\beta} = (\beta_1^\theta, \dots, \beta_K^\theta; \beta_1^S, \dots, \beta_K^S) = \mathbf{W}^{-1/2} \mathbf{M}^{-1/2} \mathbf{b}$  são uma versão normalizada do vector próprio  $\mathbf{b}$  em relação à métrica. Utilizando o fato de  $\mathbf{W}$  ser uma matriz simétrica, positiva e definida, e  $\mathbf{M}$  uma matriz diagonal, as observações podem ser projetadas em um espaço de pequena dimensão quando se calculam os PCs, designados por:

$$\mathbf{y}_l = \mathbf{C}\mathbf{M}^{-1/2}\mathbf{W}^{-1/2}\mathbf{b}_l. \quad (2.17)$$

Para cada componente, uma observação é definida como três curvas de três traçadores diferente, descritos como curvas funcionais contínuas usando uma decomposição sobre os coeficientes do modelo de expansão *B-splines*, para cada um dos gases. Ela transforma as variáveis originais em novas variáveis que são combinações lineares não correlacionadas das originais e concentram a variância do sistema. Essas novas variáveis são componentes principais e representam os modos mais significativos de variação de dados na estrutura dos solos, e permite quantificar a quantidade de variação contida pelos gases em estudo em relação a cada componente (Pauthenet *et al.*, 2017; Pauthenet *et al.*, 2019; Ramsay; Graves; Hooker, 2022).

Um modo de auxiliar na interpretação da FPCA e destacar seus principais benefícios é analisar os componentes principais com perturbações para a curva média (Ormoneit *et al.*, 2005; Dona *et al.*, 2009; Coffey *et al.*, 2011). Para isso, aumenta a deformação das curvas médias adicionando e subtraindo um múltiplo adequado sob a forma de uma constante a função média do componente principal, com objetivo de explorar relações entre diferentes padrões das curvas e outras variáveis de interesse (Warmenhoven *et al.*, 2019a; Warmenhoven *et al.*, 2019b). Para o presente estudo, cada autofunção é multiplicada por um fator de 1,5.

## 2.9 Análise de agrupamento: *Cluster* hierárquico

A análise de *cluster* visa agrupar um conjunto de dados de modo que as observações dentro do grupo sejam mais semelhantes ao que ocorre em grupos com respeito a uma métrica. O agrupamento hierárquico é um método utilizando estratégias aglomerantes ou divisórias. Esse método requer uma abordagem diferenciada entre os conjuntos de observações, para informar quando um agrupamento deve ser combinado ou quando um agrupamento deve ser dividido (Wang; Chiou; Müller, 2016).

Inicialmente, calcula-se a similaridade entre duas observações, representando o grau de correspondência entre elas ao longo de todas as características usadas na análise. Desse modo, uma medida de distância é utilizada, em que quanto menor o valor da distância, maior é a semelhança entre as observações. A medida de distância frequentemente assumida é a Euclidiana, definida por:

$$d_{ij} = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2}, \quad (2.18)$$

em que  $x_{ik}$  é o valor da variável  $k$  referente à observação  $i$  e  $x_{jk}$  representa a variável  $k$  para a observação  $j$  (Fávero *et al.*, 2009; Hair *et al.*, 2009).

Embora a distância Euclidiana seja a mais familiar, outras métricas podem ser mais apropriadas para alguns casos. A citar a distância de Chebychev, essa distância representa a maior diferença ao longo de todas as variáveis de agrupamento, dado por:

$$d_{ij} = \max_k |x_{ik} - x_{jk}|. \quad (2.19)$$

Ela é particularmente suscetível a diferenças em escalas ao longo das variáveis (Hair *et al.*, 2009; Muniswamaiah *et al.*, 2023).

A distância de Manhattan é utilizada para calcular a soma das diferenças absolutas entre as coordenadas de um par de observações, dado por:

$$d_{ij} = \sum_{k=1}^n |x_{ik} - x_{jk}|. \quad (2.20)$$

Não é baseada na distância euclidiana. No lugar disso, ela emprega a soma das diferenças absolutas das variáveis. Este procedimento é o mais simples de calcular, mas pode conduzir a agrupamentos inválidos se as variáveis forem altamente correlacionadas (Hair *et al.*, 2009; Faisal; Zamzami, 2020).

A distância de Camberra é utilizada quando é necessário encontrar a distância entre pares de pontos, em que os dados se encontram em torno da origem no espaço vetorial, definida pela seguinte fórmula (Faisal; Zamzami, 2020):

$$d_{ij} = \sum_{k=1}^n \frac{|x_{ik} - x_{jk}|}{|x_{ik}| + |x_{jk}|}. \quad (2.21)$$

A distância de Minkowski é uma forma generalizada da distância Euclidiana (quando o expoente for igual a dois) e de Manhattan (quando o expoente for igual a um), ela é definida por (Li; Zhang, 2011):

$$d_{ij} = \left( \sum_{k=1}^n |x_{ik} - x_{jk}|^p \right)^{1/p}. \quad (2.22)$$

Existem duas subdivisões possíveis para o agrupamento hierárquico, que são: aglomerativos, em que cada sujeito começa representando seu próprio grupo e, a partir deste momento, novos grupos são realizados por similaridade; divisíveis, iniciam como um único grupo, e por divisão sucessivas vão sendo formados outros grupos. O método Ward é um agrupamento (tipo aglomerativo) que está fundamentado na variância dos grupos e busca aumentar a similaridade associada a cada agrupamento, definido como:

$$ESS_k = \sum_{i=1}^n x_i^2 - \frac{1}{n} (\sum_{i=1}^n x_i)^2, \quad (2.23)$$

em que  $k$  é o agrupamento,  $n$  é o número de elementos do agrupamento  $k$  e  $x_i$  é o  $i$ -ésimo elemento do agrupamento  $k$  (Bussab; Miazaki; Andrade, 1990; Fávero *et al.*, 2009; Santana; Silva, 2019).

Trazendo para o contexto funcional, a análise de agrupamento é possível ser realizada após a redução de dimensão, onde é frequentemente realizada pela projeção de dados funcionais em um conjunto de funções de base de menor dimensão, como os componentes principais, no qual a maioria das características dos dados funcionais são construídas. Desta forma, a análise de agrupamento é então realizada nas primeiras pontuações dos componentes principais visando identificar grupos homogêneos (Shang, 2014; Jacques; Preda, 2014a).

## **3 METODOLOGIA**

### **3.1 Descrição da área de estudo**

A área de estudo está localizada na Unidade de Pesquisa e Desenvolvimento de Brotas, da Agência Paulista de Tecnologia dos Agronegócios (APTA), na região central do Estado de São Paulo (22°17'12''S e 48°07'35''W), a 60 km da cidade de São Carlos. Esta região de pesquisa foi selecionada por causa da sua similaridade com as áreas de cultivo da cana-de-açúcar no Brasil, como as condições do solo, declividade, presença de pasto e possibilidade de uso contínuo durante dois ciclos da cana-de-açúcar. As áreas adjacentes consistem em uma mata nativa (Cerrado) em processo avançado de regeneração, localizada ao lado; e uma lavoura da cana-de-açúcar à montante.

Conforme a classificação climática Köppen-Geiger, o clima da região caracteriza como subtropical úmido (Cwa), apresenta chuva em todos os meses do ano, com maior concentração no verão e redução no inverno. A precipitação média anual na região é de, aproximadamente, 1200 mm, a temperatura média é de 21,8 °C (Trevisan; Moschini; Guerrero, 2017) a altitude média é de 650 m acima do nível do mar, e a declividade média é 4%.

De acordo com a classificação Brasileira de Solos, o solo da área de estudo é do tipo Latossolo Vermelho-Amarelo (Anexo A). Este tipo de solo geralmente é bem drenado, profundo e uniforme em características de cor, textura, estrutura e profundidade (Embrapa, 2018).

### **3.2 Desenho experimental**

O modelo experimental estabelecido para a realização da pesquisa abrangeu três matrizes de uso do solo: pastagem extensiva (PE), pastagem intensiva (PI) e cana-de-açúcar (C). Para cada uso da terra, foram construídas cinco parcelas de 0,25 ha (50 × 50 m), totalizando 15 unidades experimentais, com uma câmara estática inserido em cada unidade para a pastagem extensiva e intensiva e duas câmaras inseridas em cada unidade da cana-de-açúcar, uma referente a linha da cana-de-açúcar (LC) e a outra inserida na entrelinha da cana-de-açúcar (EC). Para a demarcação dos tratamentos e delimitação das parcelas, um mapa base de

declividade do terreno foi confeccionado com o objetivo de que o manejo do solo em cada tratamento exercesse o mínimo de influência um no outro. Dessa forma, os tratamentos foram alocados seguindo a inclinação da área de estudo (4%). A pastagem extensiva foi alocada na parte mais alta da área de estudo, a pastagem intensiva no meio e a cana-de-açúcar na parte inferior.

O delineamento experimental permitiu que os processos de escoamento superficial e lixiviação dos tratamentos com maior aplicação de agroquímicos (C e PI, respectivamente) não afetassem o tratamento PE. Aqui, definimos como pastagens extensivas aquelas em que não há adição de fertilizantes minerais ou orgânicos, manejo do solo e que prevalecem no Brasil (Bento *et al.*, 2021). A alocação das parcelas considerou também a proximidade de pelo menos uma parcela de cada tratamento com a área do Cerrado adjacente. Para garantir o isolamento e a locomoção entre os tratamentos, foi construída uma estrada entre eles. O delineamento da área experimental está representado na Figura 1.

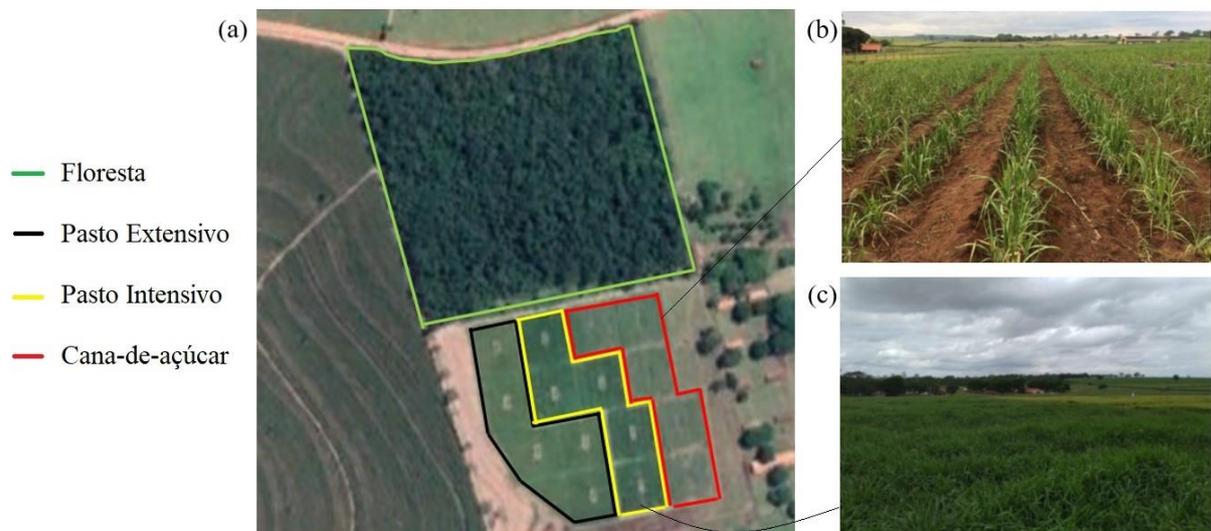


Figura 1 - (a) Delimitação da área experimental: de vegetação florestal (verde); parcelas do pasto extensivo (preto); parcelas do pasto intensificado (amarelo); parcelas da cana-de-açúcar (vermelho). (b) Cana-de-açúcar dois meses após o plantio. (c) Pastagem intensificada (verde escuro) dois meses após o plantio. Fonte: adaptado do segundo relatório científico do projeto: Consequências ambientais da conversão pastagem-cana-de-açúcar e intensificação de pastagens. Instituição Sede do projeto: Centro de Energia Nuclear na Agricultura (CENA).

### 3.3 Manejo do solo

Após serem definidas as unidades experimentais, iniciou-se os procedimentos de manejo do solo nas áreas do PI e C. A pastagem já estabelecida na área experimental (*Brachiaria decumbens*) foi mantida no tratamento PE, uma vez que não foi realizado manejo do solo neste tratamento, considerado como controle. Nos tratamentos PI e C, foram realizadas as etapas convencionais de preparo do solo – gradagem calagem (calcário com 70% de poder relativo neutralizante total - 2 Mg ha<sup>-1</sup>) e aragem (do tipo Aiveca). Após o preparo do solo, plantou-se o pasto (*Brachiaria brizantha* cv *marandu*) no tratamento PI e a cana-de-açúcar (variedade IAC SP 97-4039) no tratamento C. A escolha da espécie *Brachiaria brizantha* cv *marandu*, deu-se por ser bastante utilizada nas pastagens brasileiras e exibir melhor produtividade e qualidade nutricional (Euclides *et al.*, 2014; Cezário *et al.*, 2015).

Posteriormente, foi realizada a aplicação de fertilizante (NPK) nos tratamentos PI e C, seguida da aplicação do inseticida Regent® 800 W G (a.s. fipronil, BASF S.A.) no tratamento C, e do herbicida DMA® 806 BR (a.s. 2,4-D, Dow AgroSciences Industrial Ltda.) nos tratamentos PI e C (Tabela 1).

Tabela 1 - Adubação e variedades utilizados durante a intensificação dos agroecossistemas, com a conversão de pastagem extensiva para pastagem intensiva e canavial, nas áreas experimentais.

<b>Cana-de-açúcar</b>	<b>RB86-7515</b>
N (kg ha <sup>-1</sup> de N) <sup>δ</sup>	60
P (kg ha <sup>-1</sup> de P <sub>2</sub> O <sub>5</sub> ) <sup>‡</sup>	150
K (kg ha <sup>-1</sup> de K <sub>2</sub> O) <sup>γ</sup>	120
Fipronil (g ha <sup>-1</sup> ) <sup>ξ</sup>	500
2,4-D (L ha <sup>-1</sup> ) <sup>‡</sup>	1,5
<b><i>Brachiaria Brizantha</i></b>	<b>cv. Marandú</b>
N (kg ha <sup>-1</sup> de N) <sup>δ</sup>	40
P (kg ha <sup>-1</sup> de P <sub>2</sub> O <sub>5</sub> ) <sup>‡</sup>	60
K (kg ha <sup>-1</sup> de K <sub>2</sub> O) <sup>γ</sup>	30
2,4-D (L ha <sup>-1</sup> ) <sup>‡</sup>	1,5

<sup>δ</sup> na forma de ureia; <sup>‡</sup> na forma de supersimples; <sup>γ</sup> na forma de KCl; <sup>ξ</sup> nome comercial Regent® 800 WG; <sup>‡</sup> nome comercial DMA® 806 BR; <sup>†</sup> nome comercial GAMIT®360 CS; <sup>ρ</sup> nome comercial Boral® 500 SC; <sup>‡</sup> nome comercial Volcane®. Fonte: adaptado do segundo relatório científico do projeto: Consequências ambientais da conversão pastagem-cana-de-açúcar e intensificação de pastagens. Instituição Sede do projeto: Centro de Energia Nuclear na Agricultura (CENA).

Esses agrotóxicos foram selecionados por estarem entre os principais utilizados na cultura da cana-de-açúcar no Brasil (Freitas *et al.*, 2023). Os fertilizantes utilizados foram: nitrogênio (N) na forma de ureia -  $\text{CH}_4\text{N}_2\text{O}$  (45% de N), fosfato (P) na forma de superfosfato simples (18% de  $\text{P}_2\text{O}_5$ ) e potássio (K) na forma de cloreto de potássio KCl (60% de  $\text{K}_2\text{O}$ ) (Tabela 1).

Além disso, a pastagem de gado (raça Nelore) para o consumo da forragem produzida foi simulada nos tratamentos PE e PI. A capacidade de suporte das pastagens foi estimada considerando a unidade animal (UA) de 450 kg e o consumo diário de 4% do peso vivo para não comprometer a produtividade e resiliência da pastagem (Euclides *et al.*, 2014; Dias-Filho, 2014).

Ressalta-se que o manejo do solo conduzido nos tratamentos seguiu as recomendações das normas brasileiras para plantio da cana-de-açúcar. Além disso, os valores de aplicação dos fertilizantes químicos foram calculados a partir de análise físico-química (Anexo A) prévia do solo da área de estudo (Projeto temático FAPESP nº 2015/18790-3) e valores de agrotóxicos foram calculados de acordo com as recomendações dos fabricantes para o plantio da cana-de-açúcar. As etapas de manejo do solo conduzidas durante o período do experimento estão enumeradas na Tabela 2.

Tabela 2 - Etapas de manejo do solo realizadas nos tratamentos pastagem extensiva (PE), pastagem intensiva (PI) e cana-de-açúcar (C) durante o primeiro ciclo da cana-de-açúcar.

Etapa	Tratamento	Período
Marcação inicial da área experimental	Todos	31/08/2017
Delineamento dos tratamentos e das parcelas	Todos	16/04/2018
Gradagem	PI e C	23/10/2018
Calagem (Cálcario 70% PNRT)	PI e C	31/10/2018
Aragem (tipo Aiveca)	PI e C	27/11/2018
Aplicação de NPK	PI e C	29/11/2018
Plantio do pasto e cana	PI e C	04/12/2018
Aplicação do agrotóxico fipronil	C	04/12/2018
Aplicação do agrotóxico 2,4-D	C	26/01/2019
Pastagem do gado	PE e PI	11/04/ a 04/05/ 2019
Aplicação de 2,4-D	PI	01/08/2019

---

Fonte: Adaptada de Freitas (2023).

### 3.4 Amostragem para análises das emissões de CO<sub>2</sub>, CH<sub>4</sub> e N<sub>2</sub>O

As amostras dos gases CH<sub>4</sub>, CO<sub>2</sub> e N<sub>2</sub>O foram coletadas através das câmaras estáticas fixadas na superfície do solo em cada uma das parcelas dos usos do solo: pasto extensivo, pasto intensivo e cana-de-açúcar (linha e entre linha), segundo o método proposto por Davidson *et al.* (2002). As câmaras foram fixadas entre as operações de preparo do solo, após esse período foram fixadas e não foram removidas novamente de forma a minimizar alterações no material orgânico da superfície e assegurar boa vedação. Em seu interior, a aplicação de fertilizantes foi feita de acordo com a área delimitada por ela quando inserida no solo, sendo proporcional a demanda de fertilizante na área total das parcelas de cada tratamento.

Nos dias de coleta, quatro amostras de ar foram coletadas de cada câmara, representando as concentrações em quatro pontos de tempos (0, 12, 24 e 36 minutos). As coletas foram realizadas no período da manhã. De acordo com Garcia-Montiel *et al.* (2001), as câmaras foram vedadas com as tampas de PVC no primeiro tempo de coleta e, prioritariamente, somente durante o período de amostragens. Para o cálculo de volume da câmara, em todos os dias de coleta, após as câmaras serem tampadas, três alturas foram medidas em pontos previamente fixados. Um orifício sobre a tampa da câmara foi confeccionado para manter a pressão interna igual à pressão atmosférica. Juntamente com as coletas das concentrações dos GEEs foram medidos a temperatura do ar e do solo, para cada dia de amostragem.

De acordo com procedimento descrito por Siqueira-Neto *et al.* (2011), foram retiradas amostras de gás através da utilização de seringas plásticas BD (Cremer S.A., Blumenau, Santa Catarina, Brasil) de 60 mL, que foram imediatamente transferidas para frascos de vidro de 30 mL, previamente higienizados, identificados, vedados e evacuados com septos de borracha estilo rolha (Bellco Glass, Vineland, NJ, EUA).

Após o término da coleta na área experimental, os frascos foram levados para o laboratório, onde foram analisados por cromatografia gasosa. As amostras foram analisadas em período máximo de um mês após a data de coleta. Para o cálculo das concentrações dos gases, por comparação das áreas dos picos onde ocorre a integração, foram utilizados padrões previamente preparados (Scott-Marrin – Riverside, CA, USA).

### 3.5 Procedimentos de análises

A ideia inicial do procedimento é verificar se através do modelo proposto é possível representar o fluxo dos gases estudados. Para isso, o resultado gerado pelo modelo funcional será comparado com os resultados do modelo linear e exponencial. Através do modelo linear, o fluxo é calculado baseado na inclinação da reta. O modelo exponencial, é um método não linear, que leva em consideração a curvatura exponencial dos dados (Roland, 2020). E o modelo proposto, que é o funcional. Ele leva em consideração a representação contínua dos dados em relação ao tempo.

A comparação dos fluxos obtidos pelos modelos será realizada por meio do coeficiente de correlação de concordância de Lin (CCC), ele quantifica a concordância entre duas medidas, utilizada para variáveis contínuas. Além do valor do coeficiente de concordância, também é possível gerar outros valores através do método CCC. O fator de correlação de tendência (C.b) é uma medida de quão longe uma linha de melhor ajuste está de um ângulo de 45° através da origem, ou seja, o tanto que a linha de CCC se descola em relação a linha de tendência do modelo (Stevenson *et al.*, 2022). O coeficiente correlação de Pearson ( $\rho$ ), é usado para medir o grau de associação entre os modelos e o coeficiente de determinação ( $R^2$ ), calcula a porcentagem de variação na resposta que é explicada pela correlação entre os modelo. Para essa análise, foram definidos:  $0 < CCC \leq 0,6$ , ruim/fraca;  $0,6 < CCC \leq 8,5$ , moderado;  $8,5 < CCC \leq 0,95$ , substancial;  $CCC > 0,95$ , excelente. E foram definidos:  $0 < \rho \leq 0,6$ , fraca;  $0,6 < \rho \leq 0,8$ , moderado;  $0,8 < \rho \leq 0,9$  forte e  $\rho > 0,9$  muito forte (Steichen; Cox, 2002; Silva, 2022).

Para reconstruir os dados das concentrações discretas dos GEEs como funções contínuas e suaves em relação ao tempo, foi utilizada a suavização B-spline de ordem 4 (grau + 1) (Bande *et al.*, 2020). Para obter uma estimativa de dados funcionais via suavização, usou-se o método de validação cruzada generalizada (GCV) com uma medida de penalidade de rugosidade ( $PENNSE_\lambda$ ). Esse método fornece o menor valor de GCV para dados funcionais baseado em uma lista dos números de funções bases ( $K$ ) e o valor de lambda ( $\lambda$ ), necessário para minimizar o erro aleatório das funções. Após a reestruturação dos dados de concentração dos gases, será realizado um *boxplot* funcional, como uma ferramenta gráfica exploratória informativa para visualização de dados funcionais e detecção de possíveis valores extremos (Dai *et al.*, 2020). O *boxplot* funcional produz gráficos de caixa funcionais que permite ordenar uma amostra de curvas do centro para fora e, assim, introduz uma medida para definir quantis funcionais,

mediana funcional, a centralidade ou periferia de uma observação e os outliers (Sun; Genton; Nychka, 2012; Ramsay; Graves; Hooker, 2022). Diferentemente do *boxplot* tradicional, as informações relativas a mediana e os quartis são representados por curvas ao invés de informações numéricas pontuais.

