AGRICULTURAL SCIENCE & TECHNOLOGY

Open Access

Article

Optimizing Greenhouse Gas Emissions Studies with Functional Data Analysis in Tropical Land Uses

Mickaelle M. de Almeida Pereira, Luiz Antonio Martinelli, Janaina Braga do Carmo, and Paulo J. Duarte-Neto*

ABSTRACT: Inadequate agricultural management and soil cover changes are key drivers of greenhouse gas (GHG) emissions. Traditional statistical methods for measuring GHG fluxes often fail to account for the variability and randomness inherent to these processes. Functional data analysis (FDA) offers a novel approach to model and optimize exploratory analyses by addressing continuous variation (linear or nonlinear) and randomness over time. This study highlights the advantages of the FDA in assessing GHG emissions across three land-use treatments—extensive pasture, intensive pasture, and sugar cane—over a 182-day sampling period. FDA demonstrates superior versatility compared to conventional models, capturing the dynamic and continuous nature of gas exchange between soil and atmosphere. It allows for comprehensive uni- and multivariate analyses, effectively representing linear and nonlinear behaviors in gas fluxes. The results emphasize FDA's potential as a robust tool for studying GHG emissions in tropical agricultural systems.

KEYWORDS: *functional curves, derivatives, GHGs, land use change, carbon dioxide, nitrous oxide, methane*

INTRODUCTION

The emission of greenhouse gases (GHGs) from agricultural activities has recently been of great concern, as different land use management practices have different impacts on the composition of the microbial community, influencing the role of soil as a source or sink of the main GHG, such as carbon dioxide (CO_2), methane (CH_4) and nitrous oxide (N_2O).[1−3] Therefore, obtaining knowledge about GHG emissions across different land uses is fundamental for planning effective adaptation and mitigation strategies in future scenarios of global climatological systems.[4]

A static chamber is a widely used tool for collecting GHG concentrations from both forest and cultivated soils because of its low cost. In this equipment, concentrations are measured between 3 and 4 time points, and the flux is measured based on the temporal variation of gas concentration inside the chamber.[5−7] More often, the concentration variation in time is described as a simple linear model, where the flux represents the slope of the linear regression line between concentration and the time elapsed since the beginning of the measurements.[8,9] However, applying a linear model to concentration data can lead to systematic errors through hidden nonlinearity, and a series of factors can systematically influence changes in gas concentrations, leading to the development of a curvilinear change in concentration.[10]

To minimize this effect, several authors have proposed alternatives for measuring the flux, for example, using the exponential model, the Hutchinson/Mosier (HM) model, the nonstationary diffusive flow estimator (NDFE), the HM modified and the Quad method.[8,11] However, nonlinear models may not take into account random variation in concentration data.[12] In this sense, a method that considers at the same time the continuous variation and randomness of the gas concentration data is needed.

In this sense, one statistical family that has great potential for studying the variations of a given variable in relation to a continuum, be it time, space, or frequency, is Functional Data Analysis (FDA). It is a well-applied method of data modeling, especially those constructed from differential equations. Although FDA is a relatively new methodology, it has several statistical foundations (hypothesis testing and data exploration methods, among others) that can be used directly in observations without losing its functionality.

The main characteristics of the FDA are (i) representing data using functions to simplify manipulations and highlight the main characteristics of the data, allowing analysis of patterns and variations in these representations; (ii) working with observations that do not need to be equally spaced; (iii) working with both linear and nonlinear data sets; (iv) using derivatives and other functional values for the analyzes.[13] There is the applicability of the FDA in several areas of study, such as kinesiology,[14] climatology,[15] education,[16] environment,[17,18] neurology,[19−21] oceanography,[22] pediatrics,[23] demographic forecasting,[24,25] among others. However, there are

Received: December 16, 2024
Revised: June 5, 2025
Accepted: June 5, 2025

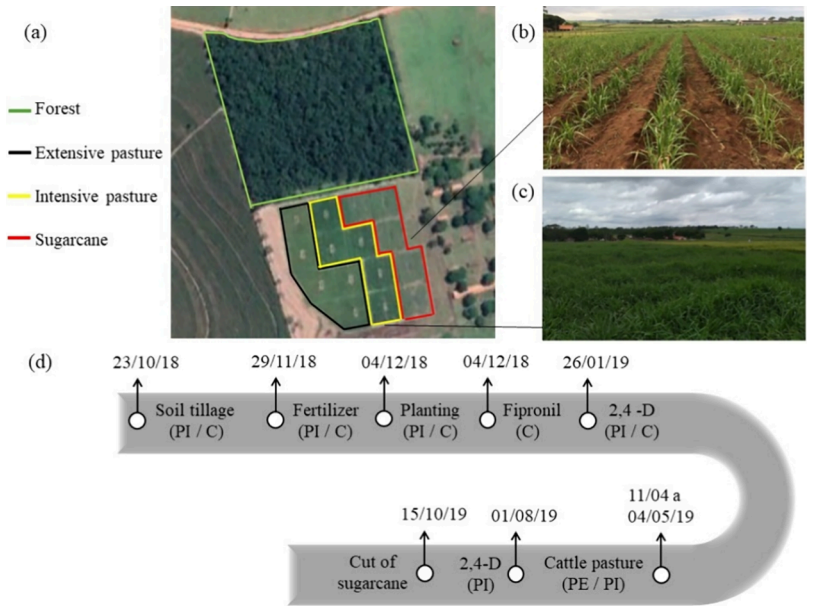

Figure 1. (a) Delimitation of the experimental area: forest vegetation (green); extensive pasture plots (black); intensified pasture plots (yellow); and sugar cane plots (red). (b) Sugar cane two months after planting. (c) Intensified pasture (dark green) two months after planting. (d) Flowchart representing the experimental design and soil management events.

still no studies aiming to apply the FDA to greenhouse gas emissions.

Therefore, the present study aims to present the advantages of using functional data analysis in the assessment of greenhouse gas (GHG) emissions from different land uses, exploring a data set that includes GHG fluxes measured during a 182-day sampling period across three land uses: extensive pasture, intensive pasture, and sugar cane. The analysis of GHG emissions from these land uses is justified as they represent three of the main types of land cover/use in Brazil, with a significant impact on agriculture. The intensification of pastures and the conversion of pastures to sugar cane involve intensive management with fertilizers and pesticides, leading to potential environmental damage, such as soil degradation, loss of biodiversity, and increased GHG emissions. Given the expansion of these land covers, understanding their environmental impacts is essential to predict consequences, mitigate negative effects, and guide more sustainable land use policies.[26−28]

Each of these land uses has distinct characteristics that directly affect greenhouse gas (GHG) fluxes. Typically, CO_2 is observed with a more linear behavior, reflecting its continuous and regular release. On the other hand, N_2O may exhibit fluctuations, alternating between regular and erratic behavior, mainly depending on soil management practices that influence microbial activity and denitrification processes. Regarding CH_4, emissions show greater variability due to the complexity of the microbial processes involved, such as methanogenesis, which can be highly sensitive to environmental conditions and soil interventions.[29,30] The diversity of these GHG fluxes is important for evaluating the versatility of FDA, which is capable of handling continuous variation and randomness in the data, providing a more robust modeling of emissions under different management conditions. Furthermore, the use of both univariate and multivariate approaches enables a more comprehensive analysis, considering both the individual effects of each land use type and their interactions.

MATERIALS AND METHODS

Study Area, Experimental Design, and Soil Management. The field sampling was carried out on an experimental farm in the city of Brotas, São Paulo, Brazil, operated by the São Paulo Agribusiness Technology Agency This choice was motivated by the area's similarity to sugar cane growing environments in Brazil, including characteristics such as the type of soil (red-yellow latosol) and the slope of the land (with a gentle slope of 4%).

The experimental model established for the research included three land use matrices: extensive pasture (EP), intensive pasture (IP), and sugar cane (C). For each land use, five 0.25 ha (50 × 50 m) plots were constructed (Figure 1), totaling 15 experimental units, with a static chamber inserted in each unit for extensive and intensive grazing and two chambers inserted in each sugar cane unit, one in the sugar cane row (CR) and the other in the sugar cane inter-row (CI). The arrangement of treatments followed the natural inclination of the research area. Consequently, the EP treatment was positioned at the highest elevation, the IP treatment occupied an intermediate position, and the C treatment was located at the lowest part of the terrain.

This experimental arrangement was designed to prevent the processes of surface runoff and leaching resulting from the more intensive applications of agrochemicals in the C and IP treatments from affecting the EP treatment.[31]

The pasture already established in the experimental area (*Brachiaria decumbens*) was maintained in the EP treatment since no soil management was carried out in this treatment, considered as the control. In IP and C treatments, the conventional soil preparation steps were carried out: (a) harrowing, liming (limestone with 70% Total Relative Neutralizing Power −2 mg ha^{-1}) and plowing (Aiveca); (b) planting pasture (*B. brizantha* cv marandu) in treatment IP and sugar cane (variety IAC SP 97-4039) in treatment C; (c) The application of nitrogen (N), phosphorus (P), and potassium (K) fertilizers at concentrations of 40, 26, 25 kg ha^{-1} in IP treatment, and 60, 65, 100, and 155, 18, 71 kg ha^{-1} in treatment C; (d) the application of pesticides based on the doses recommended by the manufacturer for the pre-emergent phase of sugar cane crops −500 g of Regent 800WG/ha (as fipronil, BASF SA), and 1.5 L of DMA 806BR/ha (as 2.4-D, Dow AgroSciences Industrial Ltda.); (e) cattle pasture (Nelore breed) for consumption of forage produced with pasture support capacity considering an animal unit (AU) of 450 kg and daily consumption of 4 of live weight.[32,33] The species *B.*

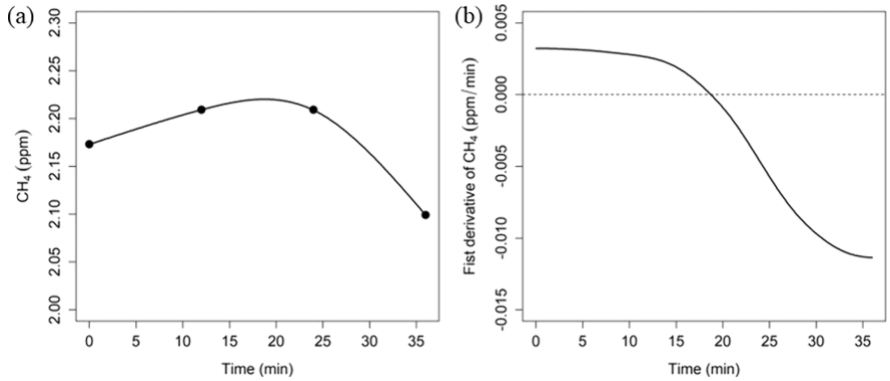

Figure 2. (a) Example of the smoothing of methane gas concentration in the extensive pasture treatment. (b) First derivative of smoothing.

brizantha cv marandu was chosen because it is widely used in Brazilian pastures and has better productivity and nutritional quality,[34] and these pesticides were selected because they are among the main ones used in sugar cane cultivation in Brazil.[35]

Gas samples were collected in each of the five experimental plots to quantify the fluxes of CO_2, CH_4, and N_2O in each of the land use plots: extensive pasture, intensive pasture, and sugar cane (row and inter-row). Samples were collected in the morning throughout the 182-day sampling period, which began in December 2018 and ended in May 2019. In total, approximately 3,360 samples were collected during 42 sampling events. The gas samples were collected in situ in cylindrical PVC chambers installed in the field, using a well-established method developed by Davidson et al.[36] The chambers were fixed between soil preparation operations, after this period they were fixed and were not removed again in order to minimize changes in the organic material on the surface and ensure good sealing. Inside, fertilizers were applied according to the area delimited by the chamber it was inserted into the soil, proportional to the fertilizer demand in the total area of the plots of each treatment.

