

**Mickaelle Maria de Almeida Pereira**

**Análise de dados funcionais aplicada à modelagem de gases de efeito estufa**

**Recife**

**19 de fevereiro de 2020**



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

**Análise de dados funcionais aplicada à modelagem de gases de efeito estufa**

Dissertação julgada adequada para obtenção do título de Mestre em Biometria e Estatística Aplicada, defendida e aprovada por unanimidade em 19/02/2020 pela comissão examinadora.

**Área de concentração: Biometria e Estatística Aplicada**

**Orientador: Prof. Dr. Luiz Antônio Martinelli**

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

**Recife**

**19 de fevereiro de 2020**

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

**Análise de dados funcionais aplicada à modelagem de gases de efeito estufa**

Mickaelle Maria de Almeida Pereira

Dissertação julgada adequada para obtenção do título de Mestre em Biometria e Estatística Aplicada, defendida e aprovada por unanimidade em 19/02/2020 pela comissão examinadora.

Orientador:

---

**Prof. Dr. Luiz Antônio Martinelli**  
Orientador

Banca examinadora:

---

**Prof. Dr. Antonio Samuel Alves da  
Silva**

---

**Prof. Dr. Lucian Bogdan Bejan**

---

**Prof. Dr. José Ramon Barros  
Cantalice**

*Este trabalho é dedicado aos meus avós José Vicente e Maria de Lourdes e minha mãe  
Maria Marcia. E ao meu esposo Lucimario Gois.*

# Agradecimentos

Agradeço a Deus e a nossa senhora do Perpétuo Socorro.

A minha mãe Maria Marcia e aos meus avós Maria de Lourdes e José Vicente por todo amor, carinho e dedicação, ao meu esposo Lucimario Gois, as minhas tias Maria, Mariza, Mirian, Maria José, Maciana, aos meus sogros Luci Gois e Cláudio Mario, meus tios Pedro, Paulo, Maurício, Murilio, Zé Ricardo, Moreira, Moura e Sergio. Ao meu pai adotivo Adilson Melo e meus primos/irmãos Adones Melo, Maurício Junior e Bruna Costa. Aos meus afilhados Miguel, Laura, Davi e Gustavo. E aos demais primos, tios e tias.

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

As minhas irmãs de coração que a vida me deu Quercia Eloi e Leika Irabele. Aos meus amigos de departamento: Jucarlos Rufino, Kerolly Kedma, Fábio Sandro, Joelma Mayara, Denise Almeida, Laura Alves, Jackson Amorim, Luciano Pereira, Bárbara Carvalho, Elielma Santana, Raiane, Glauce Guerra, Jonas Vasconcelos, Diego Alves e todos os demais que compõem nosso departamento. Aos meus amigos de graduação Valdenise Maciel, Eduardo Lucas e Roberto Oliveira e o Professor Doutor Dâmocles Aurélio e o Professor Mestre Elizângelo Lopes. E a minha primeira professora Dona Matilde Maria Ferreira.

Agradeço a todo o corpo docente do DEINFO pelo compartilhamento dos seus conhecimentos e em particular aos professores Frank Sinatra, Tiago Alessandro e Moacyr Cunha. Ao secretário do programa Marco Antônio.

A Capes, pelo concessão de apoio financeiro e a UFRPE, que me proporcionou cursar o mestrado em Biometria e Estatística Aplicada.

*“Nada do que foi será  
De novo do jeito que já foi um dia  
Tudo passa, tudo sempre passará”  
(Lulu Santos, 1983).*

# Resumo

Mudanças na estrutura do solo influenciam nas emissões dos gases na atmosfera contribuindo para as mudanças climáticas. Uma das formas para um melhor entendimento dessas emissões é medir o fluxo do gás baseando-se na variação temporal da concentração em função do tempo. Entretanto, as análises tradicionais não levam em considerações possíveis características comuns que podem estar presente no fenômeno. No caso de modelos lineares, não é levado em consideração a natureza contínua das concentrações. No caso dos modelos não lineares, pode não ser levado em consideração a variação aleatória das concentrações. Nesse sentido, a análise de dados funcionais (FDA) representa as observações discretas em forma de uma função, que passa a ser uma única observação, levando em consideração tanto a variação contínua quanto a aleatoriedade dos dados. Diante disso, este trabalho tem como objetivo avaliar a aplicabilidade da análise de dados funcionais no estudo das emissões dos principais gases de efeito estufa ( $CO_2$ ,  $CH_4$  e  $N_2O$ ) nos diferentes usos da terra. Para transformação dos dados brutos em funções, utilizou-se a suavização *B-splines*. As derivadas das funções foram utilizadas para calcular a variação da concentração em função do tempo, ou seja, os fluxos, pois a natureza das emissões do gás é justamente a variação temporal da concentração. A análise de componentes principais funcional (FPCA) foi realizada para examinar os principais modos de variação que estão presentes nas observações. Aplicou-se a análise de variância funcional (FANOVA) a fim de verificar se existe diferença significativa, ao nível de 5%, entre as médias funcionais das emissões do  $CO_2$ ,  $CH_4$  e  $N_2O$  nos diferentes sistema de manejo. Após a aplicação dos métodos estatísticos funcionais, verificou-se que o fluxo médio diário obteve grande variabilidade. O primeiro e segundo FPCA capturaram, cumulativamente, mais de 90% desta variação total presente nos dados. A partir da FANOVA, verificou-se que existe diferença significativa entre as médias funcionais analisadas para o  $CO_2$  e  $N_2O$ . Já para o  $CH_4$  constatou que não existe diferença entre as emissões do pasto intensivo, linha da cana de açúcar e entre linha da cana de açúcar. Portanto, a FDA levou em consideração a natureza contínua do sistema fotoquímico e representou o processo desses gases de uma maneira mais completa e precisa.

**Palavras-chave:** dados funcionais, dióxido de carbono, óxido nitroso, metano.

# Abstract

Changes in soil structure influence gas emissions into the atmosphere contributing to climate change. One way to better understand these emissions is to measure the gas flow based on the temporal variation of the concentration in function of time. However, traditional analyses do not lead to possible considerations common features that may be present in the phenomenon. In the case of linear models, the continuing nature of concentrations is not taken into account. In the case of the models non-linear, random variation in concentrations may not be taken into account. In this sense, functional data analysis (FDA) represents the discrete observations in form of a function, which becomes a single observation, taking into account both the vario continues and the randomness of the data. In light of this, this work aims to evaluate the applicability of functional data analysis in the study of emissions of the main greenhouse gases ( $CO_2$ ,  $CH_4$  and  $N_2O$ ) in the different uses of the land. To transform the raw data into functions, B-splines smoothing was used. The derivatives of the functions were used to calculate the variation of the concentration in function of time, i.e. the flows, since the nature of gas emissions is precisely the temporal variation of concentration. Functional Principal Component Analysis (FPCA) was carried out to examine the main modes of variation that are present in the observations. Analysis of functional variance (FANOVA) was applied in order to check whether there is significant difference, at the level of 5%, between the functional averages of  $CO_2$ ,  $CH_4$  and  $N_2O$  emissions in the different management systems. After the application of statistical methods functional, it was found that the average daily flow obtained great variability. The first and according to FPCA cumulatively captured more than 90% of this total present variation in the data. From FANOVA, it was found that there is a significant difference between the functional averages analysed for  $CO_2$  and  $N_2O$ . For  $CH_4$  it was found that there is no difference between pasture intensified, sugarcane line and between sugar cane. Therefore, the FDA has taken into consideration the continuing nature of the system photochemical and represented the process of these gases in a more complete and precise way.

**keywords:** functional data, carbon dioxide, nitrous oxide, methane.

# Lista de Figuras

- Figura 1 – Imagem de satélite, fornecida pela plataforma Google Earth, que mostra a área de estudo, localizada no município de Brotas, Estado de São Paulo. Fonte: Próprio autor. . . . . 16
- Figura 2 – A) Cana-de-açúcar dois meses após o plantio ; B) Pastagens intensiva (verde escuro) e extensiva (verde claro) dois meses após o plantio da pastagem em Brotas. . . . . 17
- Figura 3 – Imagem de satélite, fornecida pela plataforma Google Earth, acessada em 18/08/2019. Em verde: delimitação da área de vegetação florestal, em preto: delimitação das parcelas de pastagem convencional, em amarelo: delimitação das parcelas de pastagem intensiva, em vermelho: delimitação das parcelas de cana-de-açúcar. . . . . 18
- Figura 4 – Exemplo da câmara estática utilizada para a coleta das concentrações dos gases (COLLIER, et al, 2014). . . . . 19
- Figura 5 – Suavização das concentrações do  $CO_2$  nos diferentes usos do solo. (A) pasto extensivo, (B) pasto intensivo, (C) linha da cana de açúcar e (D) entre linha da cana de açúcar. . . . . 22
- Figura 6 – Suavização das concentrações do  $N_2O$  nos diferentes usos do solo. (A) pasto extensivo, (B) pasto intensivo, (C) linha da cana de açúcar e (D) entre linha da cana de açúcar. . . . . 23
- Figura 7 – Suavização das concentrações do  $C_4H$  nos diferentes usos do solo. (A) pasto extensivo, (B) pasto intensivo, (C) linha da cana de açúcar e (D) entre linha da cana de açúcar. . . . . 24
- Figura 8 – Boxplot funcional das suavizações correspondente a emissão do  $CO_2$  para (A) pasto extensivo, (B) pasto intensivo, (C) linha da cana de açúcar e (D) entre linha da cana de açúcar. A curva preta, expressa a mediana funcional. A região rosa, retrata o envelopamento da região central que indica o acúmulo de 50% das funções. As curvas azuis extremas, representam o máximo e mínimo não periférico. As curvas vermelhas, são os outliers. . . . . 26

- Figura 9 – Boxplot funcional das suavizações correspondente a emissão do  $N_2O$  para (A) pasto extensivo, (B) pasto intensivo, (C) linha da cana de açúcar e (D) entre linha da cana de açúcar. A curva preta, expressa a mediana funcional. A região rosa, retrata o envelopamento da região central que indica o acúmulo de 50% das funções. As curvas azuis extremas, representam o máximo e mínimo não periférico. As curvas vermelhas, são os outliers. . . . . 27
- Figura 10 – Boxplot funcional das suavizações correspondente a emissão do  $CH_4$  para (A) pasto extensivo, (B) pasto intensivo, (C) linha da cana de açúcar e (D) entre linha da cana de açúcar. A curva preta, expressa a mediana funcional. A região rosa, retrata o envelopamento da região central que indica o acúmulo de 50% das funções. As curvas azuis extremas, representam o máximo e mínimo não periférico. As curvas vermelhas, são os outliers. . . . . 28
- Figura 11 – Emissões dos fluxos correspondente ao  $CO_2$  nos diferentes usos do solo. (A) pasto extensivo, (B) pasto intensivo, (C) linha da cana de açúcar e (D) entre linha da cana de açúcar. . . . . 31
- Figura 12 – Emissões dos fluxos correspondente ao  $N_2O$  nos diferentes usos do solo. (A) pasto extensivo, (B) pasto intensivo, (C) linha da cana de açúcar e (D) entre linha da cana de açúcar. . . . . 32
- Figura 13 – Emissões dos fluxos correspondente ao  $CH_4$  nos diferentes manejos do solo. (A) pasto extensivo, (B) pasto intensivo, (C) linha da cana de açúcar e (D) entre linha da cana de açúcar. . . . . 33
- Figura 14 – Fluxo médio diário do  $CO_2$ : Médias funcionais - curvas sólidas. Médias funcionais adicionando e subtraindo seu desvio padrão - curvas pontilhadas. (A) pasto extensivo, (B) pasto intensivo, (C) linha da cana de açúcar e (D) entre linha da cana de açúcar. . . . . 34
- Figura 15 – Fluxo médio diário do  $N_2O$ : Médias funcionais - curvas sólidas. Médias funcionais adicionando e subtraindo seu desvio padrão - curvas pontilhadas. (A) pasto extensivo, (B) pasto intensivo, (C) linha da cana de açúcar e (D) entre linha da cana de açúcar. . . . . 35
- Figura 16 – Fluxo médio diário do  $CH_4$ : Médias funcionais - curvas sólidas. Médias funcionais adicionando e subtraindo seu desvio padrão - curvas pontilhadas. (A) pasto extensivo, (B) pasto intensivo, (C) linha da cana de açúcar e (D) entre linha da cana de açúcar. . . . . 36
- Figura 17 – Análise descritivas referentes aos fluxos médios diário de: (A)  $CO_2$ , (B)  $N_2O$  e (C)  $CH_4$  para os diferentes usos do solo. . . . . 38

- Figura 18 – Primeiro componente principal (painéis da esquerda) e segundo componente principal (painéis da direita) das emissões de: (A)  $CO_2$ , (B)  $N_2O$  e (C)  $CH_4$ , sob os diferentes usos do solo. Curvas médias (preto) das emissões dos gases de efeito estufa adicionando (vermelho) e subtraindo (azul) os múltiplos aplicada a função média do componente principal. . . . . 40
- Figura 19 – Biplot dos componentes funcionais principais das emissões de: (A)  $CO_2$ , (B)  $N_2O$  e (C)  $CH_4$ , sob os diferentes usos da terra pasto extensivo (manejo 1), pasto intensivo (manejo 2), linha da cana de açúcar (manejo 3) e entre linha da cana de açúcar (manejo 4). . . . . 42
- Figura 20 – Primeiro componente principal funcional médio diário, mais ou menos os desvios padrão referentes ao  $CO_2$  nos diferentes manejos do solo. (A) pasto extensivo, (B) pasto intensivo, (C) linha da cana de açúcar e (D) entre linha da cana de açúcar. . . . . 43
- Figura 21 – Segundo componente principal funcional médio diário, mais ou menos os desvios padrão referentes ao  $CO_2$  nos diferentes manejos do solo. (A) pasto extensivo, (B) pasto intensivo, (C) linha da cana de açúcar e (D) entre linha da cana de açúcar. . . . . 44
- Figura 22 – Segundo componente principal funcional médio diário, mais ou menos os desvios padrão referentes ao  $N_2O$  nos diferentes manejos do solo. (A) pasto extensivo, (B) pasto intensivo, (C) linha da cana de açúcar e (D) entre linha da cana de açúcar. . . . . 45
- Figura 23 – Segundo componente principal funcional médio diário, mais ou menos os desvios padrão referente ao  $N_2O$  nos diferentes usos do solo. (A) pasto extensivo, (B) pasto intensivo, (C) linha da cana de açúcar e (D) entre linha da cana de açúcar. . . . . 46
- Figura 24 – Primeiro componente principal funcional médio diário, mais ou menos os desvios padrão referente ao  $CH_4$  nos diferentes usos do solo. (A) pasto extensivo, (B) pasto intensivo, (C) linha da cana de açúcar e (D) entre linha da cana de açúcar. . . . . 47
- Figura 25 – Segundo componente principal funcional médio diário, mais ou menos os desvios padrão referente ao  $CH_4$  nos diferentes usos do solo. (A) pasto extensivo, (B) pasto intensivo, (C) linha da cana de açúcar e (D) entre linha da cana de açúcar. . . . . 48
- Figura 26 – Análise de variância funcional para as médias das emissões de: (A)  $CO_2$ , (B)  $N_2O$  e (C)  $CH_4$ , sob os diferentes usos do solo: pasto extensivo (manejo 1), pasto intensivo (manejo 2), linha da cana de açúcar (manejo 3) e entre linha da cana de açúcar (manejo 4). . . . . 50

# Lista de tabelas

Tabela 1 – Adubação e variedades utilizados durante a intensificação dos agroecosistemas, com a conversão de pastagem extensiva para pastagem intensiva e canavial, nas áreas experimentais. . . . .	17
--	----

## Lista de abreviaturas e siglas

FDA	Análise de Dados Funcionais
GEE	Gases de Efeito Estufa
$CO_2$	Dióxido de Carbono
$CH_4$	Metano
$N_2O$	Óxido nitroso
C	Carbono
N	Nitrogênio
PCA	Análise de Componente Principal
FPCA	Análise de Componente Principal Funcional
FANOVA	Análise de Variância Funcional
CV	Validação Cruzada
GVC	Validação Cruzada Generalizada
PEX	Pasto Extensivo
PIN	Pasto Intensivo
L	Linha da Cana de Açúcar
EL	Entre Linha da Cana de Açúcar
$H / L_1^2$	Espaço de Hilbert
ppm	partícula por milhão
$\lambda$	parâmetro de suavização
$H_0$	Hipótese nula
$H_a$	Hipótese alternativa
$\xi(t)$	Função peso

# Sumário

<b>1</b>	<b>Introdução</b>	<b>1</b>
<b>2</b>	<b>Objetivos</b>	<b>3</b>
2.1	Objetivo Geral	3
2.2	Objetivos Específicos	3
<b>3</b>	<b>Revisão de Literatura</b>	<b>4</b>
3.1	Análise de Dados Funcionais	4
3.2	Suavização	4
3.2.1	Spline	5
3.2.2	B-splines	6
3.2.3	Erro Observacional	7
3.2.4	Método de Validação Cruzada Generalizada (GVC)	8
3.3	Estatísticas descritiva básicas para dados funcionais	9
3.3.1	Média e variância funcionais	9
3.3.2	Covariância e correlação funcionais	9
3.4	Boxplot funcional (FBP)	9
3.5	Derivativos	10
3.6	Análise de Componentes Principais Funcional (FPCA)	11
3.7	Análise de variância funcional (FANOVA)	12
3.8	Mudanças climáticas	14
<b>4</b>	<b>Metodologia</b>	<b>15</b>
4.1	Caracterização da área de estudo	15
4.1.1	Plantio de cana e braquiária nas parcelas experimentais	16
4.2	Descrição dos dados	18
4.3	Procedimentos de análise	19
<b>5</b>	<b>Resultados e Discussão</b>	<b>21</b>
5.1	Curvas funcionais do $CO_2$ , $N_2O$ e $CH_4$	21
5.2	Análise descritivas	25
5.3	Derivativos	29
5.4	Análise de componentes principais funcionais	39
5.5	Análise de variância funcional	49
<b>6</b>	<b>Conclusão</b>	<b>51</b>

**Referências Bibliográficas . . . . . 52**

# 1 Introdução

O efeito estufa baseia-se na absorção de parte da radiação infravermelha emitida pela superfície por gases (ANDERSON; HAWKINS; JONES, 2016) tais quais, dióxido de carbono ( $CO_2$ ), metano ( $CH_4$ ) e óxido nitroso ( $N_2O$ ), aumentando a temperatura do planeta. Contudo, o aumento na concentração atmosférica dos chamados gases geradores do efeito estufa ( $CO_2$ ,  $CH_4$  e  $N_2O$ ), daqui por diante, por facilidade, denominados “gases do efeito estufa” (GEE), vêm aumentando devido, principalmente, a queima de combustíveis fósseis e biomassa e atividades agrícolas e industriais, acarretando o aumento das temperaturas e, conseqüentemente, os padrões climáticos globais (SILVA; PAULA, 2009).

A atividade agrícola e as mudanças na cobertura do solo são fatores importantes para o aumento dos GEE, pois várias práticas comumente utilizadas alteram os principais processos biogeoquímicos que regulam a dinâmica do carbono (C) e do nitrogênio (N) no solo (BENTO et al., 2018; OERTEL et al., 2016). As principais fontes de emissões antropogênicas de  $CO_2$  são as queimas de combustíveis fósseis (carvão, petróleo e gás), biomassa e as mudanças na cobertura e no uso da terra (KRAM et al., 2000). Também o uso de fertilizantes nitrogenados em solos agrícolas e atividades das indústrias petroquímicas, são as principais fontes de  $N_2O$  para a atmosfera (RECK; HOAG, 1997; SIQUEIRA NETO et al., 2009). O  $CH_4$  é produzido pela oxidação anaeróbia da matéria orgânica, sendo produzido no processo de digestão de ruminantes, em aterros sanitários, pela queima de biomassa e áreas naturalmente ou artificialmente inundadas (DENMEAD, 2008).

Nas últimas décadas, foram desenvolvidas e aperfeiçoadas metodologias capazes de coletar e determinar os fluxos dos GEE oriundos tanto de solos florestais como de solos cultivados. Dentre estes métodos, as câmaras estáticas têm recebido destaque, pois os fluxos são medidos baseando-se na variação temporal da concentração do gás no interior das câmaras (DENMEAD, 2008; KELLER; KAPLAN; WOFYSY, 1986).

A escolha do modelo matemático para se estimar os fluxos de GEE a partir da variação temporal das concentrações na câmara estática constitui-se em uma das principais fontes de incerteza existentes no método (LEVY et al., 2011), visto que 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). Alguns métodos estatísticos são utilizados na estimativa dos fluxos dos GEE, por exemplo, são aplicados modelos lineares, exponenciais e polinomiais, correlacionando-

se as concentrações medidas na câmara estática em relação ao tempo decorrido desde o começo das medidas. De acordo com Pirk et al. (2016), o modelo linear presume uma mudança constante de concentração ao longo do tempo, fato que ocorre somente em alguns casos. O modelo exponencial se baseia no pressuposto de que a curvatura é proporcional à diferença de concentração, mas a priori não assume que nenhum processo seja responsável pela curvatura. Além disso, as medidas podem ser calculadas usando os modelos de Hutchinson/Mosier (H/M), o estimador de fluxo difusivo não estacionário (NDFE), o método de H/M modificado e o método Quad. Todos eles têm como objetivo comum eliminar a suposição que a relação entre concentração e tempo é linear (PARKIN; VENTEREA; HARGREAVES, 2012).

Assim, nota-se que, basicamente, todas as definições de fluxos são uma tentativa de expressar a variação da concentração em função do tempo. Porém, aplicar um modelo linear nos dados de concentração pode haver um erro sistemático por meio da não linearidade oculta, além disso, 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. Sendo assim, uma família estatística que expressa um grande potencial para estudar as variações de uma determinada variável em relação ao contínuo (tempo, espaço, frequência, entre outros) é a análise de dados funcionais (FDA). A ideia básica da FDA é expressar observações discretas decorrentes de séries temporais na forma de uma função (para criar dados funcionais), função esta que representaria toda a função medida como uma observação única (ULLAH; FINCH, 2013).

A FDA vem se destacando em diversas áreas de pesquisa, por exemplo, educação (FORTUNA; MATURO, 2019), quimiometria (GRANATO et al., 2018), medicina (HAPP; GREVEN, 2018; PFEIFFER et al., 2002; ULLAH; FINCH, 2010), economia (WANG et al., 2008), criminologia (BERK, 2008), previsões demográfica (HYNDMAN; SHANG, 2009; HYNDMAN; ULLAH, 2007), climatologia (GAO; NIEMEIER, 2008), entre outras. Seus objetivos são (i) representar os dados de forma a facilitar futuras análises; (ii) exibir os dados de forma a destacar suas principais características; (iii) estudar as principais fontes de variação e determinar comportamentos padronizados entre diferentes curvas; (iv) explicar a variação existente na variável de resposta através da informação sobre as diversas variáveis independentes, e (v) comparar dois ou mais conjuntos de dados em relação a determinados tipos de variações (RAMSAY; SILVERMAN, 2002).