Em seguida a suavização, será calculado os fluxos dos GEEs por meio das derivadas funcionais. Por definição, o fluxo é a taxa de variação das concentrações em relação ao tempo. Desta forma, a função suavizada  $r_i(t)$ , descreve a posição da concentração em movimento no instante  $t$ . Isso quer dizer que a derivada dessa função  $D^m r_i(t)$  (para designar a  $m$ -ésima derivada da função  $r_i(t)$ ), reflete intrinsecamente o fluxo dos gases, pois a natureza da emissão dos gases é justamente a variação da concentração ao longo do tempo. Assim, analisar o padrão de variação dessas derivadas significa analisar o padrão de variação dos fluxos dos gases. Também será estimado o fluxo acumulado, usando a integral da curva funcional do fluxo médio diário. A integral será realizada através da função desenvolvida por Bi e Kuesten (2013).

Após a comparação dos modelos, será analisado se através dos modos de variação dos gases é possível explicar o impacto dos usos dos gases em diferentes solos. Nesse sentido, os modos de variação para os gases serão analisados a partir de uma perspectiva univariada, onde cada gás é analisado separadamente e multivariada, os gases serão analisados de forma conjunta usando FPCA. Após a aplicação da FPCA para o caso multivariado, consideram-se as primeiras pontuações (*scores*) dos componentes principais como variáveis aleatórias. Desta forma, será aplicado uma análise não supervisionada por meio da análise de *cluster* com o método Ward em relação a esse conjunto de dados, com o objetivo de identificar a separação dos usos dos solos através dos grupos formados.

Os *scores* dos PCs também serão utilizados para realizar uma análise supervisionada por meio da análise de classificação, com objetivo de explicar se através da variabilidade dos fluxos dos gases é possível classificar os diferentes usos do solo. Em relação ao algoritmo de aprendizado de máquina, foi utilizada a função SVC (*Support Vector Classification*) cujo algoritmo é baseado em Chang e Lin (2011). Os dados foram organizados considerando os três componentes principais como variáveis independentes e os usos dos solos (linha, entrelinha, pasto intensivo e pasto extensivo) foram considerados como variáveis dependentes (*target*). Para seção de treinamento, os dados foram divididos aleatoriamente em 70% (dos dados totais) em conjunto de treinamento e 30% conjunto de teste. Os dados de treinamento foram utilizados para geração dos hiperparâmetros. O fluxograma do modelo de classificação pode ser visto na Figura 2.

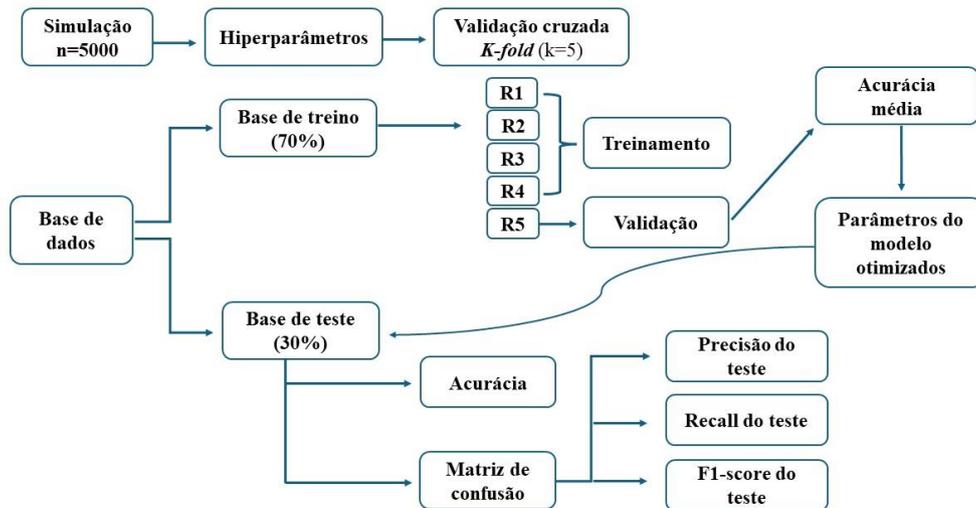


Figura 2 - Fluxograma do modelo de classificação. Fonte: o autor.

Os melhores hiperparâmetros serão aqueles que obtiverem melhor performance de acordo com uma determinada métrica. Para o presente trabalho, foi considerado a acurácia para determinação dos parâmetros. A acurácia representa o número de instâncias classificadas corretamente pelo modelo dividido pelo número total de instâncias. Após a determinação dos hiperparâmetros, o modelo é validado no conjunto de teste. Em termos computacionais, os parâmetros foram gerados a partir do método *RandomizedSearchCV* do *sklearn.model*, onde os parâmetros são gerados aleatoriamente a partir de uma distribuição de probabilidade. Tanto os parâmetros utilizados como as distribuições podem ser visualizadas no Apêndice B. Nessa etapa, os dados de treinamento são divididos em  $k$  conjuntos de dados, através do método de validação cruzada (*k-fold CV*). O modelo é então treinado considerando  $k - 1$  conjuntos e avaliado no conjunto restante. Logo, o processo é repetido  $k$  vezes (Salazar *et al.*, 2022). A cada ciclo, o método armazena diversos resultados da validação cruzada. Por fim, é gerada a matriz de confusão e outras métricas com os melhores parâmetros obtidos no conjunto de treinamento.

Uma boa forma para análise da performance de classificação é uso da matriz de confusão. Através dela é possível analisar o número de instâncias  $X$  que são classificadas com  $Y$  (Géron, 2021). Assim, pode-se verificar quantas vezes o rótulo  $X$  foi confundido com o rótulo  $Y$ . Em relação ao uso do solo, por exemplo, é possível verificar quantas vezes pasto intensivo foi classificado pasto extensivo. Uma matriz de confusão é organizada em rótulo real e rótulo previsto, sendo as linhas reservadas aos rótulos reais e as colunas reservadas ao rótulo previsto.

Logo, quanto maior os valores da diagonal principal, melhor o modelo. Uma matriz de confusão considerada ideal teria apenas valores diferentes de zero presente na diagonal principal.

Além da análise visual, também é possível formular diferentes métricas com o auxílio da matriz de confusão, uma vez que é possível extrair da mesma bastante informação. Assim, além da acurácia usada no treinamento, também podemos definir outras métricas que trarão diferentes *insights* em relação ao modelo de classificação. Nesse sentido, a primeira métrica utilizada é a precisão que representa a acurácia das previsões positivas (Géron, 2021). Cálculo da métrica pode ser visualizado na equação abaixo.

$$Precisão = \frac{TP}{TP+FP}, \quad (3.1)$$

na fórmula acima, TP é o número de verdadeiros positivos e FP representa o número de verdadeiros negativos. Em relação ao modelo geral (para número de classes maior que 2), basta para cada classe em relação à coluna dividir o número da diagonal principal da matriz de confusão pelo total da coluna do rótulo que se deseja calcular a precisão. Embora informativa, a precisão sozinha não funciona de maneira completa, pois o modelo ignoraria os falsos negativos do nosso modelo. Para uma análise mais completa, a precisão normalmente é utilizada com a métrica *Recall* também comumente chamada de sensibilidade (Géron, 2021). O *recall* é dado pela seguinte equação:

$$Recall = \frac{TP}{TP+FN}, \quad (3.2)$$

em que FN representa os falsos negativos. No caso do cálculo do *Recall*, basta calcular a razão da diagonal principal do rótulo sob estudo pela soma total da linha. Por último, é possível combinar as duas métricas acima em uma única métrica conhecida  $F_1$  dada na equação abaixo (Géron, 2021).

$$F_1 = 2 \frac{Precisão * Recall}{Precisão + Recall}. \quad (3.3)$$

Como  $F_1$  é uma média harmônica, o valor de  $F_1$  só obterá valores altos se a Precisão e o *Recall* forem altos. Para a análise estatística dos dados, foi utilizado *software* estatísticos RStudio (TEAM, 2022), que é um conjunto integrado de ferramentas computacionais que possibilita a manipulação de dados, análise estatística e produção de gráficos. Para a classificação, os resultados foram gerados através do *Jupyter* notebook usando a linguagem programação Python (versão 3.11).

## 4 RESULTADOS E DISCUSSÕES

### 4.1 Comparação entre os modelos funcional, linear e exponencial sobre os diferentes usos do solo

Na primeira etapa, que é mesurar os fluxos do  $\text{CH}_4$ ,  $\text{CO}_2$ , e  $\text{N}_2\text{O}$  não foram levados em consideração as variáveis climáticas (precipitação, temperatura, pressão, umidade) no cálculo dos modelos, as coletas das concentrações foram medidas no período da manhã para todos os dias de amostragem. Dessa forma, o valor dos fluxos através do modelo linear é dado a partir da reta de regressão do modelo, o modelo exponencial foi através do modelo HM e, para as curvas funcionais, o fluxo é calculado através da derivada da representação funcional.

Para exemplificar as diferenças entre os diferentes modelos testados, foram escolhidas situações em que há mudança de linearidade de maneira gradativa nas emissões dos gases. Na Figura 3a foi um momento de medição em que gerou um padrão linear, já a Figura 3d está a situação que apresentou maior não linearidade da variação da concentração.

Uma inspeção visual da Figura 3a sugere que não há diferença entre os valores dos fluxos calculado pelo modelo linear e exponencial, quando os dados de concentração do gás têm um alto grau de linearidade. Já o modelo funcional, o fluxo estimado é ligeiramente diferente dos outros dois modelos. Contudo, à medida que esse comportamento muda se tornando não linear, os valores têm uma diferença bastante notável.

Ao analisar os padrões similares a Figura 3c, 3d, constatou-se que o fluxo mensurado pelo modelo exponencial sempre é elevado, esses padrões são bastante encontrados no  $\text{CH}_4$ . Alguns valores dos fluxos calculados pelo modelo exponencial podem ser até quatro vezes maior que os valores dos fluxos obtidos pelo modelo funcional. Uma vez que a curva funcional passa pelos pontos de concentração (ou bem próximos aos pontos) em todas as amostras, elas usam a própria variação do gás em todo intervalo de tempo, conseguindo ter uma boa representação para o fluxo.

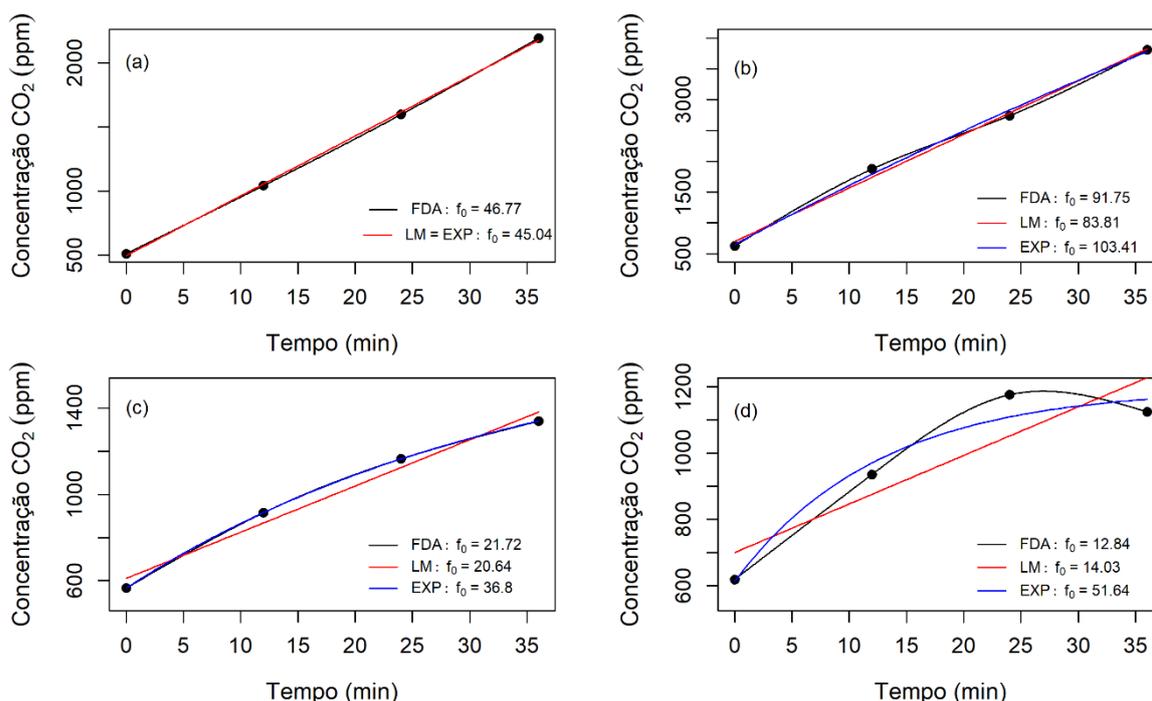


Figura 3 - Exemplos de situações em que existe mudança de linearidade de maneira gradativa nas emissões das concentrações (pontos) do CO<sub>2</sub> para o pasto extensivo, coletadas através da câmara estática fechada inserida no solo. O fluxo ( $f_0$ ) é medido a partir da reta de regressão para o modelo linear (LM) (linha vermelha). O modelo HM é utilizado para mensurar o valor do fluxo para a curva exponencial (EXP) (linha azul) e o valor do fluxo para o modelo funcional (FDA) (linha preta) é mensurado a partir da derivada da função. Fonte: autor.

Como relatam Parkin e Ventura (2010), quando a taxa de variação for constante (Figura 3a), pode-se calcular o fluxo utilizando uma regressão linear, porém quando as observações não forem constantes, ou seja, curvilínea (Figura 3d), a regressão linear não é apropriada para a mensuração do fluxo dos gases. Como também analisado em Silva *et al.* (2015), não apenas uma abordagem de modelo de regressão linear deve ser usada para estimar os fluxos de GEE. Dessa forma, o modelo funcional é apropriado para mensurar tanto a taxa de variação dos gases quando forem constantes, quanto curvilíneas para os três gases.

As Tabela 3, Tabela 4 e Tabela 5 mostram as medidas de ajuste e as medidas de concordância em relação aos fluxos calculados através do modelo funcional, modelo linear e o modelo exponencial, para os quatro usos da terra dos três gases em estudo.

Na Tabela 3, a partir da comparação entre modelos linear vs. funcional para o CH<sub>4</sub>, verifica-se que existe uma correlação positiva moderada ( $0,7 < \rho < 0,8$ ) para a linha e entrelinha da cana de açúcar, e uma correlação positiva forte para o pasto extensivo e intensivo ( $\rho = 0,824$  e  $\rho = 0,822$ , respectivamente). Há correlação positiva fraca ( $\rho < 0,49$ ) para os quatro usos da terra do modelo exponencial vs. funcional. Para todos os usos do solo, houve

uma variação do modelo pequena ( $R^2 < 0,68$ ), especialmente o modelo exponencial vs. Funcional ( $0,14 < R^2 < 0,23$ ). O modelo linear vs. funcional exibiu um grau de concordância moderado para todos os usos da terra, com uma variação entre 0,67 a 0,82.

Tabela 3 - Medidas de ajuste em relação aos modelos linear, exponencial e funcional para os CH<sub>4</sub> para cada uso da terra.

<b>Modelos</b>	<b>Uso do solo</b>	<b><math>\rho</math></b>	<b><math>R^2</math></b>	<b>CCC</b>	<b>C.b</b>
<b>Linear vs. Funcional</b>	Pasto extensivo	0,824	0,678	0,816	0,989
	Pasto intensivo	0,822	0,674	0,758	0,921
	Linha da cana-de-açúcar	0,732	0,533	0,703	0,959
	Entrelinha da cana-de-açúcar	0,728	0,527	0,676	0,929
<b>Exponencial vs. Funcional</b>	Pasto extensivo	0,482	0,229	0,390	0,805
	Pasto intensivo	0,398	0,153	0,296	0,743
	Linha da cana-de-açúcar	0,423	0,174	0,237	0,559
	Entrelinha da cana-de-açúcar	0,382	0,140	0,268	0,703

$\rho$  = coeficiente de correlação; CCC = coeficiente de correlação de concordância;  $R^2$  = coeficiente de determinação; C. b = Fator de correlação de tendência. Fonte: o autor.

Em relação ao modelo exponencial vs. funcional, a concordância foi fraca para os quatro usos do solo ( $CCC < 0,4$ ). O fator de correlação de tendência (C.b) apresentou pequeno deslocamento do modelo linear vs. funcional (Apêndice A1) em todos os usos do solo. Porém, houve um grande deslocamento a respeito da linha de tendência para todos os casos em estudo do modelo exponencial vs. funcional, as retas se distanciaram bastante (Apêndice A2), ficando concorrentes entre si. O que era esperado, já que houve uma fraca concordância entre esses dois modelos.

A comparação dos modelos para gás carbono está descrita na Tabela 4. O modelo linear vs. funcional exibiu uma correlação positiva muito forte e um excelente grau de concordância, para todos os usos da terra. A respeito do modelo exponencial vs. funcional, o pasto extensivo e a entrelinha apresentaram correlação positiva moderada com um grau moderado e fraca de concordância, respectivamente. Para o pasto intensivo e a linha da cana possuem uma correlação positiva forte e um grau substancial de concordância.

A correlação de tendência praticamente não exibiu nenhum deslocamento a respeito da linha de tendência do modelo linear *vs.* funcional (Apêndice A3), todos os seus valores foram acima de 0,99. Já para o modelo exponencial *vs.* funcional houve um pequeno deslocamento entre a linha de tendência e a reta de regressão (Apêndice A4), destacando-se apenas a entrelinha da cana-de-açúcar do modelo exponencial *vs.* funcional ( $C.b = 0,875$ ). A variação do modelo foi pequena sobre a entrelinha, comparando o modelo exponencial *vs.* funcional. Para os outros casos houve um variação moderada e grande.

Tabela 4 - Medidas de ajuste em relação aos modelos linear, exponencial e funcional para os CO<sub>2</sub> para cada uso da terra.

<b>Modelos</b>	<b>Uso do solo</b>	<b><math>\rho</math></b>	<b><math>R^2</math></b>	<b>CCC</b>	<b>C.b</b>
<b>Linear <i>vs.</i> Funcional</b>	Pasto extensivo	0,967	0,936	0,967	0,998
	Pasto intensivo	0,989	0,981	0,989	0,999
	Linha da cana-de-açúcar	0,978	0,957	0,977	0,999
	Entrelinha da cana-de-açúcar	0,965	0,932	0,964	0,998
<b>Exponencial <i>vs.</i> Funcional</b>	Pasto extensivo	0,792	0,625	0,777	0,980
	Pasto intensivo	0,885	0,782	0,852	0,962
	Linha da cana-de-açúcar	0,878	0,771	0,846	0,962
	Entrelinha da cana-de-açúcar	0,628	0,391	0,550	0,875

$\rho$  = coeficiente de correlação; CCC = coeficiente de correlação de concordância;

$R^2$  = coeficiente de determinação; C. b = Fator de correlação de tendência. Fonte: o autor.

A comparação dos modelos para óxido nitroso está descrita na Tabela 5. O modelo linear *vs.* funcional exibiu uma correlação positiva muito forte e um excelente grau de concordância, em relação ao pasto intensivo e a entrelinha da cana-de-açúcar. Já o pasto extensivo e a linha da cana, apresentaram correlação positiva forte e um grau substancial de concordância. Sobre o modelo exponencial *vs.* funcional, o pasto intensivo e a entrelinha da cana-de-açúcar apresentaram correlação positiva muito forte e grau substancial de concordância. A linha da cana-de-açúcar teve uma correlação forte e um grau moderado de concordância. Apenas o pasto extensivo foi o que mais diferenciou, seu grau de concordância foi ruim e sua correlação foi moderada.

A correlação de tendência não exibiu praticamente nenhum deslocamento a respeito da linha de tendência do modelo linear vs. funcional (Apêndice A5), todos os seus valores foram acima de 0,99. Mas para a outra comparação exibiram deslocamentos entre as linhas de tendência e a reta de regressão (Apêndice A6), principalmente para o pasto extensivo. A variação do modelo foi bem pequena para o pasto extensivo na comparação do modelo exponencial vs. funcional. Para os demais casos houve uma variação moderada e grande. O modelo linear exibiu valores mais próximos do modelo funcional do que o exponencial, apresentando uma melhor correlação e concordância entre esses dois modelos para todos os gases e usos da terra.

Tabela 5 - Medidas de ajuste em relação aos modelos linear, exponencial e funcional para os N<sub>2</sub>O para cada uso da terra.

<b>Modelos</b>	<b>Uso do solo</b>	<b><math>\rho</math></b>	<b><math>R^2</math></b>	<b>CCC</b>	<b>C.b</b>
<b>Linear vs. Funcional</b>	Pasto extensivo	0,890	0,791	0,883	0,992
	Pasto intensivo	0,990	0,988	0,990	0,999
	Linha da cana-de-açúcar	0,867	0,751	0,867	0,999
	Entrelinha da cana-de-açúcar	0,990	0,981	0,987	0,996
<b>Exponencial vs. Funcional</b>	Pasto extensivo	0,601	0,356	0,564	0,940
	Pasto intensivo	0,959	0,919	0,937	0,977
	Linha da cana-de-açúcar	0,829	0,685	0,788	0,951
	Entrelinha da cana-de-açúcar	0,938	0,881	0,881	0,938

$\rho$  = coeficiente de correlação; CCC = coeficiente de correlação de concordância;  $R^2$  = coeficiente de determinação; C. b = Fator de correlação de tendência. Fonte: o autor.

Observa-se também que as medidas exibem melhor ajuste e concordância para os gases que tem um comportamento mais estável, como é o caso do gás dióxido de carbono e o óxido nítrico. Para o gás metano as medidas têm um menor poder explicativo. O metano apresenta uma maior variabilidade nos dados, essa variabilidade está relacionada com seu comportamento instável, devido aos diferentes períodos de redução e oxidação, mostrando o quão é difícil de analisá-lo e modelá-lo, uma vez que suas emissões podem resultar de condições anaeróbicas e aeróbicas (Dowhower *et al.*, 2020).

Considerando que o modelo funcional representa melhor a variação nas estruturas lineares e não lineares dos gases, temos que o modelo linear é o que se aproxima mais dessa estimativa. Mostrando que apesar das críticas realizadas por Pedersen, Petersen e Schelde (2010) e Parkin e Venturea (2010) em relação ao modelo linear ele ainda aparenta ser um modelo razoável na estimativas de fluxo em situações lineares. No entanto, quando analisando um gás com comportamento não linear como o CH<sub>4</sub> tanto o modelo linear quanto exponencial se afasta do modelo funcional. Em termos de capacidade descritiva, o modelo funcional é o que melhor se ajusta aos pontos das concentrações, uma vez que a suavização permite que a curva passe pelos pontos observados.