DATA DESCRIPTION

On the collection days, four air samples were taken from each chamber, representing concentrations at four-time points (0, 12, 24, and 36 min). The samples were collected using 60 mL BD plastic syringes (Cremer SA, Blumenau, Santa Catarina, Brazil), which were immediately transferred to 30 mL glass bottles, previously sanitized, identified, sealed, and evacuated with stopper-style rubber septa (Bellco Glass, Vineland, NJ, USA).[37] After collection in the experimental area, the bottles were taken to the laboratory, where they were analyzed by gas chromatography. The samples were analyzed within a maximum period of one month after the collection date. According to do Carmo et al.,[37] gas concentrations in the bottles do not change during storage for up to 30 days. Previously prepared standards (Scott- Marrin – Riverside, CA, USA) were used to calculate the gas concentrations by comparing the areas of the peaks where integration occurs. Air and soil temperatures were also measured each time gas samples were taken.

Reconstruction of Continuous GHG Functions: Smoothing. Functional Data represents discrete observations in a real interval (time, space, frequency) as a smooth, continuous function. Thus, instead of single points in time, each function, produced from a series of successive observations, is analyzed as a single observation, making it possible to extract information from the entire process. In the case of GHG, the observed data are measured at four-time points t_{ij}, j = 1,2,3,4, where j is the number of observations for the i-th camera and $t \in [0,36]$ in minutes. The smoothing given by Ramsay and Silverman:[38]

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

was used to reconstruct the data on discrete GHG concentrations as smooth functions in relation to time, where $c_i(t)$ are real numbers called expansion coefficients and $\phi_k(t)$, i = 1,2, ⋯, K is the set of linearly combined basis functions. This procedure allows the evaluation of the flux at any desired value of time, as well as the analysis of the first derivative in the t argument. The estimation of the set of basis functions that smooth the function $r_i(t)$ will depend on the behavior of the data under analysis.[13] For the data under study, B-spline smoothing of order 4 (degree +1) was used. B-spline functions are a set of special spline functions, constructed from piecewise polynomials. Each interval is subdivided into a set of adjacent intervals that separate the nodes. Then, to smooth the piecewise polynomial, two adjacent polynomials are restricted to having their values and all their derivatives up to the corresponding order $m - 2$ at the point where they join.[39] Typically, B-spline smoothing is most commonly used because of its simplicity and flexibility in dealing with a wide range of semiparametric and nonparametric modeling situations.[40] It is used when observations have a nonperiodic behavior.[41]

To some extent, the basic representation is a compromise between interpolation and smoothing of the original data.[17] Measurements typically contain some measurement error or noise ε_{ij}.[14] Therefore, due to the random error present in the raw data, some degree of smoothing is essential to minimize it. One technique for filtering this error is the use of penalization or roughness measure.[41] The roughness penalization, measured by the criterion of the penalized residual sum of squares, is given by

$$\text{PENNSE}_\lambda = \sum_i (y_i - r(t_i))^2 + \lambda \int (D^2x)(t)^2 dt \quad (2)$$

The first term: $\sum_i (y_i - r(t_i))^2$, represents the sum of the squared errors between the observed values (y_i) and the fitted curve $(r(t_i))$. This term measures how well the fitted curve approximates the observed data. The second term: $\lambda \int (D^2x)(t)^2 dt$, adds a penalization to smooth the curve. It measures the "roughness" of the curve by calculating the integral of the square of the second derivative (D^2x), which represents the curvature.[41] In this way, PENNSE_λ seeks a

balance between fitting the data well and maintaining a smooth curve.

The degree of smoothing and the number of basis functions, KKK, will be determined through generalized cross-validation (GCV). The goal of GCV is to simultaneously optimize the smoothing parameter (λ) and the number of basis functions.[41,42] For this, a range of values for λ and the number of basis functions is defined, and simulations are run in the software. At the end of the process, the function returns the smallest GCV value, along with the corresponding list of the number of basis functions and the λ value necessary for smoothing. This procedure aims to ensure a proper fit to the data while maintaining a good generalization capability.[43]

Fifteen basis functions were used for CH_4, CO_2 and N_2O, all with a degree smoothing (λ) equal to 1.5. For instance, Figure 2a shows the adjustment of the smoothing of the methane gas concentration relative to extensive pasture, with the raw discrete data measured at four-time points. After this transformation, a complete and continuous function of gas emissions was obtained, allowing statistical analysis and preserving the variability in the original data.

Variation Pattern of Functions: Derivatives. For the function to be smooth, there must be a certain number of derivatives corresponding to it, for example, when smoothing observations using the B-splines technique of order m, we have continuous derivatives up to order $m - 2$. This is a property that displays additional information present in the estimated functions. This information is generally not available in applications of classical statistical techniques.[44] Therefore, the FDA makes it possible to use rates of change, such as speed and acceleration, to extract information from functional curves.[13] Thus, using a functional model, one or more derivatives associated with time can be obtained. Thus, as countless orders of derivatives can exist, the notation is used $D^m r_i(t)$ to designate the m-th derivative of the function $r_i(t)$.[45]

In this way, the smoothed function (Figure 2a) describes the position of the concentration in moving at time. This means that the derivative of this function (Figure 2b) reflects intrinsically the flux of gases since the nature of the emission gas is precisely the variation in concentration over time. Therefore, analyzing the variation pattern of these derivatives means analyzing the variation pattern of gas fluxes. One of the advantages of using functional data is the measurement of the instantaneous flux using the derivatives. Thus, it is possible to analyze the variability of gas fluxes concerning different land uses.

Calculation of the Flow of GHGs Using Linear and Exponential Models. In addition to measuring the flux of GHGs through the functional model, which takes into account the continuous representation of data over time, the gas fluxes were also calculated using two of the most used models. In the linear model, the flux is calculated based on the slope of the straight line. In turn, the exponential model was developed for when the rate of variation of the concentration of gases in the free space of the chamber is not constant; this methodology is based on the exponential method and is known as the Hutchinson and Mosier (HM) method,[9] a nonlinear method that takes into account the exponential curvature of the data.[46] Thus, the value of the fluxes using the linear model was determined from the slope of the regression line, the fluxes from the exponential model were given by the HM model, and for the functional curves, the flux was calculated from the derivative of the functional representation.

The fluxes obtained by the three models were compared using Lin's concordance correlation coefficient (CCC), which quantifies the agreement between two measurements, used for continuous variables. Thus, let X and Y be two variables, the CCC is determined as follows:

$$\mathrm{CCC} = 1 - \frac{E[(X-Y)^2]}{E_I[(X-Y)^2]} = \frac{2\sigma_{XY}}{\sigma_X^2 + \sigma_Y^2 + (\mu_X - \mu_Y)^2} \tag{3}$$

where $\mu_X = E(X)$, $\mu_Y = E(Y)$, $\sigma_X^2 = \mathrm{Var}(X)$, $\sigma_Y^2 = \mathrm{Var}(Y)$, and $\sigma_{XY} = \mathrm{Cov}(X, Y)$, and $E_I[\,]$ is hope, assuming independence between two observers.[47,48] In addition to the value of the agreement coefficient, it is also possible to generate other values using the CCC method.[49] For this analysis, the following were defined: $0 < \mathrm{CCC} \leq 0.6$, poor; $0.6 < \mathrm{CCC} \leq 0.85$, moderate; $0.85 < \mathrm{CCC} \leq 0.95$, substantial; $\mathrm{CCC} > 0.95$, excellent.[49−51]

Dimensions of Function Variation: Functional Principal Component Analysis (FPCA) Univariate and Multivariate Case. Functional principal component analysis (FPCA) is used to explore not only the variation in the CH_4, CO_2, and N_2O emission curves, but also the dimensions/modes in which they vary. That is, the variations in the curves of these gases are dominated by certain main patterns that can be summarized in a small number of representative modes with little loss of original information.[17,18] Through the modes of variation of gases, it may be possible to explain the impact of the uses of gases in different soils. In this sense, the modes of variation for the gases were analyzed here from a univariate perspective, where each gas was analyzed separately, and a multivariate perspective, where the gases were analyzed jointly using FPCA.

Concerning the univariate functional context, each principal component is established through an eigenfunction ($\xi(t)$), not defined in the same time interval as the functional data, in this case $t \in [0,36]$. Each eigenfunction determines the main characteristics of the variability of the curves for each gas separately. The principal component (PC) scores of the individuals in the sample are the zi values are given by

$$z_i = \int \xi(t) r_i(t) dt \tag{4}$$

The purpose of FPCA is to find the weighting function $\xi 1(t)$ that maximizes the variation of scores subject to the restriction $\int \xi_1 t^2 dt = 1$. The unitary restriction on the sum of squares in the weights is fundamental for the problem to be accurate. Without this restriction, the mean square values of the linear combination could become randomly large. The next principal components (second, third, and higher order) are solved similarly, but with additional restrictions.[45]

The extension of univariate FPCA to multivariate functional data is of high practical relevance for revealing the joint variation of different variables.[19] The objective is to analyze the simultaneous variations of the CH_4, CO_2, and N_2O profiles as a function of time for the four land uses. For each temporal position, an observation is defined as three average profiles to each gas, described as continuous functional curves through a decomposition on a B-spline on the coefficients of the B-splines smoothing model, for each of the gases. It transforms the original variables into new variables that are uncorrelated linear combinations of the originals and concentrates the variance of the system. These new variables are principal components and represent the most significant modes of data