Ainda não existem estudos que visem a aplicação da FDA em emissões de gases do efeito estufa, sendo assim, este é o principal objetivo deste trabalho, como descrito abaixo.

## 2 Objetivos

### 2.1 Objetivo Geral

Avaliar a aplicabilidade da análise de dados funcionais no estudo das emissões dos principais gases de efeito estufa.

### 2.2 Objetivos Específicos

- Avaliar e determinar o fluxo dos GEE como um dado funcional;
- Observar propriedade intrínseca do modelo a partir das funções ajustadas;
- Avaliar a aplicabilidade de métodos estatísticos funcionais aos dados das emissões dos gases de efeito estufa.

## 3 Revisão de Literatura

### 3.1 Análise de Dados Funcionais

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. O termo funcional refere-se à organização dos dados e não a sua maneira explícita, porque na prática as observações são medidas de forma discreta. A Análise de Dados Funcionais (*Functional Data Analysis*) conhecida como FDA foi introduzida no meio acadêmico por Ramsay e Dalzell (1991). Essa técnica representa dados cuja  $i$ -ésima observação é uma função real,  $r_i(t), i = 1, \dots, n, t \in T$ , onde  $T$  é um intervalo real definido em um espaço de função integráveis ao quadrado no conjunto compacto  $I$ , conhecido como espaço de Hilbert ( $L^2_I$  ou  $H$ ), com produto interno  $\langle f, g \rangle = \int_I fg$  e norma  $\|f\|_2 = (\int_I f^2)^{1/2}$ .

Ramsay e Dalzell (1991) exibiram alguns motivos práticos em estudar os dados a partir de uma interpretação funcional,

- Métodos de interpolação e suavização conseguem oferecer representações funcionais de conjuntos finitos de observações;
- É mais natural imaginar questões de modelagem em um aspecto funcional;
- Os objetivos de uma análise podem ser funcionais por natureza. Assim, pode se observar além da função inteira seus derivativos e outros valores funcionais.

Além das vantagens citadas acima, uma outra vantagem, como pode ser verificada em Ramsay, Bock e Gasser (1995), é a possibilidade de se trabalhar com observações que não precisam ter espaçamentos iguais.

### 3.2 Suavização

A FDA vai tratar cada  $r_i$  como uma função e não como um único valor em um ponto particular. Para transformar dados discretos em funções suaves e contínuas, é necessário de alguma técnica de suavização.

Deve-se trabalhar com qualquer tipo de função, tanto as mais simples como as mais inesperadas e complicadas. Sendo assim, precisa-se de uma técnica para produção de

funções que funcione com parâmetros acessíveis de calcular e ao mesmo tempo possam moldar todas as possíveis curvas. Desta forma, para reestruturar os dados observados como funções contínuas e suaves,  $r(\cdot)$ , em relação ao tempo,  $(t)$ , com o intuito de poder avalia-las em qualquer valor desejado de  $t$ , bem como, analisar as  $m$ -ésimas derivadas no argumento  $t$ ,  $D^m r(t)$ , precisa-se utilizar técnicas de interpolação e / ou suavização, como expansão de base (RAMSAY; SILVERMAN, 1997). Essa expansão de base é calculada por:

$$r(t) = \sum_{i=1}^K c_i(t) \phi_i(t) \quad (3.1)$$

em que  $c_i(t)$  são números reais chamados de coeficientes de expansão e os  $\phi_k(i)$ ,  $i = 1, 2, \dots, K$  é o conjunto de funções bases combinados linearmente. Para estimar o conjunto de funções bases que suavizam a função  $r(\cdot)$ , utiliza-se a suavização B-splines, Fourier, Wavelets entre outras. No entanto, a escolha da suavização vai depender do subjacente comportamento dos dados em análise (RAMSAY; SILVERMAN, 2005).

### 3.2.1 Spline

Os polinômios são bastantes usados em estudos para aproximar funções, devido a sua formação simples e obtenção de boas propriedades de aproximação. Um estudo extensivo feito por Schumaker (1981) propôs a definição para as funções *Splines*.

Seja um intervalo  $T = [a, b]$ , em que a função esta 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 estará sendo suavizada, é utilizado para aproximação em cada subintervalo. Esse método gera uma função de aproximação polinomial por partes  $s(\cdot)$ , ou seja,  $s(x) = p_i(x)$  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  $x_0$  e  $x_{k+1}$  são os nós exteriores e  $x_1, \dots, x_k$  os nós interiores.

Normalmente, as partes de polinômios,  $p_i(x)$ , são produzidos independentemente das demais. Por essa razão a aproximação não produz uma função contínua em  $T$ . Isso não é aceitável, caso se queira aproximar uma função suave. Sendo assim, necessita-se que os polinômios estejam ligados suavemente entre os nós interiores e ainda contenham certa quantidades de derivadas contínuas nessa uniões. Portanto, chama-se de função *spline* uma função polinomial por parte, suave e derivável.

Um *spline* de ordem  $m$  com  $x_i, i = 1, \dots, k$  nós interiores é toda função determinada como:

$$s(t) = \sum_{i=0}^{m-1} \theta_i t^i + \sum_{i=1}^k \delta_i (t - x_i)^{m-1}, \quad (3.2)$$

onde os  $\theta_i$  e os  $\delta_i$  são coeficientes  $\in \mathbb{R}$ ,  $\{x_1, \dots, x_k\}$  são os nós interiores e  $\{1, t, t^2, \dots, t^{m-1}, (t-x_1)^{m-1}, \dots, (t-x_k)^{m-1}\}$  são funções bases. Logo, verifica-se através da equação 3.2 que toda função *spline* é uma combinação linear de  $m+k$  funções bases.

Assim, pode-se destacar algumas propriedades importantes dos *splines* que são:

- A função  $s(t)$  é um polinômio por partes de grau  $m-1$  para cada um dos subintervalos definido na forma de  $[x_i, x_{i+1})$ ,  $i = 0, 1, \dots, k$ ;
- As funções *splines* polinomiais são relativamente suaves, com derivadas contínuas até a ordem  $m-2$ ;
- *Splines* polinomiais são mais fáceis de armazenar e manusear computacionalmente.

Uma extensão dos *splines* conhecido como B-*splines* também foi proposto por Schumaker (1981), em que os B-*splines* compõem uma base de espaços *splines*.

### 3.2.2 B-splines

Como as funções *splines* têm suas propriedades definidas pelos polinômios, então as somas, as subtrações e as multiplicações de uma função *spline* geram uma função *spline*, i.e. qualquer combinação linear dessas funções bases é uma função *spline*. Logo, a suavização B-*spline* se deriva da suavização *spline* descrita na seção anterior. Normalmente, 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 (ULLAH; FINCH, 2013). A mesma é utilizada quando as observações tem um comportamento não periódico (RAMSAY; HOOKER; GRAVES, 2009).

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  $x_{k+1} \leq \tau_{m+k+2} \leq \tau_{m+k+3} \leq \dots \leq \tau_{k+2m}$  (SCHUMAKER, 1981).

Desta forma, Boor (2001) desenvolveu um dos mais populares sistemas de suavização do B-*splines* (RAMSAY; SILVERMAN, 2005), criando um algoritmo que possui uma relação de recorrência para calculá-los. Então, consegue calcular B-*spline* de qualquer

ordem a partir de um B-spline de ordem menor. Sendo assim, uma sequência crescente com o  $i$ -ésimo B-splines de ordem  $m$  e nós  $\tau_i = \{\tau_i\}$  pode ser estimado como:

$$B_{i,m}(t) = \frac{t - \tau_i}{\tau_{i+m-1} - \tau_i} B_{i,m-1}(t) + \frac{\tau_{i+m} - t}{\tau_{i+m} - \tau_{i+1}} B_{i+1,m-1}(t), \quad (3.3)$$

em que:

$$B_{i,m}(t) = \begin{cases} 1 & \text{se } \tau_i \leq t \leq \tau_{i+m}, \\ 0, & \text{caso contrário.} \end{cases} \quad (3.4)$$

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

$$s(t) = \sum_{i=1}^{K=m+k} c_i B_{i,m}(t), \quad (3.5)$$

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

$$\hat{r}(t) = \sum_{i=1}^K \hat{c}_i B_i(t) \quad (3.6)$$

correspondente à equação 3.1.

### 3.2.3 Erro Observacional

Quando uma função apresenta uma certa quantidade de derivadas, dizemos que existe uma suavidade em relação à função. Isso se dá porque é uma das propriedades da função  $r$ , mas não é tão evidente quando trabalha-se com dados brutos  $y = (y_1, \dots, y_n)$ , devido à existência do erro observacional ou ruído encontrado na medição dos dados. Assim, uma função estimada é escrita a partir das observações como:

$$y_i = r(t_i) + \epsilon_i \quad (3.7)$$

Em que  $\epsilon_i$  é conhecido como ruído, perturbação ou erro, fornecendo uma variabilidade (rugosidade) aos dados brutos. Um dos problemas para retratar as observações como funções pode ser a tentativa de filtrar esse ruído da maneira mais eficaz possível (RAMSAY; SILVERMAN, 2005). Uma técnica de filtrar o erro é utilizando uma penalidade de aspereza ou medida de rugosidade. Essa abordagem usa um número grande de funções de base, provavelmente se estendendo a uma função de base por observação, mas, ao mesmo tempo, impondo suavidade penalizando alguma medida que afasta-se de algum comportamento "suave" da linha de base.

Defini-se a medida de rugosidade representado-a pela integral da  $m$ -ésima derivada quadrática de uma função  $r$  no valor de argumento  $t$  também conhecida como a curvatura total, representada como:

$$PEN_m(r) = \int [D^m r(t)]^2 dt, \quad (3.8)$$

em que  $D^m$  é a  $m$ -ésima derivada.

Em seguida, minimiza-se um critério de ajuste que reduz a variabilidade da curva contra a falta de ajuste dos dados, adicionando a curvatura total a soma dos erros quadrados, indicado como:

$$F(\mathbf{c}) = \sum_i [y_i - r(t_i)]^2 + \lambda \int [D^m r(t)]^2 dt \quad (3.9)$$

onde,  $\lambda$  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  $PEN_m(r)$  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).

### 3.2.4 Método de Validação Cruzada Generalizada (GVC)

As suavizações *splines* são bastante conhecidas por proporcionar curvas que suavizam dados discretos e ruidosos. Assim, foi desenvolvido um método prático por Craven e Wahba (1979) chamado de validação cruzada generalizada (GCV) com objetivo de estimar a quantidade ideal de suavização dos dados.

Inicialmente, foi criada como uma alternativa mais simples do procedimento de validação cruzada (CV), evitando a necessidade de refazer várias vezes. No entanto, foi considerado mais segura do que a CV, no sentido de possuir uma menor tendência na suavização (RAMSAY; SILVERMAN, 2005). O método é definido como:

$$GCV(\lambda) = \left( \frac{n}{n - df(\lambda)} \right) \left( \frac{SSE}{n - df(\lambda)} \right) \quad (3.10)$$

em que  $n$  é o número de observações,  $df(\lambda) = trace[H(\lambda)]$  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 minimização da soma dos erros quadrados. Portanto, a ideia básica da GCV é encontrar o valor do parâmetro de suavização que minimize  $GCV(\lambda)$ .

A validação cruzada generalizada é empregada como método que otimizar a quantidade de variáveis selecionadas em relação à previsão das observações. E o tamanho do modelo designado pode ser considerado como um parâmetro de suavização que equilibra a

proximidade do ajuste e da complexidade (JANSEN, 2015). A GVC vai otimizar tanto o valor de  $\lambda$  quanto o número de funções bases que serão necessárias para obter uma melhor suavização.

Mais detalhes sobre GCV podem ser vistos em Craven e Wahba (1979), capítulo 5 de Ramsay e Silverman (2005) e Ramsay, Hooker e Graves (2009).

### 3.3 Estatísticas descritiva básicas para dados funcionais

As estatísticas descritivas clássicas como média, variância, covariância e correlação para dados univariados, também aplicam-se aos dados funcionais.

#### 3.3.1 Média e variância funcionais

Seja  $r_i, i = 1, \dots, N$  uma amostra de curvas ou funções ajustadas aos dados. A função média é definida por:

$$\bar{r}(t) = N^{-1} \sum_i r_i(t) \quad (3.11)$$

Similarmente, a função variância é dada por:

$$Var_r(t) = (N - 1)^{-1} \sum_i [r_i(t) - \bar{r}(t)]^2 \quad (3.12)$$

e o desvio padrão ( $\sigma$ ) é igual a raiz quadrada da função variância.

#### 3.3.2 Covariância e correlação funcionais

A função que condensa a dependência de registros entre distintos valores de argumento e calculada para todos os  $t_1$  e  $t_2$  é conhecida como função de covariância, expressa por:

$$Cov_r(t_1, t_2) = (N - 1)^{-1} \sum_i [r_i(t_1) - \bar{r}(t_1)][r_i(t_2) - \bar{r}(t_2)] \quad (3.13)$$

Esta associada à função correlação:

$$Corr_r(t_1, t_2) = \frac{Cov_r(t_1, t_2)}{\sqrt{Var_r(t_1)Var_r(t_2)}} \quad (3.14)$$

### 3.4 Boxplot funcional (FBP)

O boxplot funcional (FBP) ordena dados funcionais por meio da profundidade de banda. López-Pintado e Romo (2009) propuseram o conceito de uma nova definição de

profundidade para observações funcionais com base na representação gráfica das curvas. Dada uma coleção de funções, ela estabelece a centralidade de uma observação e fornece uma ordem natural central-externa das curvas da amostra. Assim, a função mediana ou uma função média aparada, podem ser definidas a partir dessa definição de profundidade.

O FBP permite solicitar uma amostra de curvas do centro para o exterior e, portanto, introduz uma medida para definir a centralidade ou a periferia de uma observação. É possível calcular as profundidades de banda de todas as curvas de amostra e ordená-las de acordo com os valores decrescentes de profundidade. Para os dados funcionais cada observação é uma função real  $r_i(t), i = 1, \dots, n, t \in T$ , onde  $T$  é um intervalo em  $\mathbb{R}$ . Denota-se  $r_{[i]}(t)$  a curva de amostra associada ao  $i$ -ésimo maior valor de profundidade de banda. Então  $r_{[1]}(t), \dots, r_{[n]}(t)$  pode ser visto como estatística da ordem, com  $r_{[1]}(t)$  sendo a curva mais profunda (mais central) ou simplesmente a curva mediana, e  $r_{[n]}(t)$  sendo a curva mais periférica. A implicação é que uma classificação menor está associada a uma posição mais central em relação às curvas da amostra. As estatísticas da ordem induzidas pela profundidade da banda, comece na curva de amostra mais central e mova-se para fora em todas as direções (SUN; GENTON, 2011).

Para boxplots funcionais, com base na ordenação externa central induzida pela profundidade de banda para dados funcionais, as estatísticas descritivas são: o envelope da região central de 50% curva mediana e o envelope não periférico máximo. Além disso, os outliers em potencial podem ser detectados em um boxplot funcional pela regra empírica de 1,5 vezes a região central de 50%, análoga à regra para boxplots clássicos. A região externa é obtida inchando a região interna (o envelope) 1,5 vezes a altura da região central de 50%. Quaisquer curvas que atravessam a região externa são sinalizadas como possíveis discrepantes (SUN; GENTON; NYCHKA, 2012).

Para mais detalhes de como calcular o boxplot funcional ou o boxplot funcional modificado podem ser visto em López-Pintado e Romo (2009), Sun e Genton (2011), Martin-Barragan, Lillo e Romo (2016).

### 3.5 Derivativos

Como exibido na sessão 2.2.1, para a função ser suave é necessário que exista um certo número de derivativos correspondente a ela, por exemplo, quando se suaviza as observações a partir da técnica *B-splines* de ordem  $m$ , têm-se derivadas contínuas até a ordem  $m - 2$ . Isso é uma propriedade que exhibe informações a mais presentes nas funções estimadas.

Essas informações geralmente não estão disponíveis em aplicações de técnicas

estatísticas clássicas (MAS; PUMO, 2009). Sendo assim, a FDA possibilita usar as taxas de mudanças, como a velocidade e aceleração, para extrair informações das curvas funcionais. Os modelos das taxas de mudanças de um sistema é constantemente conhecido como a análise da dinâmica de um sistema (RAMSAY; SILVERMAN, 2005).

Deste modo, um modelo dinâmico contém uma ou mais derivadas associadas ao tempo. Assim, como inúmeras ordens de derivadas podem existir, usa-se a notação  $D^m r(t)$  para designar a  $m$ -ésima derivada da função  $r(t)$ . Isso é mais pertinente do que usar um símbolo separado para cada derivada, como na notação tradicional:

$$\frac{d^m r}{dt^m}$$

A maneira mais habitual do modelo dinâmico é uma equação que uni duas ou mais ordens de derivadas. Dessa forma, com base em  $r(t)$  podemos calcular suas derivadas:

$$D^m r(t) = \sum_{i=1}^{m-2} c_i(t) D^i \phi_i(t) \quad (3.15)$$

Tem-se  $n$  funções de coeficiente  $c_i(t)$  que determinam as equações, mas em aplicações específicas, pode-se querer fixar os valores de algumas delas (RAMSAY; SILVERMAN, 2002).

### 3.6 Análise de Componentes Principais Funcional (FPCA)

A análise de componentes principais (PCA) é considerado como um dos primeiros métodos a serem adaptados aos dados funcionais (DAUXOIS; POUSSE; ROMAIN, 1982; SHANG, 2014), conhecida como análise de componentes principais funcional (FPCA). A mesma é empregada para detectar as principais fontes de variabilidade, determinar o valor de sua importância e diminuir a dimensão dos dados, além de conceder informações sobre a estrutura de covariância dos dados funcionais (KRAUS, 2015).

Esse estudo diminui as dimensões de um conjunto de dados onde existe um amplo número de variáveis inter-relacionadas, ao mesmo tempo, ele ainda retém a maior variação total existente. Essa diminuição é realizada através da transformação dos dados para outro conjunto de variáveis, ou componentes principais, que não são mais inter-relacionados (ULLAH; FINCH, 2013).

Em relação ao contexto funcional, 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 escores dos componentes principais dos indivíduos da amostra são os valores  $z_i$  dados por:

$$z_i = \int \xi(t)r_i(t)dt \quad (3.16)$$

A finalidade do PCA é achar a função de ponderação  $\xi_1(t)$  que maximize a variação das pontuações do  $z_i$  sujeita à restrição:

$$\int \xi_1(t)^2 dt = 1 \quad (3.17)$$

A restrição unitária da soma dos quadrados nos pesos é fundamental para que o problema torne-se preciso. Sem essa restrição, os valores dos quadrados médios da combinação linear poderiam ficar aleatoriamente grandes. Para os passos seguintes, busque novamente os modos de variação mais significativos no entanto, necessita-se que os  $\xi_i$  definidos sejam ortogonais aos pesos identificados anteriormente, para que eles estejam indicando algo novo. Assim, a segunda função do componente  $\xi_2(t)$  é estabelecida para maximizar a variação das pontuações do componente principal sujeitas à restrição 3.17 e adicionando a restrição:

$$\int \xi_2(t)\xi_1(t)dt = 0 \quad (3.18)$$

Os próximos componentes principais (terceira, quarta e ordem superior) são resolvidos do mesmo modo, porém com restrições adicionais. De modo geral, para o  $j$ -ésimo componente tem-se:

$$\int \xi_j(t)\xi_1(t)dt = \int \xi_j(t)\xi_2(t)dt = \dots = \int \xi_j(t)\xi_{j-1}(t)dt = 0 \quad (3.19)$$

A fórmula acima se certificará que todos os componentes principais encontrados sejam mutuamente ortogonais (RAMSAY; SILVERMAN, 2002).

### 3.7 Análise de variância funcional (FANOVA)

A Análise de Variância Funcional (FANOVA) oferece, em relação aos dados funcionais, uma opção semelhante ao método de análise de variância clássica. Considere o conjunto de  $z$  funções aleatórias independentes  $R_{ij}(t)$ ,  $i = 1, \dots, z$ ,  $j = 1, \dots, n_i$  definidas em um período limitado e fechado  $I = [a, b]$ . Em que  $n \in \mathbb{N}$  é o número de observações no  $i$ -ésimo grupo. Esses conjuntos podem diferenciar em relação as funções médias, ou seja, assumimos que  $R_{ij}(t)$  são processos estocásticos com função média  $\mu_i(t)$ , e função de covariância  $\gamma(s, t)$ ,  $s, t \in I$ . Também assumimos  $R_{i,j} \in L_I^2$ , em que  $L_I^2$  está definido no espaço de Hilbert. O interesse é testar a seguinte hipótese nula:

$$H_0 : \mu_1(t) = \dots = \mu_z(t), t \in I \quad (3.20)$$

A hipótese alternativa ( $H_a$ ) é a negação da  $H_0$ . Para  $H_a$ , existe pelo menos duas funções médias que não são iguais. Esse problema é conhecido como ANOVA unidirecional para dados funcionais. Para a FANOVA, as funções médias são frequentemente decompostas como  $\mu_i(t) = \mu_0(t) + \alpha_i(t)$ ,  $i = 1, 2, \dots, z$ , em que  $\mu_0(t)$  é a grande função média e  $\alpha_i(t)$  são as  $z$  funções de efeito principal, de modo que 3.20 é geralmente escrito de maneira equivalente a se testar as funções de efeito principal são todas 0, isto é,  $H_0 : \alpha_1(t) = \dots = \alpha_z(t) = 0$ ,  $t \in I$  (GÓRECKI; SMAGA, 2019; ZHANG; LIANG, 2014).

Existe alguns testes para calcular 3.20, praticamente todos são baseados na estatística do teste  $F$ . Em um estudo extensivo produzido por Górecki, Krzyśko e Wołyński (2015), foi proposto o teste de permutação (FP) e foi observado que este teste obteve melhor desempenho em ternos de controle de tamanho e potência, ao ser comparado aos outros testes existentes, quando a grade de pontos de tempo é curta, ou seja, quando o intervalo de tempo em que as observações foram medidas é pequeno. Desta forma, como o intervalo de tempo analisado no presente estudo é relativamente pequeno, foi utilizado o teste FP para testar  $H_0$ .