## 4.2 Representação dos dados pelo modelo funcional e sua derivação

Para a reestruturação dos dados discretos da concentração dos gases, utilizou-se a suavização *B-spline*, pois os dados exibem um comportamento não periódico. Com o auxílio da GVC, foi determinada 15 funções bases com parâmetro de suavização ( $\lambda$ ) igual a 1,5. Após essa transformação, obteve a representação da função completa e contínua da concentração das emissões dos gases de efeito estufa, permitindo as análises estatísticas e conservando a variabilidade nos dados originais. Dessa forma, cada curva funcional em cinza da Figura 4 (painel superior), refere-se à suavização dos dados brutos da concentração em quatro pontos de tempo em minutos ( $t \in (0, 36)$ ), obtidos através das câmaras estáticas fechadas instaladas no solo. Sendo 141 funções em relação ao pasto extensivo, 143 para o pasto intensivo, 150 para a linha da cana-de-açúcar e 147 para entrelinha da cana-de-açúcar, totalizando 580 observações.

Na Figura 4 (painel superior) também foram adicionados o *boxplot* funcional (linhas tracejadas em azul) referentes aos quantis 2,5, 25, 75, 97,5, a mediana funcional (linha vermelha) e a média funcional (linha preta) para todos os três gases em estudo. Observa-se que as médias funcionais tiveram um comportamento diferente entre os gases. Para o CH<sub>4</sub> apresentou uma leve inclinação decrescente em todo intervalo de tempo, sua mediana também exibiu o mesmo comportamento, e elas ficaram próximas. Já o CO<sub>2</sub>, a média teve uma inclinação crescente, bem como sua mediana. No início do intervalo elas ficaram sobrepostas, mas se distanciaram no final. Entre a curva dos quantil 75 a 97,5, existe uma grande variação para os três gases, a destacar o N<sub>2</sub>O, visto que sua média funcional está entre esse intervalo. De acordo com Sun e Genton (2011), esse comportamento pode acontecer quando a região central

(entre o quantil 25 e 75) é estreita, ou seja, essa região tem menos variabilidade. Todos os conjuntos de dados apresentaram assimetria.

Na Figura 3 (painel intermediário), observa-se os gráficos de desvio padrão em relação aos três gases. É possível observar que de maneira geral o desvio padrão funcional apresenta um tendência de aumento em relação ao tempo, destacando-se principalmente o comportamento do  $\text{CO}_2$  e  $\text{N}_2\text{O}$ . Em relação ao  $\text{CH}_4$ , dado a não linearidade já destacada, inicialmente o desvio padrão funcional decresce de zero a cinco minutos e depois apresenta um crescimento com o tempo.

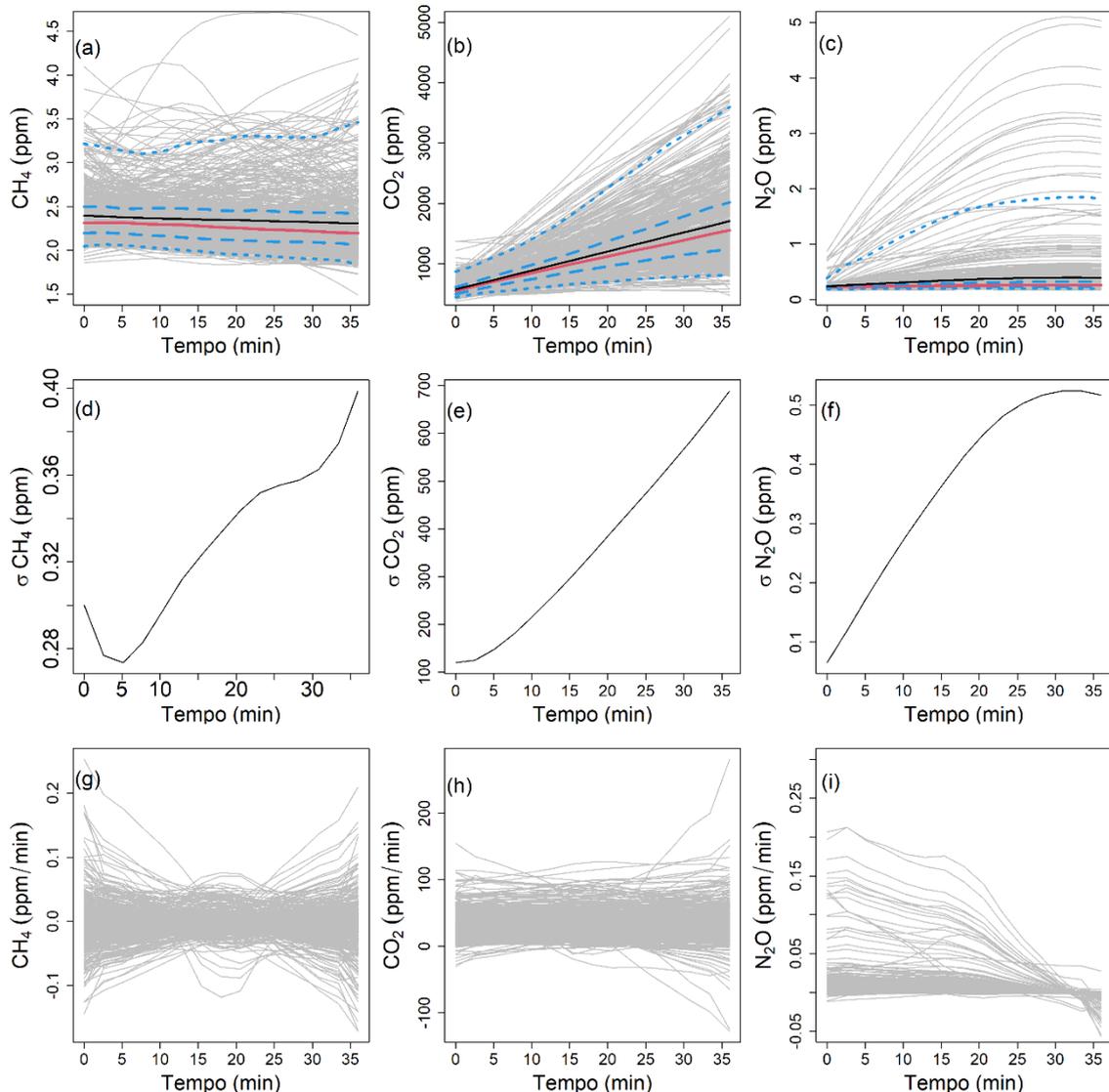


Figura 4 - Curvas funcionais (painel superior) representando as concentrações dos gases  $\text{CH}_4$ ,  $\text{CO}_2$  e  $\text{N}_2\text{O}$  para quatro usos do solo. Linha pontilhada azul, representa quantis 2,5; 25; 75; 97,5. A curva em vermelho corresponde a mediana, e em preto a média. Painel intermediário representa o desvio padrão funcional referente as concentrações dos gases  $\text{CH}_4$ ,  $\text{CO}_2$  e  $\text{N}_2\text{O}$  para quatro usos do solo. Curvas derivativas (painel inferior) das concentrações suavizadas dos gases (fluxo). Fonte: o autor.

Observando o comportamento do  $N_2O$ , percebe-se que o mesmo exibe um leve crescimento entre o intervalo de zero a 24 minutos, em seguida ele se torna constante no intervalo de  $t \in (24, 36)$  minutos. Investigando esse comportamento separadamente para cada uso da terra (Figura 5), verifica-se que esse padrão se repete para o pasto intensivo, linha e entrelinha da cana-de-açúcar. Esses usos da terra tiveram o mesmo tipo de preparação do solo (gradagem, calagem e aragem), além do mesmo tipo de fertilizantes, diferenciam-se apenas a quantidade de insumos, como podem ser vistos na Tabela 2. Desta forma, eles apresentaram um padrão similar entre suas taxas de concentração. Apenas o pasto extensivo (Figura 4a) não apresenta nenhuma variação, nem crescente nem decrescente.

Como as amostras dos gases são feitas simultaneamente, através da câmara estática, o intervalo de tempo para suas coletas (em média 40 minutos) se baseia na padronização de técnicas e procedimentos que já são estabelecidos na literatura, como pode ser visto em Matthias, Blackmer e Bremmer (1980), Bouwman, Boumans e Batjes (2002b), Davidson *et al.* (2002), Rochette, Eriksen-Hamel (2008), Parkin e Venterea (2010), Collier *et al.* (2014) e Zanatta *et al.* (2014). Entretanto, o tempo de coleta do  $N_2O$  pode ser menor, dado que houve uma estabilidade após a terceira coleta, como visto na Figura 5.

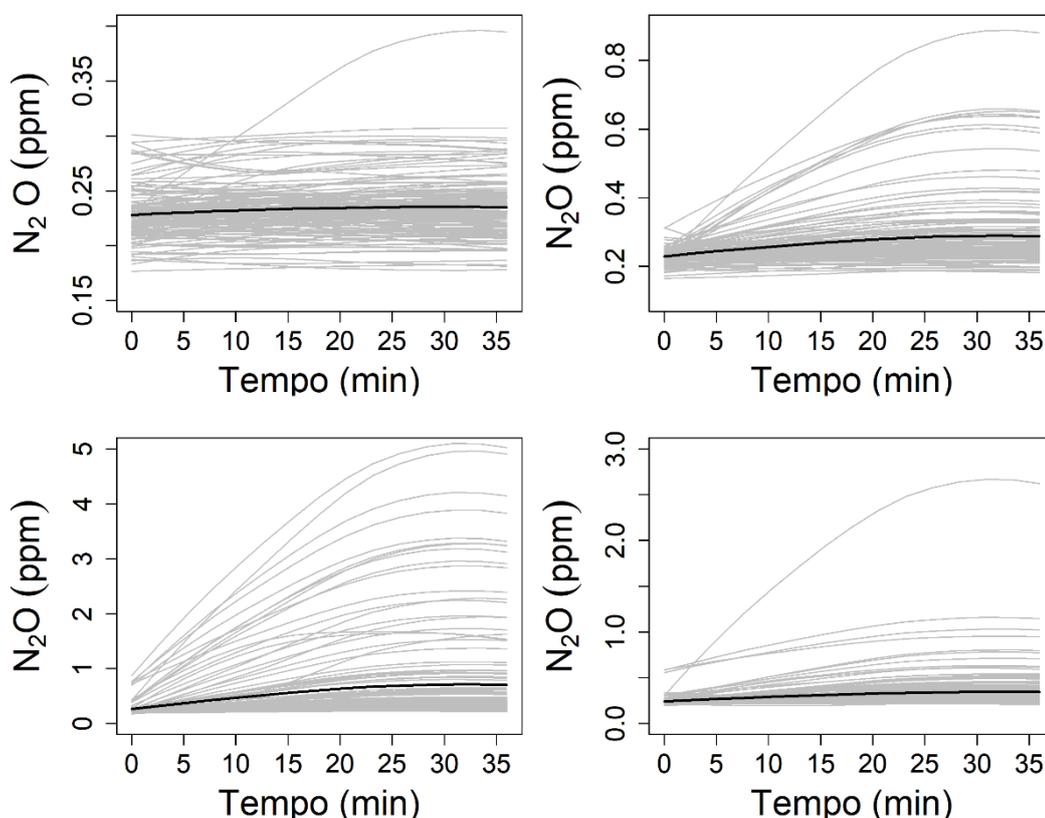


Figura 5 - Curvas funcionais representando as concentrações do gás  $N_2O$  para quatro usos do solo: (a) Pasto extensivo; (b) pasto intensivo; (c) linha da cana-de-açúcar; (d) entrelinha da cana-de-açúcar. A curva em preto a média funcional. Fonte: o autor.

Essa estabilidade também pode ser verificada em relação ao comportamento da derivada funcional (Figura 4i), em que há o ponto crítico após os 30 minutos. Ponto crítico representa o valor da derivada quando esta começa a flutuar perto de zero, nesse caso, o valor do tempo o qual corresponde a estabilidade dos fluxos. Para validar o valor do ponto crítico foi realizado o teste estatístico *t* de Student para uma amostra, com o intuito de testar a hipótese nula ( $h_0$ ) se a média do fluxo do  $N_2O$  é igual a zero em 30 minutos, caso contrário será a hipótese alternativa ( $h_a$ ).

No Apêndice C, exibe o intervalo de confiança em 95% em relação ponto crítico no tempo do fluxo do gás  $N_2O$ . Observa-se que o valor do fluxo estabiliza a partir de 30 min (Figura 4i), uma vez que os limites do intervalo de confiança flutuam próximos de 30 min e no minuto 32 o *p*-valor é significativo ( $p < 0.05$ ). Dessa maneira, se diminuir o tempo de coleta do  $N_2O$  até os 30 minutos e distribuindo os quatro pontos de tempo nesse intervalo, pode-se obter uma representação mais confiável da emissão desse gás além de diminuir o esforço amostral.

As derivadas das funções dos gases estão representadas na Figura 4 (painel inferior). Cada função exibe o comportamento do fluxo ao longo do tempo, ou seja, cada função representa a trajetória do fluxo mensurado a partir da curva funcional de cada coleta. Para o  $CH_4$ , a trajetória é bem variável, com emissão e absorção do fluxo em uma mesma coleta. O  $CO_2$  apresenta comportamento constante e positivo, em todo o intervalo de tempo e o  $N_2O$  exibe uma trajetória decrescente se aproximando de zero no final do intervalo. Adicionalmente, em relação ao  $N_2O$ , é possível observar a adequação da janela temporal de 30 minutos, uma vez que a estabilização do fluxo desse gás ocorre após 25 minutos para os quatro usos da terra (Figura 5). Isso pode ser observado na Figura 4i, que mostra o comportamento da derivada funcional.

Percebe-se que, com as derivadas obtidas através da FDA, pode-se analisar todo o comportamento do fluxo local e global. O fluxo local é o fluxo que pode ser medido em qualquer instante de tempo (onde,  $t \in [0, 36]$  minutos). Já o fluxo global é considerado o fluxo médio. Com esses fluxos, verifica-se os momentos de grandes variações, além de verificar os momentos dos fluxos positivos, referem-se a emissão dos gases para a atmosfera, e valores dos fluxos negativos, representando a captação do gás pelo solo (Debouk; Altimir; Sebastià, 2018). Estudar as derivadas é um ponto fundamental da pesquisa sobre dados funcionais, por exemplo, Gao (2007) analisando as derivadas dos ciclos diurnos de ozônio no sul da Califórnia encontrou resultados que eram intrínsecos das funções da variação do ozônio e conseguiu observar que

existia uma inibição mais fraca do ozônio pela manhã durante a semana e uma acumulação mais rápida e mais longa de ozônio no domingo.

### 4.3 Emissões das médias diárias do CH<sub>4</sub>, CO<sub>2</sub> e N<sub>2</sub>O

Os fluxos médio diários, calculados pelo modelo funcional, dos gases para os quatro usos da terra estão disponibilizados na Figura 6. As maiores emissões de fluxo médio diário do metano estão concentradas no pasto extensivo (Figura 6a), em que 26 das datas amostradas apresentaram valores positivos, isto é, houve emissão de metano em mais de 60% dos dias de amostragem. Uma explicação para esse comportamento são as mudanças induzidas pela compactação na estrutura do solo, pois afeta essencialmente as características da difusão dos gases nesse ambiente, criando condições de anaerobiose e, portanto, influenciando o potencial do solo para absorção (Vasconcelos *et al.*, 2019; Lakshani *et al.*, 2022). Resultados comparáveis, em região que exibem a mesma espécie da pastagem e o mesmo tipo de solo do presente trabalho, foram encontrados por Siqueira-Neto (2006), Frazão (2007) e Carvalho (2010), em que as maiores emissões do CH<sub>4</sub> estiveram no pasto (estabelecido na área experimental) quando comparados com o plantio.

Em relação aos outros três uso da terra (Figura 6d, 6g, 6j), cerca de 80% dos dias de coletas exibiram valores negativos, indicando que o solo funcionou como sumidouro de CH<sub>4</sub>. Esse comportamento era esperado visto que, potencialmente, os solos atuam como um sumidouro de CH<sub>4</sub> onde as condições aeróbicas favorecem a metanotrofia (Gebert; Groengroeft; Pfeiffer, 2011). Resultados semelhantes foram descritos em Siqueira-Neto *et al.* (2020), em que apenas houve emissão do CH<sub>4</sub> na pastagem, nas demais mudança de uso da terra (vegetação nativa, terras de cultivo) em estudo houve captação desse gás. Vasconcelos *et al.* (2019) também registraram fluxos de CH<sub>4</sub> mais elevados no sistema pastagem, já em áreas de cultivo apresentaram absorção.

Os fluxos médio diários do dióxido de carbono (Figura 6b, 6e, 6h, 6k) foram altamente variáveis ao longo dos experimentos para todos os tipos de uso da terra, porém sempre com valores positivos. Há alguns picos de emissão no começo do experimento, correspondendo ao mês de dezembro de 2018 e início de janeiro de 2019 em relação a linha e entrelinha da cana-de-açúcar, após esse pico, os fluxos de CO<sub>2</sub> diminuíram, mas permaneceram variáveis até o final dos experimentos. Observa-se que as emissões médias para o pasto extensivo foram maiores do que as demais, mesmo não apresentando nenhum preparo do solo. Essa situação

pode ocorrer quando há uma adoção de práticas de gestão adequadas na preparação e cultura do solo (Funk *et al.*, 2015; da Costa *et al.*, 2022).

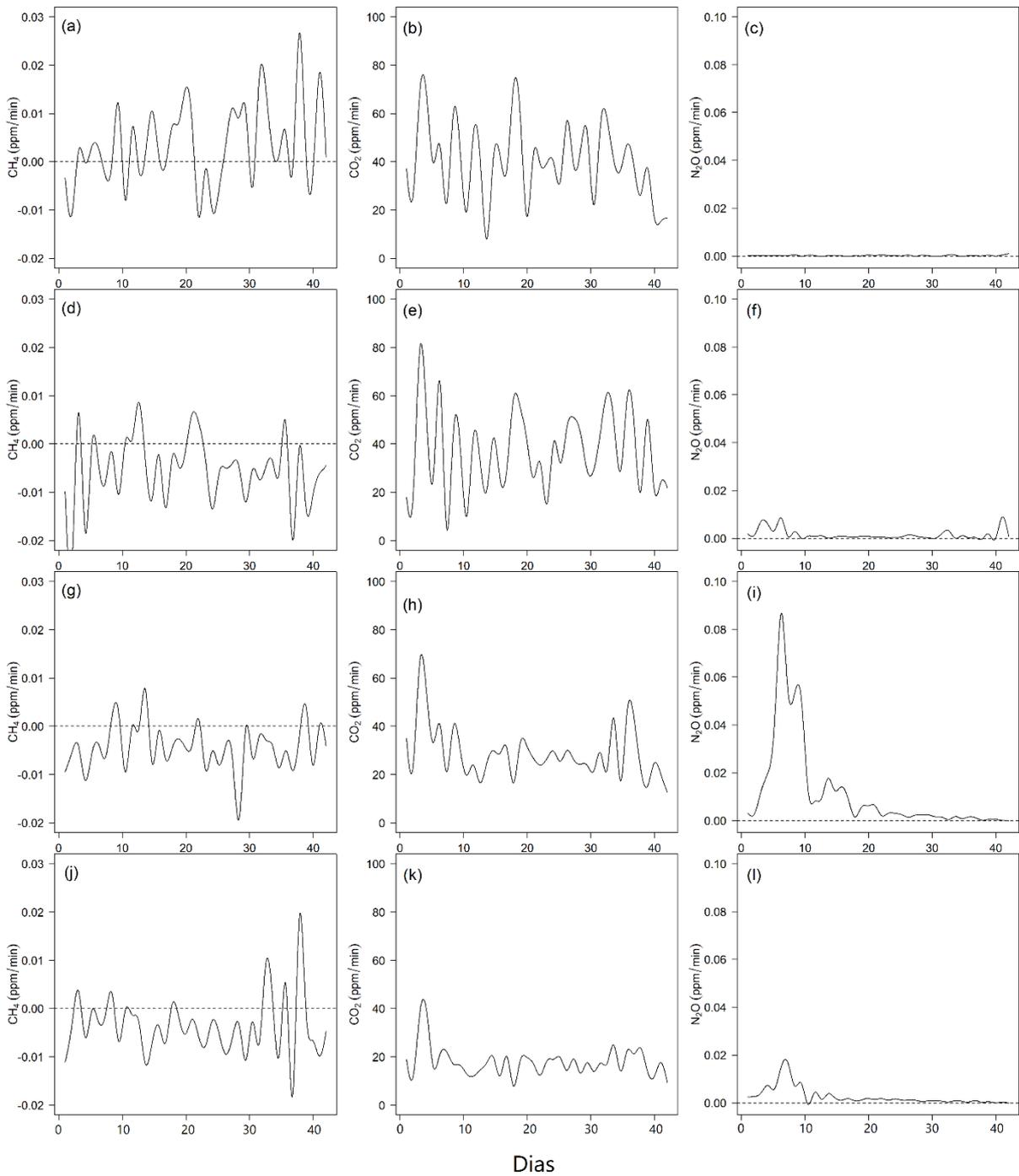


Figura 6 - Fluxo médio diário do  $\text{CH}_4$ ,  $\text{CO}_2$  e  $\text{N}_2\text{O}$ , em relação aos 42 dias de coletas para: pasto extensivo (a) - (c); pasto intensivo (d) - (f); linha da cana-de-açúcar (g) - (i); entrelinha da cana-de-açúcar (j) - (l). Fonte: o autor.

A respeito do fluxo médio diário do  $N_2O$ , foram observados valores positivos para todos os usos do solo (Figura 6c, 6f, 6i, 6l). Para o pasto extensivo, as emissões foram bem baixas, mas esse resultado está dentro do esperado em pastagens extensivas que não são fertilizadas com nitrogênio no Brasil, como relatados em Carmo *et al.* (2012) e Bento *et al.* (2021). Com exceção do pasto extensivo, houve altos picos de emissão entre o terceiro e décimo dia de coleta, que correspondem entre os dias 14 de dezembro de 2018 e 04 de janeiro de 2019, respectivamente. Nos outros dias, os valores diminuíram de forma rápida, ficando próximos a zero.

Os picos de maiores emissões dos fluxos do  $N_2O$  estão no manejo agrícola (linha e entrelinha da cana-de-açúcar). Esse aumento pode estar relacionado adubação nitrogenada, fertilização e ao cultivo. Como também observado em Bento *et al.* (2018), os fluxos de óxido nitroso aumentaram principalmente nos tratamentos envolvendo cultivo mais aplicação de fertilizantes. Martins *et al.* (2015) e Lopes *et al.* (2017) afirmam que índices de emissões de  $N_2O$  intensificam após fertilização nitrogenada, uma vez que sua aplicação aumenta a sensibilidade da respiração do solo (Oertel *et al.*, 2016). Muerer *et al.* (2016), em uma revisão crítica das emissões dos fluxos de  $N_2O$  sob diferentes usos do solo no Brasil, também obteve resultados similares, em que os tipos de uso da terra (pasto, terra de cultivo) tiveram emissões mais elevada de fluxos em terras agrícolas sob tratamento de preparo do solo e em rotações de  $N_2O$  dos solos, mas as emissões foram geralmente baixas.