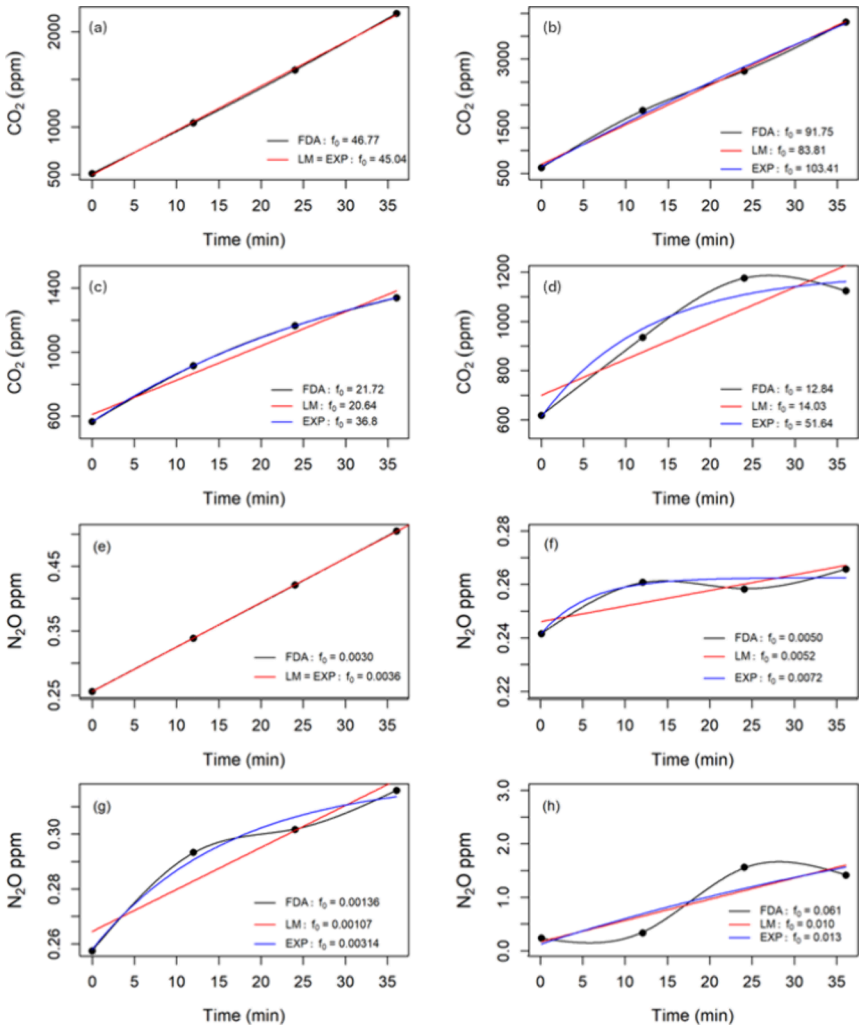

Figure 3. Examples of situations, where CO_2 (a, b, c, d) and N_2O (e, f, g, h) emissions show a gradual change in the linearity of concentrations over time. Each point represents the collection of gas concentration within a 0−36 min interval, carried out using a closed static chamber in the morning period. All four figures presented are examples of measurements taken from extensive pasture to CO_2 and sugar cane to N_2O but on different days. For the analysis of the fluxes (f_0) of these data, they are evaluated using the linear model (LM), the HM model, and the functional model (FDA). For the LM model, f_0 is measured from the regression line (red line). For the HM model, f_0 is measured from the exponential curve (EXP) (blue line), and for the functional model (FDA), f_0 is calculated from the derivative of the function (black line).

variation in soil structure and allow quantification of the amount of variation contained by each gas for each component.[22,52,53]

To modify the shape of the average curves and analyze different patterns, we add and subtract a specific value from the mean function of the principal component. This value is defined based on previous studies,[22,53,54] corresponding to the square root of the eigenvalues multiplied by the eigenfunction. This approach allows for a better exploration of the relationships between the curves and other variables of interest.[55,56]

Cluster Analysis. In the functional context, cluster analysis has been carried out by projecting functional data onto a set of lower-dimensional basis functions, such as the univariate or multivariate functional principal components, on which most of the characteristics of functional data are built.[57,58] Cluster analysis aims to group a set of data in such a way that the observations within the group are more similar to what occurs in the groups referring to a metric.[59]

Initially, the similarity between two observations is calculated, representing the degree of correspondence between them in all the characteristics used in the analysis. In this way, a distance measure is used, where the lower the distance value, the greater the similarity between the observations. Five different distance metrics were tested in the present work (Euclidean, Chebychev, Manhanttan, Canberra and Minkowiski), however, the Euclidean distance was chosen, given that it was the metric that presented the best separation to both land use and the influence of soil management. The assumed Euclidean distance is defined by $d_{ij} = \left(\sum_{k=1}^{p}(x_{ik} - x_{jk})^2\right)^{1/2}$ where x_{ik} is the value of the variable k referring to the observation i and x_{jk} represents the variable k for the observation j.[60,61]

Land-Use Classification Analysis: Support Vector Classification (SVC). PC scores were used to perform a classification concerning land uses, the machine learning algorithm using the Support Vector Classification (SVC) function. The SVC algorithm is based on Chang and Lin.[62] The data was organized considering the three main components, daily air and soil temperature as independent variables and the land uses (extensive pasture, intensive pasture, sugar cane row and inter-row) were considered as

dependent (target) variables. For the training section, data was randomly divided into a 70% training set and a 30% testing set.

The training data was used to generate the hyperparameters using the cross-validation method (k-fold CV). In this step, the training data (70% of the total data) is divided into k data sets. The model is then trained considering $k - 1$ sets and evaluated on the remaining set. Therefore, the process is repeated k times.[63] The best hyperparameters will be those that perform best according to a given metric. For the present study, the accuracy was considered to determine the parameters. Accuracy represents the number of observations correctly classified by the model divided by the total number of observations. After determining the hyperparameters, the model is validated on the test set.

A good way to analyze classification performance is to use the confusion matrix. This makes it possible to analyze the number of instances X that are classified as Y.[64] For example, it is possible to check how many times intensive pastures were classified as extensive pastures. In addition to visual analysis, it is also possible to formulate different metrics with the help of the confusion matrix, since a lot of information can be extracted from it. The first metric used is precision, which represents the accuracy of positive predictions. For a more complete analysis, precision is usually used with the recall metric, also commonly referred to as sensitivity. Finally, it is possible to combine the two above metrics into a single known F1 score metric. As F1 is a harmonic mean, the value of F1 will only obtain high values if Precision and Recall are high.[64]

For the statistical analysis of the data, RStudio statistical software[65] was used, which is an integrated set of computer tools that allows data manipulation, statistical analysis and the production of graphs. The following packages were used: gasfluxes, to calculate the flow in the linear and exponential model,[49] epiR, to measure the correlation coefficient of agreement between the flows of the models,[46] fda, for univariate functional statistical analysis,[53] fda.oce, to carry out multivariate functional statistical analysis[54] and the accumulated flow was calculated using the function developed by Bi and Kuesten.[66] For classification, the results were generated using the Jupyter notebook using the Python programming language (version 3.11).

RESULTS AND DISCUSSION

Comparison among Functional, Linear, and Exponential Models on Different Land Uses. To illustrate the differences between the fitting behavior of the three models evaluated, cases where there was a gradual change in the linearity of gas emissions were chosen and shown in Figure 3. A visual inspection of Figure 3a,e suggests that there is no difference between the flow estimated by the linear and exponential models when the gas concentration data has a high degree of linearity. As for the functional model, the estimated flow is slightly different from the other two models. However, as this behavior changes, becoming nonlinear, the values show a very large difference. When analyzing patterns similar to those in Figure 3c,d,f,g, it was found that the flux measured by the exponential model is always high, and these patterns are often found in CH4. Some flow values calculated by the exponential model can be up to four times higher than the flow values obtained by the functional model. In Figure 3h, which represents N_2O, the functional model estimated a higher flux value compared to the linear and exponential models. It can be seen that the line fitted by the linear model and the curve of the exponential model show very similar behavior, practically overlapping and passing through only one of the observation points. This indicates that, for this set of data, both models had difficulty capturing the real variation in gas flow. In contrast, the functional model fitted the concentration points more accurately over time, reflecting the dynamics of the emission. As the functional curve passes through the concentration points (or very close to the points) in all the samples, they use the gas variation itself over the time interval, obtaining a good representation of the flow.

As reported by Parkin et al.,[9] when the rate of change is constant (Figure 3a), the flow can be calculated using linear regression, but when the observations are not constant, i.e. curvilinear (Figure 3d,h), linear regression is not suitable for measuring the gas flow. As also observed by Silva et al.,[67] the linear model is appropriate, but the estimation of GHG fluxes should not be restricted to this model alone, mainly due to cases of nonlinearity. In turn, the functional model is flexible to measure both the rate of change of the three gases when they are constant and curvilinear.

In Table 1, based on the comparison between the estimation of linear models vs functional representation for CH_4, it appears that there is a moderate positive correlation ($0.7 < \rho < 0.8$) for the sugar cane row and inter-row, and a strong positive correlation for extensive and intensive pasture ($\rho = 0.824$,and $\rho = 0.822$, respectively). There is a weak positive correlation ($\rho < 0.49$) for the four land uses of the exponential model vs functional. For all land uses, there was a small model variation

Table 1. Adjustment Measures in Relation to Linear, Exponential, and Functional Models for CH_4, CO_2,and N_2O in Relation to Extensive Pasture (EP), Intensive Pasture (IP), Sugar Cane Row (CR) and Sugar Cane Inter-row (CI)[a]

gas	models	land use	ρ	R^2	CCC
CH_4	linear vs functional	EP	0.824	0.678	0.816
		IP	0.822	0.674	0.758
		CR	0.732	0.533	0.703
		CI	0.728	0.527	0.676
	exponential vs functional	EP	0.482	0.229	0.390
		IP	0.398	0.153	0.296
		CR	0.423	0.174	0.237
		CI	0.382	0.140	0.268
CO_2	linear vs functional	EP	0.967	0.936	0.967
		IP	0.989	0.981	0.989
		CR	0.978	0.957	0.977
		CI	0.965	0.932	0.964
	exponential vs functional	EP	0.792	0.625	0.777
		IP	0.885	0.782	0.852
		CR	0.878	0.771	0.846
		CI	0.628	0.391	0.550
N_2O	linear vs functional	EP	0.890	0.791	0.883
		IP	0.990	0.988	0.990
		CR	0.867	0.751	0.867
		CI	0.990	0.981	0.987
	exponential vs functional	EP	0.601	0.356	0.564
		IP	0.959	0.919	0.937
		CR	0.829	0.685	0.788
		CI	0.938	0.881	0.881

[a] ρ = correlation coefficient; CCC = agreement correlation coefficient; R^2 = coefficient of determination.