A estatística do teste FP é considerada como:

$$\frac{\int_I SSR_n(t)dt/(l-1)}{\int_I SSE_n(t)dt/(n-1)} \quad (3.21)$$

em que,

$$SSR_n(t) = \sum_{i=1}^l n_i (\bar{R}_i(t) - \bar{R}(t))^2$$

$$SSE_n(t) = \sum_{i=1}^l \sum_{j=1}^{n_i} (R_{ij}(t) - \bar{R}_i(t))^2$$

O denominador da equação 3.21 representa variabilidade interna da amostra, isto é, a variação dentro dos sujeitos. Já o numerador é responsável pela variabilidade entre os sujeitos, ou seja, a variação externa. As representações funcionais ( $R_{ij}$ ) são representadas por um número finito de funções básicas  $\phi_m \in L_2(I)$ ,  $m = 1, \dots, K$ , da forma de:

$$R_{ij}(t) \approx \sum_{m=1}^K c_{ijm} \phi_m(t), \quad t \in I \quad (3.22)$$

em que,  $c_{ijm}$  são as variáveis aleatórias,  $\text{Var}(c_{ijm}) < \infty$  e  $K$  é razoavelmente grande. Desta forma, a função média geral da amostra e as funções médias dos grupo da amostra podem ser escritas por:

$$\bar{R}(t) = n^{-1} \sum_{i=1}^k \sum_{j=1}^{n_i} c_{ijm} \phi_m(t), \quad \bar{R}_i(t) = n_i^{-1} \sum_{j=1}^{n_i} c_{ij} \phi_m(t) \quad (3.23)$$

O tamanho de  $K$  é escolhido para cada observação com intuito de existir um valor modal, mínimo, máximo ou médio comum para todas os  $R_{ij}(t)$ . Existe duas maneiras de selecionar o valor de  $K$ . A primeira é utilizar algum critério de informação, por exemplo, BIC, eBIC, AIC ou AICc. A segunda é utilizar a representação básica dos dados, que já foi calculada anteriormente (GÓRECKI; SMAGA, 2019).

### 3.8 Mudanças climáticas

O clima da Terra é limitado pelo seu balanço energético. A radiação solar visível absorvida pela atmosfera, oceano e superfície terrestre aquece a terra. A radiação infravermelha emitida pela atmosfera, oceano e superfície terrestre resfria a terra. A composição da atmosfera pode influenciar o clima alterando a quantidade de radiação solar absorvida ou refletida de volta ao espaço. Os gases de efeito estufa absorvem e emitem radiação infravermelha. Sua capacidade de fazer isso varia de acordo com sua estrutura molecular, e sua vida útil na atmosfera varia de acordo com suas propriedades químicas. Assim, as emissões de diferentes gases de efeito estufa podem ter impactos muito diferentes no clima da Terra. As mudanças climáticas devido ao aumento da concentração de gases de efeito estufa têm implicações significativas para o bem-estar humano e para o mundo natural (KIRK-DAVIDOFF, 2018).

A mudança climática é um dos maiores desafios da atualidade. Ela é uma modificação lenta e gradual das condições climáticas médias e, portanto, um fenômeno difícil de detectar e avaliar com precisão (WEBER, 2016). Seus impactos, afetam a agricultura, áreas florestais, áreas costeiras, recursos naturais, recursos hídricos, atividades econômicas, saúde, entre outras, atingindo tanto a sociedade como o meio ambiente de uma maneira global (SILVA; GUETTER, 2015).

## 4 Metodologia

### 4.1 Caracterização 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 ( $22^{\circ}17'12''S$  e  $48^{\circ}7'35''W$ ). A altitude média nessa área é de 650 m acima do nível do mar, com temperatura média anual de  $22^{\circ}$ , e média anual de precipitação de aproximadamente 1400 mm. Segundo a classificação de Köppen, o clima é subtropical do tipo Cwa, com inverno seco e verão chuvoso. Na área de estudo, com pastagem convencional, foram estabelecidas 15 parcelas, com aproximadamente 50 x 50 m, totalizando 0,25 ha cada (Figura 1). Cinco dessas parcelas continuaram como pastagem convencional, cinco foram convertidas em canaviais, e as cinco restantes foram intensificadas com a utilização de insumos agrícolas e, futuramente, com aumento da lotação. A vegetação florestal, presente há mais de 50 anos e que será utilizada para comparativo, está localizada à montante da área cultivada (Figura 1). Ambas as áreas possuem características similares em relação à topografia, ao tipo de solo e ao regime pluviométrico.

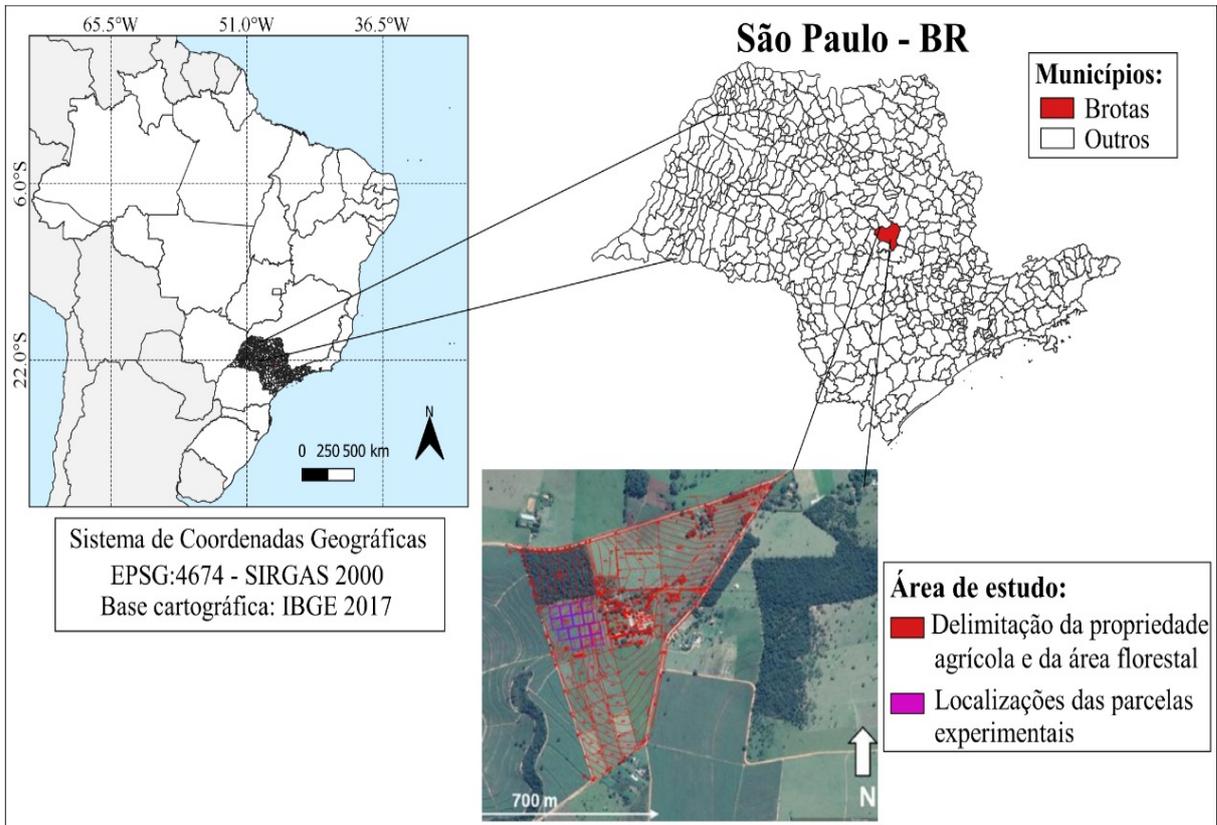


Figura 1 – Imagem de satélite, fornecida pela plataforma Google Earth, que mostra a área de estudo, localizada no município de Brotas, Estado de São Paulo. Fonte: Próprio autor.

#### 4.1.1 Plantio de cana e braquiária nas parcelas experimentais

As mudas de cana e as sementes de braquiária foram plantadas em dezembro de 2018. A calagem, seguida de adubação foi feita após análise do solo. A Tabela 1 resume os insumos aplicados nas respectivas áreas. A figura 2-A mostra a cana-de-açúcar dois meses após o plantio, e a figura 2-B as pastagens extensiva e intensiva dois meses após o plantio em Brotas. A figura 3 mostra a área experimental e a alocação das parcelas experimentais.

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.

Área experimental BROTAS	
<b>Cana-de-açúcar</b>	<b>RB86-7515</b>
Calcário 70% PNRT (Mg ha <sup>-1</sup> )	2
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>f</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>β</sup>	1,5
<b><i>Brachiaria brizantha</i></b>	<b>cv.Marandú</b>
Calcário (70% PNRT)	2
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>f</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>β</sup>	1,5

δ na forma de ureia; f na forma de supersimples;

γ na forma de KCl; ξ nome comercial Regente ®800 WG;

β nome comercial DMA ®806 BR.

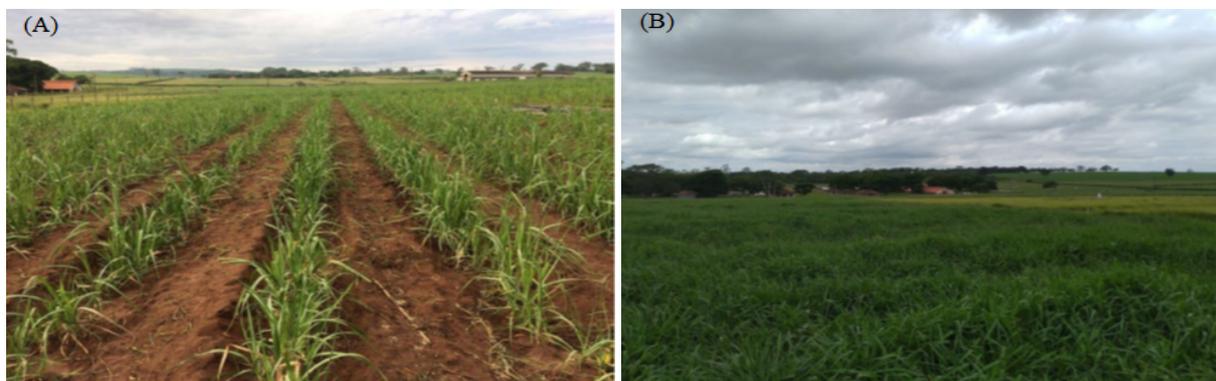


Figura 2 – A) Cana-de-açúcar dois meses após o plantio ; B) Pastagens intensiva (verde escuro) e extensiva (verde claro) dois meses após o plantio da pastagem em Brotas.



Figura 3 – Imagem de satélite, fornecida pela plataforma Google Earth, acessada em 18/08/2019. Em verde: delimitação da área de vegetação florestal, em preto: delimitação das parcelas de pastagem convencional, em amarelo: delimitação das parcelas de pastagem intensiva, em vermelho: delimitação das parcelas de cana-de-açúcar.

## 4.2 Descrição dos dados

A base de dados utilizada para a análise da emissão dos gases  $CO_2$ ,  $CH_4$  e  $N_2O$  foi disponibilizada pelo Dr. Leonardo Machado Pitombo. As amostras de gases foram coletadas através das câmaras estáticas (Figura 4) fixadas na superfície do solo, segundo o método proposto por Davidson et al. (2002). As coletas foram realizadas no período da manhã. Nos dias de coleta quatro amostras de ar foram coletadas de cada câmara, representando as concentrações em quatro pontos de tempos (1, 4, 8 e 12 minutos). De acordo com Garcia-Montiel et al. (2001), as câmaras serão vedadas no primeiro tempo de coleta e, prioritariamente, somente durante o período de amostragens. 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. 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



Figura 4 – Exemplo da câmara estática utilizada para a coleta das concentrações dos gases (COLLIER, et al, 2014).

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). O período da coleta foi de dezembro de 2018 a maio de 2019.

### 4.3 Procedimentos de análise

Inicialmente, é necessário reestruturar as observações discretas das concentrações dos GEE em relação aos sistemas de manejo. Para isso, utiliza-se a função `optim.basis` para obter uma estimativa de dados funcionais via representação de base usando o método de validação cruzada generalizada (GCV) com uma penalidade de rugosidade. Essa função vai fornecer o menor valor de GCV para dados funcionais baseado em uma lista dos números de funções bases (`num.basis`) e o valor de  $\lambda$  (`lambda`). Em seguida, aplica-se a função `create.bspline.basis` para criar uma base de dados funcionais a partir da suavização *B-spline*. As funções *B-spline* são um conjunto de funções especiais de spline, construídas a partir de polinômios por partes de ordem 4. Cada intervalo é subdividido em um conjunto de intervalos adjacentes que separam os nós. Então, para suavizar o polinômio

por partes, dois polinômios adjacentes são restritos a ter seus valores e todas as suas derivadas até a ordem  $m - 2$  correspondente no ponto em que se juntam. Por fim, a função `smooth.basis` é empregada para construir um objeto de dados funcionais suavizando os dados usando uma penalidade de rugosidade. Desta forma, as observações discretas são ajustadas a um conjunto de curvas suaves, cada uma definida por uma expansão em termos de funções básicas (BANDE et al., 2019).

Em seguida, realizou-se um boxplot funcional (FBP), através da função `boxplot.fd`, como uma ferramenta exploratória informativa para obter uma melhor compreensão das emissões suavizadas dos GEE. O FBP produz gráficos de caixa funcionais que permite solicitar uma amostra de curvas do centro para o exterior e, portanto, introduz uma medida para definir quantis funcionais, mediana funcional, a centralidade ou periferia da caixa e os outliers (RAMSAY et al., 2018; SUN; GENTON; NYCHKA, 2012).

Após o FBP, foram calculados os fluxos dos GEE. Por definição, o fluxo dos GEE é a taxa de variação das concentrações em relação ao tempo. Essa taxa de variação pode ser calculada através das derivadas usando a função `deriv.fd`. Então, uma das vantagens de utilizar dados funcionais é a mensuração do fluxo instantâneo através das derivadas. Assim, é possível analisar a variabilidade dos fluxos dos gases em relação aos diferentes manejos do solo.

A análise de componentes principais funcional foi realizada para examinar os principais modos de variação que estão presentes nas observações. Estes modos de variação são chamados de componentes principais (PC). Para analisar a FPCA, é aplicada a função `pca.fd` (RAMSAY et al., 2018). Essa função integra a função peso combinado com a observação funcional encontrando a pontuação dos scores para as funções individuais. Um modo de auxiliar na interpretação no 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, adiciona e subtrai múltiplos a função média do componente principal (RAMSAY; SILVERMAN, 1997; WARMENHOVEN et al., 2019). Os múltiplos são as raízes quadradas dos valores dos scores vezes o valor da aceleração dos scores aplicado nos pontos de amostragem.

Finalizando, aplicou-se a função `fanova.tests` (GÓRECKI; SMAGA, 2019), para análise variância funcional, a fim de verificar se existe diferença significativa, ao nível de 5%, entre as médias do pasto extensivo, pasto intensivo, linha da cana de açúcar e entre linha da cana de açúcar para o  $CO_2$ ,  $N_2O$  e  $CH_4$ . Todas as análises foram realizadas no software R (R CORE TEAM, 2019). Os pacotes utilizadas para o procedimentos de análise foram `fda` (RAMSAY et al., 2018), `fda.usc` (BANDE et al., 2019), `fdANOVA` (GÓRECKI; SMAGA, 2019) e `ggplot2`.

## 5 Resultados e Discussão

### 5.1 Curvas funcionais do $CO_2$ , $N_2O$ e $CH_4$

Estudos sobre os impactos da mudança do uso do solo devem ser aplicados em emissões de gases como o dióxido de carbono, óxido nitroso e metano para obter uma melhor compreensão sobre as alterações climáticas globais. As Figuras 5, 6 e 7 representam a reestruturação das concentrações das emissões desses gases nos diferentes sistemas de uso da terra. Cada função refere-se à suavização dos dados brutos da concentração em quatro pontos de tempo, obtidos através das câmeras estática fechada instaladas no solo. Foram utilizadas 32 funções bases para o  $CO_2$ , 34 funções bases para  $N_2O$  e 35 funções para  $CH_4$ , todas com grau de suavização ( $\lambda$ ) igual a 1, produzindo um nível considerável de precisão e suavidade. Após essa transformação, obteve uma função completa e contínua da concentração das emissões dos gases de efeito estufa, permitindo as análises estatísticas e conservando a variabilidade nos dados originais.

Nota-se um comportamento bem distinto entre as emissões das concentrações dos GEE. O  $CO_2$  (Figura 5), apresenta-se com emissões crescentes em todo intervalo de tempo, em relação aos quatro usos da terra. As concentrações suavizadas do  $N_2O$  (Figura 6) apresentam um comportamento constante para o pasto extensivo, já a linha da cana de açúcar as funções são crescentes. Para o pasto intensivo e entre linha da cana de açúcar existem funções crescentes e também constantes durante o período de coleta dos dados. As curvas funcionais das concentrações do  $CH_4$  (Figura 7) para os usos da terra mostram-se com um comportamento caótico, em comparação aos outros gases em estudo.

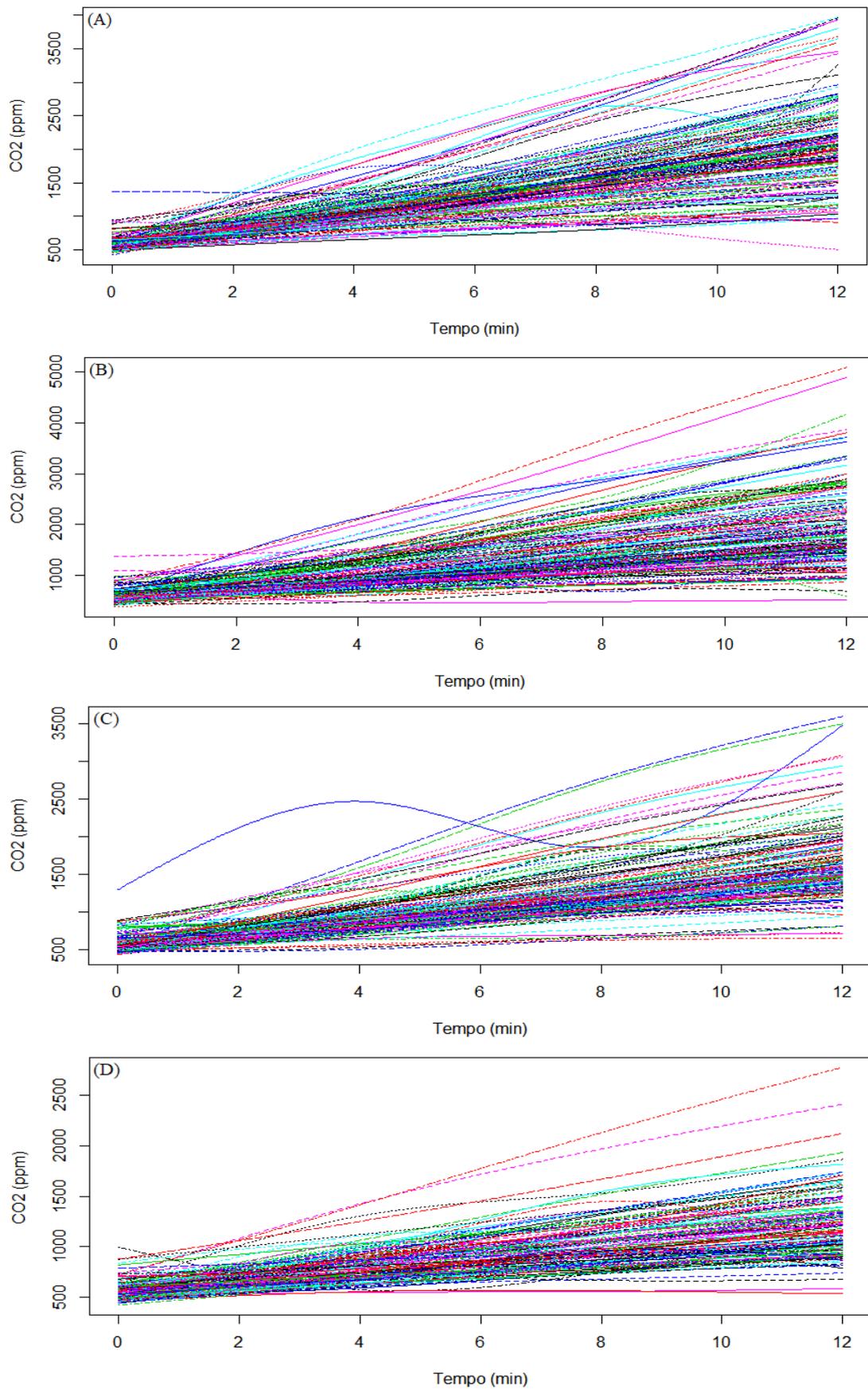


Figura 5 – Suavização das concentrações do  $CO_2$  nos diferentes usos do solo. (A) pasto extensivo, (B) pasto intensivo, (C) linha da cana de açúcar e (D) entre linha da cana de açúcar.

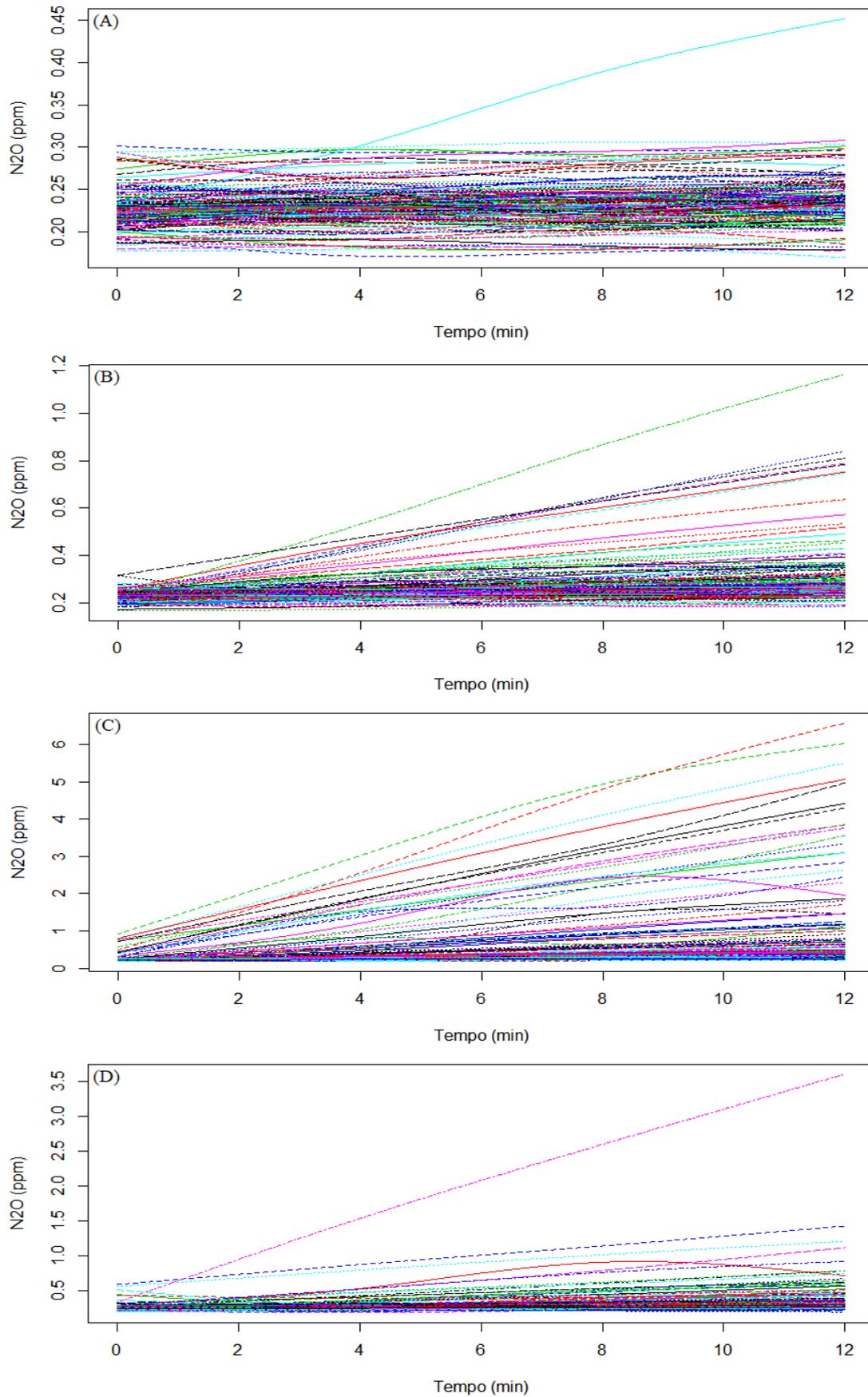


Figura 6 – Suavização das concentrações do  $N_2O$  nos diferentes usos do solo. (A) pasto extensivo, (B) pasto intensivo, (C) linha da cana de açúcar e (D) entre linha da cana de açúcar.