As emissões acumuladas dos gases para cada uso do solo foram calculadas a partir da integral da curva funcional dos fluxos médio diários (Figura 6). Na Tabela 6 estão os valores das emissões acumuladas do fluxo do  $CH_4$ ,  $CO_2$  e  $N_2O$ . Verifica-se que para o  $CH_4$  o fluxo acumulado foi positivo em relação ao pasto extensivo, para os outros usos da terra o fluxo cumulativo foi negativo. Em relação ao  $CO_2$  o fluxo acumulado foi maior no pasto extensivo seguido do pasto intensivo. Já o  $N_2O$  o uso do solo que exibiu melhor fluxo acumulado foi o pasto extensivo.

Tabela 6 - Emissões acumuladas dos fluxos do CH<sub>4</sub>, CO<sub>2</sub> e N<sub>2</sub>O em relação ao pasto extensivo (PE), pasto intensivo (PI), linha da cana-de-açúcar (LC) e entrelinha da cana-de-açúcar (EC).

Uso do solo	Gases		
	CH <sub>4</sub> (ppm/min)	CO <sub>2</sub> (ppm/min)	N <sub>2</sub> O (ppm/min)
PE	0,145	1651,72	0,00696
PI	-0,223	1541,05	0,0636
LC	-0,174	1189,87	0,477
EC	-0,141	736,94	0,107

#### 4.4 Análise de componentes principais funcional (FPCA) para os casos univariado e multivariado

Os resultados iniciais da FPCA são para o caso univariado, em que a FPCA explica a maior variabilidade possível existente nas observações para cada gás analisado em relação aos seus usos da terra. A Figura 7 mostra a média para o primeiro e segundo componente principal funcional, adicionando (marcador vermelho) e subtraindo (marcadores azuis) as autofunções aos perfis médios. Para as emissões do CH<sub>4</sub> (Figura 7a), CO<sub>2</sub> (Figura 7b) e N<sub>2</sub>O (Figura 7c), os dois primeiros PCs cumulativamente representaram mais de 90% da variação em todos os casos.

Como descrito por Schmutz *et al.* (2020), em seu estudo de caso sobre FDA multivariada para dados de poluição em cidades francesas, os marcadores podem ser interpretados como a primeira fonte de variação da média geral, e a primeira fonte é uma variação de amplitude em relação à média. No estudo, a soma dos múltiplos expressa os marcadores positivos nesse componente, tendendo a ter as maiores emissões dos gases, isto é, quanto mais crescente as emissões, maiores são os fluxos. A subtração exhibe os marcadores reversos nesse componente, implicando em fluxos com valores menores.

Os perfis médios dos PCs para o CH<sub>4</sub> exibem um comportamento decrescente com uma larga variação de amplitude nos marcadores. O primeiro PC do CO<sub>2</sub> descreve um deslocamento crescente, tanto para as marcadores positivos como para os marcadores negativos, isto é, para as maiores e menores emissões do gás existe um crescimento em todo o intervalo de tempo em comparação a curva média. Em relação ao N<sub>2</sub>O, a média é crescente, há um deslocamento

crecente para os marcadores positivos e decrescente para os marcadores negativos em todo o intervalo de tempo. Tanto para o  $\text{CO}_2$  como para o  $\text{N}_2\text{O}$  a variação dos marcadores no início da amostragem é pequena, eles ficam próximos a média, conforme o tempo da amostragem aumenta sua variação também amplia, distanciando-se da média.

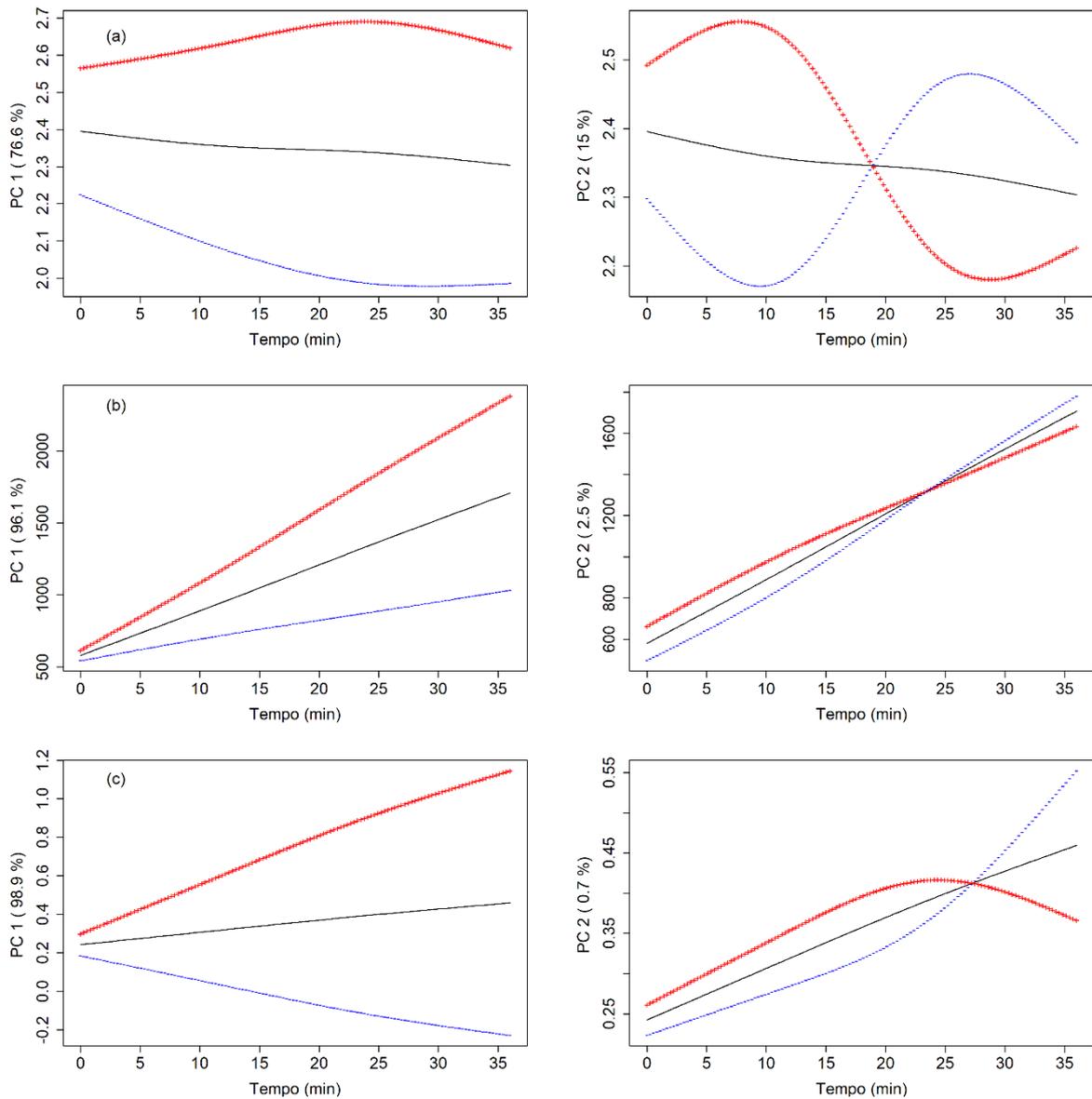


Figura 7 - Primeiro componente principal (painéis da esquerda) e segundo componente principal (painéis da direita) das emissões de: (a)  $\text{CH}_4$ , (b)  $\text{CO}_2$  e (c)  $\text{N}_2\text{O}$ , em relação aos quatro usos do solo. As curvas mostram o perfil médio (linha sólida) e os efeitos da adição (+) e da subtração (-) de autofunções. Para os três PC, cada autofunção é multiplicada por um fator de 1,5 para aumentar a deformação das curvas (+) e (-). Fonte: o autor.

A Figura 8 retrata uma técnica gráfica que representa simultaneamente as relações entre as variáveis (PC1 e PC2) os e as observações (*scores* dos componentes). No qual o valor de cada observação é uma tabela de dupla entrada da FPCA.1 vs FPCA.2. Essa análise é conhecida como Biplot, sua funcionalidade é poder organizar as observações em um formato que permite verificar formações de grupos. Em relação CH<sub>4</sub> (Figura 8a), o biplot dos *scores* dos PCs não foi possível verificar nenhuma separação de grupos a respeito dos uso da terra, os valores ficaram bem entrelaçados. Analisando o biplot para o CO<sub>2</sub> (Figura 8b), nota-se que esse gás foi quem melhor conseguiu diferenciar os usos da terra, separando o pasto extensivo (amarelo) da cana-de-açúcar (linha e entre linha, verde e azul, respectivamente), em que o pasto extensivo possuem os maiores valores das emissões e a entrelinha os menores valores, já os valores para o pasto intensivo ficou mesclado entre os outros três usos. A respeito do N<sub>2</sub>O (Figura 8c), observa-se que as emissões da linha da cana-de-açúcar se distanciaram das demais emissões, todavia não obteve uma boa separação para os outros usos do solo.

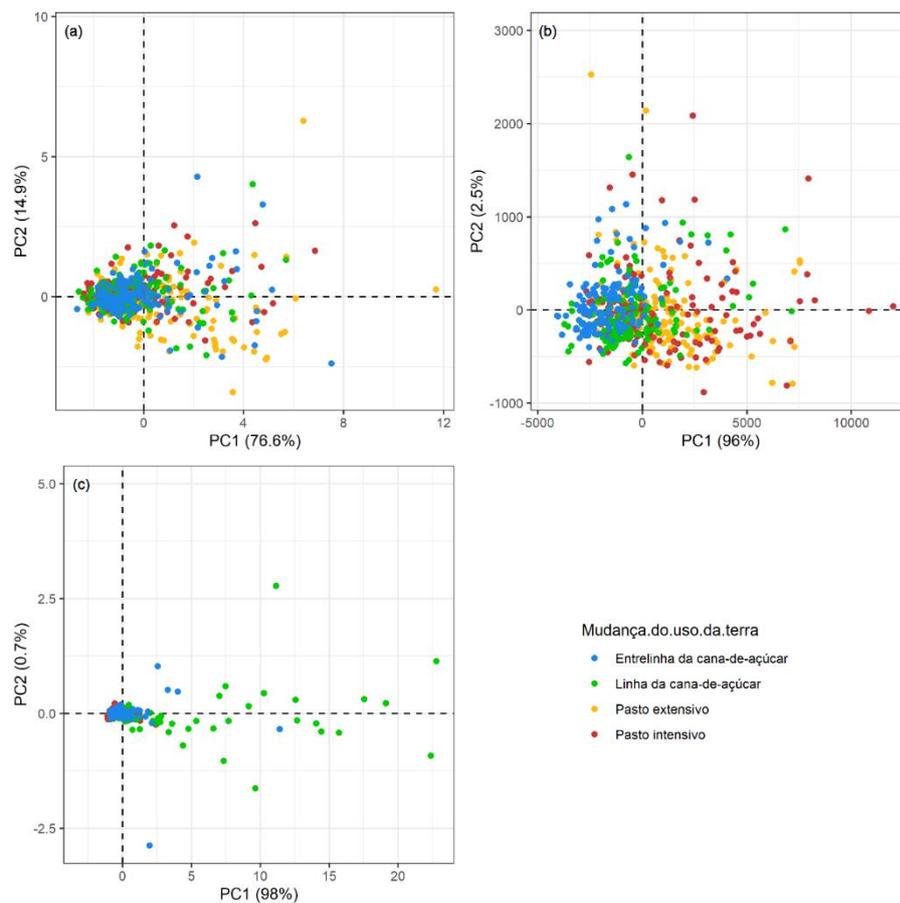


Figura 8 - Biplot dos *scores* dos componentes principais funcional das emissões de: (a) CH<sub>4</sub>, (b) CO<sub>2</sub> e (c) N<sub>2</sub>O, em relação aos usos da terra: pasto extensivo (amarelo), pasto intensivo (vermelho), linha da cana-de-açúcar (verde) e entrelinha da cana-de-açúcar (azul). Fonte: o autor.

Como examinado, apesar da FPCA univariada explicar mais de 90% da variação presente nos dados e ver como essa variação se comporta em relação a curva média dos PCs (Figura 7), ela não conseguiu obter uma boa classificação na separação por uso do solo, para os três gases analisados. Dessa maneira, buscou-se outra abordagem da FPCA em que os gases são analisados de forma conjunta com seus usos da terra. Essa abordagem é a FPCA para o caso multivariado, ela permitiu avaliar os gases simultaneamente, ganhando uma melhor visualização na compreensão do fenômeno.

As representações das emissões dos três componentes dos perfis médios simultaneamente do CH<sub>4</sub>, CO<sub>2</sub> e N<sub>2</sub>O (Figura 9), adicionando (marcador vermelho) e subtraindo (marcadores azuis) as autofunções aos perfis médios. Um total de 93,55% da variação é capturado pelos três primeiros componentes principais funcionais multivariado (MFPCA). O primeiro componente principal (PC) reteve 35,79 da variação total presente nos dados, dessa porcentagem, o CO<sub>2</sub> obteve a maior contribuição (60%), seguido do N<sub>2</sub>O (33,77%) e o CH<sub>4</sub> (6,23%). O primeiro PC do CO<sub>2</sub> descreve um deslocamento crescente, tanto para os marcadores positivos como para os marcadores negativos, isto é, para as maiores e menores emissões existe um crescimento em todo o intervalo de tempo em comparação a curva média. O segundo componente exibiu 33,98%, sendo o N<sub>2</sub>O com a maior porcentagem de variação, em seguida o CH<sub>4</sub> e CO<sub>2</sub> (53,02%, 34,14% e 12,84%, respectivamente). Em relação ao PC2, o N<sub>2</sub>O, apresentou deslocamento crescente para os marcadores positivos e decrescente para os marcadores negativo em todo o intervalo de tempo.

O terceiro componente, com uma menor redução explicativa, 23,78% da variação, com o gás metano apresentando a maior variação (59,56%), posteriormente o CO<sub>2</sub> (27,38%) e por fim o N<sub>2</sub>O (13,06%). O gás metano foi o único que apresentou uma média de PC decrescente, exibindo dois padrões diferentes nos marcadores, em que trajetória dos marcadores positivos tendem a crescer e depois decrescer (concavidade para baixo) e os marcadores negativo tendem a decrescer e depois crescer (concavidade para cima).

Comparando os resultados para o caso univariado e multivariado, notou-se que os comportamentos da média e suas perturbações foram semelhantes com os resultados da FPCA univariado (Figura 7). Contudo, quando utilizado os scores do PC multivariado, obteve-se melhor desempenho na formação de grupos, tanto para uma análise não supervisionada como no caso da análise de *cluster*, como para uma análise supervisionada como a classificação. Esses resultados podem ser observados nas próximas sessões.

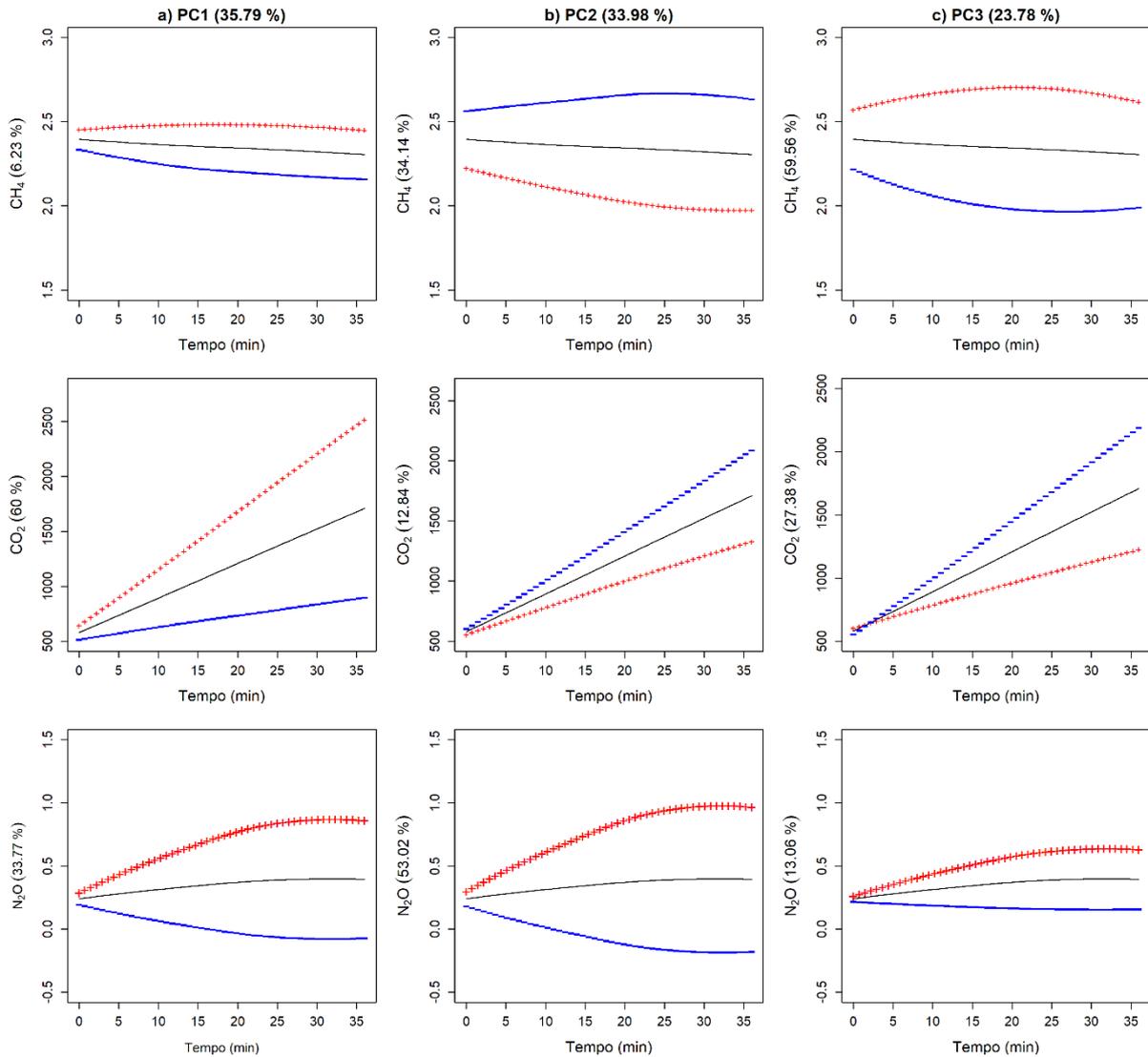


Figura 9 - Representação das emissões dos três componentes dos perfis médios do CH<sub>4</sub>, CO<sub>2</sub> e N<sub>2</sub>O. As curvas mostram o perfil médio (sólido) e os efeitos da adição (+) e da subtração (-) de autofunções. As porcentagens exibidas aos títulos dos cabeçalhos são a quantidade de variação explicada por cada componente. As porcentagens no rótulo do eixo vertical são as quantidades de variação explicada por cada variável. Para os três PCs, cada autofunção é multiplicada por um fator de 1,5 para aumentar a deformação das curvas (+) e (-). Fonte: o autor.

## 4.5 Análise de agrupamento

Para análise de *cluster*, referente aos *scores* dos componentes principais multivariado (considerando a análise simultânea dos gases), foram testadas todas as métricas de distâncias descritas na sessão 2.9 do Capítulo 2, verificando que a distância Euclidiana foi a mais

adequada, dado que foi a que apresentou a melhor separação em relação tanto ao usos da terra quanto de influência do manejo do solo.

Na Figura 10, verifica-se que foram formados quatro *clusters*. Das 134 observações do *cluster 1* (cor amarela), 89% das observações representam as emissões do pasto e 11% a emissões da cana-de-açúcar, sendo rotulado como grupo pasto. Houve valores positivos para quase todos os *scores* do PC1 e PC3, indicando emissões mais altas de CO<sub>2</sub> e CH<sub>4</sub>, respectivamente, e negativos para o PC2, indicando emissões mais baixas de N<sub>2</sub>O. Como visto na Figura 6, os valores médios do fluxo diário mais altos foram do CH<sub>4</sub> e CO<sub>2</sub> para o pasto extensivo seguido do intensivo. Por outro lado, para o N<sub>2</sub>O as emissões do pasto extensivo apresentaram os valores médios mais baixos.

O *cluster 2* (cor vermelha) têm 67 observações das quais, 55% representam o pasto e 45% a cana-de-açúcar. Apresentaram valores positivos em relação ao PC1 e valores negativos para o PC2 e PC3. Esse grupo é o grupo intermediário, pois as emissões tanto para os pastos como da cana-de-açúcar foram similares em relação aos três gases. Em relação ao *cluster 3* (cor verde), as emissões da linha e entrelinha apresentaram grande similaridade, concentrando-se 80% de suas emissões nesse grupo. Por esse motivo, ele foi composto pelo maior quantitativo de observações (362), sendo 68% a cana-de-açúcar e 32% de pasto. Por apresentar em sua grande maioria emissões da cana-de-açúcar, esse grupo foi categorizado como grupo da cana-de-açúcar. O N<sub>2</sub>O exibiu maior contribuição para formação desse grupo, pois os valores dos *scores* para o PC2 foram quase todos positivos, mostrando que são as maiores emissões do gás.

Por fim, o *cluster 4* (cor azul) apresentou apenas 17 observações, todas em relação a cana-de-açúcar. Essas observações ocorreram entre os dias de amostragem de 17 a 26 de dezembro e 02 de janeiro, com altos valores positivos das emissões para o CO<sub>2</sub> e N<sub>2</sub>O e baixos emissões do CH<sub>4</sub>. Como observado em Jacques e Preda (2014b), em presença de importantes variações de amplitude e fase entre curvas funcionais, uma análise de *cluster* forma grupo devido a estas variações. Para os nossos resultados, esse grupo foi bastante importante visto que é avaliado como um *cluster* formado pelos picos de emissões dos gases CO<sub>2</sub> e N<sub>2</sub>O observados na cana-de-açúcar, corroborando com o resultado da Figura 6h, 6i, 6k, 6l, em que mostra os picos dos fluxos médios diários, logo após o solo passar por algum processo de manejo.