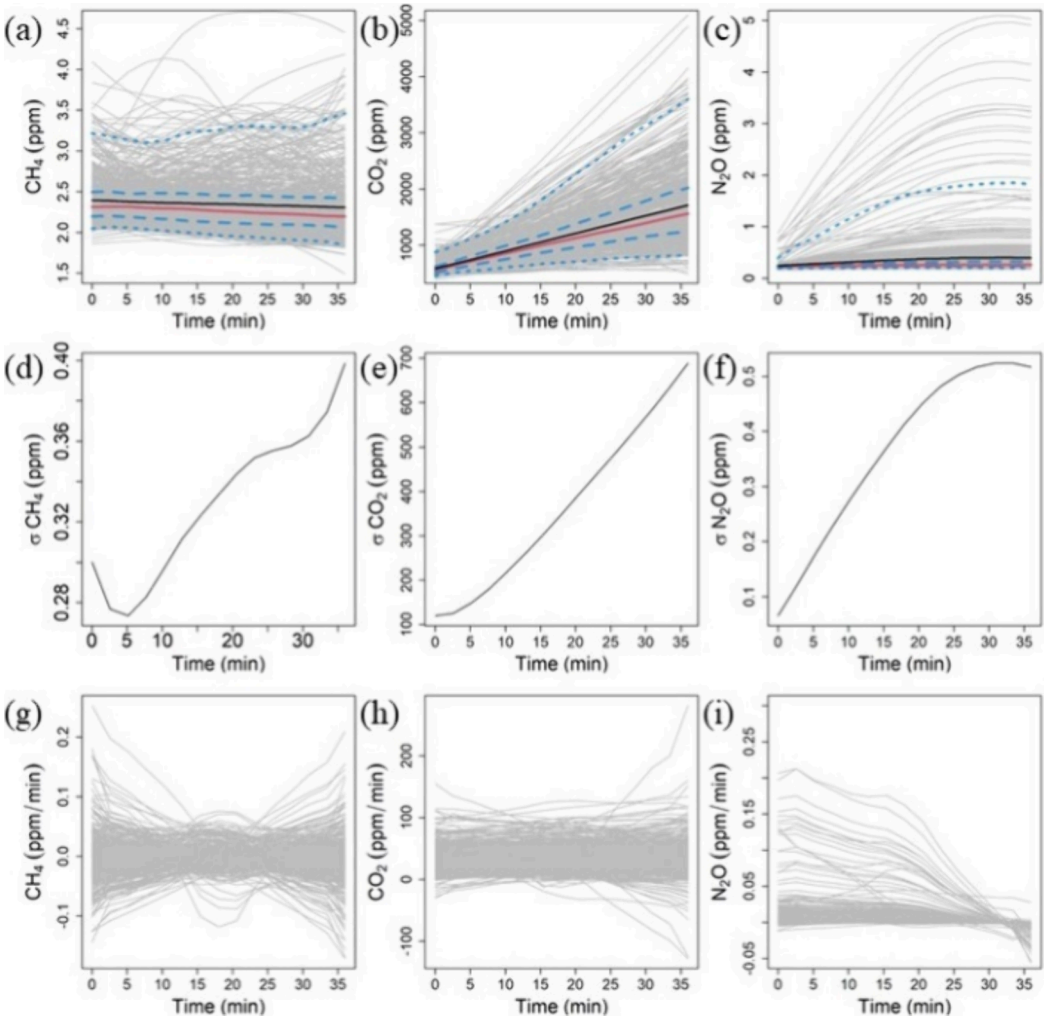

Figure 4. Functional curves (top panel) representing the concentrations of CH_4, CO_2, and N_2O gases for four land uses. Blue dotted line represents quantiles 2.5, 25, 75, and 97.5. The curve in red corresponds to the median and in black the mean. Intermediate panel represents the standard deviation referring to the concentrations of CH_4, CO_2, and N_2O gases for four land uses. Derivative curves (bottom panel) of smoothed gas concentrations (flux).

(R^2 < 0.68), especially the exponential model vs functional (0.14 < R^2 < 0.23). The linear vs functional model exhibited a moderate degree of agreement for all land uses, with a variation between 0.67 and 0.82. In relation to the exponential vs functional model, agreement was weak for the four land uses (CCC < 0.4).

The linear model vs functional for CO_2 showed a very strong positive correlation and an excellent degree of agreement, for all land uses. Regarding the exponential vs functional model, extensive pasture and inter-row showed a moderate positive correlation with a moderate and weak degree of agreement, respectively. Intensive pasture and sugar cane row have a strong positive correlation and a substantial degree of agreement. The model variation was small between the lines, comparing the exponential vs functional model. For the other cases, there was a moderate and large variation.

The linear vs functional model for N_2O exhibited a very strong positive correlation and an excellent degree of agreement, in relation to intensive pasture and sugar cane inter-row. Extensive pasture and sugar cane rows showed a strong positive correlation and a substantial degree of agreement. About the exponential vs functional model, the intensive pasture and sugar cane inter-row showed a very strong positive correlation and a substantial degree of agreement. The sugar cane row had a strong correlation and a moderate degree of agreement. Only extensive pasture was the one that differentiated the most, its degree of agreement was poor and its correlation was moderate.

The variation in the model was very small for extensive pasture when comparing the exponential model with the functional one. For the other cases, moderate to large variations were observed. The linear model showed values closer to the functional model than to the exponential model, offering better correlation and concordance between these two models for all gases and land uses. Additionally, the measurements showed better fit and concordance for gases with more stable behavior, such as carbon dioxide and nitrous oxide. As for methane, the adjustment indicators (correlation, concordance, and R^2) demonstrated a lower capacity to clarify the relationship between the models. This is due to the greater variability in the data, caused by methane's unstable behavior, which is influenced by different production (methanogenesis) and oxidation (methanotrophy) periods, as well as by the action of microorganisms and chemical elements in the processes involved.[68] These characteristics make methane a

gas difficult to analyze and model, as its emissions can result from both anaerobic and aerobic conditions.[3]

In terms of descriptive capacity, the functional model is the one that best fits the concentration points since smoothing allows the curve to pass through the observed points. Therefore, the functional model could be considered the best representation of the variation in the linear and nonlinear structures of gases among the three models. The linear model is the one that comes closest to the functional estimates. This shows that despite the criticisms made by Pedersen et al.[12] and Parkin et al.[9] concerning the linear model, it still appears to be a reasonable model for estimating flow in linear situations. However, when analyzing a gas with nonlinear behavior such as CH_4, both the linear and exponential models depart from the functional model.

Representation of Data by the Functional Model and Its Derivation. B-spline smoothing was used to restructure the discrete gas concentration data, as the data exhibits nonperiodic behavior. With the help of GVC, 15 base functions were determined with a smoothing parameter (λ) equal to 1.5. After this transformation, the representation of the complete and continuous function of the concentration of greenhouse gas emissions was obtained, allowing statistical analyses, and preserving the variability in the original data. Thus, each functional curve in gray in Figure 4 (top panel) refers to the smoothing of the raw concentration data at four-time points in minutes ($t \in (0,36)$), obtained through closed static chambers installed in the ground. With 141 functions for extensive pasture, 143 for intensive pasture, 150 for the sugar cane row and 147 for the sugar cane inter-row, totaling 580 observations.

In Figure 4 (top panel), it is observed that the functional averages had a different behavior between the gases. For CH_4, it showed a slight decreasing slope throughout the time interval, its median also exhibited the same behavior, and they were close. In the case of CO_2, the mean had an increasing slope, as well as its median. At the beginning of the break they were overlapping, but they distanced themselves at the end. Between the quantile curves 75 and 97.5, there is a large variation for the three gases, especially N_2O, since its functional average is between this range.

According to Sun and Genton,[69] this behavior can happen when the central region (between quantiles 25 and 75) is narrow, meaning this region shows less variability. All data sets showed asymmetry. It is possible to observe that, in general, the functional standard deviation presents an increasing tendency in relation to time (Figure 4, intermediate panel), particularly highlighting the behavior of CO_2 and N_2O. Concerning CH_4, given the highlighted nonlinearity, the functional standard deviation initially decreases from zero to 5 min and then increases over time.

The derivatives of the gas functions are shown in Figure 4 (lower panel). Each function displays the behavior of the flux over time, meaning each function represents the trajectory of the flux measured from the functional curve of each sample. In the case of CH_4, an irregular trajectory is observed, characterized by alternations between emission and absorption within a single sampling period. CO_2, on the other hand, presents a more stable behavior, maintaining a positive and constant flux throughout the analyzed interval. N_2O, in turn, follows a decreasing trend, approaching zero as time progresses.

It can be seen that, with the derivatives obtained through FDA, it is possible to analyze the behavior of the flux both locally and globally. The local flux is the one that can be measured at any given moment (where $t \in [0, 36]$ minutes), while the global flux represents the average flux. With these fluxes, it is possible to identify moments of large variations, as well as periods of positive fluxes, which correspond to gas emissions into the atmosphere, and negative fluxes, representing gas absorption by the soil.[70] The study of derivatives is a crucial point in functional data research. For example, Gao[71] when analyzing the derivatives of ozone's daily cycles in Southern California, found results that were intrinsically linked to the variation functions of ozone, observing weaker ozone inhibition in the mornings during the week and faster and more prolonged ozone accumulation on Sunday mornings.

In addition to the variations observed between gases, the specific behavior of N_2O over time reveals patterns that may be directly related to the type of land use and soil preparation. It is observed that N_2O shows a slight increase between 0 and 24 min, followed by stabilization in the interval of $t \in (24, 36)$ minutes. Investigating this behavior separately for each land use (Figure 5), it is noted that this pattern repeats for intensive

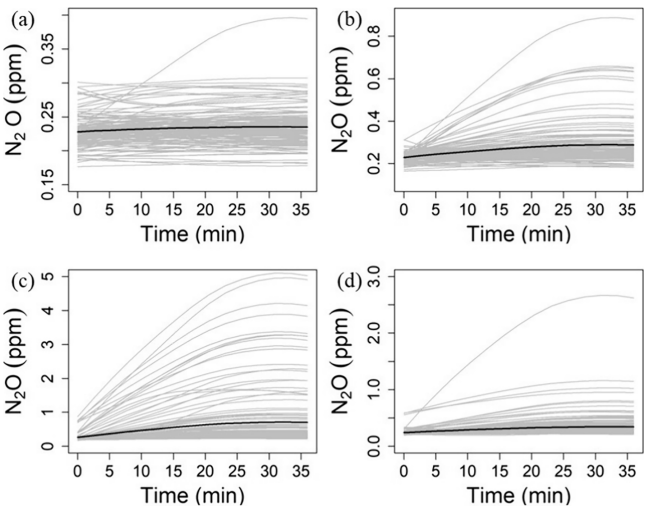

Figure 5. Functional curves representing N_2O gas concentrations for four land uses: (a) extensive pasture; (b) intensive pasture; (c) sugar cane row; and (d) sugar cane inter-row. The curve in black is the functional average.

grazing, between-row cane, and row crops. These land uses underwent the same preparation (plowing, liming, and harrowing) and received the same fertilizers, differing only in the amount of inputs used, as illustrated in the flowchart of Figure 1d. Extensive grazing (Figure 5a) is different from the other land uses by not showing significant variation over time.

The stability observed in the functional curves of N_2O concentrations after 24 min is evidenced by the behavior of the functional derivative (Figure 4i), which shows a critical point after 30 min. This critical point occurs when the derivative begins to oscillate near zero, indicating the stabilization of the fluxes. To validate this critical point, a Student's t test for a sample was conducted, aiming to test the null hypothesis that the average N_2O flux is equal to zero after 30 min. Table 2 presents the 95% confidence intervals for the N_2O flux, based on the Student's t-distribution. It is observed that the flux value

Table 2. Value of Confidence Intervals (95%) Based on the Student's *t* Distribution in Relation to the Critical Point in Time, for N_2O Flux in the Four Land Uses

time (min)	confidence interval (95%)
27	0,0013 < μ < 0,0023
28	0,001 < μ < 0,0018
29	0.00079<μ < 0,0014
30	0,00049 < μ < 0,00097
31	0,00019 < μ < 0,00055
32	−0,00011 < μ^a < 0,00021

[a]Significant *p*-value ($p < 0.05$).

stabilizes after 30 min, with the confidence interval limits close to this time and a significant *p*-value at minute 32 ($p < 0.05$).