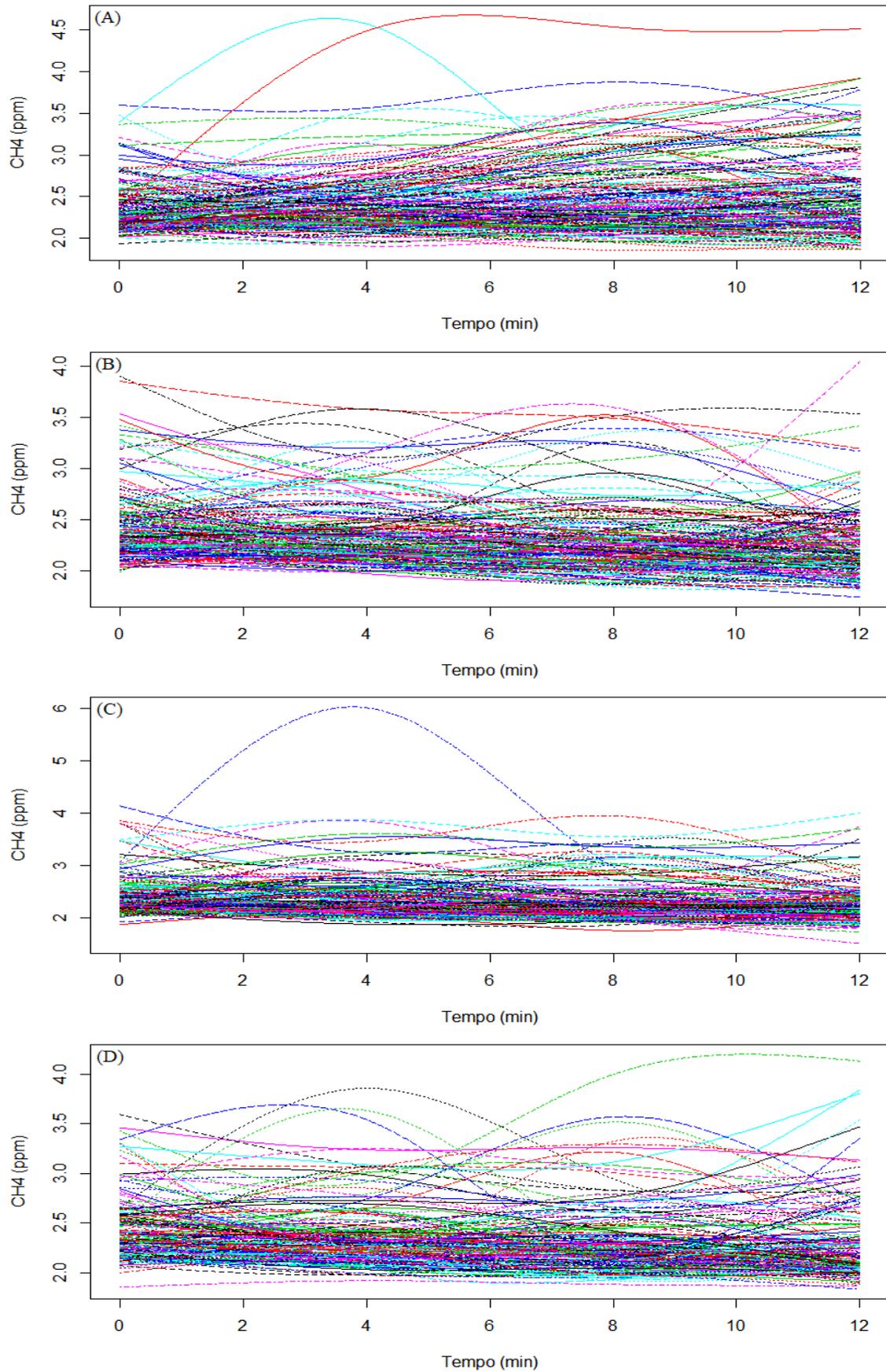


Figura 7 – Suavização das concentrações do  $C_4H$  nos diferentes usos do solo. (A) pasto extensivo, (B) pasto intensivo, (C) linha da cana de açúcar e (D) entre linha da cana de açúcar.

## 5.2 Análise descritivas

As medianas funcionais, os mínimos, os máximos e os outliers são apresentados nas Figuras 8, 9 e 10 construídos a partir do boxplot funcional (FBP) com as suavizações das emissões para todos os gases nos diferentes manejo do solo. Esses gráficos são importantes porque permitem analisar toda a variabilidade e detectam possíveis discrepâncias (NGO et al., 2015).

Uma inspeção visual na Figura 8 exhibe diferenças entre as emissões do gás carbono. O  $CO_2$  apresentou inclinações positivas para todas as medianas funcionais para os quatro usos da terra, com os maiores valores para o pasto extensivo e pasto intensivo. Já os menores valores foram para linha e entre linha da cana de açúcar. Além disso, o pasto extensivo (PEX), pasto intensivo (PIN) e linha da cana de açúcar (L) apresentam uma maior quantidade de outliers em relação a entre linha da cana de açúcar (EL). Essas curvas, que representam os outliers, demonstraram grande discrepância em relação a mediana. O envelopamento da região central (região rosa) do pasto extensivo e da linha da cana de açúcar estão centralizados entre a função mínima e função máxima, enquanto o envelopamento do pasto intensivo e entre linha estão mais próximo função mínima. Isso mostra que existe uma assimetria para a PIN e EL do que os outros dois manejos.

A Figura 9 apresenta diferenças entre as emissões do óxido nitroso. O  $N_2O$  apresentou inclinações crescentes para a mediana funcional da linha da cana de açúcar (Figura 9-C) e entre linha da cana de açúcar (Figura 9-D). Já para o pasto extensivo e pasto intensivo, a função mediana não apresentou nenhum tipo de inclinação. Dos GEE analisados, o  $N_2O$  exibiu um maior quantitativo de outliers para todos os manejos em estudo. Se destacando os dias 14, 15, 17 e 19 de dezembro com um percentual 53,8% do total de outliers existentes. Segundo Sun e Genton (2011), isso acontece quando a região central é estreita. Assim, essa região têm menos variabilidade, levando a uma porcentagem relativamente grande de outliers. Apenas o PEX teve o envelopamento centralizado entre a região não periférica. Isso indica que as observações se distribuem simetricamente em torno da mediana.

Na Figura 10, não é visualizado diferença entre os valores de emissão do metano. A mediana para os usos da terra em relação ao  $CH_4$  mostraram características equivalentes. Não exibiu nenhuma inclinação na mediana funcional, se tornando praticamente constante em todos os manejos. O envelopamento para todos os sistemas de manejo estão próximas da função mínima, implicando em uma assimetria das observações em torno da mediana. Vale destacar que o pasto extensivo (Figura 10-A) apresenta menos outliers em comparação aos outros três grupos. Isso se dá porque o envelopamento da região central dessa figura tem maior dispersão do que os outros três manejos.

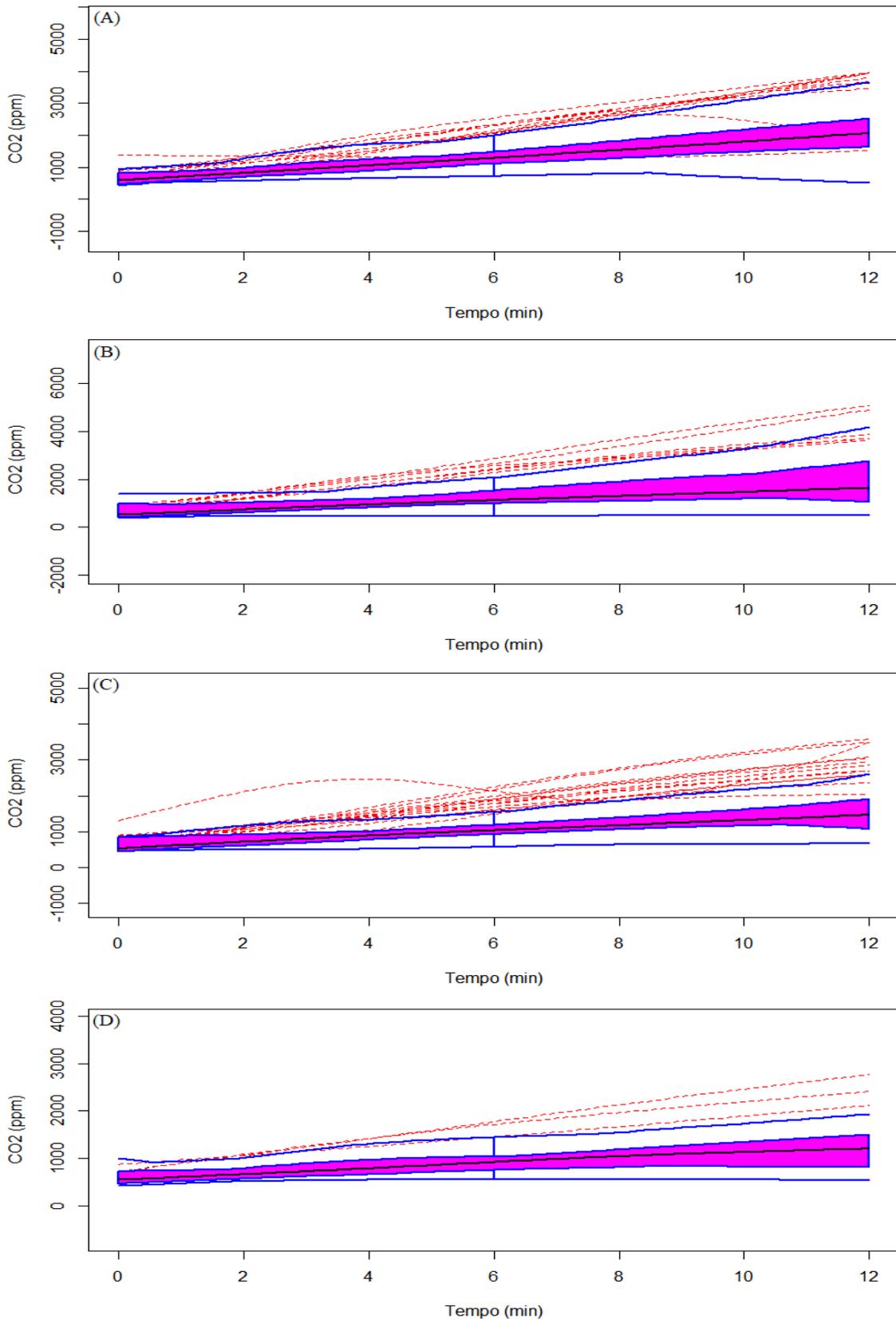


Figura 8 – Boxplot funcional das suavizações correspondente a emissão do  $CO_2$  para (A) pasto extensivo, (B) pasto intensivo, (C) linha da cana de açúcar e (D) entre linha da cana de açúcar. A curva preta, expressa a mediana funcional. A região rosa, retrata o envelopamento da região central que indica o acúmulo de 50% das funções. As curvas azuis extremas, representam o máximo e mínimo não periférico. As curvas vermelhas, são os outliers.

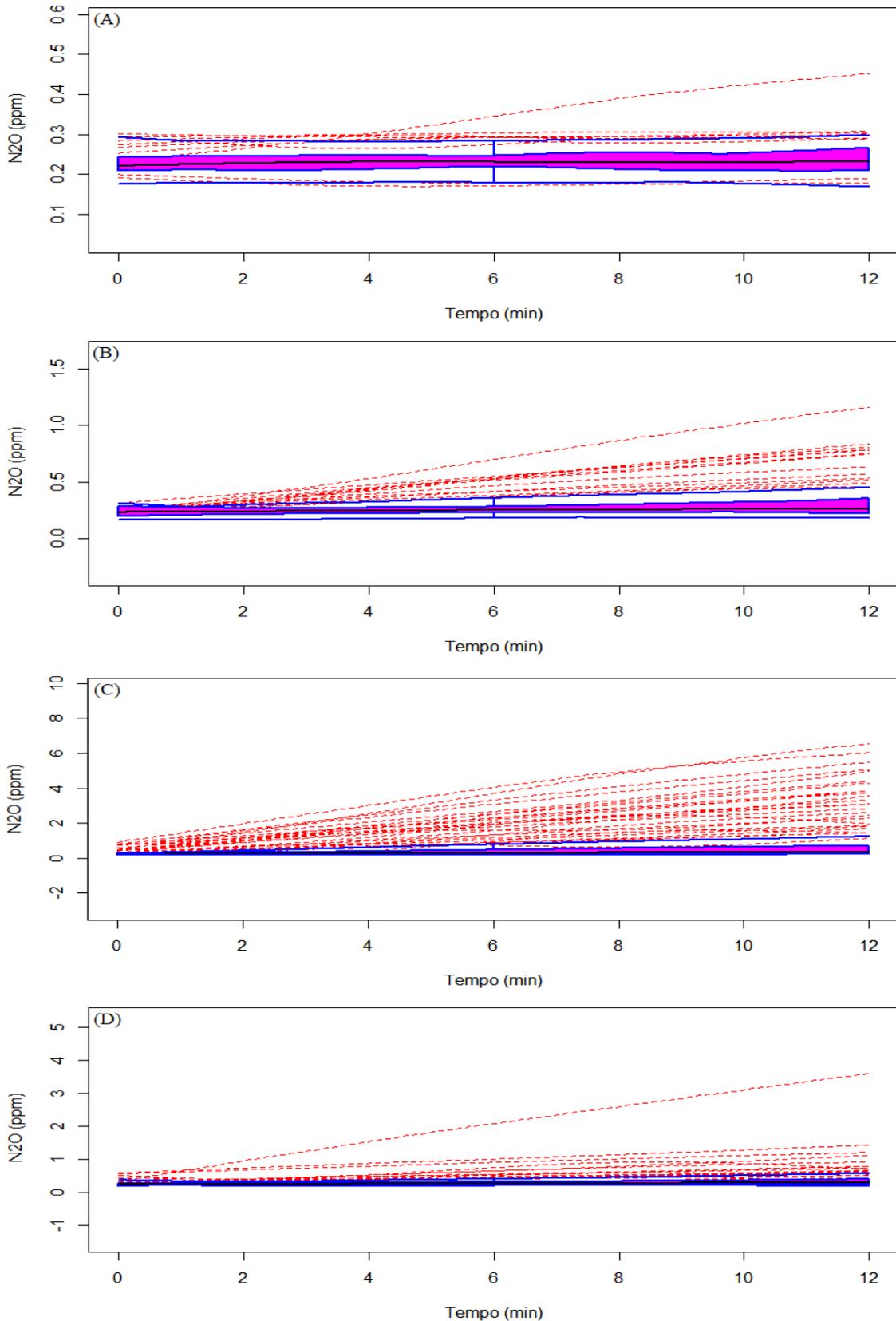


Figura 9 – Boxplot funcional das suavizações correspondente a emissão do  $N_2O$  para (A) pasto extensivo, (B) pasto intensivo, (C) linha da cana de açúcar e (D) entre linha da cana de açúcar. A curva preta, expressa a mediana funcional. A região rosa, retrata o envelopamento da região central que indica o acúmulo de 50% das funções. As curvas azuis extremas, representam o máximo e mínimo não periférico. As curvas vermelhas, são os outliers.

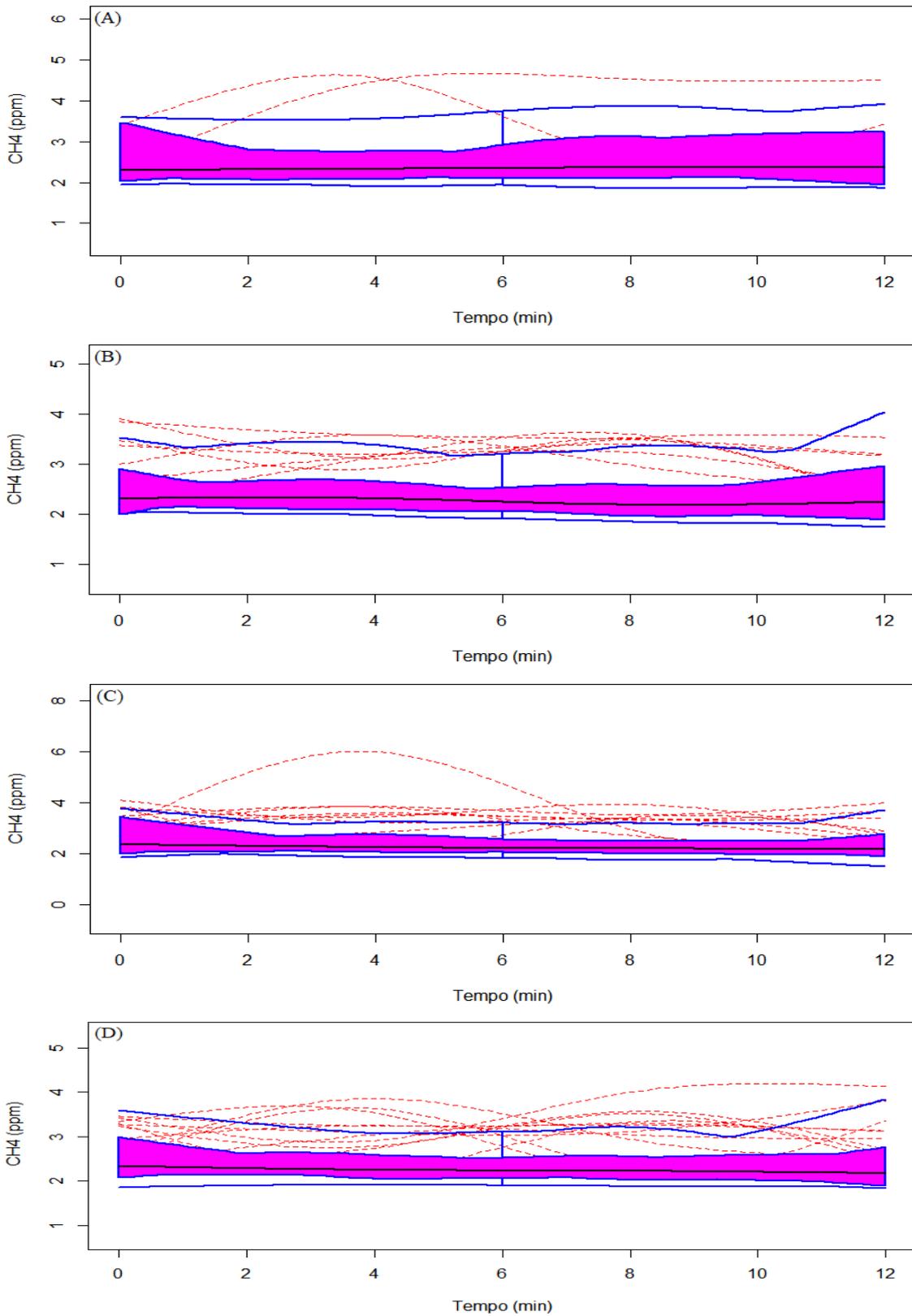


Figura 10 – Boxplot funcional das suavizações correspondente a emissão do  $CH_4$  para (A) pasto extensivo, (B) pasto intensivo, (C) linha da cana de açúcar e (D) entre linha da cana de açúcar. A curva preta, expressa a mediana funcional. A região rosa, retrata o envelopamento da região central que indica o acumulo de 50% das funções. As curvas azuis extremas, representam o máximo e mínimo não periférico. As curvas vermelhas, são os outliers.

### 5.3 Derivativos

A função suavizada descreve a posição da concentração em movimento no instante  $t$ . Isso quer dizer que as derivadas dessas funções, que estão sendo analisadas no referente trabalho, refletem intrinsecamente o fluxo dos gases, pois a natureza da emissão dos gases é justamente a variação da concentração ao longo do tempo. Então, analisar o padrão de variação dessas derivadas significam analisar o padrão de variação dos fluxos dos gases.

As derivadas das funções dos gases estão representadas nas Figuras 11, 12 e 13. Cada função exibe a variação da concentração ao longo do tempo, ou seja, cada curva funcional representa o fluxo mensurado a partir da concentração suavizada de cada coleta. As Figuras 14, 15 e 16 mostram as médias dos fluxos diários mais/menos os desvios padrão durante os 42 dias de coletas para os quatro usos da terra. 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,  $t \in [0, 12]$ ). Já o fluxo global é considerado o fluxo médio. Com esses fluxos, verifica-se os momentos de grandes variações, além dos momentos de transição do fluxo positivo para negativo conhecido como influxo, quando há uma absorção do gás no solo (LIU et al., 2010), ou fluxo negativo para positivo chamado de efluxo, quando o solo emite o gás (POMATTI, 2016). 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.

Verifica-se que os fluxos do gás carbono (Figura 11) para todos os quatro sistemas de manejo foram positivos, variando entre 0 ppm a 200 ppm para todo intervalo de tempo, mas com pequenas alterações em algumas funções. Na Figura 14, observa-se que as maiores emissões dos fluxos médio diários do  $CO_2$ , para todos os tipos de manejo, foi no começo do experimento correspondendo ao mês de dezembro de 2018 e início de janeiro de 2019. Houve grandes amplitudes de variações nas emissões do PEX e PIN, com altos desvios padrão indicando grande dispersão nas funções em relação a média e também apresentaram valores discrepantes entre os dias de coleta.