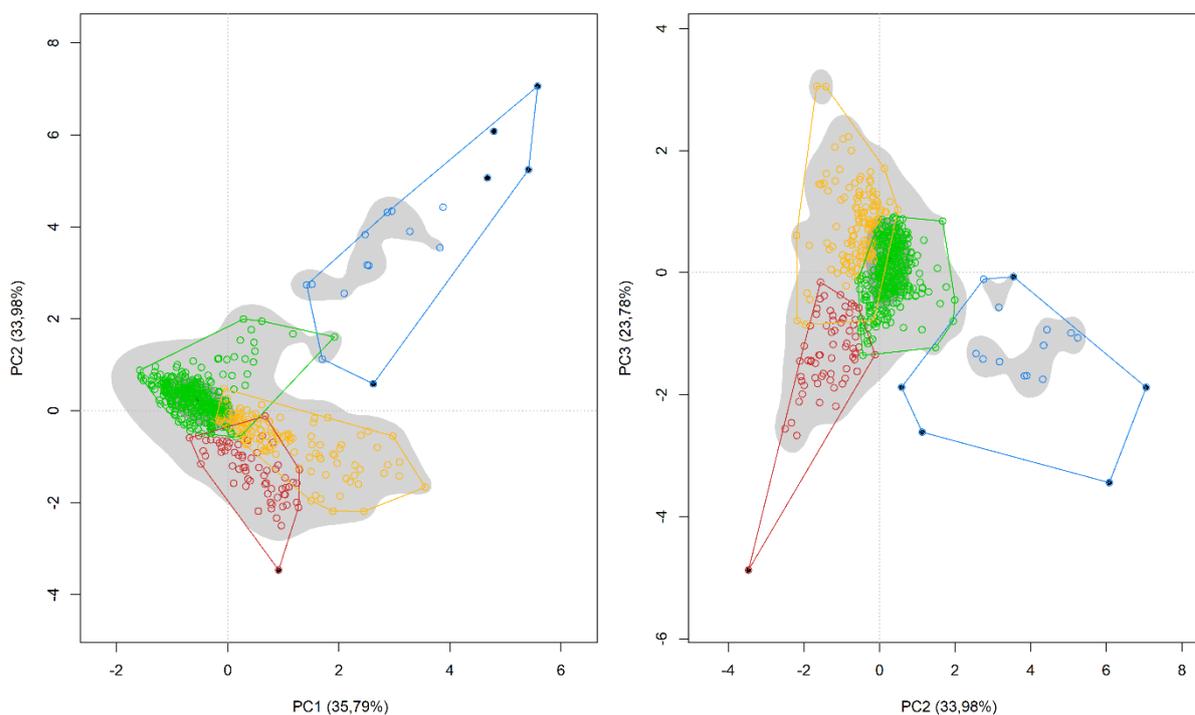


Figura 10 - Análise de *Cluster* baseado nos três componentes principais para o caso multivariado. Esquemas de cores: amarelo grupo 1 (pastagem); vermelho grupo 2 (intermediário); verde grupo 3 (cana-de-açúcar); azul grupo 4 (picos das emissões do CO<sub>2</sub> e N<sub>2</sub>O provocados pelo manejo no solo). Fonte: o autor.

## 4.6 Análise de classificação do uso dos solos

Os resultados da acurácia obtida na validação cruzada durante o treinamento da análise de classificação com os três componentes principais para o caso multivariado utilizando o método SVC, encontra-se no Apêndice D. A acurácia média obtida na validação cruzada durante o treinamento e acurácia obtida pelo teste referente ao modelo de classificação estão na Tabela 7, enquanto as métricas de avaliação do modelo de classificação estão na Tabela 8. Foram considerados diferentes cenários para verificação de diferentes influências na modelagem dos gases. Após a definição dos parâmetros para cada cenário, descrito na Tabela 7, o modelo foi avaliado nos dados do teste. Para ilustrar a performance do modelo nos dados de teste, foram geradas as matrizes de confusão na Figura 12. Através da matriz de confusão é possível analisar os resultados da classificação de maneira mais detalhada, pois é possível observar os erros de classificação em relação aos usos do solo. Por exemplo, das observações que deveriam ser classificadas como pasto extensivo e não foram, é possível identificar quantas foram classificadas em outro uso da terra. Supondo que a maioria dos erros fossem classificados

como pasto intensivo, isso implicaria que o modelo não diferencia bem o pasto extensivo do intensivo.

O primeiro cenário (Tabela 7) foi a classificação apenas com os PCs como variáveis independentes e quatro usos da terra como variáveis dependentes, obtendo uma baixa acurácia do teste de apenas 53%. Quando examinada sua matriz de confusão (Figura 12a) e as métricas de avaliação do modelo de classificação (Tabela 8), a precisão do modelo variou entre 52% e 57% para os usos da terra, sendo o EC o tratamento que obteve maior identificação das observações corretamente ( $recall = 66\%$ ) seguido do PE ( $recall = 65\%$ ), já o tratamento LC foi o que menos acertou em sua classificação ( $recall = 40\%$ ), seguido do PI ( $recall = 43\%$ ).

Dado que fatores ambientais, principalmente temperatura (ar e solo) e umidade do solo, influenciam as emissões dos fluxos do CH<sub>4</sub>, CO<sub>2</sub> e N<sub>2</sub>O (Besen *et al.*, 2018; da Silva *et al.*, 2019), para o segundo cenário (Tabela 7), além dos três componentes principais, foram incorporadas às variáveis independentes a temperatura do solo e do ar no momento da coleta de dados. A temperatura do solo controla o consumo biológico de oxigênio, alterando o crescimento das comunidades microbianas (Wang *et al.*, 2021). Com a adição das temperaturas, aumentou tanto a acurácia do teste (Tabela 7) como a precisão do modelo (Tabela 8). A acurácia foi de 64% e a precisão do modelo variou entre 51% e 82% para os usos da terra. Na matriz de confusão para o cenário 2 (Figura 12b), verifica-se novamente que o EC foi o tratamento que obteve maior identificação das observações corretamente ( $recall = 80\%$ ), seguido do PE ( $recall = 74\%$ ). Já o tratamento LC foi o que menos acertou em sua classificação ( $recall = 45\%$ ), seguido do PI ( $recall = 57\%$ ).

Como observado na Figura 11, as temperaturas médias diárias do ar em relação a linha e entrelinha da cana-de-açúcar são bem próximas uma da outra. Essa aproximação entre as temperaturas também pode ser vista em relação as temperaturas médias diárias do solo para os mesmos tratamentos. Além disso, devido as características do experimento, dado que LC e EC estão na mesma parcela experimental, esses dois tratamentos também foram considerados como um mesmo grupo em um cenário adicional.

Para o terceiro cenário (Tabela 7), foi considerado os PCs e três usos do terra (PE, PI e C). A acurácia do teste foi de 65%, sendo 12% maior do que o cenário 1. A precisão do modelo variou entre 55% e 91% para os usos da terra. Observa-se que na matriz de confusão para o cenário 3 (Figura 12c), o tratamento C obteve maior identificação das observações corretamente

(*recall* = 90%), seguido do PE (*recall* = 63%) e o PI foi o tratamento que teve a menor classificação correta (*recall* = 21%).

Tabela 7 - Acurácia média obtida na validação cruzada durante o treinamento e acurácia obtida pelo teste referente ao modelo de classificação.

Cenários	Descrição do Cenário	Média da acurácia do treinamento	Acurácia do teste
1	PCs com 4 grupos.	0,57	0,53
2	PCs mais a Temperatura com 4 grupos.	0,65	0,64
3	PCs com 3 grupos.	0,70	0,65
4	PCs mais a Temperatura com 3 grupos.	0,78	0,74
5	PCs no período de 06/12/2018 a 25/01/2019 com 4 grupos.	0,65	0,60
6	PCs no período de 06/12/2018 a 25/01/2019 mais a temperatura com 4 grupos.	0,70	0,70
7	PCs no período de 06/12/2018 a 25/01/2019 com 3 grupos.	0,77	0,73
8	PCs no período de 06/12/2018 a 25/01/2019 mais a temperatura com 3 grupos.	0,82	0,77

Fonte: o autor.

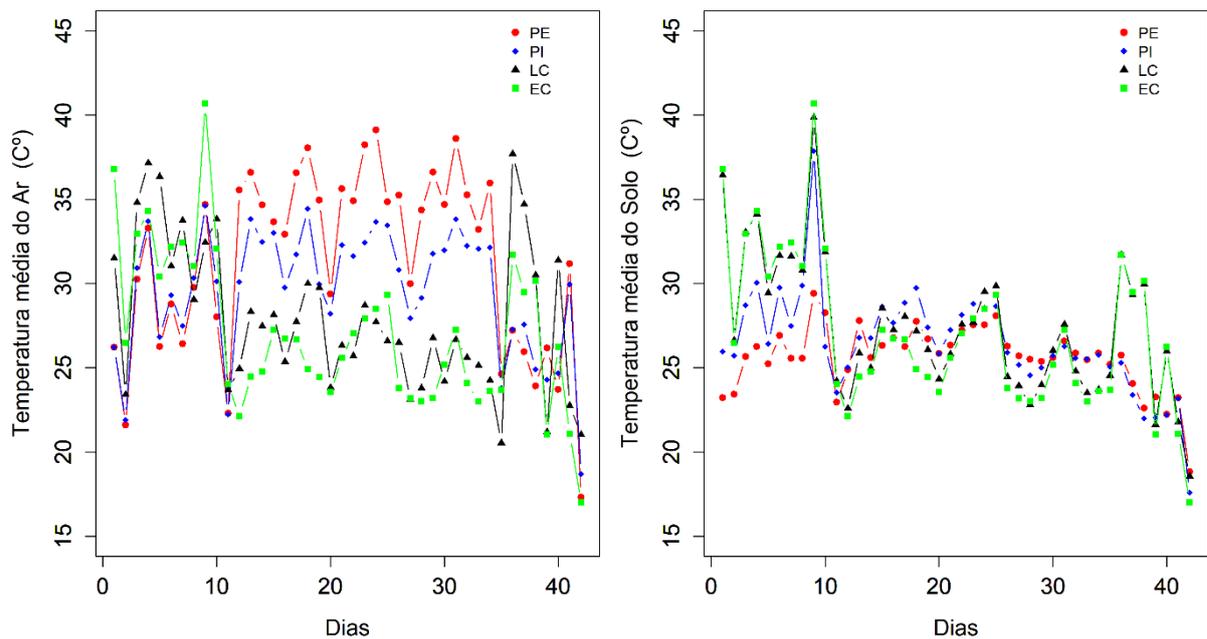


Figura 11 - Temperatura média diária do Ar (painel esquerdo) e do Solo (painel direito) (C°) para cada uso da terra: pasto extensivo (PE), pasto intensivo (PI), linha da cana-de-açúcar (LC) e entrelinha da cana-de-açúcar (EC). Fonte: o autor.

Como visto no cenário 2, quando adicionamos as temperaturas no modelo de classificação ele exibe um melhor desempenho. Assim, no quarto cenário (Tabela 7), foi considerado os PCs, as temperaturas e três usos do terra, obtendo uma acurácia moderada (74%), com a precisão do modelo variando entre 67% e 78% para os usos da terra (Tabela 8). Verifica-se que, na matriz de confusão para o cenário 4 (Figura 12d), o tratamento que exibiu maior proporção de classificação correta foi a cana-de-açúcar ( $recall = 82\%$ ), o pior foi o pasto intensivo ( $recall = 55\%$ ) e o pasto extensivo foi o intermediário ( $recall = 77\%$ ).

Por fim, foi adicionado um componente temporal apenas aos dados coletados de 06/12/2018 até 25/01/2019, a escolha da redução dos dados se deu por duas razões, a primeira foi sobre os espaçamentos dos dias das coletas, entre essas datas houve um grande quantitativo de amostras, e a segunda razão foi que a partir do dia 26 de janeiro de 2019, houve a aplicação do agrotóxico 2,4-D apenas na cana-de-açúcar, como descrito na Tabela 2, então evidencia-se dois períodos em que há diferenças em relação à preparação do manejo dos solos.

Tabela 8 - Métricas de avaliação do modelo de classificação para cada uso do solo. A descrição dos cenários foi escrita anteriormente na Tabela 7.

Cenários	Uso do solo	Precisão do teste	Recall do teste	F1-score do teste
1	PE	0,52	0,65	0,58
	PI	0,57	0,43	0,49
	LC	0,55	0,40	0,46
	EC	0,52	0,66	0,58
2	PE	0,82	0,74	0,78
	PI	0,59	0,57	0,58
	LC	0,51	0,45	0,48
	EL	0,65	0,80	0,71
3	PE	0,55	0,63	0,59
	PI	0,91	0,21	0,34
	C	0,67	0,90	0,77
	PE	0,70	0,77	0,73
4	PI	0,67	0,55	0,60
	C	0,78	0,82	0,80
	PE	0,80	0,75	0,77
	PI	0,55	0,54	0,55
5	LC	0,52	0,57	0,54
	EC	0,56	0,56	0,56
	PE	0,85	0,82	0,84
	PI	0,68	0,61	0,64
6	LC	0,63	0,66	0,64
	EC	0,61	0,66	0,63
	PE	0,84	0,66	0,74
	PI	0,82	0,36	0,50
7	C	0,69	1	0,82
	PE	0,76	0,78	0,77
	PI	0,78	0,46	0,58
8	PI	0,78	0,46	0,58
	C	0,77	0,95	0,85

Fonte: o autor.

Nessa perspectiva, os próximos cenários (5, 6, 7 e 8) foram construídos com o intervalo de amostragem menor. O quinto, sexto e sétimo cenário (Tabela 7) tiveram uma acurácia do teste menor que o cenário 4, com 60%, 70% e 73%, respectivamente. O oitavo cenário foi o que apresentou a maior acurácia do teste com 77% (Tabela 7), nesse cenário foi considerado os PCs, as temperaturas, três usos do solo com os dados com o intervalo de tempo menor. A precisão do modelo (Tabela 8) ficaram bem próximas. De 76% para o PE, 78% para o PI e 77% para a C. E a cana-de-açúcar apresentou a maior proporção de observações classificadas corretamente (*recall* = 95%), seguido do PE (*recall* = 78%) e o PI (*recall* = 46%), como podem ser vistos na Tabela 7 e na matriz de confusão (Figura 12h). No geral, a média da acurácia do treinamento sempre foi maior que o valor do teste, mas a diferença máxima foi de 5% para o cenários 3 e 8, assim, não houve *overfitting* e *underfitting* no modelo de classificação.

Os resultados evidenciam uma forte influência do fatores externos (de manejo e ambientais) no modelo de classificação como referenciado na literatura (Bouwman; Boumans; Batjes, 2002b; Zheng *et al.*, 2023). Isso pode ser observado pelo aumento da acurácia quando tanto a temperatura do solo e do ar e o componente temporal são acrescentadas nos cenários. Além dos fatores externos assinalados acima, observa-se também o aumento da acurácia quando os solos LC e EC são considerados um único tratamento. Do ponto de vista experimental, não existe um isolamento completo entre a linha e entrelinha dado que eles estão na mesma parcela experimental.

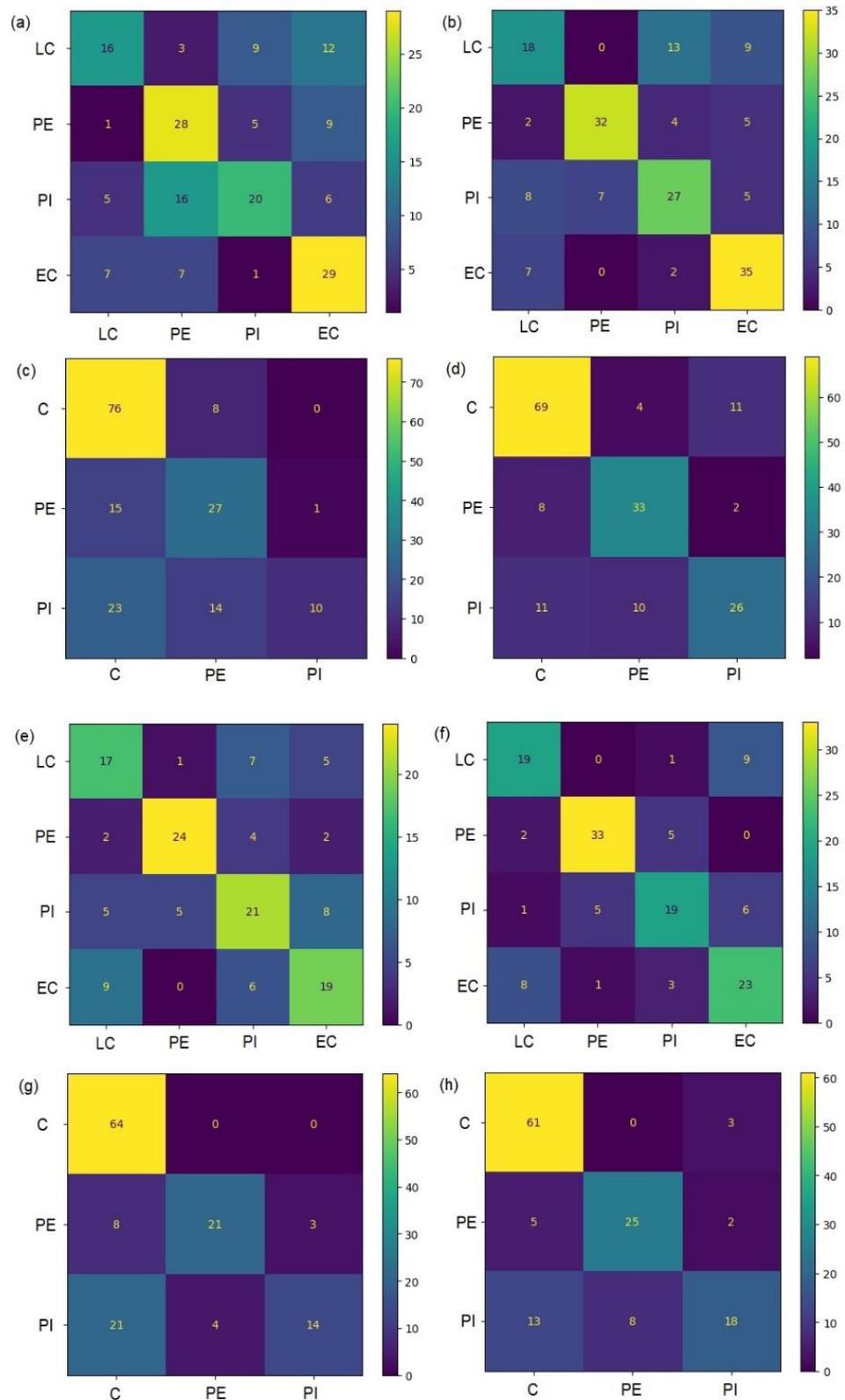


Figura 12 - Matrizes de confusão dos modelos de classificação a respeito da base de teste referentes ao cenário 1 (a), cenário 2 (b), cenário 3 (c), cenário 4 (d), cenário 5 (e), cenário 6 (f), cenário 7 (g) e cenário 8 (h). Em relação aos usos do solo pasto extensivo (PE), pasto intensivo (PI), linha da cana-de-açúcar (LC), entrelinha da cana-de-açúcar (EC) e a união das observações da linha e entrelinha da cana-de-açúcar (C). Fonte: o autor.

## 5 CONCLUSÕES

Visto que estudos sobre as estimativas das emissões dos gases de efeito estufa provenientes do solo são fundamentais para obter uma melhor compreensão sobre as alterações climáticas, faz-se necessário um estudo que consiga representar de maneira eficaz esse fenômeno. Desta forma, aplicou-se técnicas da análise de dados funcionais uni e multivariados aos dados das concentrações do dióxido de carbono, metano e óxido nitroso referentes ao pasto extensivo, pasto intensivo, linha e entrelinha da cana-de-açúcar, além de comparar os fluxos dos GEEs calculados com o modelo funcional em relação aos modelos linear e exponencial.

Em termos de capacidade descritiva, o modelo funcional é o que melhor se ajusta aos pontos das concentrações, uma vez que a suavização permite que a curva passe pelos pontos observados. Dessa forma, o método é apropriado para avaliar o fluxo dos GEEs quando as concentrações forem tanto constantes, como é o caso do  $\text{CO}_2$  e  $\text{N}_2\text{O}$ , quanto curvilíneas, como o  $\text{CH}_4$ .

Por meio das derivadas das funções, calculou-se o fluxo, tanto o local, como global, verificando que uma das vantagens de utilizar dados funcionais é a mensuração do fluxo instantâneo através das derivadas, possibilitando analisar a variabilidade dos fluxos dos gases em relação aos diferentes usos do solo. Desse modo, as derivadas funcionais detectaram que existe uma variabilidade nas emissões dos fluxos entre os dias de coleta. O que pode ser resultados da influência de fatores externos e atividade de manejo do solo. No geral, os usos do solo contribuíram como sumidouro do  $\text{CH}_4$  diferentemente do  $\text{CO}_2$ , que atuou como fonte do gás em todos os usos da terra. A respeito do  $\text{N}_2\text{O}$ , os fluxos ficaram muito próximo a zero, uma vez que foi observada elevadas emissões do gás no início das amostragens, período próximo com preparo do solo. Exibindo assim, a influência do preparo do solo nas emissões desse gás.

Quando analisados os resultados do FPCA, para o caso univariado e multivariado, apesar de resultados semelhantes nos comportamentos da média e suas perturbações, O MFPCA mostrou maior poder discriminante.

Com os *scores* do MFPCA foi realizada a análise de *cluster*, no qual separou em quatro grupos. O primeiro foi representado pelos observações dos pasto, o segundo foi o intermediário, com cerca de 50% pasto e 50% cana-de-açúcar, o terceiro é retratado a cana-de-açúcar, aproximadamente 70% das observações. Por fim, o grupo formado pelos picos de emissões do  $\text{C}_2\text{O}$  e  $\text{N}_2\text{O}$ , exibindo um momento de manejo solo que influenciou o aumento dessas emissões.

Adicionalmente, foi feita uma análise de classificação dos solos a partir dos três componentes principais gerados pelo MFPCA com a utilização do *Support Vector Classification*. Foram considerados oito cenários. A partir dos resultados, observou-se que a acurácia do teste aumenta significativamente quando variáveis externas (temperatura, tempo de coleta) são adicionadas ao modelo de classificação.

Portanto, a aplicação da FDA demonstra ser capaz de captar o comportamento do fenômeno estudado, englobando a natureza contínua do sistema. Sendo assim, representa o processo das trocas dos gases entre solo e ar atmosférico de maneira detalhada.

Em relação aos trabalhos futuros, pretende-se levar em consideração o efeito das variáveis externas ambientais (precipitação, temperaturas, umidade, entre outros) na análise dos fluxos dos gases e também aumentar o número de observações em cada câmara.