This behavior suggests that the sampling time used in the soil gas emission measurements could be adjusted to reflect the N_2O emission pattern more accurately. The sampling time for soil gas emission measurements is an important factor in ensuring result accuracy and sampling process efficiency. Gas samples are simultaneously collected using the static chamber, with a sampling interval of approximately 40 min, according to established practices in the literature, as cited in.[9,36,72−76] However, for N_2O, it was observed that flux stability is reached already after the third collection (Figures 4c,i and 5b−d). This behavior suggests that the sampling time for N_2O could be reduced to 30 min, without compromising the representativeness of the emissions. Reducing the sampling time would not only allow for a more accurate representation of the N_2O flux but also result in a reduction of sampling effort, as the analysis time would be optimized. This approach can improve result accuracy by avoiding the risk of gas saturation and ensuring that the collected data more faithfully reflects N_2O emission conditions.

Daily Average Emissions of CH_4, CO_2, and N_2O. The average daily gas fluxes calculated by the functional model are shown in Figure 6. The accumulated gas emissions for each land use were also calculated from the integral of the functional curve of the average daily fluxes. The highest average daily methane flux emissions are concentrated in the extensive pasture (Figure 6a), where 26 of the sampled dates presented positive values, i.e. methane was emitted on more than 60% of the sampling days, with a flux accumulated 0.145 ppm/min. One explanation for this behavior is the changes induced by compaction in the soil structure, as it essentially affects the characteristics of gas diffusion in this environment, creating conditions of anaerobic and, therefore, influencing the soil's potential for absorption.[77,78] Comparable results, in a region that exhibits the same pasture species and the same type of soil as in the present work, were found by Siqueira-Neto,[79] Frazão,[80] Carvalho,[81] in which the highest CH_4 emissions were in pasture (established in the experimental area) when compared to planting.

Concerning the other three land uses (Figure 6d,g,j), around 80% of the collection days exhibited negative values, indicating that the soil functioned as a CH_4 sink, with an accumulated flux of −0.223 ppm/min, −0.174 ppm/min and −0.141 ppm/min, for IP, CR, and CI, respectively. These results corroborate the findings of Siqueira-Neto et al.,[82] who observed CH_4 emissions only in pastures, with capture of the gas in other land use changes, such as in native vegetation and cultivated land. Similarly, Vasconcelos et al.[78] also found higher CH_4 flows in pastures, while agricultural areas, such as croplands, showed absorption. These observations are in line with the understanding that agricultural practices, especially those with more managed soils, act as methane sinks. However, it is important to note that the potential for CH_4 capture can vary considerably depending on local management conditions, soil type, and climate.[83] Studies such as that by Tubiello et al.[84] indicate that practices, e.g. the use of fertilizers and soil compaction, can alter CH_4 capture capacity, highlighting the importance of optimizing soil management in order to maximize its function as a carbon sink and reduce greenhouse gas emissions.

The average daily carbon dioxide fluxes (Figure 6b,e,h,k) were highly variable throughout the experiments for all types of land use, always with positive values. There are some emission peaks at the beginning of the experiment, corresponding to December 2018 and the beginning of January 2019, concerning the rows and inter-rows of sugar cane. After these peaks, CO2 fluxes decreased but remained variable until the end of the experiments. With an accumulated flow of 1651.72 ppm/min for the EP, 1541.05 ppm/min for the IP, 1189.87 ppm/min for the CR, and 736.94 ppm/min for the CI.

Extensive grazing had the highest average CO_2 fluxes, even without prior soil preparation. This result indicates that carbon emissions do not depend exclusively on management, but also on factors such as microbial activity and the availability of organic matter. In contrast, sugar cane subjected to intensive management showed relatively lower CO_2 fluxes. This suggests that soil preparation can influence the dynamics of emissions, possibly by modifying the decomposition of organic matter and microbial respiration. Studies such as those by Funk et al.,[85] Da Costa et al.[86] and Abreu et al.[87] show that appropriate management practices can mitigate CO_2 emissions by favoring carbon retention in the soil.

Regarding the average daily N_2O flux, positive values were observed for all land uses (Figure 6c,f,i,l). For extensive pasture, emissions were very low, remaining close to zero, recording an accumulated flux of 0.00696 ppm/min, but this result is within what is expected in extensive pastures that are not fertilized with nitrogen in Brazil, as reported in Fuss and Hueppi[46] and Bento et al.[88] Except for extensive pasture, there were high emission peaks between the third and tenth days of collection, which correspond to December 14, 2018, and January 4, 2019, respectively. On the days, the values decreased quickly, remaining close to zero. The accumulated flux for the IP was 0.0636 ppm/min, while the CR and CI were 0.477 ppm/min and 0.107 ppm/min, respectively.

The highest emission peaks of N_2O fluxes are in agricultural management (sugar cane row and inter-row). This increase may be related to nitrogen fertilization and the cultivation process. As also observed in Bento et al.,[89] nitrous oxide fluxes increased mainly in treatments involving cultivation plus fertilizer application.

Oertel et al.,[83] Martins et al.,[90] and Lopes et al.[91] indicated that N_2O emission rates intensify after fertilizer application, since the increased availability of nitrogen in the soil favors denitrification, the process in which N_2O is produced. In agreement, Meurer et al.,[92] in their critical review of N_2O emissions under different land uses in Brazil, also found higher N_2O fluxes in agricultural areas under crop management with fertilization, reinforcing the idea that agricultural intensification, especially in areas with high fertilizer application, is a determining factor for increased emissions. In addition,

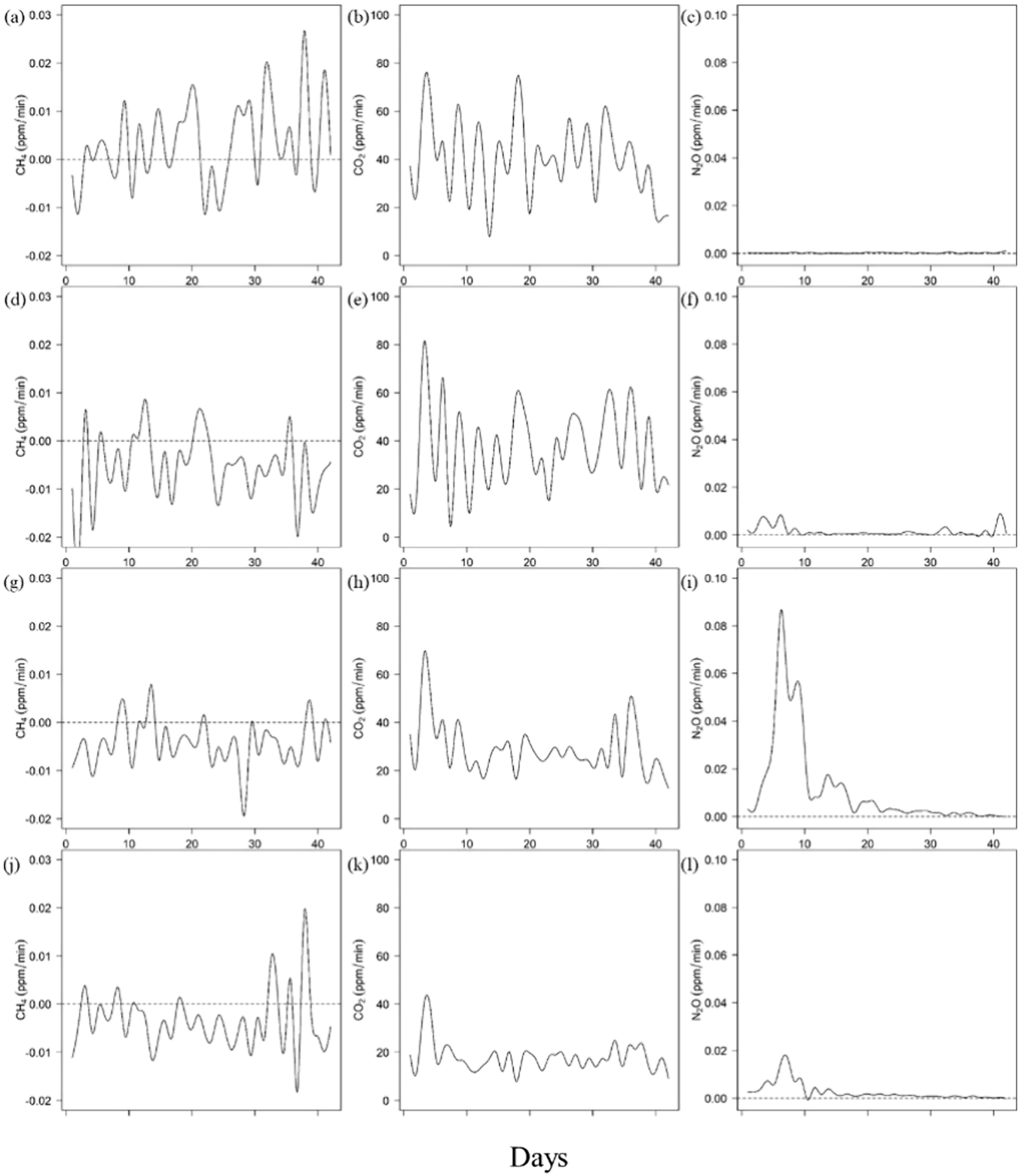

Figure 6. Average daily flux of CH_4 (first column), CO_2 (second column), and N_2O (third column), in relation to the 42 days of collection for each land use. Extensive pasture (a–c); intensive pasture (d–f); sugar cane row (g–i); sugar cane inter-row (j–l).

Canadell et al.[93] indicate that agricultural N_2O emissions have increased by more than 45% in the last four decades, mainly due to the increased use of nitrogen fertilizers and manure. This corroborates our findings of higher emissions in agricultural areas, such as sugar cane and pasture intensification. These data not only reinforce the relevance of monitoring N_2O emissions in different land use systems, but also indicate the importance of more sustainable management practices to mitigate the environmental impacts of intensive agriculture.