Os fluxos do  $N_2O$  (Figura 12) apresentaram fortes variações para o pasto extensivo e pasto intensivo, alternando-se entre efluxos e influxos. Já para linha da cana de açúcar, os fluxos exibiram menos variações, além de obter os maiores índices de efluxos em relação aos manejos anteriores. Para a entre linha ouve uma estabilidade entre  $-0,05\text{ppm}$  a  $0,1\text{ppm}$ . Através dos manejos da linha e entre linha da cana de açúcar verifica-se que as maiores emissões dos fluxos médio diários e os maiores valores de desvios padrão do  $N_2O$  (Figura

15) foram entre o terceiro ao décimo dia de coleta que correspondem aos dias 14, 15, 17, 19, 21, 26 de dezembro de 2018, 02 e 04 de janeiro de 2019 para a PIN, L e EL. Em relação aos outros dias, os valores ficaram próximos a zero com reduzida amplitude de variação nos fluxos. Já o PEX os maiores fluxos foram nas últimas duas coletas (dia 07 e 25 de maio de 2019). É notável que os maiores fluxos estão no manejo agrícola (linha da cana de açúcar). Esse aumento pode estar relacionado a adubação nitrogenada onde estudos, como Jantalia et al. (2008), Martins et al. (2015), Lopes et al. (2017), afirmam que índices de emissões de  $N_2O$  intensificam após fertilização nitrogenada, uma vez que aplicação de fertilizantes nitrogenados aumenta a sensibilidade da respiração do solo (OERTEL et al., 2016).

Como observado em Bento et al. (2018), La Scala Jr et al. (2005), Siqueira Neto et al. (2011), os altos valores apresentados no  $CO_2$  e  $N_2O$  no início do período de estudo podem estar relacionados com os altos índices de precipitação. Visto que o município de Brotas-SP apresenta maior concentração pluviométrica no verão, referentes aos meses de dezembro, janeiro e fevereiro (TREVISAN; MOSCHINI; GUERRERO, 2017). Segundo Valentini, Abreu e Faria (2015), a disponibilidade de água no solo atua nas relações bioquímicas relevantes dos microrganismos do solo, levando a um aumento nas emissões desses gases.

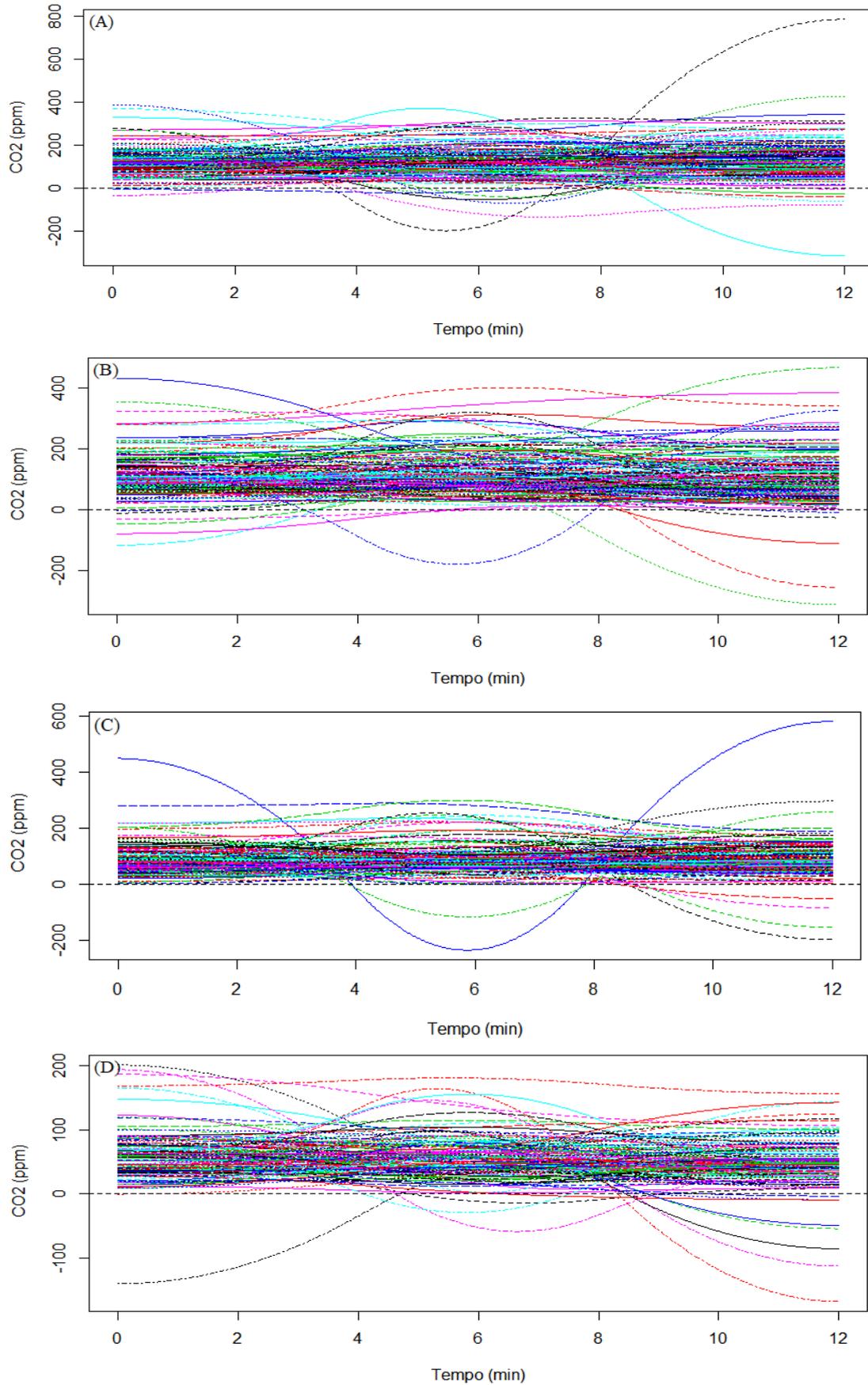


Figura 11 – Emissões dos fluxos correspondente ao  $CO_2$  nos diferentes usos do solo. (A) pasto extensivo, (B) pasto intensivo, (C) linha da cana de açúcar e (D) entre linha da cana de açúcar.

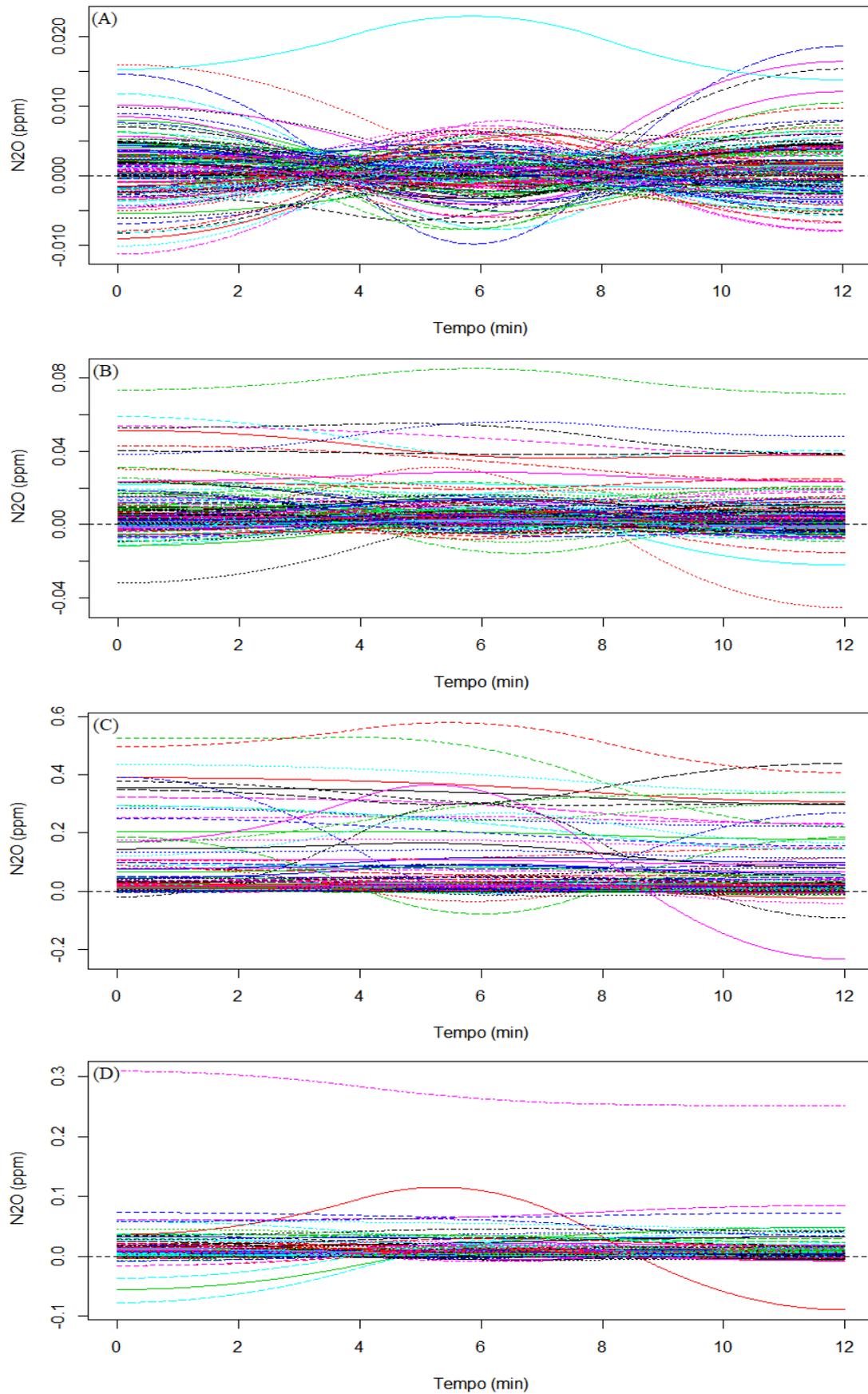


Figura 12 – Emissões dos fluxos correspondente ao  $N_2O$  nos diferentes usos do solo. (A) pasto extensivo, (B) pasto intensivo, (C) linha da cana de açúcar e (D) entre linha da cana de açúcar.

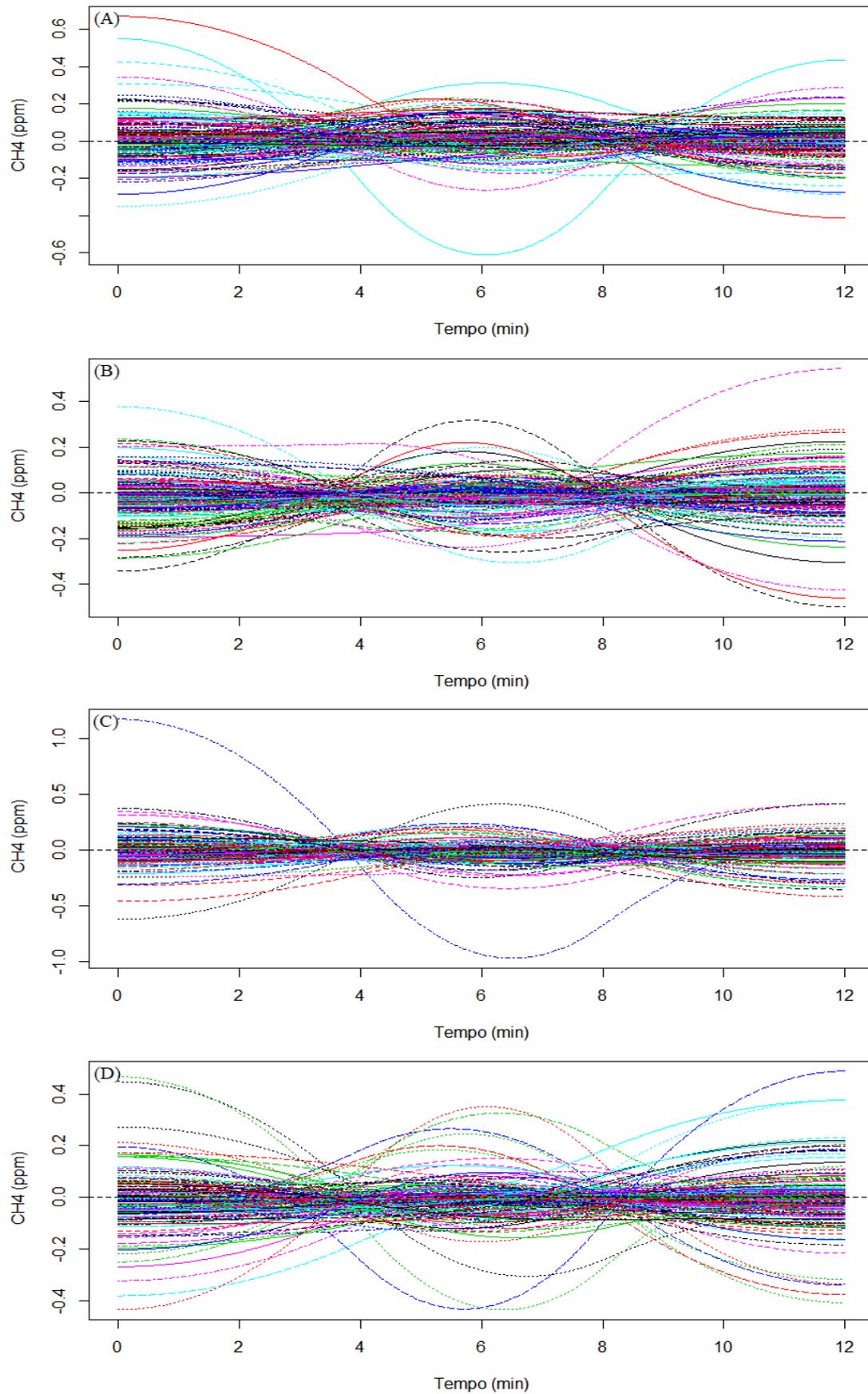


Figura 13 – Emissões dos fluxos correspondente ao  $CH_4$  nos diferentes manejos do solo. (A) pasto extensivo, (B) pasto intensivo, (C) linha da cana de açúcar e (D) entre linha da cana de açúcar.

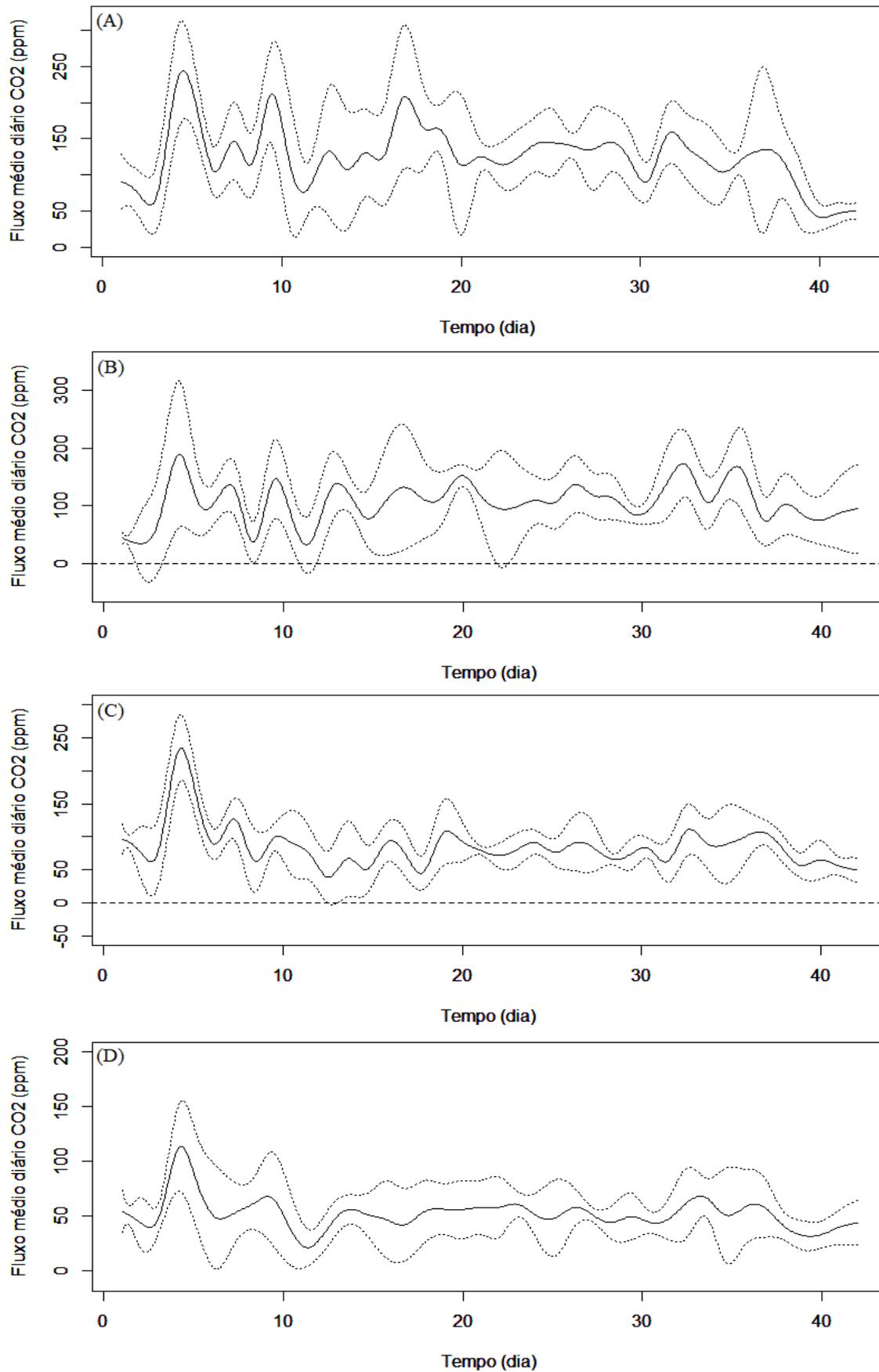


Figura 14 – Fluxo médio diário do  $CO_2$ : Médias funcionais - curvas sólidas. Médias funcionais adicionando e subtraindo seu desvio padrão - curvas pontilhadas. (A) pasto extensivo, (B) pasto intensivo, (C) linha da cana de açúcar e (D) entre linha da cana de açúcar.

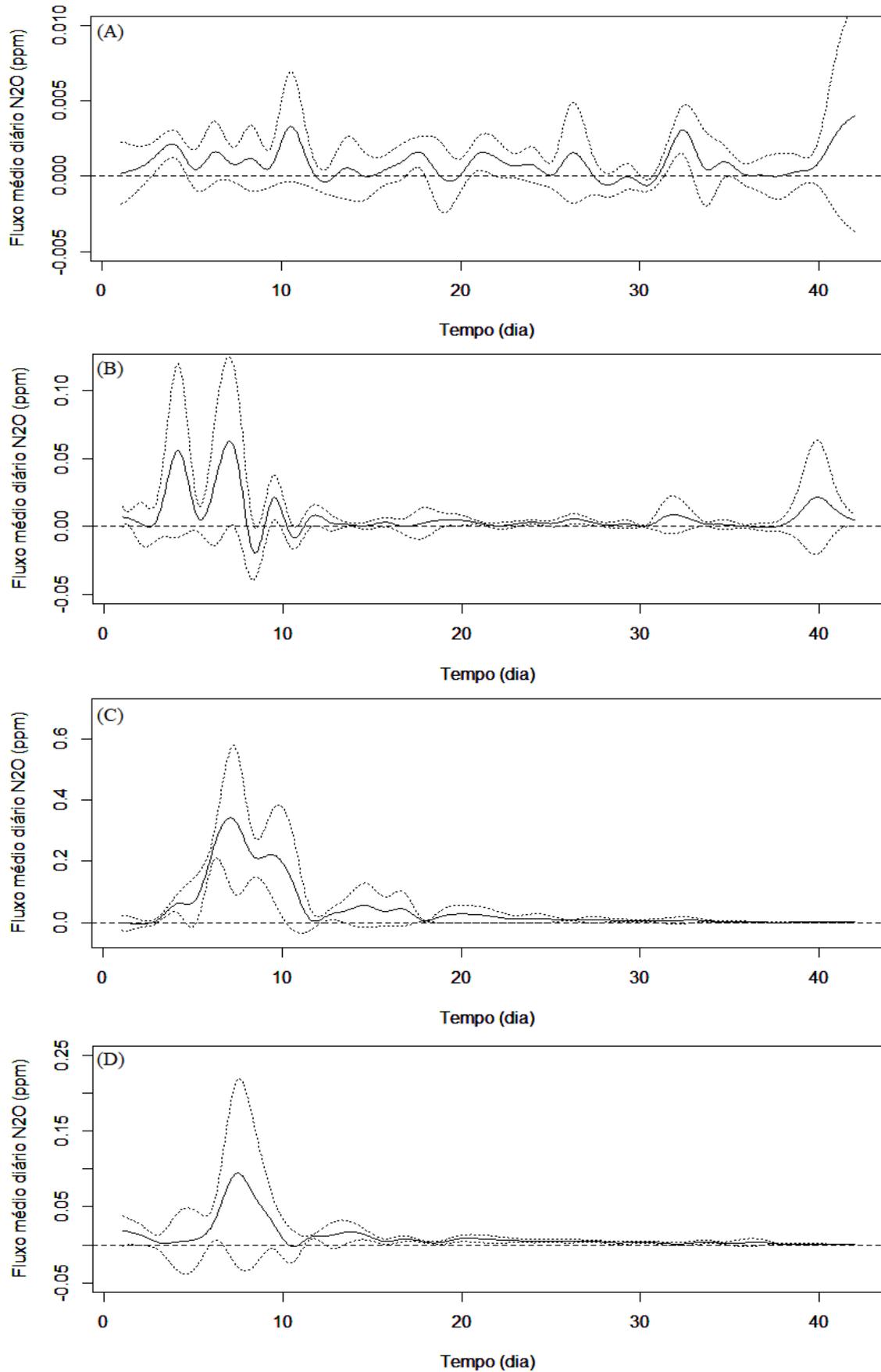


Figura 15 – Fluxo médio diário do  $N_2O$ : Médias funcionais - curvas sólidas. Médias funcionais adicionando e subtraindo seu desvio padrão - curvas pontilhadas. (A) pasto extensivo, (B) pasto intensivo, (C) linha da cana de açúcar e (D) entre linha da cana de açúcar.