## REFERÊNCIAS BIBLIOGRÁFICAS

- ADAMI, M. *et al.* Remote sensing time series to evaluate direct land use change of recent expanded sugarcane crop in Brazil. **Sustainability**, 4.4: 574-585, 2012.
- AKOGLU, H. User's guide to correlation coefficients. **Turkish journal of emergency medicine**, 18(3), 91-93, 2018.
- ALKIMIM, A.; CLARKE, K. C. Land use change and the carbon debt for sugarcane ethanol production in Brazil. **Land Use Policy**, 72: 65-73, 2018.
- ALTMAN, D. G. **Practical statistics for medical research**. CRC press, 1990.
- ANDERSON T. R.; HAWKINS E.; JONES P. D. CO<sub>2</sub>, the greenhouse effect and global warming: from the pioneering work of Arrhenius and Callendar to today's Earth System Models. **Endeavour**, 40:178–87, 2016.
- ANEIROS, G. *et al.* Recent advances in functional data analysis and high-dimensional statistics. **Journal of Multivariate Analysis**, 170, 3-9, 2019.
- BANDE, M. F. *et al.* fda.usc: Functional Data Analysis and Utilities for Statistical Computing. **R package version 2.1.0.**, 2020. <https://CRAN.R-project.org/package=fda>.
- BANDYOPADHYAY, K. K.; LAL, R. Effect of land use management on greenhouse gas emissions from water stable aggregates. **Geoderma**, 232, 363-372, 2014.
- BARUA, S. *et al.* A functional spatial analysis platform for discovery of immunological interactions predictive of low-grade to high-grade transition of pancreatic intraductal papillary mucinous neoplasms. **Cancer informatics**, 17: 1176935118782880, 2018.
- BEEGUM, S.; SUN, W.; TIMLIN, D.; WANG, Z.; FLEISHER, D.; REDDY, V. R.; RAY, C. Incorporation of carbon dioxide production and transport module into a Soil-Plant-Atmosphere continuum model. **Geoderma**, 437, 116586, 2023.
- BEGUM, K. *et al.* Modelling greenhouse gas emissions and mitigation potentials in fertilized paddy rice fields in Bangladesh. **Geoderma**, 341, 206-215, 2019.
- BENTO, C. B. Impactos do manejo do solo em área de pastagem extensiva convertida em cultivo da cana-de-açúcar e em pastagem intensiva. 94f. **Tese** (Doutorado em Biotecnologia e Monitoramento Ambiental). Programa de Pós-Graduação em Biotecnologia e Monitoramento Ambiental da Universidade Federal de São Carlos, São Carlos, SP, 2020.
- BENTO, C. B. *et al.* Effects of extensive-to-intensive pasture conversion on soil nitrogen availability and CO<sub>2</sub> and N<sub>2</sub>O fluxes in a Brazilian oxisol. **Agriculture, Ecosystems and Environment**, v. 321, n. March, p. 107633, 2021.
- BENTO, C. B., *et al.* Impacts of sugarcane agriculture expansion over low-intensity cattle ranch pasture in Brazil on greenhouse gases. **Journal of environmental management**, 206, 980-988, 2018.
- BERRENDERO, J. R., JUSTEL, A., SVARC, M. Principal components for multivariate functional data. **Computational Statistics & Data Analysis**, 55(9), 2619-2634, 2011.
- BESSEN, R. M.; RIBEIRO, R. H.; MONTEIRO, A. N. T. R.; IWASAKI, G. S.; PIVA, J. T. Práticas conservacionistas do solo e emissão de gases do efeito estufa no Brasil. **Scientia Agropecuaria**, 9(3), 429-439, 2018.

- BI, J.; KUESTEN, C. Using Functional Data Analysis (FDA) Methodology and the R Package “fda” for Sensory Time-Intensity Evaluation. **Journal of Sensory Studies**, 28.6: 474-482, 2013.
- BHATIA, S. K. *et al.* Carbon dioxide capture and bioenergy production using biological system – A review. **Renewable and Sustainable Energy Reviews**, 110, 143–158, 2019.
- BOOR, C. D. A practical guide to splines. **Appl. Math. Sci**, 2001.
- BORGES, C. D. *et al.* Can alternative N-fertilization methods influence GHG emissions and biomass production in sugarcane fields?. **Biomass and Bioenergy**, 120: 21-27, 2019.
- BOUWMAN, A. F.; BOUMANS, L. J. M.; BATJES, N. H. Modeling global annual N<sub>2</sub>O and NO emissions from fertilized fields. **Glob. Biogeochem. Cycles**, 16, 28, 2002a.
- BOUWMAN, A.F.; BOUMANS, L.J.M.; BATJES, N.H., Emissions of N<sub>2</sub>O and NO from fertilized fields: Summary of available measurement data. **Global Biogeochemical Cycles**. 16 (4), 1058, 2002b.
- BRIDGHAM, S. D. *et al.* Methane emissions from wetlands: biogeochemical, microbial, and modeling perspectives from local to global scales. **Global change biology**, 19.5: 1325-1346, 2013.
- BUSSAB, W. D. O.; MIAZAKI, É. S.; ANDRADE, D. F. D. Introdução à análise de agrupamento. In Introdução à análise de agrupamento pp. 105-105, 2015.
- BUTTERBACH-BAHL, K. *et al.* Nitrous oxide emissions from soils: how well do we understand the processes and their controls?. **Philosophical Transactions of the Royal Society B: Biological Sciences**, 368(1621), 20130122, 2013.
- CANADELL, J. G., *et al.* Global Carbon and other Biogeochemical Cycles and Feedbacks. **IPCC AR6 WGI**, Final Government Distribution, chapter 5, 2021. fhal-03336145.
- CANISARES, L. P.; CHERUBIN, M. R.; DA SILVA, L. F. S.; FRANCO, A. L. C.; COOPER, M., MOONEY, S. J.; CERRI, C. E. P. Soil microstructure alterations induced by land use change for sugarcane expansion in Brazil. **Soil Use and Management**, 36(2), 189-199, 2020.
- CARMO, J. B., *et al.* Conversion of the coastal Atlantic forest to pasture: Consequences for the nitrogen cycle and soil greenhouse gas emissions. **Agriculture, Ecosystems & Environment**, 2012, 148: 37-43.
- CARVALHO, J. L. N. Dinâmica do carbono e fluxo de gases do efeito estufa em sistemas de integração lavoura-pecuária na Amazônia e no Cerrado. 2010. **Tese** (Doutorado em Solos e Nutrição de Plantas) - Escola Superior de Agricultura Luiz de Queiroz, University of São Paulo, Piracicaba, 2010. doi:10.11606/T.11.2010.tde-19042010-164213. Acesso em: 2023-12-05.
- CEZÁRIO, A. S. *et al.* Silages of *Brachiaria brizantha* cv. Marandu harvested at two regrowth ages: Microbial inoculant responses in silage fermentation, ruminant digestion and beef cattle performance. **Animal Feed Science and Technology**, v. 208, p. 33–43, 2015.
- CHANG, C. C.; LIN, C. J. LIBSVM: a library for support vector machines. **ACM transactions on intelligent systems and technology (TIST)**, 2.3: 1-27, 2011.
- CHEN, H.; ZHU, Q. A.; PENG, C.; WU, N.; WANG, Y.; FANG, X. *et al.* Methane emissions from rice paddies natural wetlands, lakes in China: synthesis new estimate. **Global change biology**, 19(1), 19-32, 2013.

- CHERUBIN, M. R.; CARVALHO, J. L. N.; CERRI, C. E. P.; NOGUEIRA, L. A. H.; SOUZA, G. M.; CANTARELLA, H. Land use and management effects on sustainable sugarcane-derived bioenergy. **Land**, 10(1), 72, 2021.
- CHERUBIN, M. R.; KARLEN, D. L.; FRANCO, A. L.; TORMENA, C. A.; CERRI, C. E.; DAVIES, C. A.; CERRI, C. C. Soil physical quality response to sugarcane expansion in Brazil. **Geoderma**, 267, 156-168, 2016.
- CHRISTIANSEN, J. R. *et al.* Assessing the effects of chamber placement, manual sampling and headspace mixing on CH<sub>4</sub> fluxes in a laboratory experiment. **Plant and soil**, 343(1-2), 171-185, 2011.
- CIAIS, P. *et al.* Carbon and Other Biogeochemical Cycles. In: STOCKER, T. F. *et al.* editors. Climate Change 2013 the Physical Science Basis: Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. New York, NY: **Cambridge University Press**, 465–570, 2013.
- CLOUGH, T. J.; ROCHETTE, P.; THOMAS, S. M.; PIHLATIE, M.; CHRISTIANSEN, J. R.; THORMAN, R. E. Global Research Alliance N<sub>2</sub>O chamber methodology guidelines: Design considerations. **Journal of Environmental Quality**, 49(5), 1081-1091, 2020.
- COFFEY, N. *et al.* Common functional principal components analysis: A new approach to analyzing human movement data. **Human movement science**, 30(6), 1144-1166, 2011.
- COLLIER, S. M.; RUARK, M. D.; OATES, L. G.; JOKELA, W. E.; DELL, C. J. Measurement of greenhouse gas flux from agricultural soils using static chambers. **JoVE (Journal of Visualized Experiments)**, (90), e52110, 2014.
- CONAB - COMPANHIA NACIONAL DE ABASTECIMENTO. Acompanhamento da safra brasileira de cana-de-açúcar, Brasília, DF, v. 11, n. 3, novembro 2023.
- CONRAD, R. The global methane cycle: recent advances in understanding the microbial processes involved. **Environ Microbiol Rep.**, 1:285–92, 2009.
- COSKUN, D.; BRITTO, D. T.; SHI, W.; KRONZUCKER, H. J. Nitrogen transformations in modern agriculture and the role of biological nitrification inhibition. **Nature Plants**, 3(6), 1-10, 2017.
- CRAVEN, P.; WAHBA, G. Smoothing noisy data with spline functions. **Numerische mathematik**, Springer, v. 31, n. 4, p. 377–403, 1979.
- CURCEAC, S. *et al.* Short-term air temperature forecasting using nonparametric functional data analysis and SARMA models. **Environmental Modelling & Software**, 111, 394-408, 2019.
- DA COSTA, L. M., *et al.* Spatiotemporal variability of atmospheric CO<sub>2</sub> concentration and controlling factors over sugarcane cultivation areas in southern Brazil. **Environment, Development and Sustainability**, 24(4), 5694-5717, 2022.
- DA SILVA, D. A. P.; CAMPOS, M. C. C.; MANTOVANELLI, B. C.; DOS SANTOS, L. A. C.; SOARES, M. D. R.; DA CUNHA, J. M. Variabilidade espacial da emissão de CO<sub>2</sub>, temperatura e umidade do solo em área de pastagem na região Amazônica, Brasil. **Revista de Ciências Agroveterinárias**, 18(1), 119-126, 2019.
- DAI, W.; MRKVIČKA, T.; SUN, Y.; GENTON, M. G. Functional outlier detection and taxonomy by sequential transformations. **Computational Statistics & Data Analysis**, 149, 106960, 2020.

- DALCIN, G.; MANGINI, L. F. K.; DE GODOI, L. Química Ambiental e o impacto ambiental relacionado aos gases de Efeito Estufa. **Caderno Intersaberes**, 10.27: 117-134, 2021.
- DAVIDSON, E. A.; KANTER, D. Inventories and scenarios of nitrous oxide emissions. **Environmental Research Letters**, 9(10), 105012, 2014.
- DAVIDSON, E. *et al.* Minimizing artifacts and biases in chamber-based measurements of soil respiration. **Agricultural and Forest Meteorology**, Elsevier, v. 113, n. 1-4, p. 21–37, 2002.
- DEBOUK, H.; ALTIMIR, N.; SEBASTIÀ, M. T. Maximizing the information obtained from chamber-based greenhouse gas exchange measurements in remote areas. **MethodsX**, 5: 973-983, 2018.
- DENMEAD, O. T. Approaches to measuring fluxes of methane and nitrous oxide between landscapes and the atmosphere. **Plant and Soil**, 309(1-2), 5-24, 2008.
- DI SALVO, F.; RUGGIERI, M.; PLAIA, A. Functional principal component analysis for multivariate multidimensional environmental data. **Environmental and ecological statistics**, 22(4), 739-757, 2015.
- DIAS, N. M. S. **Emission of greenhouse gases in the land use change for sugarcane production in the Center-South region of Brazil, 2018**. Doctoral Thesis. Programa de Pós Graduação em Ciências. Centro de Energia Nuclear na Agricultura. Universidade de São Paulo – USP. Piracicaba, 2018-02-28.
- DIAS-FILHO, M. B. Diagnóstico das pastagens no Brasil. Belém: **Embrapa Amazônia Oriental**, 2014.
- DONA, G., *et al.* Application of functional principal component analysis in race walking: an emerging methodology. **Sports Biomechanics**, 8(4), 284-301, 2009.
- DOWHOWER, S. L. *et al.* Soil greenhouse gas emissions as impacted by soil moisture and temperature under continuous and holistic planned grazing in native tallgrass prairie. **Agriculture, Ecosystems & Environment**, 287, 106647, 2020.
- EMBRAPA. Sistema Brasileiro de Classificação de Solos 5. ed., rev. e ampl. – Brasília, DF : Embrapa, 2018.
- EUCLIDES, V. P. B. *et al.* Pasture and grazing management of *Brachiaria brizantha* (Hochst) Stapf and *Panicum maximum* Jacq. (In Portuguese, with English abstract.). **Revista Ceres**, v. 61, p. 808–818, 2014.
- FÁVERO, L. P. L. *et al.* **Análise de dados: modelagem multivariada para tomada de decisões**. Rio de Janeiro: Elsevier, 2009. Acesso em: 07 jun. 2023.
- FAISAL, M.; ZAMZAMI, E. M. Comparative analysis of inter-centroid K-Means performance using euclidean distance, canberra distance and manhattan distance. In **Journal of Physics: Conference Series** (Vol. 1566, No. 1, p. 012112). IOP Publishing, 2020, June.
- FORTUNA, F.; MATURO, F. K-means clustering of item characteristic curves and item information curves via functional principal component analysis. **Quality & Quantity**, 53(5), 2291-2304, 2019.
- FOWLER, D. *et al.* Effects of global change during the 21st century on the nitrogen cycle. **Atmos. Chem. Phys**, 15, 13849–13893, 2015.
- FRAZÃO, L. A. Conversão do Cerrado em pastagem e sistemas agrícolas: efeitos na dinâmica da matéria orgânica do solo. 2007. **Dissertação** (Mestrado em Química na Agricultura e no

Ambiente) - Centro de Energia Nuclear na Agricultura, Universidade de São Paulo, Piracicaba, 2007. doi:10.11606/D.64.2007.tde-18092007-113334. Acesso em: 2023-12-05.

FREITAS, I. B. F. *et al.* Soil management effects of extensive pastures, intensive pastures and sugarcane crops on the availability of metals and nutrients in freshwater: A realistic mesocosm approach. **Agriculture, Ecosystems & Environment**, 350, 108473, 2023.

FREITAS, I. B. F. Efeitos do manejo do solo em áreas de pastagem e cultivo de cana-de-açúcar sobre ecossistemas aquáticos continentais tropicais: uma abordagem realística em mesocosmos. 2023. Tese (Doutorado em Ciências da Engenharia Ambiental) - Escola de Engenharia de São Carlos, University of São Paulo, São Carlos, 2023. doi:10.11606/T.18.2023.tde-16112023-100200. Acesso em: 2023-12-11.

FUNK, R., *et al.* From potential to implementation: an innovation framework to realize the benefits of soil carbon. In **Soil carbon: Science, management and policy for multiple benefits**, pp. 47-59, 2015. Wallingford UK: CABI.

GALDINO, T. L. G.; SIGNOR, D.; DE MORAES, S. A. Modification of closed static chambers for collection of greenhouse gases emitted by soil. **International Journal of Environmental Science and Technology**, 1-10, 2023.

GAO, H. O. Day of week effects on diurnal ozone/NO<sub>x</sub> cycles and transportation emissions in Southern California. **Transportation Research Part D: Transport and Environment**, 12(4), 292-305, 2007.

GAO, H. O.; NIEMEIER, D. A. Using functional data analysis of diurnal ozone and NO<sub>x</sub> cycles to inform transportation emissions control. **Transportation Research Part D: Transport and Environment**, 13(4), 221-238, 2008.

GARCIA-MONTIEL, D. C. *et al.* Controls on soil nitrogen oxide emissions from forest and pastures in the Brazilian Amazon. **Global Biogeochemical Cycles**, Wiley Online Library, v. 15, n. 4, p. 1021–1030, 2001.

GAROFALO, D. F. T., NOVAES, R. M. L., PAZIANOTTO, R. A., MACIEL, V. G., BRANDÃO, M., SHIMBO, J. Z., FOLEGATTI-MATSUURA, M. I. Land-use change CO<sub>2</sub> emissions associated with agricultural products at municipal level in Brazil. **Journal of Cleaner Production**, 364, 132549, 2022.

GEBERT, J.; GROENGROEFT, A.; PFEIFFER, E. M. Relevance of soil physical properties for the microbial oxidation of methane in landfill covers. **Soil Biology and Biochemistry**, 43(9), 1759-1767, 2011.

GÉRON, A. Mãos à obra: aprendizado de máquina com Scikit-Learn. Keras & TensorFlow: Conceitos, ferramentas e técnicas para a construção de sistemas inteligentes. Rio de Janeiro: Alta Books, 2ª edição, 640p., 2021.

GIROTTO, L.; FREITAS, I.B.F.; YOSHII, M.P.C. *et al.* Using mesocosms to evaluate the impacts of pasture intensification and pasture-sugarcane conversion on tadpoles in Brazil. **Environ Sci Pollut Res** 30, 21010–21024, 2023.

GONG, Y.; LI, P.; LU, W.; NISHIWAKI, J.; KOMATSUZAKI, M. Response of soil carbon dioxide emissions to no-tillage and moldboard plow systems on Andosols in a humid, subtropical climate, Japan. **Geoderma**, 386, 114920, 2021.

GRACINDO, C. V.; LOUVANDINI, H.; RIET-CORREA, F.; BARBOSA-FERREIRA, M.; DE CASTRO, M. B. Performance of sheep grazing in pastures of *Brachiaria decumbens*, *Brachiaria brizantha*, *Panicum maximum*, and *Andropogon gayanus* with different protodioscin concentrations. **Tropical Animal Health and Production**, v. 46, n. 5, p. 733–737, 2014.

- HAIR, J. F. *et al.* Análise multivariada de dados. **Bookman editora**, 2009.
- HAPP, C.; GREVEN, S. Multivariate functional principal component analysis for data observed on different (dimensional) domains. **Journal of the American Statistical Association**, 113(522), 649-659, 2018.
- HUNTINGFORD, C. *et al.* Implications of improved representations of plant respiration in a changing climate. **Nature Communications**, 8.1: 1602, 2017.
- HUTCHINSON, G. L.; MOSIER, A. R. Improved soil cover method for field measurement of nitrous oxide fluxes. **Soil Science Society of America**, 1981, 45.2: 311-316, 1981.
- HYNDMAN, R. J.; ULLAH, M. S. Robust forecasting of mortality and fertility rates: a functional data approach. **Computational Statistics & Data Analysis**, 51(10), 4942-4956, 2007.
- IPCC. 2014. Climate change 2014: synthesis report. In: Core Writing Team Pachauri RK, Meyer LA, editors. Contribution of working group I, II and III to the fifth assessment report of the intergovernmental panel on climate change. Geneva (Switzerland): **IPCC**; p. 151.
- JACQUES, J.; PREDA, C. Functional data clustering: a survey. *Adv Data Anal Classif* 8, 231–255, 2014a.
- JACQUES, J.; PREDA, C. Model-based clustering for multivariate functional data. **Computational Statistics & Data Analysis**, 71: 92-106, 2014b.
- JANSEN, M. Generalized cross validation in variable selection with and without shrinkage. **Journal of statistical planning and inference**, 159, 90-104, 2015.
- JIANG, C. *et al.* CO<sub>2</sub> flux estimation by different regression methods from an alpine meadow on the Qinghai-Tibetan Plateau. **Adv. Atmos. Sci.** 27, 1372–1379, 2010.
- KAMRAN, M. *et al.* Assessment of greenhouse gases emissions, global warming potential and net ecosystem economic benefits from wheat field with reduced irrigation and nitrogen management in an arid region of China. **Agriculture, Ecosystems & Environment**, 341: 108197, 2023.
- KARA, L. Z. *et al.* Data-driven kNN estimation in nonparametric functional data analysis. **Journal of Multivariate Analysis**, 153, 176-188, 2017.
- KELLER, M.; KAPLAN, W. A.; WOFSEY, S. C. Emissions of N<sub>2</sub>O, CH<sub>4</sub> and CO<sub>2</sub> from tropical forest soils. *Journal of Geophysical Research: Atmospheres*, 91(D11), 11791-11802, 1986.
- KRAM, T. *et al.* Global and regional greenhouse gas emissions scenarios. **Technological Forecasting and Social Change**, Elsevier, v. 63, n. 2-3, p. 335–371, 2000.
- LAKSHANI, M. M. T. *et al.* Effects of Soil Moisture on Simulated Methane Flow Under Varying Levels of Soil Compaction. In: 12th International Conference on Structural Engineering and Construction Management: Proceedings of the ICSECM 2021. **Singapore: Springer Nature Singapore**, p. 271-282, 2022.
- LAMBERT, M.; FRÉCHETTE, J. L. Analytical techniques for measuring fluxes of CO<sub>2</sub> and CH<sub>4</sub> from hydroelectric reservoirs and natural water bodies. **Greenhouse gas emissions—fluxes and processes: hydroelectric reservoirs and natural environments**, 37-60, 2005.