Functional Principal Component Analysis (FPCA): Univariate and Multivariate Cases. In both the univariate and multivariate analyses, the FPCA was applied only to the

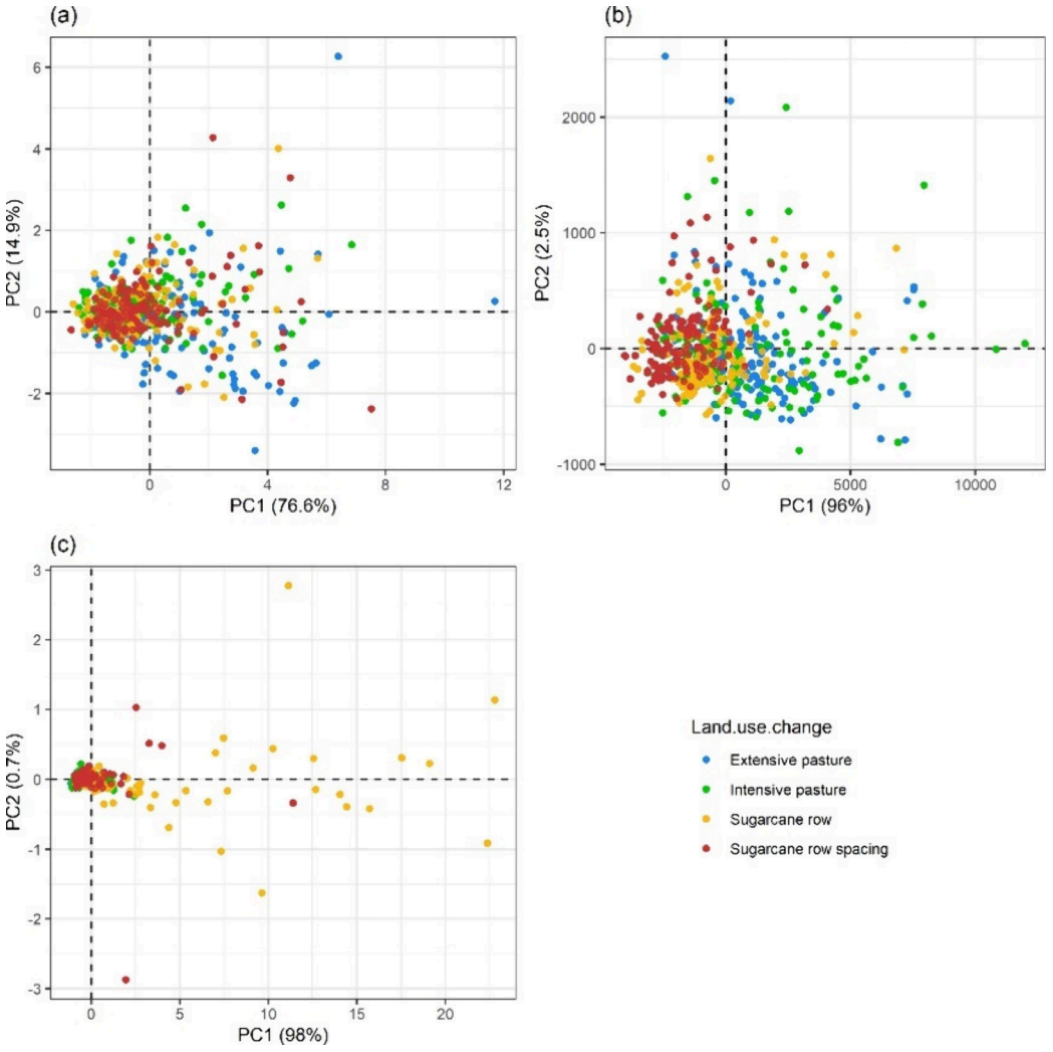

Figure 7. Biplot of the scores of the main functional components of emissions of: (a) CH_4, (b) CO_2, and (c) N_2O, in relation to land uses: extensive pasture (blue), intensive pasture (green), sugar cane row (yellow), and sugar cane inter-row (red).

functions representing the temporal variation in the concentrations of the gases CO_2, N_2O, and CH_4. It is important to note that FPCA serves a different purpose than classical PCA. Rather than aiming to reduce the number of variables, FPCA is commonly used to represent the dominant modes of variation in functional data, either for exploratory analysis or as input for subsequent functional regression or classification procedures. In the present study, FPCA supported the characterization of gas concentration variation under each land-use treatment.

The initial results of the FPCA are for the univariate case, in which the FPCA explains the greatest possible variability existing in the observations for each gas analyzed in relation to its land uses. Figure 7 shows a graphical technique that simultaneously represents the relationships between variables (PC1 and PC2) and observations (component scores). In which the value of each observation is a double-entry table of FPCA.1 vs FPCA.2. This analysis is known as biplot, and its functionality is to be able to organize observations in a format that allows group formations to be verified. It is observed that for the emissions of CH_4 (Figure 7a), CO_2 (Figure 7b) and N_2O (Figure 7c), the first two PCs cumulatively represented more than 90% of the variation in all cases.

In relation to CH_4 (Figure 7a), the biplot of the PC scores showed no separation of groups regarding land uses, the values were well intertwined. Analyzing the biplot for CO_2 (Figure 7b), it is noted that this gas was the one that best-differentiated land uses, separating extensive pasture (blue) from sugar cane (row and inter-row, yellow and red, respectively), in which extensive pasture has the highest emission values and the interline has the lowest values, while the values for intensive pasture were mixed between the other three uses. Regarding N_2O (Figure 7c), it is observed that emissions from the sugar cane row were distant from other emissions, however, a good separation for other land uses was not obtained.

As we have seen, although the univariate FPCA explains more than 90% of the variation present in the data, it failed to obtain a good classification in the separation by land use for the three gases analyzed. Therefore, another FPCA approach was sought in which gases were analyzed together with their land uses. This approach is the FPCA for the multivariate case, it allowed the evaluation of the gases simultaneously, gaining a better visualization in understanding the phenomenon.

In the multivariate case, where the three gases were analyzed simultaneously, the first three principal components accounted for 93.55% of the overall variation while also preserving the correlation structure between gases. Each gas contributed more prominently to a different component, providing a three-dimensional view of the emission profiles associated with land

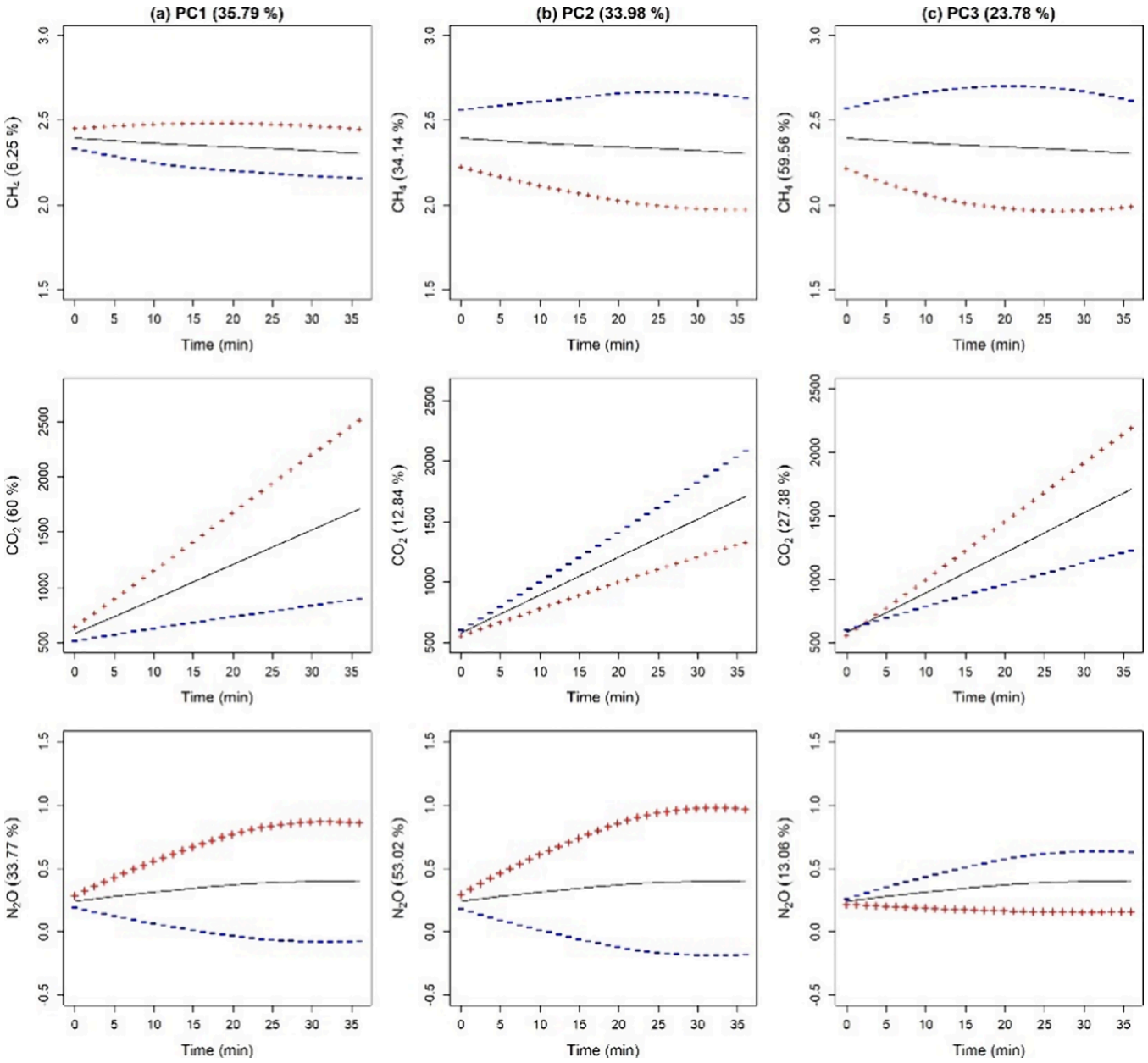

Figure 8. Representation of the emissions of the three components of the average profiles of CH_4, CO_2, and N_2O. The curves show the average profile (solid) and the effects of adding (+) and subtracting (−) eigenfunctions. The percentages displayed in the header titles are the amount of variation explained by each component. The percentages on the vertical axis label are the amounts of variation explained by each variable. For the three PCs, each eigenfunction is multiplied by a factor of 1.5 to increase the deformation of the (+) *e* curves (−).

use. In Figure 8, the representations of the emissions of the three components of the simultaneously averaged profiles of CH_4, CO_2, and N_2O are presented, adding (red marker) and subtracting (blue markers) the eigenfunctions to averaged profiles. As described by Schmutz et al.,[94] in their case study on multivariate FDA for pollution data in French cities, the markers can be interpreted as the first source of variation from the general average, and the first source is a variation in amplitude relative to the average. In the study, the sum of the multiples expresses the positive markers in this component, tending to have the highest gas emissions; that is, the more emissions increase, the greater the fluxes. Subtraction displays the reverse markers in this component, implying fluxes with smaller values.

The first principal component (PC) retained 35.79% of the total variation present in the data. Of this percentage, CO_2 had the largest contribution (60%), followed by N_2O (33.77%) and CH_4 (6.23%). The first PC for CO_2 describes an increasing shift, both for positive and negative markers, that is, for the largest and smallest emissions, there is growth throughout the time interval compared to the average curve. The second component exhibited 33.98%, with N_2O having the highest percentage of variation (53.02%), followed by CH_4 (34.14%), and CO_2 (12.84%). Concerning PC2, N_2O showed an increasing displacement for the positive markers and a decreasing shift for the negative markers throughout the time interval.