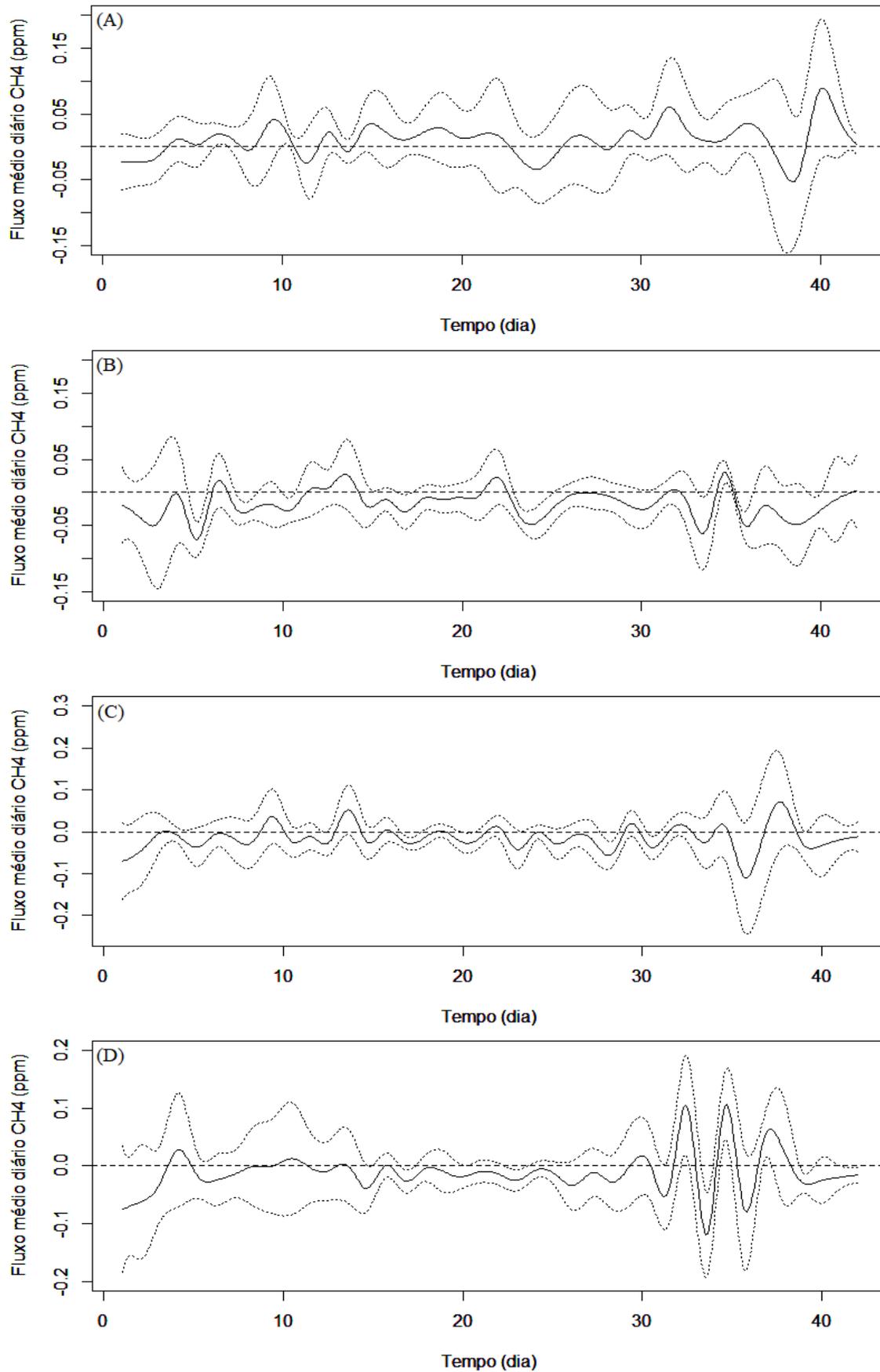


Figura 16 – Fluxo médio diário do  $CH_4$ : Médias funcionais - curvas sólidas. Médias funcionais adicionando e subtraindo seu desvio padrão - curvas pontilhadas. (A) pasto extensivo, (B) pasto intensivo, (C) linha da cana de açúcar e (D) entre linha da cana de açúcar.

Para o  $CH_4$  (Figura 13) os fluxos exibiram grande variação em todo o intervalo de tempo. Para os fluxos médio diários do  $CH_4$  (Figura 16) exibiram os maiores valores nos últimos dias de coleta, cenário inverso dos outros dois gases em estudo que obtiveram maiores valores dos fluxos nos primeiros dias de coleta. Além disso, os maiores índices de fluxo e as maiores variações do metano estão concentrado no PEX, onde quase todos os dias os valores foram positivos, isto é, houve efluxo de metano quase todos os dias de coletas. Em relação a intensificação, linha e entre linha da cana de açúcar, cerca de 80% dos dias de coletas exibiram fluxos baixos ou negativos, indicando que o solo funcionou como sumidouro de  $CH_4$ . Também existiu grande dispersão das observações para a PIN, L e EL no começo e no final das coletas.

Pode-se também observar, de maneira simplificada, a estatística descritiva do fluxo médio diário (Figura 16) dos GEE nos diferentes usos da terra. Verifica-se que os fluxos médio diários do  $CO_2$  (Figura 17-A) distribuem-se assimetricamente em torno da mediana para o pasto extensivo, pasto intensivo e a linha da cana de açúcar. Já para o uso do solo da entre linha da cana de açúcar os fluxos médio distribuem-se simetricamente. Para o  $N_2O$  (Figura 17-B), nota-se um percentual maior de outliers. Os valores dos fluxos médio diários entre os usos do solo são bem distintos, prejudicando a análise visual do pasto extensivo. Entretanto, analisa-se que para o pasto intensivo, linha e entre linha da cana de açúcar os fluxos médio distribuem-se assimetricamente em torno na mediana. Em relação ao  $CH_4$  (Figura 17-C) os dados distribuem-se simetricamente em relação a mediana para o pasto intensivo, linha e entre linha da cana de açúcar, já o pasto extensivo os dados distribuem-se assimetricamente em torno da mediana.

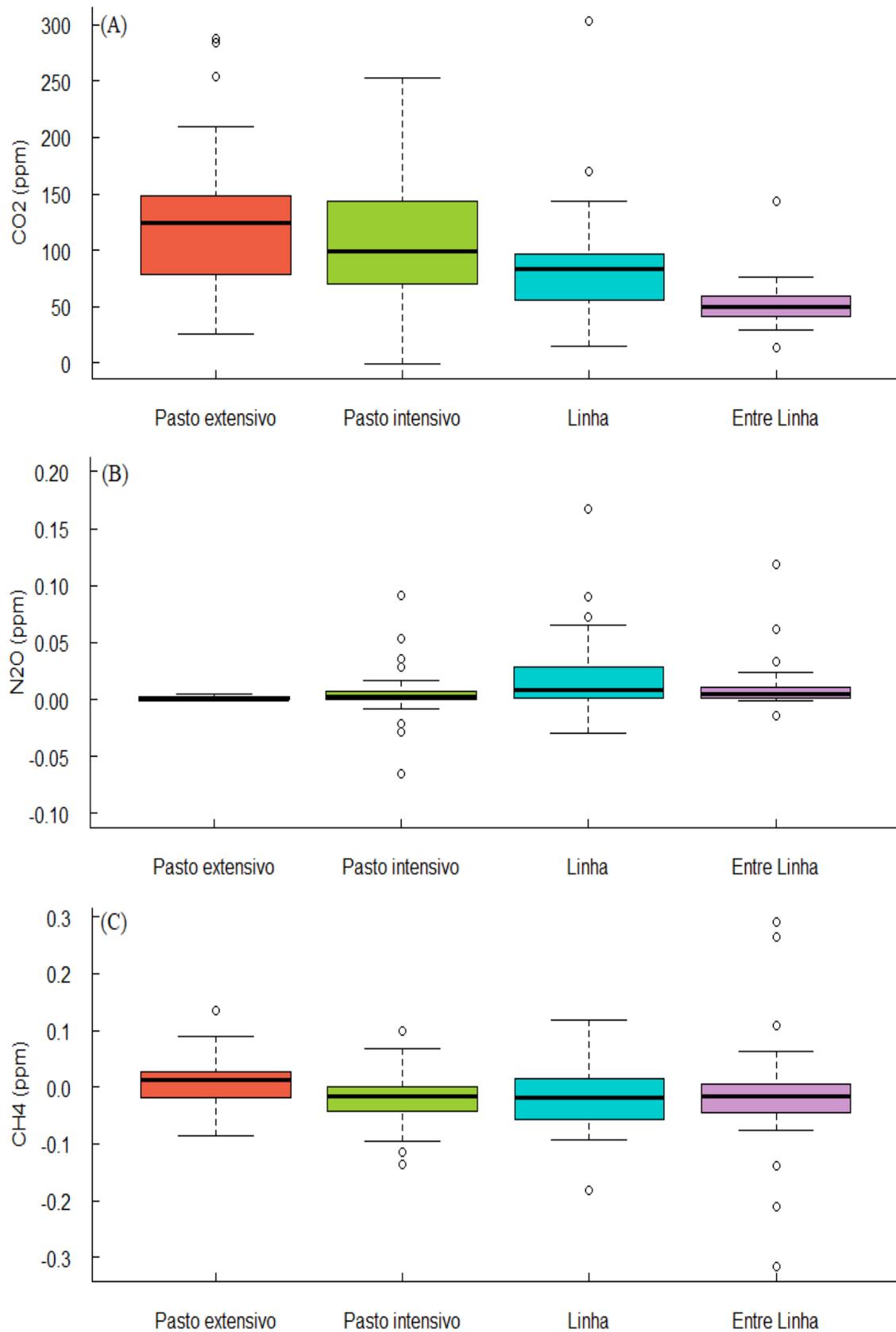


Figura 17 – Análise descritivas referentes aos fluxos médios diário de: (A)  $CO_2$ , (B)  $N_2O$  e (C)  $CH_4$  para os diferentes usos do solo.

## 5.4 Análise de componentes principais funcionais

A análise dos componentes principais funcionais (FPCA) mostra como um conjunto de observações se modifica em relação a média. Ela é utilizada para explicar a maior variabilidade possível existente nas observações, quantificando a discrepância em torno da média de cada observação.

A Figura 18 mostra os gráficos de perturbação para o primeiro e segundo componente principal funcional para as emissões do  $CO_2$ ,  $N_2O$  e  $CH_4$ . O primeiro componente principal (PC) correspondente ao  $CO_2$  (Figura 18-A) conseguiu capturar 95,8% da variação presente nos dados, já o segundo PC funcional explicou 2,3%. Para o  $N_2O$  (Figura 18-B), o primeiro e segundo componente captaram respectivamente 99% e 0,5% da variabilidade total. Os PC's do  $CH_4$  (Figura 18-C) explicaram 78,8% e 14,1% da variabilidade das observações. Os dois primeiros PC cumulativamente representam mais de 90% da variação em todos os gases. Portanto, os PC's resumem efetivamente a variabilidade amostral total e podem ser utilizados para o estudo do conjunto de dados.

A soma dos múltiplos (marcador vermelho), expressa os scores positivos nesse componente tendendo a ter as maiores emissões dos GEE ou seja, quanto mais crescente as emissões maior é o fluxo. A subtração (marcadores azuis) exibem os scores reversos nesse componente, obtendo as menores emissões dos GEE implicando em fluxos com valores menores. O primeiro PC do  $CO_2$  descreve um deslocamento crescente, tanto para as 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. Em relação ao  $N_2O$ , há um deslocamento crescente para os marcadores positivos e decrescente para os marcadores negativo em todo o intervalo de tempo, implicando em uma emissão de fluxos (positivo) e uma absorção de fluxo (negativo). As maiores emissões do  $CH_4$  exibida no primeiro PC tendem a crescer até o sexto minuto e em seguida ela começa a decrescer. As menores emissões apresentam um decaimento até o sexto minuto mantendo-se constante até o final do tempo.

O segundo PC, com um reduzido poder explicativo, possui menos variação em relação a média. A variação do PC mostra uma inversão nos marcadores, destacando-se o  $CH_4$  onde, os marcadores positivos emitiram valores diminuídos ao longo do tempo já os marcadores negativos emitiram valores aumentados no intervalo em estudo.

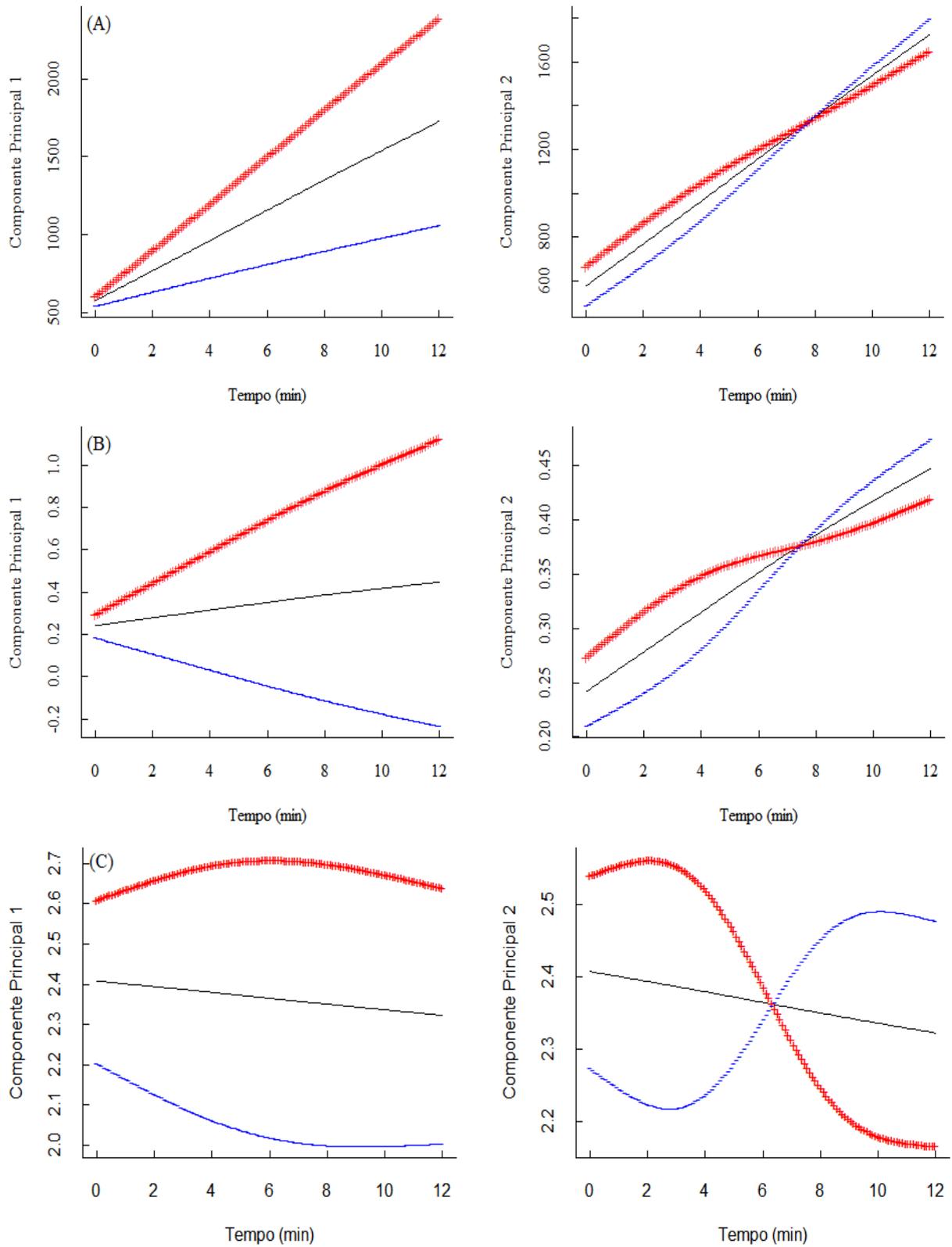


Figura 18 – Primeiro componente principal (painéis da esquerda) e segundo componente principal (painéis da direita) das emissões de: (A)  $CO_2$ , (B)  $N_2O$  e (C)  $CH_4$ , sob os diferentes usos do solo. Curvas médias (preto) das emissões dos gases de efeito estufa adicionando (vermelho) e subtraindo (azul) os múltiplos aplicada a função média do componente principal.

A Figura 19 representa graficamente os resultados da análise do PC funcional, no qual o valor de cada observação é uma tabela de dupla entrada do FPCA.1 *vs* FPCA.2. Essa análise é conhecida como biplot. O biplot vai representar a emissão dos GEE de cada coleta para os quatro sistema de manejo. Uma das funcionalidades do biplot é de poder organizar as observações em um formato que permite verificar formações de grupos.

Observa-se que, o biplot do  $CO_2$  (Figura 19-A) conseguiu separar as observações em dois grupos. O primeiro grupo corresponde ao manejo 1 (PEX), apresentando as observações que possuíram os maiores valores, ou seja, as maiores emissões. O segundo grupo corresponde ao manejo 4 (EL), exibindo as observações que têm as menores emissões do gás carbono. Porém, o biplot não conseguiu separar as observações do manejo 1 e manejo 2 (PIN) devido a sua semelhança. O maior percentual dos dados do manejo 3 (L) esta concentrada entre os dois grupos. O restante das observações estão dispersas.

Para o  $N_2O$  (Figura 19-B), há um grande quantitativo de outliers correspondendo ao manejo 3 dificultando a análise. Nota-se que existe uma aglomeração dos outros três manejos próximo à zona de estabilidade que é a região central localizada na interseção dos eixos. O biplot para o gás metano (Figura 19-C) não conseguiu formar nenhum grupo. Dentro de cada manejo, há tanto as maiores como as menores emissões. Entretanto, percebe-se que uma grande quantidade de observações no manejo 1 apresenta-se com valores superiores em comparação aos outros manejos, podendo ser considerado um grupo a parte.

Utilizando o FPCA também pode-se observar os principais padrões de variabilidade temporal diária entre as curvas médias funcionais do  $CO_2$ ,  $N_2O$  e  $CH_4$  em cada um dos sistemas de uso da terra. Para isso, necessita-se analisar os valores dos scores médios em relação aos dias de coletas, adicionando e subtraindo da média seu desvio padrão. Assim, as Figuras 20 a 25 mostram a variabilidade temporal dos GEE através do primeiro e segundo componente principal para os quatro usos da terra. O primeiro PC (Figuras 20, 22 e 24) explica praticamente toda a variação do  $CO_2$  e  $N_2O$  então, ele exhibe bastante variabilidade durante os 42 dias. Já o segundo PC (Figuras 21, 23 e 25) permanece quase estável em torno do zero. O  $CH_4$  apresentou uma menor porcentagem de variação no primeiro componente principal, produzindo uma maior variabilidade no segundo PC. Nota-se que, apenas o pasto extensivo houve mais instabilidade durante todos os 42 dias. Para os demais manejos, exibiu uma pequena variação no início e no final do período em estudo.

Percebe-se que existe uma semelhança entre a média do fluxo diário (Figuras 14, 15 e 16) e a do primeiro componente principal funcional médio diário (Figuras 20, 22 e 24). Portanto, constata-se que o FPCA conseguiu obter uma boa representação do fluxo dos gases.

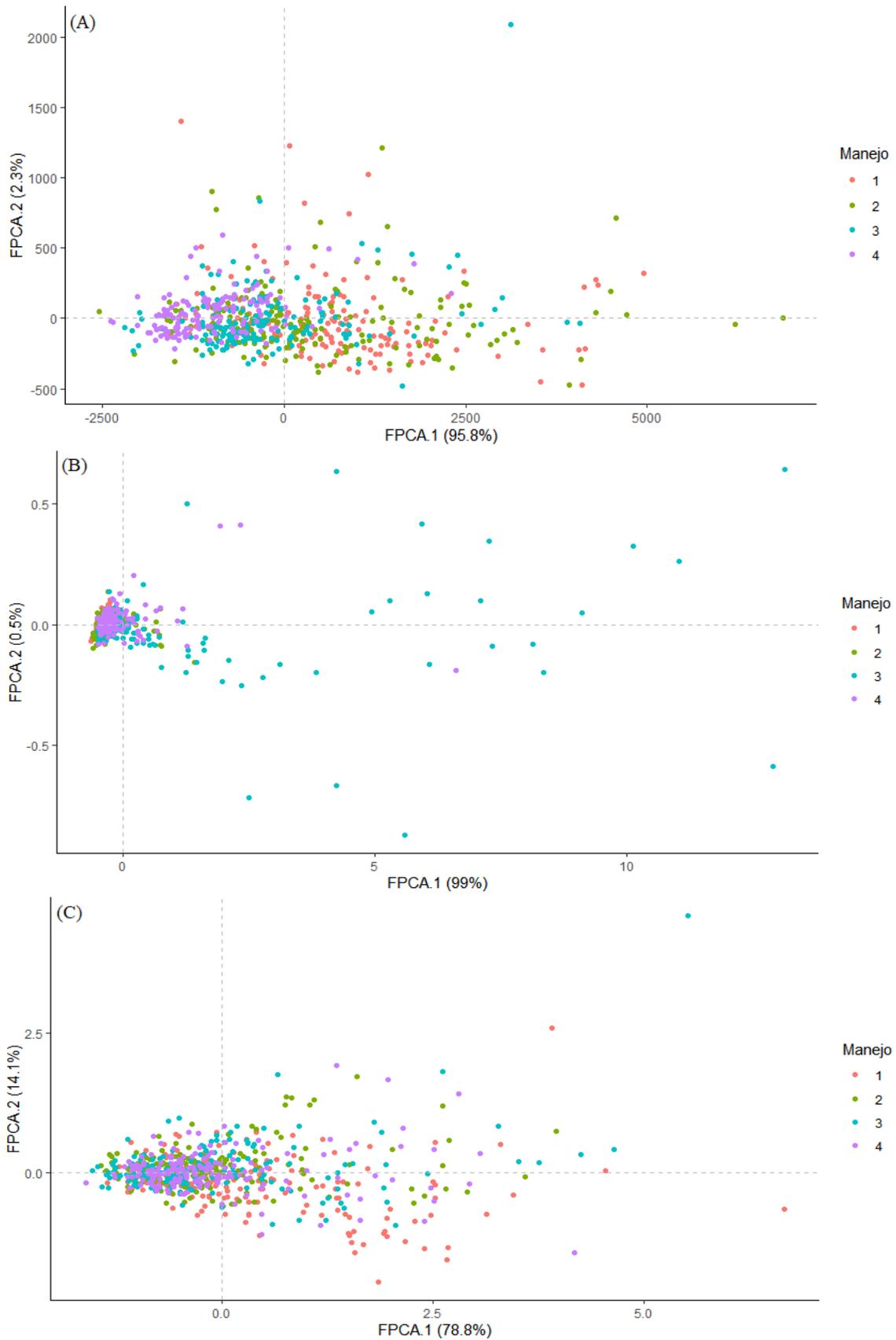


Figura 19 – Biplot dos componentes funcionais principais das emissões de: (A)  $CO_2$ , (B)  $N_2O$  e (C)  $CH_4$ , sob os diferentes usos da terra pasto extensivo (manejo 1), pasto intensivo (manejo 2), linha da cana de açúcar (manejo 3) e entre linha da cana de açúcar (manejo 4).