- LASHERAS, F. S. *et al.* Detection of outliers in pollutant emissions from the Soto de Ribera coal-fired plant using Functional Data Analysis: A case study in northern Spain. **In Proceedings**, Vol. 2, No. 23, p. 1473, 2018, November.
- LEUNG, D. Y. C.; CARAMANNA, G.; MAROTO-VALER, M. M. An overview of current status of carbon dioxide capture and storage technologies. **Renewable and Sustainable Energy Reviews**, 39: 426-443, 2014.
- LI, Z.; DING, Q.; ZHANG, W. A comparative study of different distances for similarity estimation. In *Intelligent Computing and Information Science: International Conference, ICICIS 2011, Chongqing, China, January 8-9, 2011. Proceedings, Part I* (pp. 483-488), 2011. Springer Berlin Heidelberg.
- LIN, L. I. K. A concordance correlation coefficient to evaluate reproducibility. **Biometrics**, 255-268, 1989.
- LOPES, I. M. *et al.* Emissões de N<sub>2</sub>O em Solos sob Cultivo de Cana-de-Açúcar no bioma Mata Atlântica: Efeito dos Sistemas de Colheita e da Adubação com Vinhaça. **Revista Virtual de Química**, 2017.
- LYU, Z. *et al.* Methanogenesis. **Current Biology**, 28.13: R727-R732, 2018.
- MAPBIOMAS. Área de agropecuária no Brasil cresceu 50% nos últimos 38 anos. 06 de outubro de 2023. Disponível em: <https://brasil.mapbiomas.org/2023/10/06/area-de-agropecuaria-no-brasil-cresceu-50-nos-ultimos-38-anos/>. Acesso em: 11 de dezembro de 2023.
- MARTINS, M. R. *et al.* Nitrous oxide and ammonia emissions from N fertilization of maize crop under no-till in a Cerrado soil. **Soil and Tillage Research**, 151, 75-81, 2015.
- MAS, A.; PUMO, B. Functional linear regression with derivatives. **Journal of Nonparametric Statistics**, 21(1), 19-40, 2009.
- MATTHIAS, A. D.; BLACKMER, A. M.; BREMNER, J. M. A simple chamber technique for field measurement of emissions of nitrous oxide from soils. **American Society of Agronomy, Crop Science Society of America, and Soil Science Society of America**. Vol. 9, No. 2, pp. 251-256, 1980.
- MCBRIDE, G. B. A proposal for strength-of-agreement criteria for Lin's concordance correlation coefficient. **NIWA client report: HAM2005-062**, v. 45, p. 307-310, 2005.
- MEENA, V. S. *et al.* Land use changes: Strategies to improve soil carbon and nitrogen storage pattern in the mid-Himalaya ecosystem, India. **Geoderma**, 321, 69-78, 2018.
- MEURER, K. HE, *et al.* Direct nitrous oxide (N<sub>2</sub>O) fluxes from soils under different land use in Brazil—a critical review. **Environmental Research Letters**, 11.2: 023001, 2016.
- MOSIER, A., *et al.* Methane and nitrous oxide fluxes in native, fertilized and cultivated grasslands. **Nature**, 350(6316), 330-332, 1991.
- MOSIER, A.; WASSMANN, R.; VERCHOT, L.; KING, J.; PALM, C. Methane and nitrogen oxide fluxes in tropical agricultural soils: sources, sinks and mechanisms. **Environment, Development and Sustainability**, 6, 11-49, 2004.
- MUNISWAMAIAH, M.; AGERWALA, T.; TAPPERT, C. C. Applications of Binary Similarity and Distance Measures. **arXiv preprint arXiv:2307.00411**, 2023.
- NGO, D. *et al.* An exploratory data analysis of electroencephalograms using the functional boxplots approach. **Frontiers in neuroscience**, 9, 282, 2015.

- OERTEL, C., *et al.* Greenhouse gas emissions from soils—A review. **Geochemistry**, 76(3), 327-352, 2016.
- OLIVEIRA, D. M., *et al.* Is the expansion of sugarcane over pasturelands a sustainable strategy for Brazil's bioenergy industry?. **Renewable and Sustainable Energy Reviews**, 102, 346-355, 2019.
- ORMONEIT, D., *et al.* Representing cyclic human motion using functional analysis. **Image and Vision Computing**, 23(14), 1264-1276, 2005.
- PANDIT, V.; SCHULLER, B. The many-to-many mapping between the concordance correlation coefficient and the mean square error. **arXiv preprint arXiv:1902.05180**, 2019.
- PARKIN, T. B.; VENTEREA, R. T. USDA-ARS GRACEnet project protocols, chapter 3. Chamber-based trace gas flux measurements. **Sampling protocols**. Beltsville, MD p 1-39, 2010.
- PARKIN, T. B.; VENTEREA, R. T.; HARGREAVES, S. K. Calculating the detection limits of chamber-based soil greenhouse gas flux measurements. **Journal of environmental quality**, 41(3), 705-715, 2012.
- PAUTHENET, E. *et al.* A linear decomposition of the Southern Ocean thermohaline structure. **Journal of Physical Oceanography**, 47(1), 29-47, 2017.
- PAUTHENET, E. *et al.* Seasonal meandering of the Polar Front upstream of the Kerguelen Plateau. **Geophysical Research Letters**, 45(18), 9774-9781, 2018.
- PAUTHENET, E., *et al.* The thermohaline modes of the global ocean. **Journal of Physical Oceanography**, 49.10: 2535-2552, 2019.
- PEDERSEN, A. R.; PETERSEN, S. O.; SCHELDE, K. A comprehensive approach to soil-atmosphere trace-gas flux estimation with static chambers. **European Journal of Soil Science**, 61(6), 888-902, 2010.
- PIRK, N. *et al.* Calculations of automatic chamber flux measurements of methane and carbon dioxide using short time series of concentrations. **Biogeosciences**, 13(4), 903, 2016.
- PITOMBO, L.M.; DO CARMO, J.B.; DE HOLLANDER, M.; ROSSETTO, R.; LOPEZ, M.V.; CANTARELLA, H.; KURAMAE, E.E. Exploring soil microbial 16S rRNA sequence data to increase carbon yield and nitrogen efficiency of a bioenergy crop. **GCB Bioenergy** 8, 867–879, 2016.
- R Core Team. R: A language and environment for statistical computing. **R Foundation for Statistical Computing**, Vienna, Austria, 2022. URL <https://www.R-project.org/>.
- RABOT, E.; WIESMEIER, M.; SCHLÜTER, S.; VOGEL, H. J. Soil structure as an indicator of soil functions: A review. **Geoderma**, 314, 122-137, 2018.
- RAMSAY, J. O. Functional components of variation in handwriting. **Journal of the American Statistical Association**, 95(449), 9-15, 2000.
- RAMSAY, J. O.; BOCK, R. D.; GASSER, T. Comparison of height acceleration curves in the Fels, Zurich, and Berkeley growth data. **Annals of Human Biology**, 22.5: 413-426, 1995.
- RAMSAY, J. O.; DALZELL, C. Some tools for functional data analysis. **Journal of the Royal Statistical Society: Series B (Methodological)**, Wiley Online Library, v. 53,1991.
- RAMSAY, J. O.; GRAVES, S.; HOOKER, G. fda: Functional Data Analysis. **R package version 6.0.5**, 2022. <https://CRAN.R-project.org/package=fda>.

- RAMSAY, J. O.; HOOKER, G.; GRAVES, S. Functional Data Analysis with R and MATLAB. **Springer: New York, NY, USA**, 1-207, 2009.
- RAMSAY, J. O.; SILVERMAN, B. W. Functional data analysis. **New York: Springer-Verlag**, 1997.
- RAMSAY, J. O.; SILVERMAN, B. W. Principal components analysis for functional data. **Functional data analysis**, 147-172, 2005.
- RAMSAY, J. O.; SILVERMAN, B. W. Applied functional data analysis: methods and case studies. **Springer**, 2002.
- RHODEN, I.; WELLER, D.; VOIT, A-K. Spatio-Temporal Dynamics of European Innovation—An Exploratory Approach via Multivariate Functional Data Cluster Analysis. **Journal of Open Innovation: Technology, Market, and Complexity**. 8(1):6, 2022.
- ROCHETTE, P.; ERIKSEN-HAMEL, N. S. Chamber Measurements of Soil Nitrous Oxide Flux: Are Absolute Values Reliable. **Soil Science Society of America Journal**. 72 (2), 331, 2008.
- ROLAND, F. gasfluxes: Greenhouse Gas Flux Calculation from Chamber Measurements. **R package version 0.4-4, 2020**. <https://CRAN.R-project.org/package=gasfluxes>.
- SALAZAR, J. J.; GARLAND, L.; OCHOA, J.; PYRCZ, M. J. Fair train-test split in machine learning: Mitigating spatial autocorrelation for improved prediction accuracy. **Journal of Petroleum Science and Engineering**, 209, 109885, 2022.
- SÁNCHEZ-SÁNCHEZ, M. L. *et al.* Functional principal component analysis as a new methodology for the analysis of the impact of two rehabilitation protocols in functional recovery after stroke. **Journal of NeuroEngineering and Rehabilitation**, 11(1), 1-9.J.R, 2014.
- SANTANA, L. V. R.; DA SILVA, A. S. A. Análise de agrupamento da velocidade do vento no Nordeste do Brasil. **Sigmae**, 8(2), 481-489, 2019.
- SCHMUTZ, A. *et al.* Clustering multivariate functional data in group-specific functional subspaces. **Computational Statistics**, 35.3: 1101-1131, 2020.
- SCHUMAKER, L. Spline functions: Basic theory, John Wiley and Sons. **Inc., New York**, 1981.
- SHANG, H. L. A survey of functional principal component analysis. **ASTA Advances in Statistical Analysis**, 98(2), 121-142, 2014.
- SHANG, H. L. *et al.* A multilevel functional data method for forecasting population, with an application to the United Kingdom. **International Journal of Forecasting**, 32(3), 629-649, 2016.
- SHANG, H. L.; KEARNEY, F. Dynamic functional time-series forecasts of foreign exchange implied volatility surfaces. **International Journal of Forecasting**, 38(3), 1025-1049, 2022.
- SILVA, A. J. N.; RIBEIRO, M. R.; CARVALHO, F. G.; SILVA, V. N.; SILVA, L. E. S. F. Impact of sugarcane cultivation on soil carbon fractions, consistence limits and aggregate stability of a Yellow Latosol in Northeast Brazil. **Soil and Tillage Research**, 94(2), 420-424, 2007.
- SILVA, F. S. R. Predição de Cr, Cu, Hg, Ni, Pb, Zn em solos de manguezal poluído no nordeste do Brasil por meio da espectroscopia de reflectância no infravermelho próximo. 2022. 46 f. **Dissertação** (Programa de Pós-Graduação em Ciência do Solo) - Universidade Federal Rural de Pernambuco, Recife.

- SILVA, J. P. *et al.* Biases in greenhouse gases static chambers measurements in stabilization ponds: Comparison of flux estimation using linear and non-linear models. **Atmospheric environment**, 109, 130-138, 2015.
- SIQUEIRA-NETO, M. Estoque de carbono e nitrogênio do solo com diferentes usos no Cerrado em Rio Verde (GO). 2006. **Tese** (Doutorado em Energia Nuclear na Agricultura) - Centro de Energia Nuclear na Agricultura, University of São Paulo, Piracicaba, 2006. doi:10.11606/T.64.2006.tde-11042007-113740. Acesso em: 2023-12-05.
- SIQUEIRA-NETO, M. *et al.* Emissão de gases do efeito estufa em diferentes usos da terra no bioma cerrado. **Revista Brasileira de Ciência do Solo**, SciELO Brasil, v. 35, n. 1, p. 63–76, 2011.
- SIQUEIRA-NETO, M. *et al.* Impacts of land use and cropland management on soil organic matter and greenhouse gas emissions in the Brazilian Cerrado. **European Journal of Soil Science**, 72(3), 1431–1446, 2020.
- 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. **European journal of soil science**, 54(4), 779-791, 2003.
- STEICHEN, T. J.; COX, N. J. A note on the concordance correlation coefficient. **The Stata Journal**, 2.2: 183-189, 2002.
- STEVENSON, M. *et al.* epiR: Tools for the Analysis of Epidemiological Data. **R package version 2.0.53.**, 2022. <https://CRAN.R-project.org/package=epiR>.
- SUN, Y; GENTON, M. G. Functional boxplots. **Journal of Computational and Graphical Statistics**, 20 (2), 316-334, 2011.
- SUN, Y.; GENTON, M. G.; NYCHKA, D. W. Exact fast computation of band depth for large functional datasets: How quickly can one million curves be ranked?. **Stat**, 1(1), 68-74, 2012.
- TÄUMER, J. *et al.* Linking transcriptional dynamics of CH<sub>4</sub>-cycling grassland soil microbiomes to seasonal gas fluxes. **ISME J** 16, 1788–1797, 2022.
- THORBURN, N. Researching Underage Sex Work: Dynamic Risk, Responding Sensitive, and Protecting Participants and Researchers. In: Liamputtong, P. (eds) **Handbook of Research Methods in Health Social Sciences**. Springer, Singapore, 2017. [https://doi.org/10.1007/978-981-10-2779-6\\_139-1](https://doi.org/10.1007/978-981-10-2779-6_139-1)AKOGLU, H. User's guide to correlation coefficients. **Turkish journal of emergency medicine**, 18(3), 91-93, 2018.
- THOMPSON, R. L., *et al.* Acceleration of global N<sub>2</sub>O emissions seen from two decades of atmospheric inversion. **Nature Climate Change**, 9.12: 993-998, 2019.
- TREVISAN, D. P.; MOSCHINI, L. E.; GUERRERO, J. V. R. Dinâmica Temporal do Uso e Cobertura da Terra no Município de Brotas-SP entre os Anos de 1988 e 2016. **Fronteiras: Journal of Social, Technological and Environmental Science**, 6(4), 204-219, 2017.
- TSAI, M. Y.; LIN, C. C. New model-averaged estimators of concordance correlation coefficients: simulation and application to longitudinal overdispersed Poisson data. **Communications in Statistics-Simulation and Computation**, 52.3: 961-979, 2023.
- ULLAH, S.; FINCH, C. F. Applications of functional data analysis: A systematic review. **BMC medical research methodology**, 13(1), 43, 2013.
- VALLEJOS, R. *et al.* A spatial concordance correlation coefficient with an application to image analysis. **Spatial Statistics**, 40, 100405, 2020.

VASCONCELOS, A. L. S. *et al.* Agricultura e gases de efeito estufa - estudos de casos no Brasil. **Revista Trópica: Ciências Agrárias E Biológicas**, 10(2), 2019.

VOORA, V.; BERMÚDEZ, S.; LE, H., LARREA, C.; LUNA, E. Sugar cane prices and sustainability. GLOBAL MARKET REPORT, september 2023. Disponível em: <https://www.iisd.org/system/files/2023-09/2023-global-market-report-sugar-cane.pdf>. Acesso em 12 de dezembro de 2023.

WALLING, E.; VANECKHAUTE, C. Greenhouse gas emissions from inorganic and organic fertilizer production and use: A review of emission factors and their variability. **Journal of Environmental Management**, 276, 111211, 2020.

WALSH, B. *et al.* Pathways for balancing CO<sub>2</sub> emissions and sinks. **Nature communications**, 8.1: 14856, 2017.

WANG, C. *et al.* Factors That Influence Nitrous Oxide Emissions from Agricultural Soils as Well as Their Representation in Simulation Models: A Review. **Agronomy**, 11(4):770, 2021.

WANG, J. L.; CHIOU, J. M.; MÜLLER, H. G. Functional data analysis. **Annual Review of Statistics and its application**, 3, 257-295, 2016.

WANG, N.; ZHU, X.; ZUO, Y.; LIU, J.; YUAN, F.; GUO, Z. *et al.* Microbial mechanisms for methane source-to-sink transition after wetland conversion to cropland. **Geoderma**, 429, 116229, 2023.

WARMENHOVEN, J. *et al.* Bivariate functional principal components analysis: considerations for use with multivariate movement signatures in sports biomechanics. **Sports Biomechanics**, 18(1), 10-27, 2019a.

WARMENHOVEN, J. *et al.* Considerations for the use of functional principal components analysis in sports biomechanics: examples from on-water rowing. **Sports Biomechanics**, 18(3), 317-341, 2019b.

XING, H.; WANG, E.; SMITH, C. J.; ROLSTON, D.; YU, Q. Modelling nitrous oxide and carbon dioxide emission from soil in an incubation experiment. **Geoderma**, 167, 328-339, 2011.

YERLI, C., SAHIN, U., OZTAS, T. CO<sub>2</sub> emission from soil in silage maize irrigated with wastewater under deficit irrigation in direct sowing practice. **Agricultural Water Management**, 271: 107791, 2022.

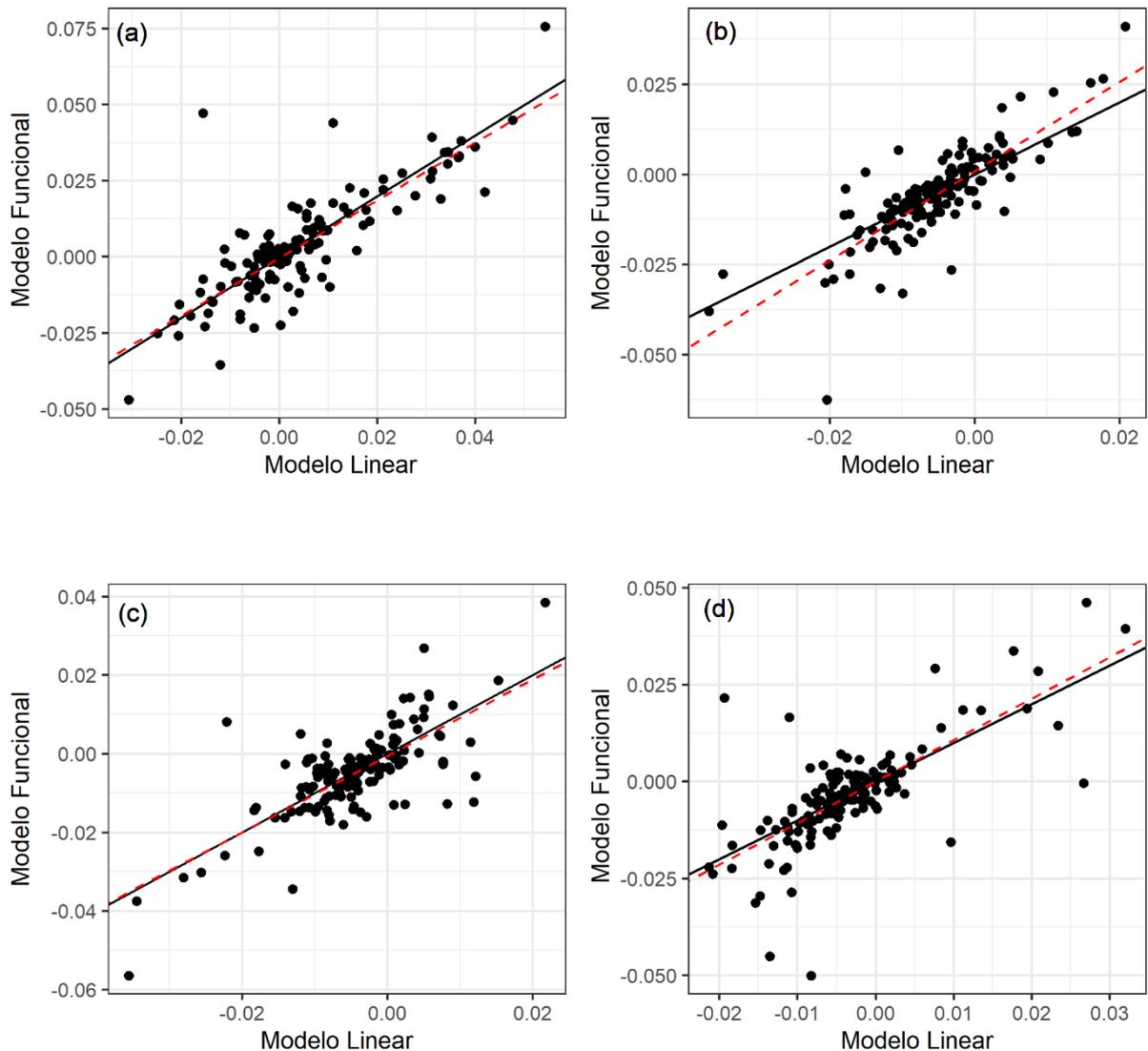
Rochette, Eriksen-Hamel (2008) ZANATTA, J. A.; ALVES, B. J. R.; BAYER, C.; TOMAZI, M.; FERNANDES, A. H. B. M.; COSTA, F. D. S.; CARVALHO, A. D. Protocolo para medição de fluxos de gases de efeito estufa do solo. **Embrapa Florestas**, 1º edição, 2014. ISSN 1980-3958.

ZHENG, J. *et al.* Deciphering nitrous oxide emissions from tropical soils of different land uses. **Science of The Total Environment**, 862: 160916, 2023.

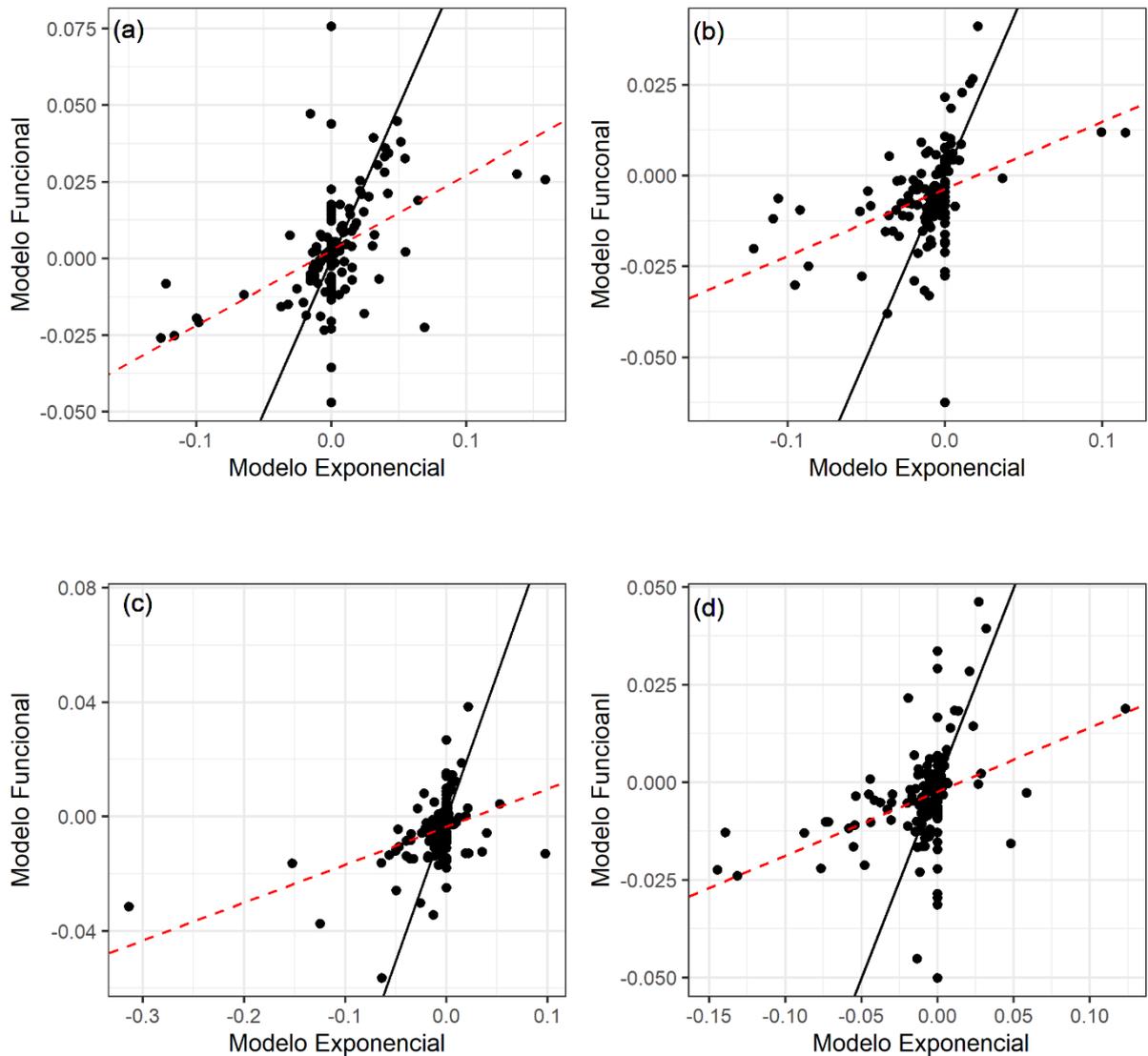
ZIA, M.; HANSEN, J.; HJORT, K.; VALDES, C. (2019). Brazil once again becomes the world's largest beef exporter. USDA ERS-Brazil Once Again Becomes the World's Largest Beef Exporter, 1. Disponível em: <https://www.ers.usda.gov/amber-waves/2019/july/brazil-once-again-becomes-the-world-s-largest-beef-exporter>. Acessado em 12 de dezembro de 2023.