The third component presented a smaller explanatory reduction, 23.78% of the variation, with methane gas

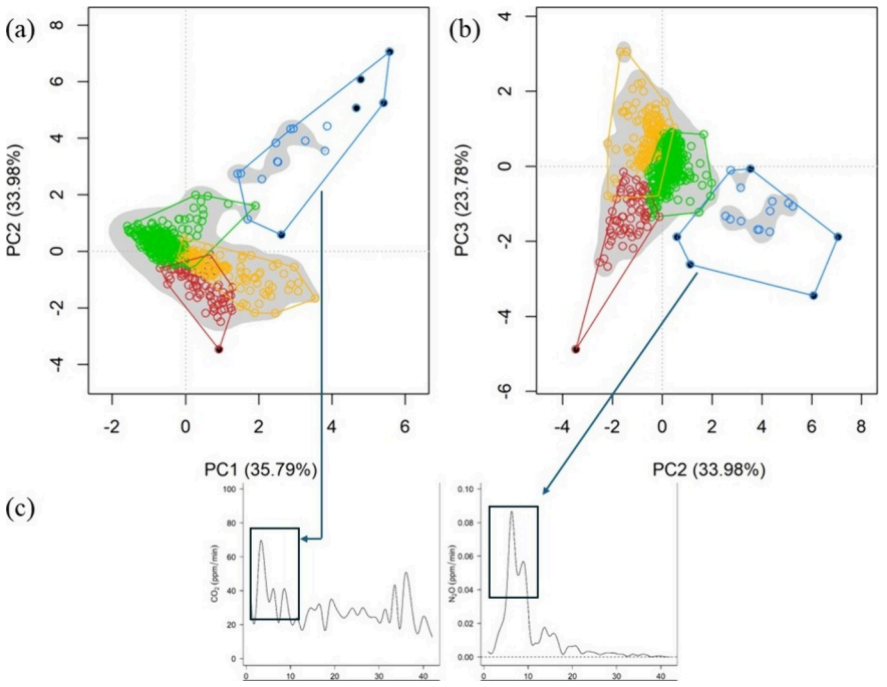

Figure 9. Cluster analysis based on the three main components for the multivariate case. Color schemes: yellow group 1 (pasture); red group 2 (intermediate); green group 3 (sugar cane); and blue group 4 (peaks in CO_2 and N_2O emissions caused by soil management).

presenting the largest variation (59.56%), then CO_2 (27.38%), and finally, N_2O (13.06%). Methane gas was the only one that showed a decreasing average PC, showing two different patterns in the markers, in which the positive markers tend to grow and then decrease (concavity downward) and the negative markers tend to decrease and then grow (concavity up).

Comparing the results for the univariate and multivariate cases, when using the multivariate PC scores, better performance was obtained in the formation of groups, both for an unsupervised analysis such as cluster analysis and a supervised analysis such as classification. These results can be observed in the next sessions.

Cluster Analysis. The cluster analysis of the scores of the multivariate principal components (considering the simultaneous analysis of gases) (Figure 9) showed that four clusters were formed. Of the 134 observations in cluster 1 (yellow color), 89% of the observations represent emissions from pasture and 11% represent emissions from sugar cane, being labeled as the pasture group. There were positive values for almost all PC1 and PC3 scores, indicating higher emissions of CO_2 and CH_4, respectively, and negative values for PC2, indicating lower emissions of N_2O. As seen in Figure 6, the values of the highest daily flux averages were CH_4 and CO_2 for extensive pasture, followed by intensive pasture. On the other hand, for N_2O, emissions from extensive pasture presented the lowest average values.

Cluster 2 (red color) has 67 observations, 55% of which represent pasture and 45% sugar cane. They presented positive values in relation to PC1 and negative values for PC2 and PC3. This group is the intermediate group, as emissions from both pastures and sugar cane were similar in relation to the three gases. Concerning cluster 3 (green color), row and inter-row emissions showed great similarity, with 80% of their emissions concentrated in this group. For this reason, it comprised the largest number of observations (362), with 68% being sugar cane and 32% pasture. Since the vast majority of emissions come from sugar cane, this group was categorized as the sugar cane group. The N_2O exhibited the greatest contribution to the formation of this group, as the score values for PC2 were almost all positive, showing that they are the highest emissions of the gas.

Finally, cluster 4 (blue) consisted of only 17 observations, all related to sugar cane. These events took place between December 17 and 26 and January 2, with high CO_2 and N_2O emissions, but low CH_4 emissions. As noted in Jacques and Preda,[95] cluster analysis groups together observations with significant variations in amplitude and phase in functional curves. This group is particularly relevant in our results, as it represents the peaks in CO_2 and N_2O emissions observed in sugar cane, corroborating Figure 6h,i,k,l, which show peaks in daily flows after soil management processes.

Land Use Classification Analysis. The accuracy obtained by the test regarding the classification model is in Table 3, along with the classification model evaluation metrics and description of the scenarios. Different scenarios were considered to verify different influences on gas modeling. After defining the parameters for each scenario, the model was evaluated on the test data. To illustrate the performance of the best models on the test data, the confusion matrices in Figure 10 were generated. Using the confusion matrix, it is possible to analyze the classification results in more detail, as it is possible to observe the classification errors in relation to land uses. For example, from the observations that should have been classified as extensive pasture and were not, it is possible to identify how many were classified as another land use. Assuming that the majority of errors were classified as intensive pasture, this would imply that the model does not differentiate well between extensive and intensive pasture.

In the first scenario (Table 3), the classification was conducted with only PCs as independent variables and four land uses as dependent variables, obtaining a low-test accuracy

Table 3. Classification Model Evaluation Metrics for Each Land Use: Extensive Pasture (EP), Intensive Pasture (IP), Sugar Cane Row (CR), and Sugar Cane Inter-row (CI)

model	description	land use	precision	recall	F1 score	accuracy
1	PCs	EP	0.52	0.65	0.58	0.53
		IP	0.57	0.43	0.49	
		CR	0.55	0.40	0.46	
		CI	0.52	0.66	0.58	
2	PCs + temperature, four groups	EP	0.82	0.74	0.78	
		IP	0.59	0.57	0.58	0.64
		CR	0.51	0.45	0.48	
		CI	0.65	0.80	0.71	
3	PCs, with three groups	EP	0.55	0.63	0.59	
		IP	0.91	0.21	0.34	0.65
		C	0.67	0.90	0.77	
4	PCs + temperature, three groups.	EP	0.70	0.77	0.73	
		IP	0.67	0.55	0.60	0.74
		C	0.78	0.82	0.80	
5	data from 06/12/2018–01/01/2019, PCs	EP	0.80	0.75	0.77	0.60
		IP	0.55	0.54	0.55	
		CR	0.52	0.57	0.54	
		CI	0.56	0.56	0.56	
6	data from 06/12/2018–01/01/2019, PCs + temperature	EP	0.85	0.82	0.84	
		IP	0.68	0.61	0.64	0.70
		CR	0.63	0.66	0.64	
		CI	0.61	0.66	0.63	
7	data from 06/12/2018–01/01/2019, PCs	EP	0.84	0.66	0.74	0.73
		IP	0.82	0.36	0.50	
		C	0.69	1	0.82	
8	data from 06/12/2018–01/01/2019, PCs + temperature	EP	0.76	0.78	0.77	
		IP	0.78	0.46	0.58	0.77
		C	0.77	0.95	0.85	

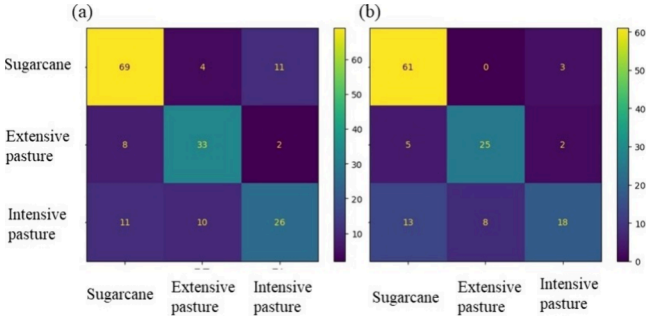

Figure 10. Confusion matrices for scenario 4 (a) and scenario 8 (b) in relation to the classification models that showed better accuracy. Soil uses extensive pasture (EP), intensive pasture (IP), and the union of row and inter-row uses of sugar cane (C).

of only 53%. Regarding the metrics used to evaluate the classification model, the model's accuracy varied between 52% and 57% for land uses, with CI being the treatment that obtained the highest identification of observations correctly (recall = 66%) followed by EP (recall = 65%), the CR treatment was the one with the least correct classification (recall = 40%), followed by IP (recall = 43%). Given that environmental factors, mainly temperature (air and soil) and soil humidity, influence emissions of CH_4, CO_2, and N_2O fluxes,[96,97] in addition to the three main components, soil and air temperature at the time of data collection were incorporated into the independent variables to the second scenario (Table 3). Soil temperature controls biological oxygen consumption, altering the growth of microbial communities.[30] With the addition of temperatures, both the accuracy of the test and the precision of the model increased (Table 3). The accuracy was 64% and the model precision varied between 51% and 82% for land uses. Regarding metrics, it appears again that CI was the treatment that obtained the highest identification of observations correctly (recall = 80%), followed by EP (recall = 74%). The CR treatment was the one with the least correct classification (recall = 45%), followed by IP (recall = 57%).

Due to the characteristics of the experiment, given that CR and CI are in the same experimental plot, these two treatments were also considered as the same group in an additional scenario. Therefore, for the third scenario (Table 3), PCs and three land uses (EP, IP, and C) were considered. The test accuracy was 65%, 12% higher than scenario 1. The model accuracy varied between 55% and 91% for land uses. It is observed that treatment C obtained the highest identification of observations correctly (recall = 90%), followed by EP (recall = 63%) and IP was the treatment that had the lowest correct classification (recall = 21%). As seen in scenario 2, when we add temperatures to the classification model it exhibits better performance. Thus, in the fourth scenario (Table 3), PCs, temperatures and three land uses were considered, obtaining moderate accuracy (74%), with model precision varying between 67% and 78% for land uses. In the confusion matrix for scenario 4 (Figure 10a), it can be seen that the treatment that exhibited the highest proportion of correct classification was sugar cane (recall = 82%), the worst was intensive pasture (recall = 55%) and extensive pasture was the intermediate one (recall = 77%).

Finally, a temporal component was added only to the data collected from December sixth of 2018 to January 25th of 2019. The choice to reduce the data was due to two reasons, the first was about the spacing of the collection days, among these dates there was a large number of samples, and the second reason was that from January 26, 2019, the pesticide 2,4-D was applied only to sugar cane, as described in Figure 1d. Then, two periods are evident in which there are differences regarding soil management preparation. From this perspective, the next scenarios (5, 6, 7, and 8) were constructed with a smaller sampling interval.

The fifth, sixth, and seventh scenarios (Table 3) had a lower test accuracy than scenario 4, with 60%, 70%, and 73%, respectively. The eighth scenario was the one that presented the highest test accuracy with 77% (Table 3), in this scenario, PCs, temperatures, and three land uses were considered with data with the smallest time interval. The accuracy of the model (Table 3) was very close, 76% for EP, 78% for IP, and 77% for C. Sugar cane had the highest proportion of correctly classified observations (recall = 95%), followed by PE (recall = 78%) and the IP (recall = 46%), as can be seen in Table 3 and the confusion matrix (Figure 10b).

The results show a strong influence of external factors (management and handling) on the classification model as referenced in the literature.[98,99] This can be observed by the increase in accuracy when both the soil and air temperature and the temporal component are added to the scenarios. In addition to the external factors noted above, an increase in accuracy is also observed when CR and CI soils are considered as a single treatment. From an experimental point of view, there is no complete isolation between the row and the inter-row since they are in the same experimental plot.

In summary, the results of the analyses demonstrated that through the derivatives of the functions, the flux was calculated, both local and global, and with the integral of the functional curves, verifying that one of the advantages of using functional data is the measurement of the instantaneous flux through derivatives, making it possible to analyze the variability of gas fluxes concerning different land uses. Therefore, the functional derivatives detected variability in flow emissions between collection days. This may result from the influence of external factors and soil management activities.