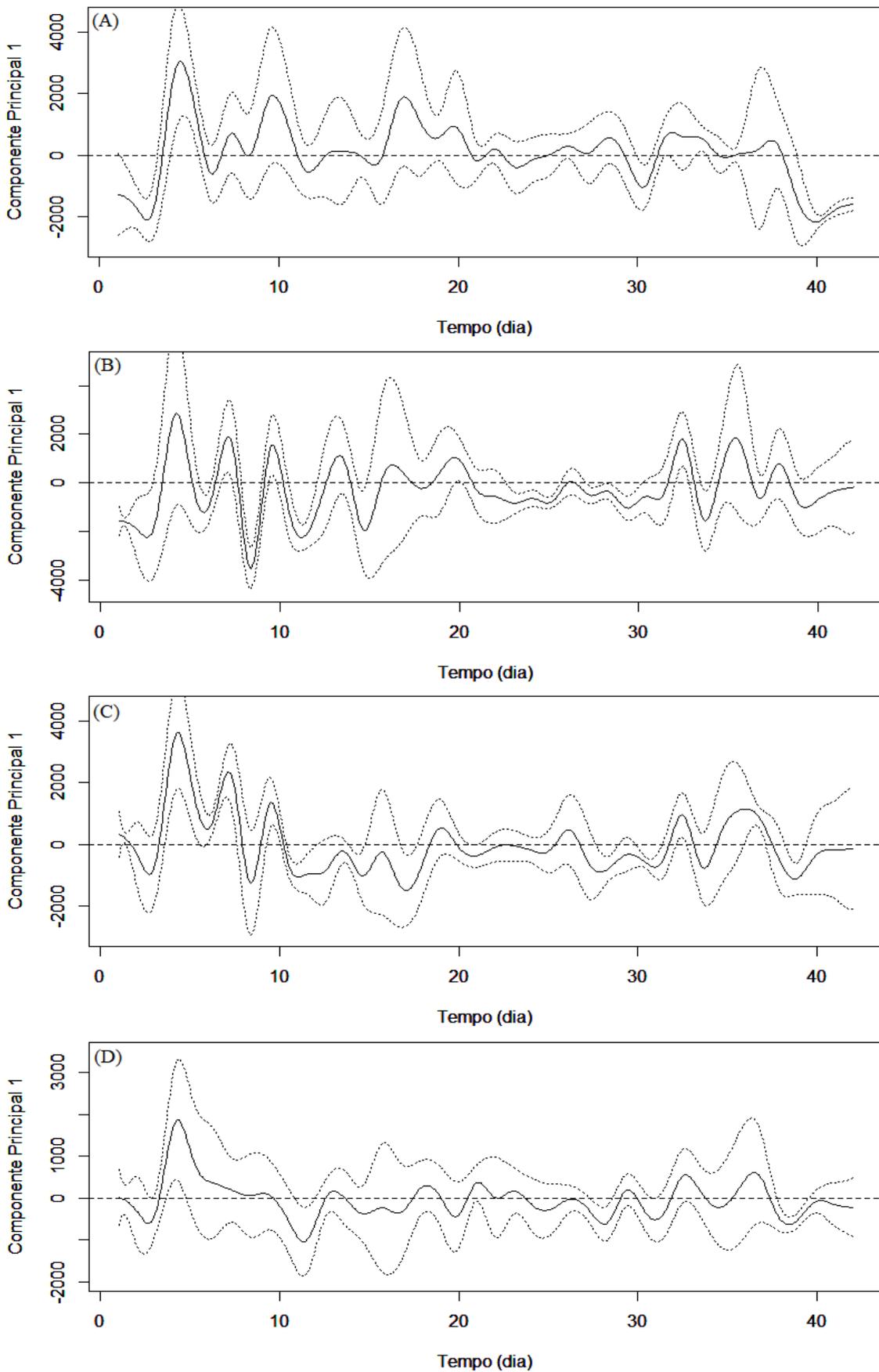


Figura 20 – Primeiro componente principal funcional médio diário, mais ou menos os desvios padrão referentes ao  $CO_2$  nos diferentes manejos do solo. (A) pasto extensivo, (B) pasto intensivo, (C) linha da cana de açúcar e (D) entre linha da cana de açúcar.

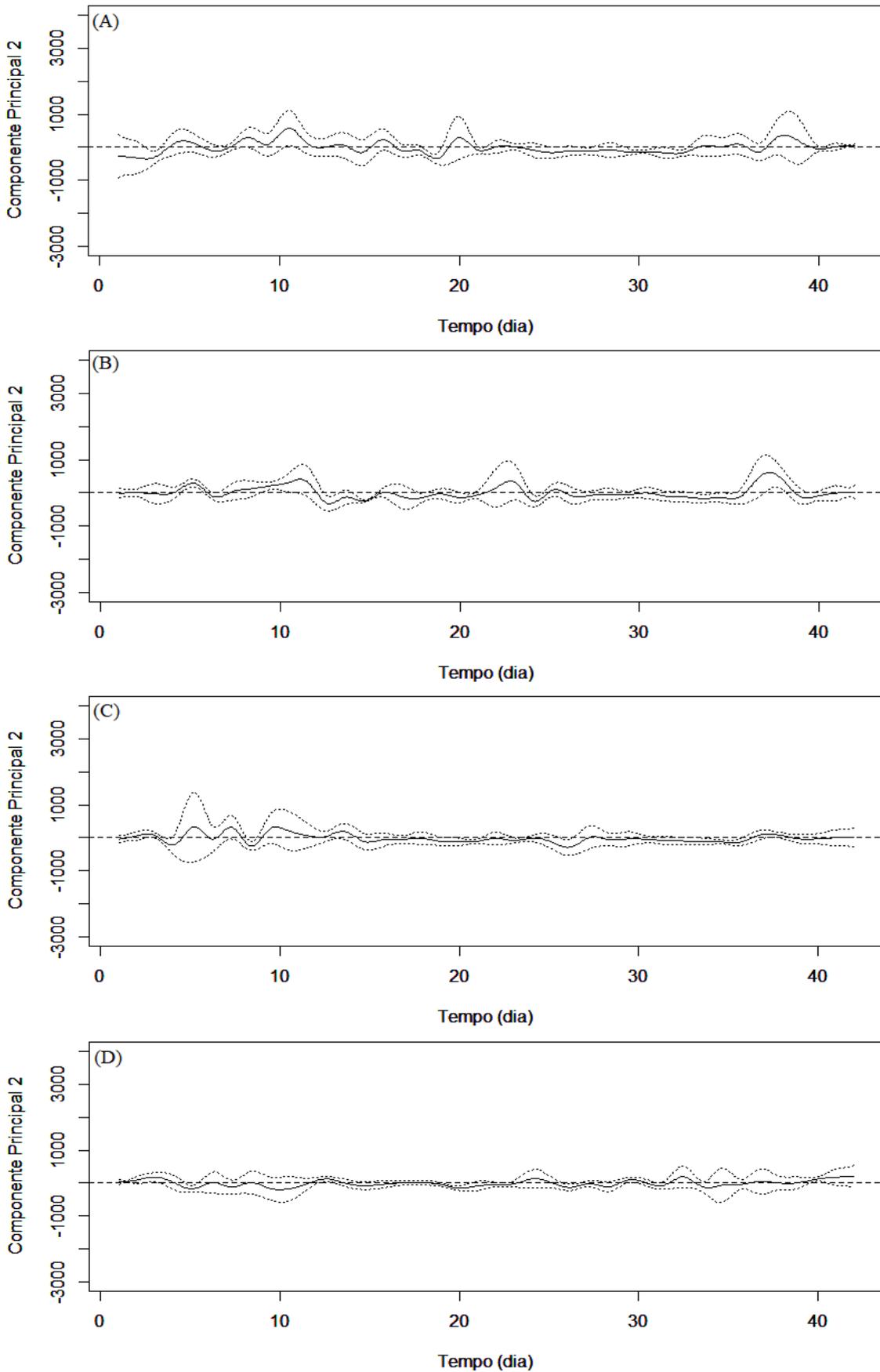


Figura 21 – Segundo componente principal funcional médio diário, mais ou menos os desvios padrão referentes ao  $CO_2$  nos diferentes manejos do solo. (A) pasto extensivo, (B) pasto intensivo, (C) linha da cana de açúcar e (D) entre linha da cana de açúcar.

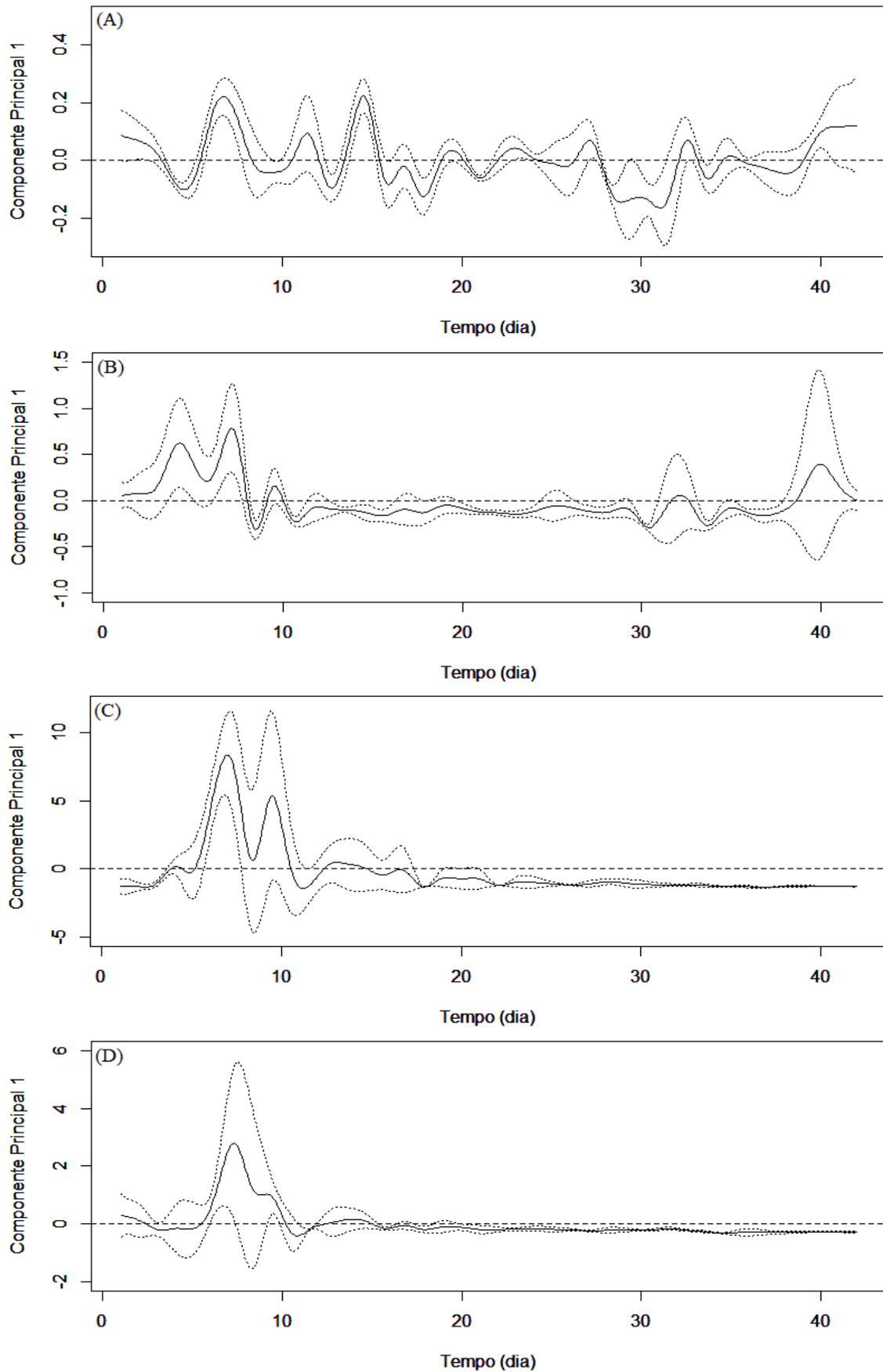


Figura 22 – Segundo componente principal funcional médio diário, mais ou menos os desvios padrão referentes ao  $N_2O$  nos diferentes manejos do solo. (A) pasto extensivo, (B) pasto intensivo, (C) linha da cana de açúcar e (D) entre linha da cana de açúcar.

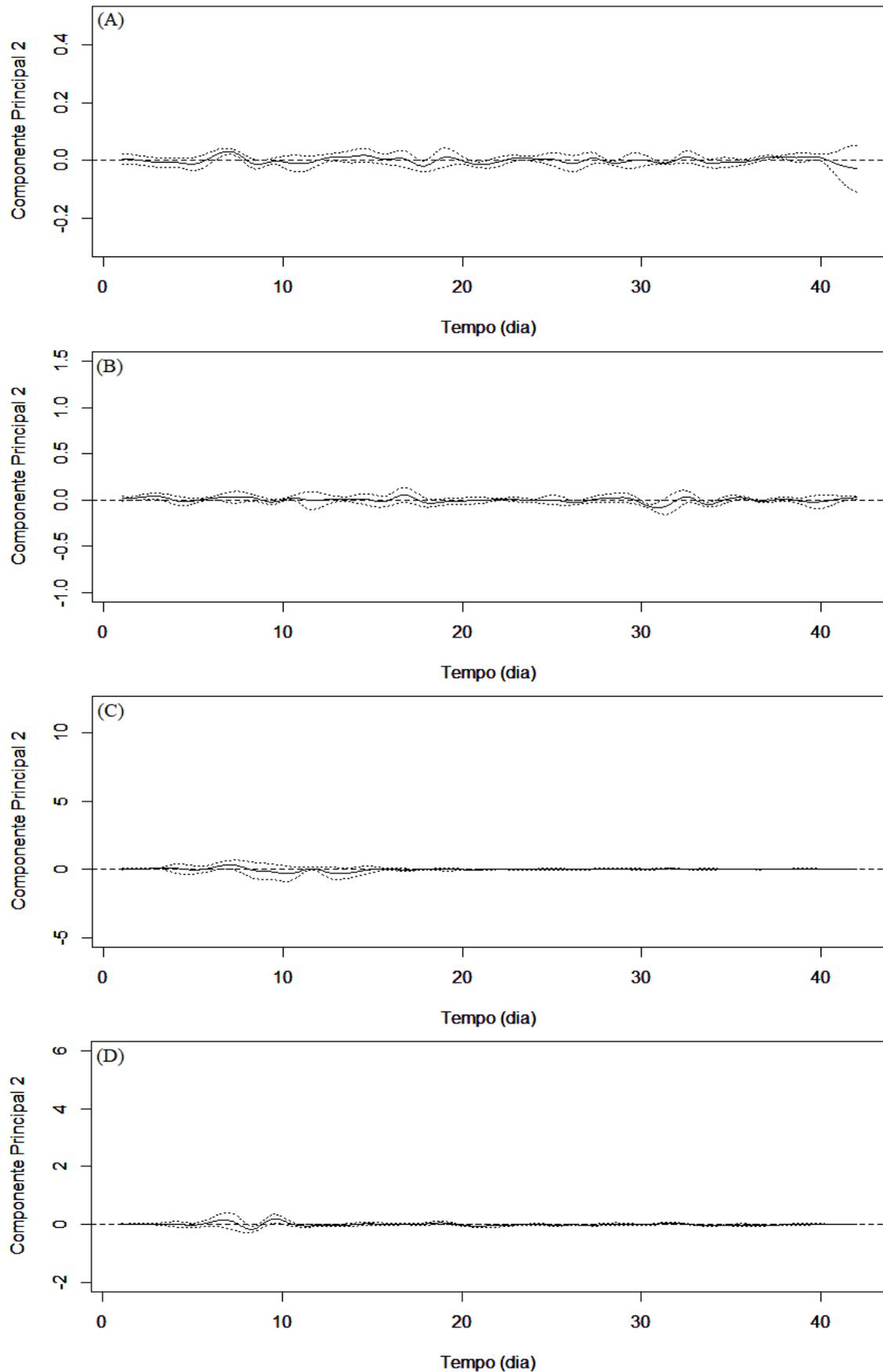


Figura 23 – Segundo componente principal funcional médio diário, mais ou menos os desvios padrão referente ao  $N_2O$  nos diferentes usos do solo. (A) pasto extensivo, (B) pasto intensivo, (C) linha da cana de açúcar e (D) entre linha da cana de açúcar.

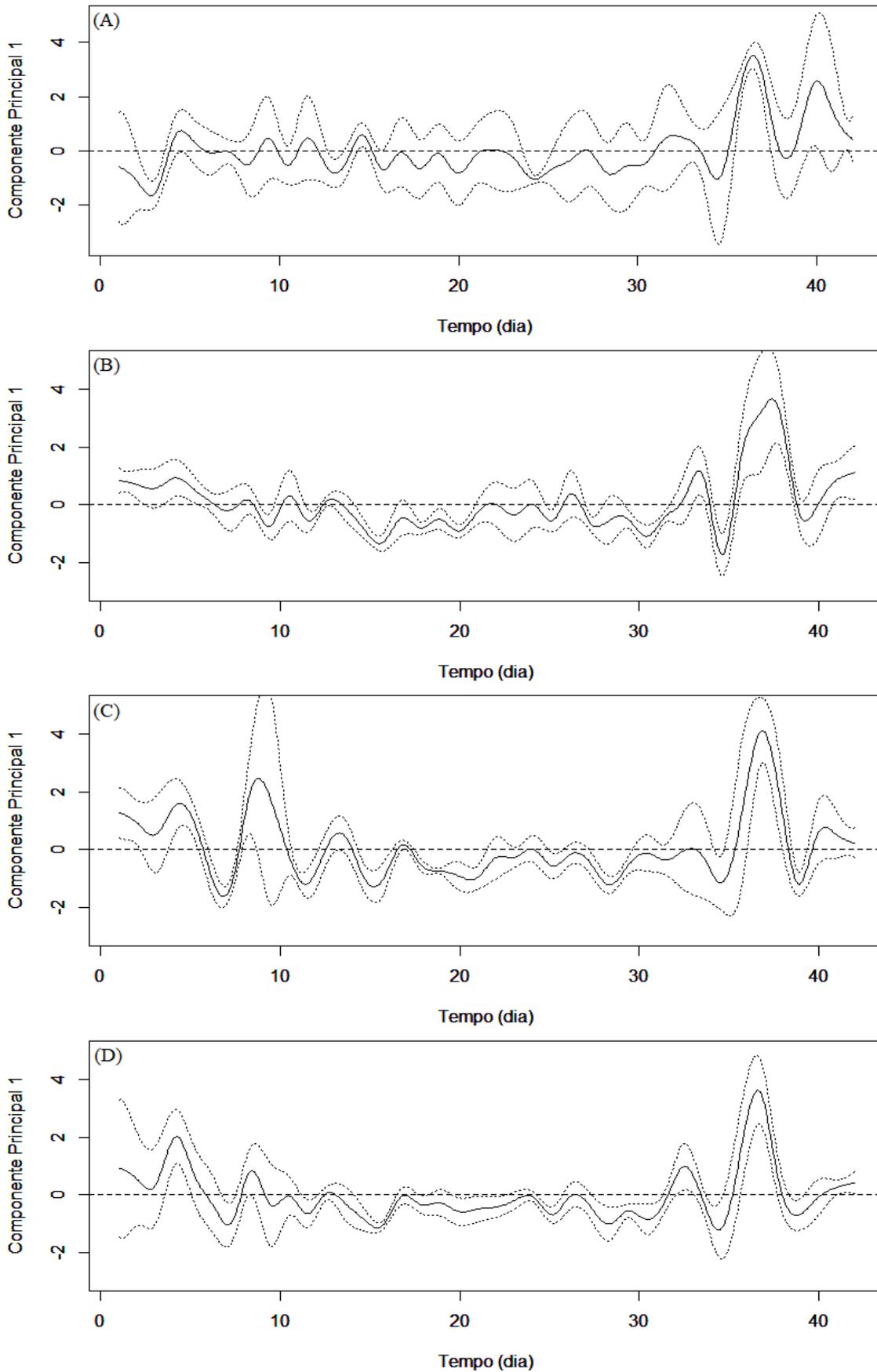


Figura 24 – Primeiro componente principal funcional médio diário, mais ou menos os desvios padrão referente ao  $CH_4$  nos diferentes usos do solo. (A) pasto extensivo, (B) pasto intensivo, (C) linha da cana de açúcar e (D) entre linha da cana de açúcar.

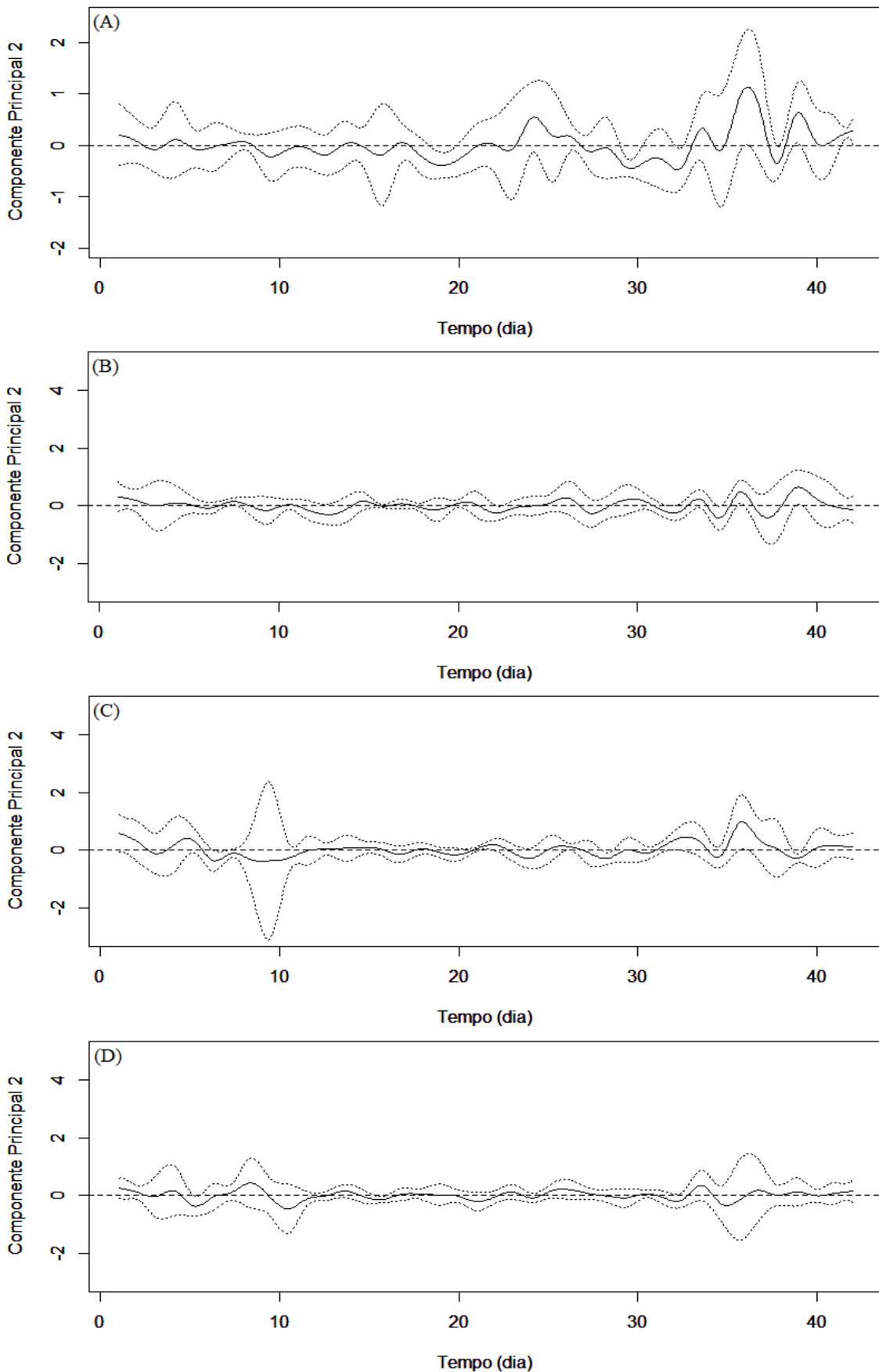


Figura 25 – Segundo componente principal funcional médio diário, mais ou menos os desvios padrão referente ao  $CH_4$  nos diferentes usos do solo. (A) pasto extensivo, (B) pasto intensivo, (C) linha da cana de açúcar e (D) entre linha da cana de açúcar.