ZOMER, R.J.; BOSSIO, D.A.; SOMMER, R. *et al.* Global Sequestration Potential of Increased Organic Carbon in Cropland Soils. **Sci Rep**, 7, 15554, 2017.

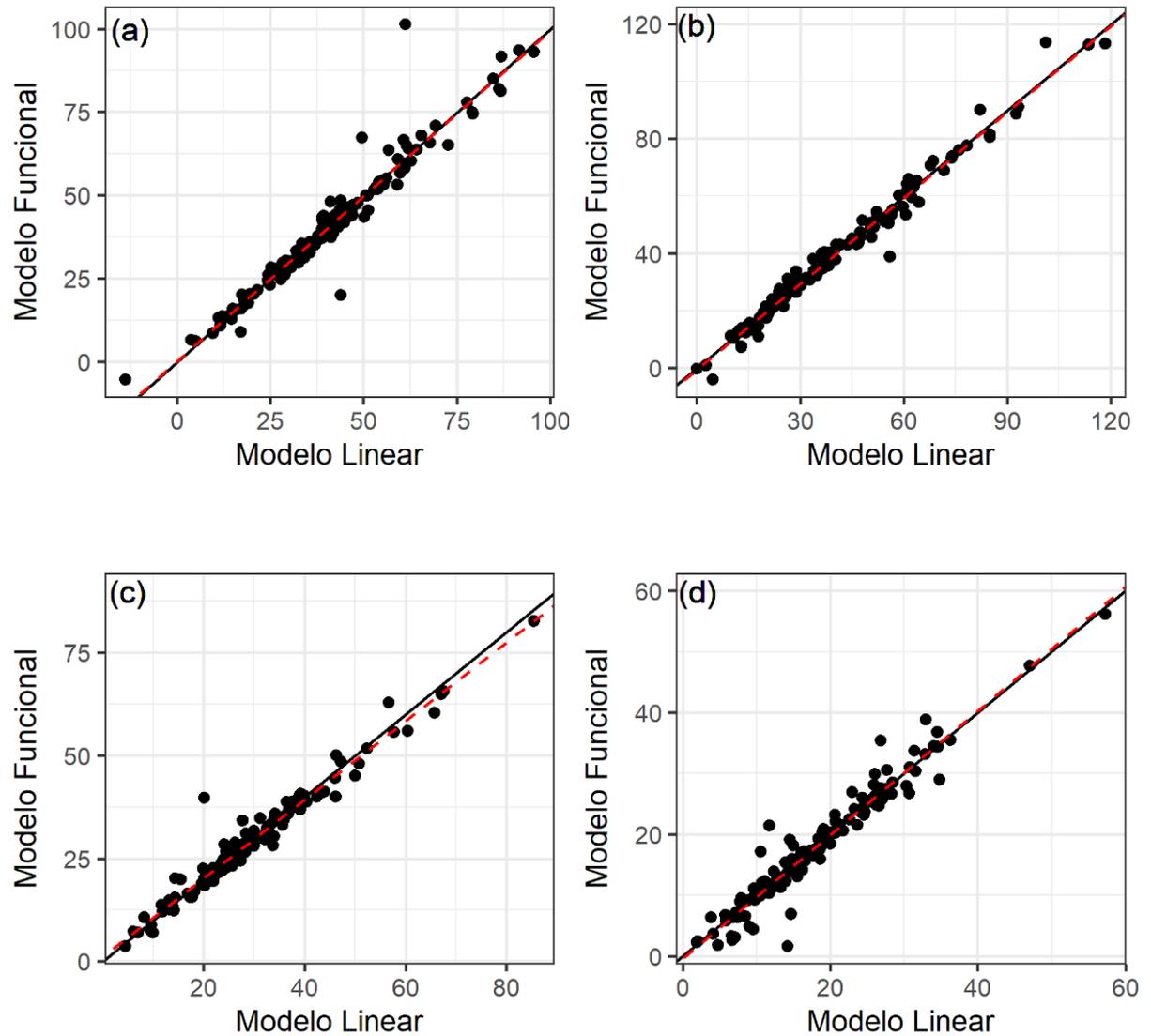
## Apêndice A



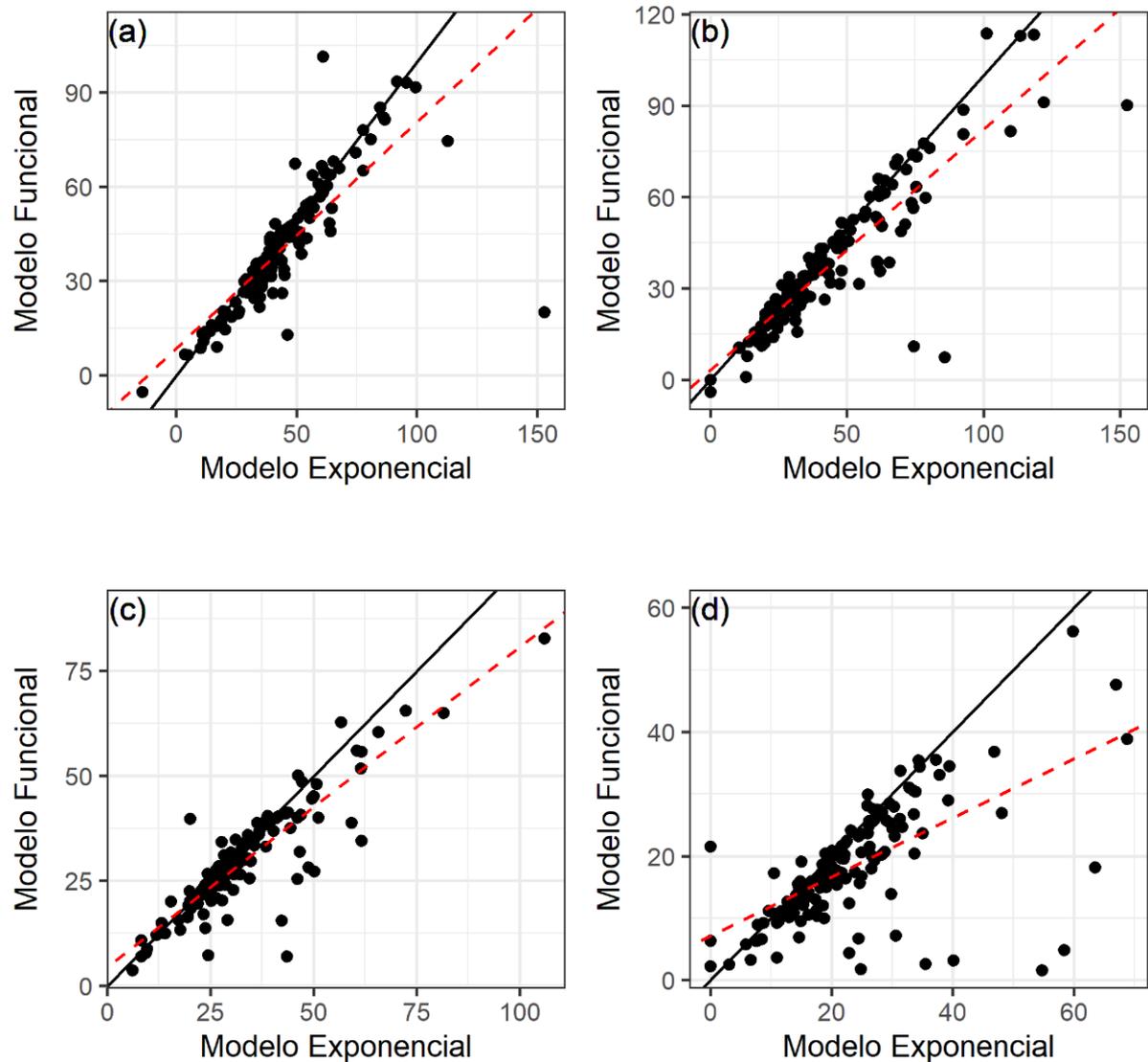
Apêndice A1 - Análise do fator de correlação de tendência, uma medida de quão longe uma linha de melhor ajuste está de um ângulo de  $45^\circ$  através da origem para a comparação do modelo linear vs. funcional do  $\text{CH}_4$ , em relação ao pasto intensivo (a), pasto intensivo (b) linha da cana-de-açúcar (c) e entrelinha da cana-de-açúcar (d). A linha tracejada vermelha é a linha de concordância e representa a concordância perfeita entre o modelo linear e o modelo funcional (declive = 1 e intercepto = 0). A linha preta representa a regressão linear entre os modelos. Fonte: o autor.



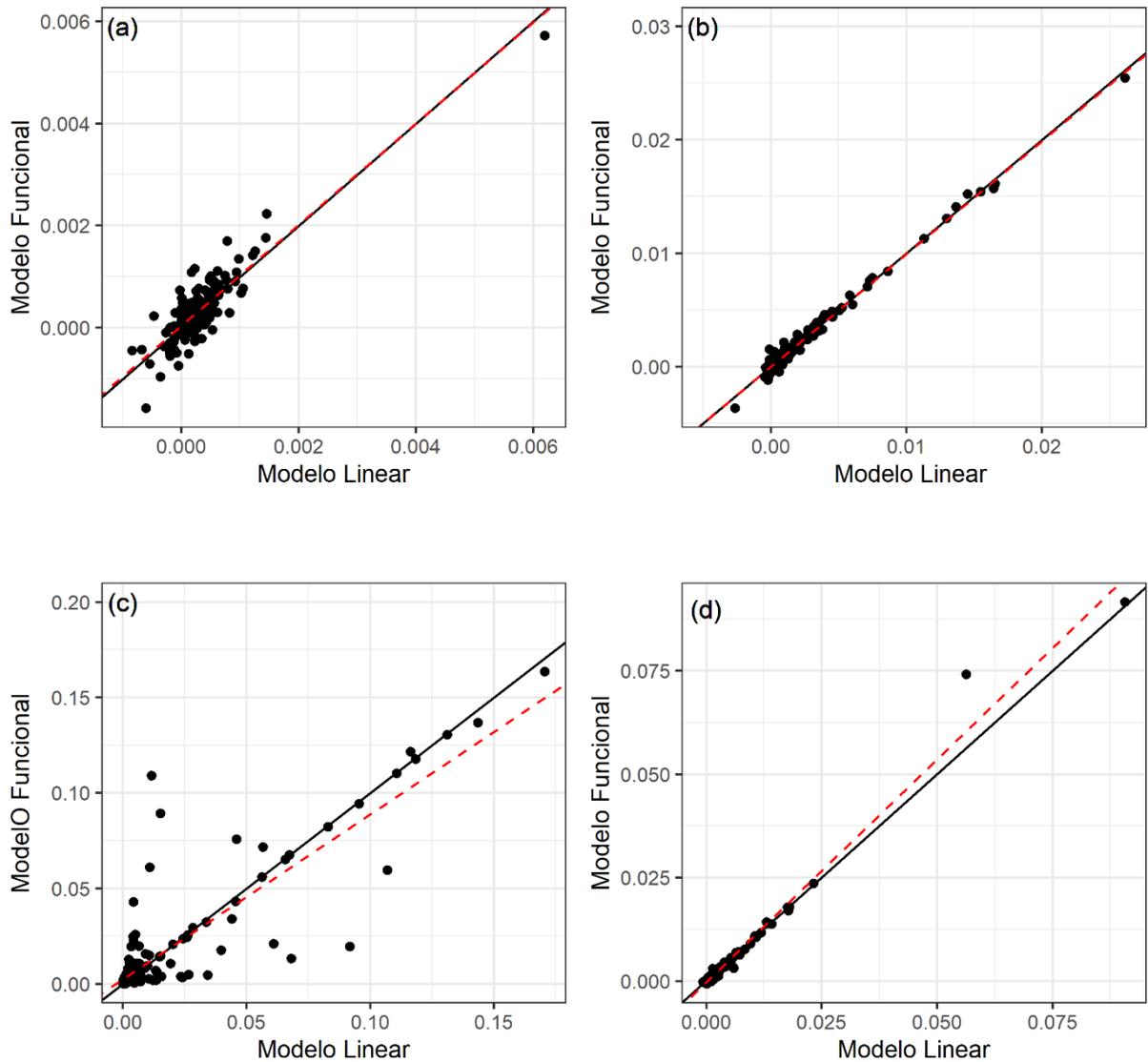
Apêndice A2 - Análise do fator de correlação de tendência, uma medida de quão longe uma linha de melhor ajuste está de um ângulo de  $45^\circ$  através da origem para a comparação do modelo exponencial vs. funcional do  $\text{CH}_4$ , em relação ao pasto intensivo (a), pasto intensivo (b) linha da cana-de-açúcar (c) e entrelinha da cana-de-açúcar (d). A linha tracejada vermelha é a linha de concordância e representa a concordância perfeita entre o modelo linear e o modelo funcional (declive = 1 e intercepto = 0). A linha preta representa a regressão linear entre os modelos. Fonte: o autor.



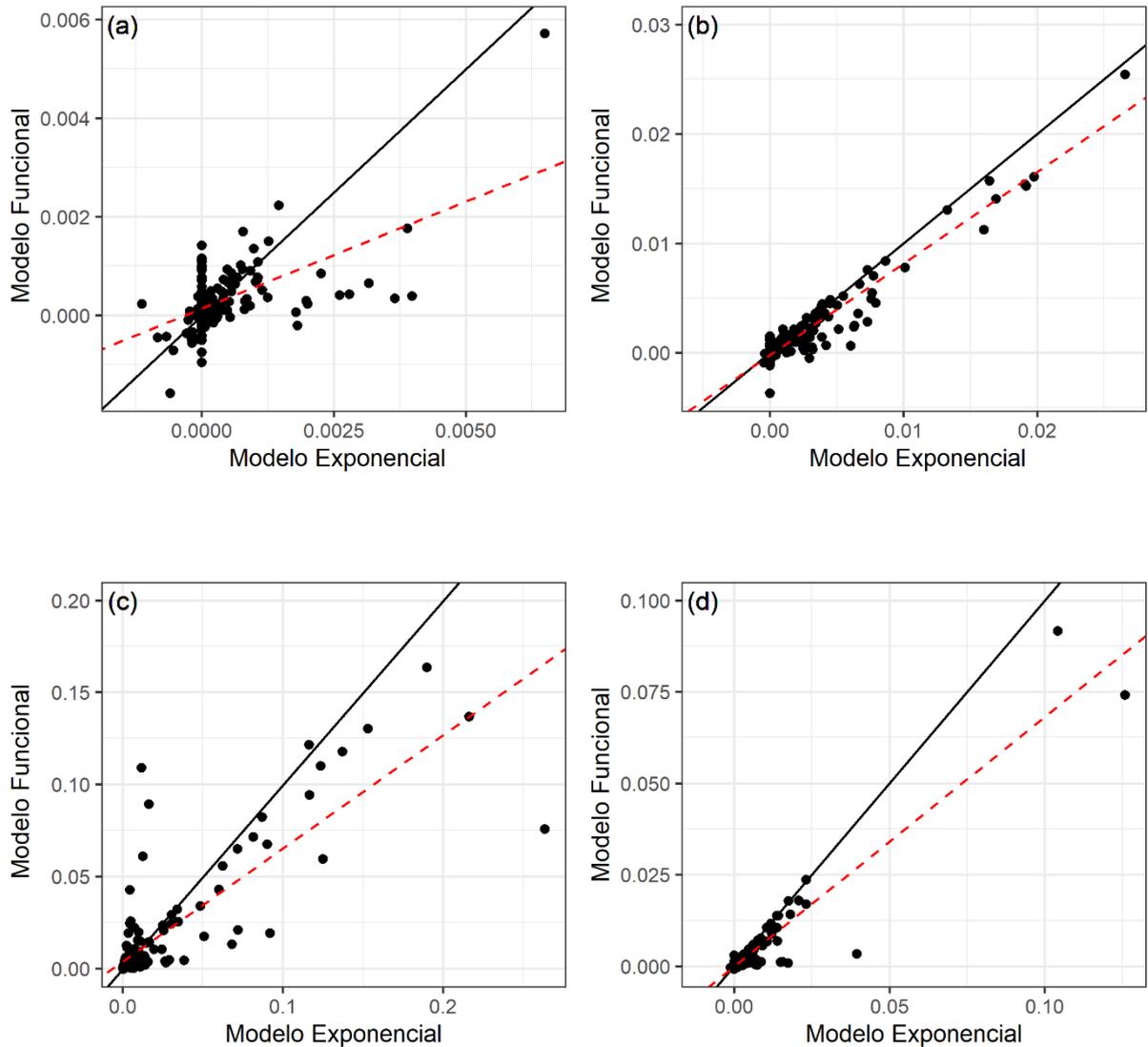
Apêndice A3 - Análise do fator de correlação de tendência, uma medida de quão longe uma linha de melhor ajuste está de um ângulo de  $45^\circ$  através da origem para a comparação do modelo linear vs. funcional do  $\text{CO}_2$ , em relação ao pasto intensivo (a), pasto intensivo (b) linha da cana-de-açúcar (c) e entrelinha da cana-de-açúcar (d). A linha tracejada vermelha é a linha de concordância perfeita entre o modelo linear e o modelo funcional (declive = 1 e intercepto = 0). A linha preta representa a regressão linear entre os modelos. Fonte: o autor.



Apêndice A4 - Análise do fator de correlação de tendência, uma medida de quão longe uma linha de melhor ajuste está de um ângulo de  $45^\circ$  através da origem para a comparação do modelo exponencial vs. funcional do  $\text{CO}_2$ , em relação ao pasto intensivo (a), pasto intensivo (b) linha da cana-de-açúcar (c) e entrelinha da cana-de-açúcar (d). A linha tracejada vermelha é a linha de concordância e representa a concordância perfeita entre o modelo linear e o modelo funcional (declive = 1 e intercepto = 0). A linha preta representa a regressão linear entre os modelos. Fonte: o autor.



Apêndice A5 - Análise do fator de correlação de tendência, uma medida de quão longe uma linha de melhor ajuste está de um ângulo de  $45^\circ$  através da origem para a comparação do modelo linear vs. funcional do  $N_2O$ , em relação ao pasto intensivo (a), pasto intensivo (b) linha da cana-de-açúcar (c) e entrelinha da cana-de-açúcar (d). A linha tracejada vermelha é a linha de concordância e representa a concordância perfeita entre o modelo linear e o modelo funcional (declive = 1 e intercepto = 0). A linha preta representa a regressão linear entre os modelos. Fonte: o autor.



Apêndice A6 - Análise do fator de correlação de tendência, uma medida de quão longe uma linha de melhor ajuste está de um ângulo de  $45^\circ$  através da origem para a comparação do modelo exponencial vs. funcional do  $N_2O$ , em relação ao pasto intensivo (a), pasto intensivo (b) linha da cana-de-açúcar (c) e entrelinha da cana-de-açúcar (d). A linha tracejada vermelha é a linha de concordância e representa a concordância perfeita entre o modelo linear e o modelo funcional (declive = 1 e intercepto = 0). A linha preta representa a regressão linear entre os modelos. Fonte: o autor.

## Apêndice B

Apêndice B - Parâmetros e distribuições consideradas no *RandomizedSearchCV*.

Parâmetros	Distribuição
Parâmetro C	expon(scale=1000)
Gamma	expon(scale=.1)
Kernel	[rbf, linear]
Class_weight	[Balanced, None]

Fonte: o autor.

## Apêndice C

Apêndice C - Valor dos intervalos de confiança (95%) a partir da distribuição t de Student em relação ao ponto de tempo que estabiliza o fluxo do N<sub>2</sub>O para os usos da terra. Valor de *p* significativo (*p* < 0.05).

Tempo (min)	Intervalo de confiança (95%)
27	$0,00134 < \mu < 0,00229$
28	$0,00107 < \mu < 0,00185$
29	$0,00079 < \mu < 0,00141$
30	$0,00049 < \mu < 0,00097$
31	$0,00019 < \mu < 0,00055$
32	$-0,00011 < \mu^* < 0,00021$

Fonte: o autor.

## Apêndice D

Apêndice D - Acurácia obtida na validação cruzada para as cinco rodadas durante o treinamento.

Cenários	Descrição do Cenário	Acurácia: Validação Cruzada				
		Rodada 1	Rodada 2	Rodada 3	Rodada 4	Rodada 5
1	PCs com 4 grupos.	0,57	0,50	0,60	0,54	0,60
2	PCs mais a Temperatura com 4 grupos.	0,61	0,65	0,64	0,73	0,64
3	PCs com 3 grupos.	0,69	0,68	0,68	0,69	0,76
4	PCs mais a Temperatura com 3 grupos.	0,83	0,83	0,71	0,80	0,72
5	PCs no período de 06/12/2018 a 25/01/2019 com 4 grupos.	0,70	0,70	0,55	0,62	0,67
6	PCs no período de 06/12/2018 a 25/01/2019 mais a temperatura com 4 grupos.	0,71	0,79	0,63	0,69	0,68
7	PCs no período de 06/12/2018 a 25/01/2019 com 3 grupos.	0,79	0,78	0,68	0,81	0,79
8	PCs no período de 06/12/2018 a 25/01/2019 mais a temperatura com 3 grupos.	0,79	0,86	0,79	0,84	0,79

Fonte: o autor.

## Anexo A

Anexo A - Caracterização do solo da área de estudo. Fonte: Freitas (2023).

<b>Parâmetros</b>	<b>Valores</b>	<b>Unidade</b>
Alumínio (Al)	3.3	mmolc dm <sup>-3</sup>
Boro (B)	0.2	mg dm <sup>-3</sup>
Cálcio (Ca)	11.7	mmolc dm <sup>-3</sup>
Cobre (Cu)	0.6	mg dm <sup>-3</sup>
Enxofre (S)	2.2	mg dm <sup>-3</sup>
Ferro (Fe)	69.0	mg dm <sup>-3</sup>
Fósforo (P)	4.2	mg dm <sup>-3</sup>
Magnésio (Mg)	4.3	mmolc dm <sup>-3</sup>
Manganês (Mn)	1.5	mg dm <sup>-3</sup>
Matéria orgânica	25	g dm <sup>-3</sup>
pH	4.3	-
Potássio (K)	1.3	mmolc dm <sup>-3</sup>
Zinco (Zn)	0.8	mg dm <sup>-3</sup>