## 5.5 Análise de variância funcional

Utilizou-se o teste FP para a análise de variância funcional. O objetivo da FANOVA é verificar se existe diferença significativa das médias das emissões dos GEE além de compreender as características dos padrões entre elas nas mudanças no uso da terra.

Para o  $CO_2$  (painel superior da Figura 26), as médias das emissões não exibem variações durante o intervalo de tempo e ainda apresentaram comportamento crescente para os quatro manejos. Os perfis de  $CO_2$  que exibem os maiores valores das emissões são o manejo 1 (PEX) e o manejo 2 (PIN) e os menores valores são o manejo 3 (L) seguido do manejo 4 (EL). Ao testar a hipótese nula ( $H_0$ ), ao nível de 5%, entre as médias funcionais das emissões do gás carbono, foi constatado que existe diferença significativa entre elas, rejeitando a  $H_0$ . Entretanto, quando analisado os perfis em pares não é possível que exista diferença significativa ( $p > 0,05$ ) entre as médias do pasto extensivo e pasto intensivo. Assim, observa-se que o resultado corresponde ao apresentado no biplot (Figura 19-A) onde não é possível separar estatisticamente esses dois manejos.

As médias das emissões do  $N_2O$  (painel intermediário da Figura 26) também não exibiram variação. Para o manejo 1, obteve comportamento relativamente constante. Comportamento bem diferente do manejo 3, onde mostrou-se média crescente com valores bem elevados em todo o intervalo de tempo. Já o manejo 2 e manejo 4, expressaram médias levemente crescente. Ao testar a  $H_0$  entre as médias das emissões para os quatro usos da terra do gás óxido nitroso, constatou que apresenta diferença estatística entre as médias funcionais, assim rejeitando  $H_0$  ( $p < 0,01$ ).

No painel inferior da Figura 26, as médias das emissões do  $CH_4$  exibiram pequenas variações. Para o PEX, ouve um crescimento no intervalo de tempo entre zero e oito minutos, continuando constante nos instantes seguintes. O PIN, L e EL houve um decaimento em todo o intervalo de tempo. Observa-se uma interação entre os quatro perfis no início do intervalo. Quando investigados as médias do  $CH_4$ , rejeitou-se  $H_0$  em comparação do pasto extensivo e os outros manejos ( $p < 0,001$ ). Entretanto, quando analisado as médias em pares verifica-se que não rejeita-se  $H_0$ , isto é, não há diferença estatisticamente entre as médias do pasto intensivo, linha e entre linha da cana de açúcar. Portanto nota-se que esse resultado também foi corroborado pelo o biplot (Figura 19-C) em que o mesmo não distinguiu os grupos em análise.

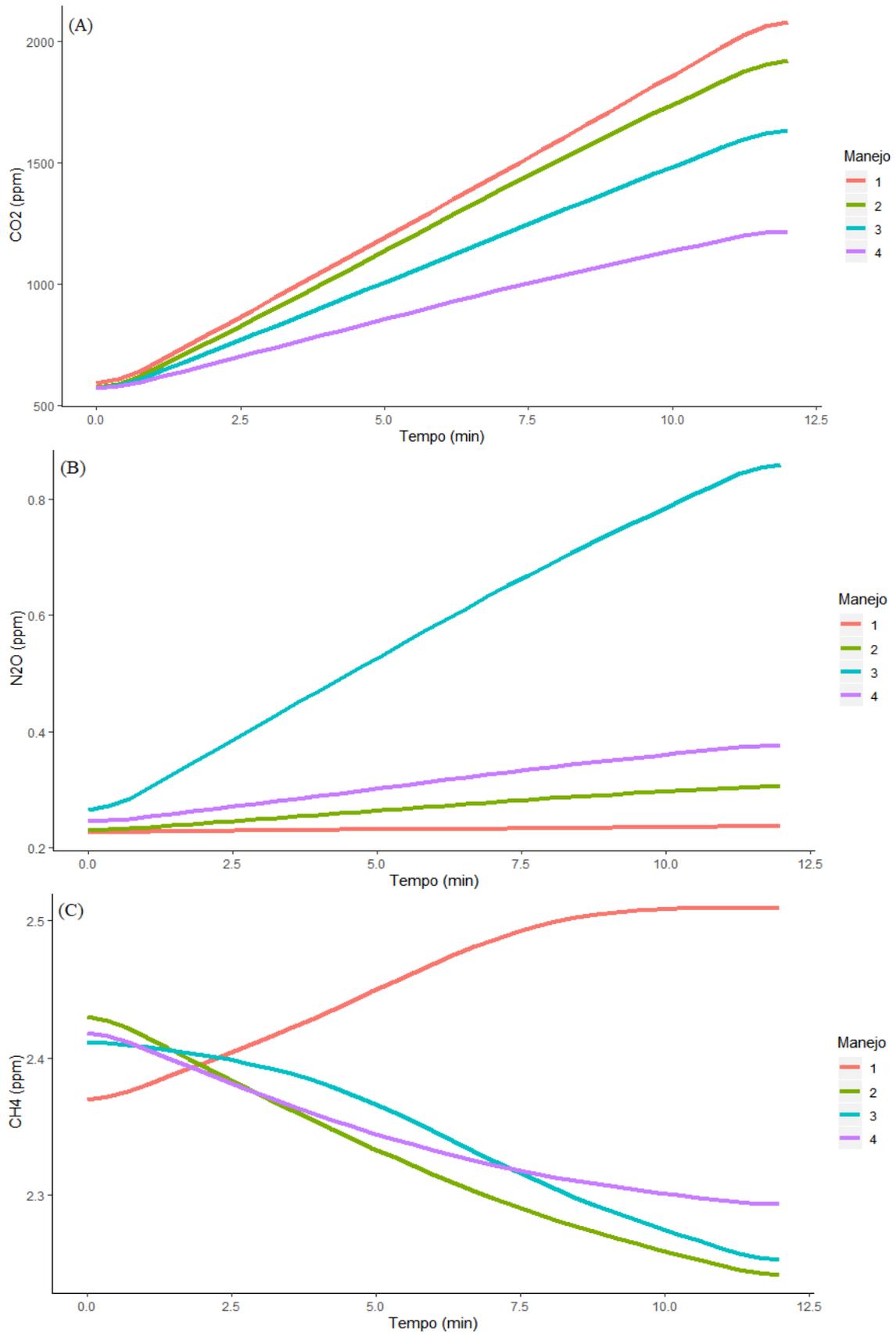


Figura 26 – Análise de variância funcional para as médias das emissões de: (A)  $CO_2$ , (B)  $N_2O$  e (C)  $CH_4$ , sob os diferentes usos do solo: pasto extensivo (manejo 1), pasto intensivo (manejo 2), linha da cana de açúcar (manejo 3) e entre linha da cana de açúcar (manejo 4).

## 6 Conclusão

Nesse estudo, aplicou-se técnicas da análise de dados funcionais aos dados das concentrações, proveniente do solo, do dióxido de carbono, óxido nitroso e metano referentes ao pasto extensivo, pasto intensivo, linha da cana de açúcar e entre linha da cana de açúcar.

Os resultado das análises demonstraram que existe uma larga variabilidade nas emissões, tanto em relação aos usos da terra como para os três gases. Essa variabilidade esta presente, principalmente, entre os dias de coleta. Através das derivadas calculou-se o fluxo, tanto o fluxo local como global, verificando que o solo atuou como fonte do gás carbono em todos os sistemas de uso da terra. As emissões dos fluxos do óxido nitroso só aumentam quando este recebeu uma intervenção no solo, como a adubação, esse aumento foi sobre a linha da cana de açúcar e entre linha da cana de açúcar. Apesar dos fluxos de  $N_2O$  terem um comportamento mais estável, após a intervenção, apresentando valores próximos a zero, em média, o solo atuou como fonte desse gás. Já no  $CH_4$ , o solo contribuiu como sumidouro do gás para o pasto intensivo, linha e entre linha da cana de açúcar.

A análise de componentes principais funcional, verificou a trajetória das emissões dos gases em relação a média, constatando que as emissões do  $CO_2$  são sempre crescente em todo intervalo de tempo em relação a média. As emissões do  $N_2O$  são crescente e decrescente, implicando que o solo atua como uma fonte e um sumidouro desse gás. As emissões do  $CH_4$  obteve uma média decrescente em todo intervalo de tempo, implicando que esse gás contribuiu mais como um sumidouro do que fonte de emissão.

Considerando os gases isoladamente para cada tipo de manejo, constatou-se que o manejo do pasto extensivo e pasto intensivo não apresentou diferença estatística para o gás carbono. Já o gás óxido nitroso apresentou diferença estatística entre as médias funcionais em todos os tipos de uso da terra. Para o gás metano, não existe diferença estatística entre as médias do pasto intensivo, linha e entre linha da cana de açúcar.

Portanto, a análise de dados funcionais conseguiu representar a modelagem dos gases de efeito estufa nos diferentes ambientes estudados, sem perder suas característica iniciais, levando em consideração a natureza contínua do sistema fotoquímico e representando o processo desses gases de uma maneira mais completa e precisa.

## Referências Bibliográficas

- 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**, Elsevier, v. 40, n. 3, p. 178–187, 2016.
- BANDE, M. F. et al. Package fda. usc. 2019.
- 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**, Elsevier, v. 206, p. 980–988, 2018.
- BERK, R. A. **Statistical learning from a regression perspective**. [S.l.]: Springer, 2008. v. 14.
- BOOR, C. D. A practical guide to splines. 2001. **Appl. Math. Sci**, 2001.
- COFFEY, N. et al. Common functional principal components analysis: A new approach to analyzing human movement data. **Human movement science**, Elsevier, v. 30, n. 6, p. 1144–1166, 2011.
- COLLIER, S. M. et al. Measurement of greenhouse gas flux from agricultural soils using static chambers. **Journal of visualized experiments: JoVE**, MyJoVE Corporation, n. 90, 2014.
- CRAVEN, P.; WAHBA, G. Smoothing noisy data with spline functions. **Numerische mathematik**, Springer, v. 31, n. 4, p. 377–403, 1979.
- DAUXOIS, J.; POUSSE, A.; ROMAIN, Y. Asymptotic theory for the principal component analysis of a vector random function: some applications to statistical inference. **Journal of multivariate analysis**, Elsevier, v. 12, n. 1, p. 136–154, 1982.
- 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.
- DENMEAD, O. Approaches to measuring fluxes of methane and nitrous oxide between landscapes and the atmosphere. **Plant and Soil**, Springer, v. 309, n. 1-2, p. 5–24, 2008.
- DONA, G. et al. Application of functional principal component analysis in race walking: an emerging methodology. **Sports Biomechanics**, Taylor & Francis, v. 8, n. 4, p. 284–301, 2009.
- FORTUNA, F.; MATURO, F. K-means clustering of item characteristic curves and item information curves via functional principal component analysis. **Quality & Quantity**, Springer, v. 53, n. 5, p. 2291–2304, 2019.

GAO, H. O. Day of week effects on diurnal ozone/nox cycles and transportation emissions in southern california. **Transportation Research Part D: Transport and Environment**, Elsevier, v. 12, n. 4, p. 292–305, 2007.

GAO, H. O.; NIEMEIER, D. A. Using functional data analysis of diurnal ozone and nox cycles to inform transportation emissions control. **Transportation Research Part D: Transport and Environment**, Elsevier, v. 13, n. 4, p. 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.

GÓRECKI, T.; KRZYŚKO, M.; WOŁYŃSKI, W. Classification problems based on regression models for multi-dimensional functional data. **Statistics in Transition new series**, Zakład Wydawnictw Statystycznych GUS, v. 16, n. 1, 2015.

GÓRECKI, T.; SMAGA, Ł. fdanova: an r software package for analysis of variance for univariate and multivariate functional data. **Computational Statistics**, Springer, v. 34, n. 2, p. 571–597, 2019.

GRANATO, D. et al. Use of principal component analysis (pca) and hierarchical cluster analysis (hca) for multivariate association between bioactive compounds and functional properties in foods: A critical perspective. **Trends in Food Science & Technology**, Elsevier, v. 72, p. 83–90, 2018.

HANUSHEVYCH, K.; SRIVASTAVA, V. Coalbed methane: places of origin, perspectives of extraction, alternative methods of transportation with the use of gas hydrate and nanotechnologies. **Mining of Mineral Deposits**, National Mining University, 2017.

HAPP, C.; GREVEN, S. Multivariate functional principal component analysis for data observed on different (dimensional) domains. **Journal of the American Statistical Association**, Taylor & Francis, v. 113, n. 522, p. 649–659, 2018.

HYNDMAN, R. J.; SHANG, H. L. Forecasting functional time series. **Journal of the Korean Statistical Society**, Elsevier, v. 38, n. 3, p. 199–211, 2009.

HYNDMAN, R. J.; ULLAH, M. S. Robust forecasting of mortality and fertility rates: a functional data approach. **Computational Statistics & Data Analysis**, Elsevier, v. 51, n. 10, p. 4942–4956, 2007.

JANSEN, M. Generalized cross validation in variable selection with and without shrinkage. **Journal of statistical planning and inference**, Elsevier, v. 159, p. 90–104, 2015.

JANTALIA, C. P. et al. Fluxes of nitrous oxide from soil under different crop rotations and tillage systems in the south of brazil. **Nutrient Cycling in Agroecosystems**, Springer, v. 82, n. 2, p. 161–173, 2008.

KELLER, M.; KAPLAN, W. A.; WOFSY, 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**, Wiley Online Library, v. 91, n. D11, p. 11791–11802, 1986.

- KIRK-DAVIDOFF, D. The greenhouse effect, aerosols, and climate change. In: **Green Chemistry**. [S.l.]: Elsevier, 2018. p. 211–234.
- 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.
- KRAUS, D. Components and completion of partially observed functional data. **Journal of the Royal Statistical Society: Series B (Statistical Methodology)**, Wiley Online Library, v. 77, n. 4, p. 777–801, 2015.
- La Scala Jr, N. et al. Soil CO<sub>2</sub> efflux following rotary tillage of a tropical soil. **Soil and Tillage Research**, Elsevier, v. 84, n. 2, p. 222–225, 2005.
- LEVY, P. et al. Quantification of uncertainty in trace gas fluxes measured by the static chamber method. **European Journal of Soil Science**, Wiley Online Library, v. 62, n. 6, p. 811–821, 2011.
- LIU, S. et al. Effects of water regime during rice-growing season on annual direct N<sub>2</sub>O emission in a paddy rice–winter wheat rotation system in southeast China. **Science of the Total Environment**, Elsevier, v. 408, n. 4, p. 906–913, 2010.
- LOPES, I. 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.
- LÓPEZ-PINTADO, S.; ROMO, J. On the concept of depth for functional data. **Journal of the American Statistical Association**, Taylor & Francis, v. 104, n. 486, p. 718–734, 2009.
- MARTIN-BARRAGAN, B.; LILLO, R.; ROMO, J. Functional boxplots based on epigraphs and hypographs. **Journal of Applied Statistics**, Taylor & Francis, v. 43, n. 6, p. 1088–1103, 2016.
- 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**, Elsevier, v. 151, p. 75–81, 2015.
- MAS, A.; PUMO, B. Functional linear regression with derivatives. **Journal of Nonparametric Statistics**, Taylor & Francis, v. 21, n. 1, p. 19–40, 2009.
- NGO, D. et al. An exploratory data analysis of electroencephalograms using the functional boxplots approach. **Frontiers in Neuroscience**, Frontiers, v. 9, p. 282, 2015.
- OERTEL, C. et al. Greenhouse gas emissions from soils—a review. **Geochemistry**, Elsevier, v. 76, n. 3, p. 327–352, 2016.
- OLIVEIRA, A. D. de et al. Emissão de óxido nitroso e metano em florestas plantadas e vegetação nativa do cerrado. In: IN: CONGRESSO BRASILEIRO DE AGROMETEOROLOGIA, 19., 2015, LAVRAS . . . . **Embrapa Cerrados-Artigo em anais de congresso (ALICE)**. [S.l.], 2015.

- ORMONEIT, D. et al. Representing cyclic human motion using functional analysis. **Image and Vision Computing**, Elsevier, v. 23, n. 14, p. 1264–1276, 2005.
- PARKIN, T.; VENTEREA, R.; HARGREAVES, S. Calculating the detection limits of chamber-based soil greenhouse gas flux measurements. **Journal of environmental quality**, The American Society of Agronomy, Crop Science Society of America, and Soil . . . , v. 41, n. 3, p. 705–715, 2012.
- 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**, Wiley Online Library, v. 61, n. 6, p. 888–902, 2010.
- PFEIFFER, R. M. et al. Two approaches to mutation detection based on functional data. **Statistics in Medicine**, Wiley Online Library, v. 21, n. 22, p. 3447–3464, 2002.
- PIRK, N. et al. Calculations of automatic chamber flux measurements of methane and carbon dioxide using short time series of concentrations. **Biogeosciences**, Copernicus GmbH, v. 13, n. 4, p. 903–912, 2016.
- POMATTI, S. R. Avaliação de nitrogênio mineral e gases de efeito estufa em sistemas de manejo de solos e rotação de culturas. 2016.
- R Development Core Team. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2019. Disponível em: <<http://www.R-project.org>>.
- RAMSAY, J.; BOCK, R.; GASSER, T. Comparison of height acceleration curves in the fels, zurich, and berkeley growth data. **Annals of Human Biology**, Taylor & Francis, v. 22, n. 5, p. 413–426, 1995.
- RAMSAY, J.; HOOKER, G.; GRAVES, S. **Functional Data Analysis with R and MATLAB. Use R, 1-207**. [S.l.]: Springer, 2009. 208 p. ISSN 978-0-387-98184-0.
- RAMSAY, J.; SILVERMAN, B. Principal components analysis for functional data. **Functional data analysis**, Springer, p. 147–172, 2005.
- RAMSAY, J. et al. Package fda. Citeseer, 2018.
- 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, n. 3, p. 539–561, 1991.
- RAMSAY, J. O.; SILVERMAN, B. **Functional data analysis**. Springer, 1997.
- RAMSAY, J. O.; SILVERMAN, B. W. **Applied functional data analysis: methods and case studies**. [S.l.]: Springer, 2002.
- RECK, R. A.; HOAG, K. J. A comparison of greenhouse gas mitigation options. **Energy**, Elsevier, v. 22, n. 2-3, p. 115–120, 1997.
- RONQUIM, C. C. et al. Dinâmica da mudança de uso e cobertura da terra e adequação ambiental dos municípios de bonito (ms) e brotas (sp). In: IN: SIMPÓSIO DE GEOTECNOLOGIAS NO PANTANAL, 7., 2018, JARDIM, MS. ANAIS . . . . **Embrapa Territorial-Artigo em anais de congresso (ALICE)**. [S.l.], 2018.

SCHADE, J. D.; BAILIO, J.; MCDOWELL, W. H. Greenhouse gas flux from headwater streams in new hampshire, usa: Patterns and drivers. **Limnology and Oceanography**, Wiley Online Library, v. 61, n. S1, p. S165–S174, 2016.

SCHUMAKER, L. Spline functions: Basic theory, john willey and sons. **Inc., New York**, 1981.

SHANG, H. L. A survey of functional principal component analysis. **AStA Advances in Statistical Analysis**, Springer, v. 98, n. 2, p. 121–142, 2014.

SILVA, M. E. S.; GUETTER, A. K. Mudanças climáticas regionais observadas no estado do paran . **Terra livre**, v. 1, n. 20, p. 111–126, 2015.

SILVA, R. W. da C.; PAULA, B. L. de. Causa do aquecimento global: antropog nica versus natural. **Terr  Didatica**, v. 5, n. 1, p. 42–49, 2009.

SIMONOFF, J. S. **Smoothing methods in statistics**. [S.l.]: Springer Science & Business Media, 1996.

SIQUEIRA NETO, M. et al. Rota o de culturas no sistema plantio direto em tibagi (pr). i-sequestro de carbono no solo. **Revista Brasileira de Ci ncia do Solo**, Sociedade Brasileira de Ci ncia do Solo, v. 33, n. 4, p. 1013–1022, 2009.

Siqueira Neto, M. et al. Emiss o de gases do efeito estufa em diferentes usos da terra no bioma cerrado. **Revista Brasileira de Ci ncia do Solo**, SciELO Brasil, v. 35, n. 1, p. 63–76, 2011.

SUN, Y.; GENTON, M. G. Functional boxplots. **Journal of Computational and Graphical Statistics**, Taylor & Francis, v. 20, n. 2, p. 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**, Wiley Online Library, v. 1, n. 1, p. 68–74, 2012.

TREVISAN, D. P.; MOSCHINI, L. E.; GUERRERO, J. V. R. Din mica temporal do uso e cobertura da terra no munic pio de brotas-sp entre os anos de 1988 e 2016. **Fronteiras: Journal of Social, Technological and Environmental Science**, v. 6, n. 4, p. 204–219, 2017.

ULLAH, S.; FINCH, C. F. Functional data modelling approach for analysing and predicting trends in incidence rates—an application to falls injury. **Osteoporosis international**, Springer, v. 21, n. 12, p. 2125–2134, 2010.

ULLAH, S.; FINCH, C. F. Applications of functional data analysis: A systematic review. **BMC medical research methodology**, BioMed Central, v. 13, n. 1, p. 43, 2013.

VALENTINI, C. M. A.; ABREU, J. G. de; FARIA, R. A. P. G. de. Respira o do solo como bioindicador em  reas degradadas. **Revista Internacional de Ci ncias**, v. 5, n. 2, p. 127–142, 2015.

WANG, S. et al. Modeling price dynamics in ebay auctions using differential equations. **Journal of the American Statistical Association**, Taylor & Francis, v. 103, n. 483, p. 1100–1118, 2008.

WARMENHOVEN, J. et al. Considerations for the use of functional principal components analysis in sports biomechanics: examples from on-water rowing. **Sports Biomechanics**, Taylor & Francis, v. 18, n. 3, p. 317–341, 2019.

WEBER, E. U. What shapes perceptions of climate change? new research since 2010. **Wiley Interdisciplinary Reviews: Climate Change**, Wiley Online Library, v. 7, n. 1, p. 125–134, 2016.

ZHANG, J.-T.; LIANG, X. One-way anova for functional data via globalizing the pointwise f-test. **Scandinavian Journal of Statistics**, Wiley Online Library, v. 41, n. 1, p. 51–71, 2014.