Land uses under study contributed most of the time as a sink for CH_4, unlike CO_2 and N_2O, which acted as a source of the gases in all land uses. Regarding N_2O, the fluxes were very close to zero, since high emissions of the gas were observed at the beginning of the sampling, a period close to soil preparation. Thus, showing the influence of soil preparation on emissions of this gas. In terms of descriptive capacity, the functional model is the one that best fits the concentration points, since smoothing allows the curve to pass through the observed points. Therefore, the method is appropriate for evaluating the fluxes of GHGs when the flux is both constant, as is the case with CO_2 and N_2O, and curvilinear, such as CH_4. Moreover, it accurately represents the process of gas exchange between soil and atmospheric air. These findings have important implications for environmental management and the development of public policies aimed at mitigating the impacts of GHG emissions in agriculture.

However, despite the advancements in the comparison between the models, this study has some limitations. Notably, climatic variables such as temperature, humidity, and precipitation, which can influence GHG fluxes, were not considered. Additionally, the temporal limitation of the study, with data collected over only six months, may have prevented the capture of broader seasonal patterns throughout the year. To overcome these limitations, future research could integrate climate models to assess how environmental variables impact the estimation of GHG fluxes in different land uses. Hybrid approaches that combine the Functional Regression Model with machine learning techniques, such as random forests or support vector machines, could offer a more nuanced interpretation of the data, particularly when nonlinear relationships are predominant.

Furthermore, studies monitoring gas fluxes over time during different seasons of the year would allow for the evaluation of seasonal patterns and enhance the robustness of the analyses, providing a deeper understanding of gas dynamics in relation to land use changes, such as the conversion of pastures to sugar cane cultivation. Applying the methodology to different ecosystems could also help to understand regional variations in emissions.

AUTHOR INFORMATION

Corresponding Author

Paulo J. Duarte-Neto − *Programa de Pós-Graduação em Biometria e Estatística Aplicada and Departamento de Estatística e Informática, Universidade Federal Rural de Pernambuco, Recife, Pernambuco 52171-900, Brazil;* orcid.org/0000-0002-2592-7297; Email: paulo.duartent@ufrpe.br

Authors

Mickaelle M. de Almeida Pereira − *Programa de Pós-Graduação em Biometria e Estatística Aplicada, Universidade Federal Rural de Pernambuco, Recife, Pernambuco 52171-900, Brazil*

Luiz Antonio Martinelli − *Programa de Pós-Graduação em Biometria e Estatística Aplicada, Universidade Federal Rural de Pernambuco, Recife, Pernambuco 52171-900, Brazil; Laboratório de Ecologia Isotópica, Centro de Energia Nuclear na Agricultura, Piracicaba, São Paulo 13400-970, Brazil*

Janaina Braga do Carmo − *Departamento de Ciências Ambientais, Universidade Federal de São Carlos, Sorocaba, São Paulo 18052-780, Brazil*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsagscitech.4c00810

Funding

The Article Processing Charge for the publication of this research was funded by the Coordenacao de Aperfeicoamento de Pessoal de Nivel Superior (CAPES), Brazil (ROR identifier: 00x0ma614). This work was partially funded by the Coordination for the Improvement of Higher Education Personnel - CAPES [Financing Code 001]; the São Paulo State Research Support Foundation - FAPESP [Process 2015/18790−3]; National Council for Scientific and Technological Development − CNPq [no. 310320/2020-8].

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors would like to thank the Coordination for the Improvement of Higher Education Personnel, the Research Support Foundation of the State of São Paulo. The authors are grateful to the National Council for Scientific and Technological Development for providing research grants to P.J.D.-N.

ABBREVIATIONS

GHGs,greenhouse gases; CO2,carbon dioxide; CH4,methane; N2O,nitrous oxide; FDA,functional data analysis; H/M,Hutchinson/Mosier model; NDFE,nonstationary diffusive flow estimator; EP,extensive pasture; IP,intensive pasture; C,sugar cane; CR,sugar cane row; CI,sugar cane inter-row; N,nitrogen; P,phosphorus; K,potassium; AU,animal unit; GCV,generalized cross-validation; λ,smoothing parameter; CCC,concordance correlation coefficient; FPCA,functional principal component analysis; MFPCA,multivariate functional principal components; PC,principal component; SVC,support vector classification

REFERENCES

(1) Bandyopadhyay, K. K.; Lal, R. Effect of land use management on greenhouse gas emissions from water stable aggregates. *Geoderma* **2014**, *232*, 363−372.

(2) Begum, K.; Kuhnert, M.; Yeluripati, J. B.; Ogle, S. M.; Parton, W. J.; Williams, S. A.; Pan, G.; Cheng, K.; Ali, M. A.; Smith, P. Modeling greenhouse gas emissions and mitigation potentials in fertilized paddy rice fields in Bangladesh. *Geoderma* **2019**, *341*, 206−215.

(3) Dowhower, S. L.; Teague, W. R.; Casey, K. D.; Daniel, R. Soil greenhouse gas emissions as impacted by soil moisture and temperature under continuous and holistic planned grazing in native tallgrass prairie. *Agriculture, Ecosystems & Environment* **2020**, *287*, No. 106647.

(4) Meena, V. S.; Mondal, T.; Pandey, B. M.; Mukherjee, A.; Yadav, R. P.; Choudhary, M.; Singh, S.; Bisht, J. K.; Pattanayak, A. Land Use Changes: Strategies to improve the soil carbon and nitrogen storage

pattern in the Himalayan ecosystem in India. *Geoderma* **2018**, *321*, 69−78.
(5) Denmead, O. T. Approaches to measuring fluxes of methane and nitrous oxide between landscapes and the atmosphere. *Plant and Soil* **2008**, *309* (1−2), 5−24.
(6) Keller, M.; Kaplan, W. A.; Wofsy, S. C. Emissions of N2O, CH4 and CO2 from tropical forest soils. *Journal of Geophysical Research: Atmospheres* **1986**, *91* (D11), 11791−11802.
(7) Pumpanen, J.; Kolari, P.; Ilvesniemi, H.; Minkkinen, K.; Vesala, T.; Niinistö, S.; Lohila, A.; Larmola, T.; Morero, M.; Pihlatie, M.; Janssens, I.; Yuste, J. C.; Grünzweig, J. M.; Reth, S.; Subke, J. A.; Savage, K.; Kutsch, W.; Ostreng, G.; Ziegler, W.; Anthoni, P.; Lindroth, A.; Hari, P. Comparison of different chamber techniques for measuring soil CO2 efflux. *Agricultural and Forest Meteorology* **2004**, *123* (3−4), 159−176.
(8) Christiansen, J. R.; Korhonen, J. F.; Juszczak, R.; Giebels, M.; Pihlatie, M. Assessing the effects of chamber placement, manual sampling and headspace mixing on CH4 fluxes in a laboratory experiment. *Plant and soil* **2011**, *343* (1−2), 171−185.
(9) Parkin, T. B.; Venterea, R. T.; Follett, R. F. Chamber-based trace gas flux measurements. *Sampling Protoc.* **2010**, *3*, 3−39.
(10) Pirk, N.; Mastepanov, M.; Parmentier, F. J. W.; Lund, M.; Crill, P.; Christensen, T. R. Calculations of automatic chamber flux measurements of methane and carbon dioxide using short time series of concentrations. *Biogeosciences* **2016**, *13* (4), 903.
(11) Parkin, T. B.; Venterea, R. T.; Hargreaves, S. K. Calculating the detection limits of chamber-based soil greenhouse gas flux measurements. *Journal of Environmental Quality* **2012**, *41* (3), 705−715.
(12) 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* **2010**, *61* (6), 888−902.
(13) Ramsay, J. O.; Silverman, B. W. Principal components analysis for functional data. *Funct. Data Anal.* **2005**, 147−172.
(14) Coffey, N.; Harrison, A. J.; Donoghue, O. A.; Hayes, K. Common functional principal components analysis: A new approach to analyzing human movement data. *Human movement science* **2011**, *30* (6), 1144−1166.
(15) Sun, Y.; Genton, M. G. Adjusted functional boxplots for spatial-temporal data visualization and outlier detection. *Environmetrics* **2012**, *23* (1), 54−64.
(16) Fortuna, F.; Maturo, F. K-means clustering of item characteristic curves and item information curves via functional principal component analysis. *Quality & Quantity* **2019**, *53* (5), 2291−2304.
(17) Gao, H. O.; Niemeier, D. A. Using functional data analysis of diurnal ozone and NOx cycles to inform transportation control emissions. *Transportation Research Part D: Transport and Environment* **2008**, *13* (4), 221−238.
(18) Di Salvo, F.; Ruggieri, M.; Plaia, A. Functional principal componente analysis for multivariate multidimensional environmental data. *Environmental and ecological statistics* **2015**, *22* (4), 739−757.
(19) Happ, C.; Greven, S. Multivariate functional principal component analysis for data observed on different (dimensional) domains. *Journal of the American Statistical Association* **2018**, *113* (522), 649−659.
(20) Ngo, D.; Sun, Y.; Genton, M. G.; Wu, J.; Srinivasan, R.; Cramer, S. C.; Ombao, H. An exploratory data analysis of electroencephalograms using the functional boxplots approach. *Front. Neurosci.* **2015**, *9*, 282.
(21) Ramsay, J. O. Functional components of variation in handwriting. *Journal of the American Statistical Association* **2000**, *95* (449), 9−15.
(22) Pauthenet, E.; Roquet, F.; Madec, G.; Nerini, D. A linear decomposition of the Southern Ocean thermohaline structure. *Journal of Physical Oceanography* **2017**, *47* (1), 29−47.
(23) Ramsay, J. O.; Bock, R. D.; Gasser, T. Comparison of height acceleration curves in the Fels, Zurich, and Berkeley growth data. *Annals of Human Biology* **1995**, *22* (5), 413−426.
(24) Hyndman, R. J.; Ullah, M. S. Robust forecasting of mortality and fertility rates: a functional data approach. *Comput. Stat. Data Anal.* **2007**, *51* (10), 4942−4956.
(25) Shang, H. L.; Smith, P. W.; Bijak, J.; Wisniowski, A. A multilevel functional data method for forecasting population, with an application to the United Kingdom. *International Journal of Forecasting* **2016**, *32* (3), 629−649.
(26) Canisares, L. P.; Cherubin, M. R.; da Silva, L. F. S.; Franco, A. L. C.; Cooper, M.; Mooney, S. J.; Cerri, C. E. P. Soil microstructure alterations induced by land use change for sugarcane expansion in Brazil. *Soil Use and Management* **2020**, *36* (2), 189−199.
(27) MAPBIOMAS *Área de agropecuária no Brasil cresceu 50% nos últimos 38 anos*. 06 de outubro de 2023. Disponível em: https://brasil.mapbiomas.org/2023/10/06/area-de-agropecuaria-no-brasil-cresceu-50-nos-ultimos-38-anos/. Acesso em: 11 de dezembro de 2023.
(28) Girotto, L.; Freitas, I. B. F.; Yoshii, M. P. C.; Goulart, B. V.; Montagner, C. C.; Schiesari, L. C.; Espíndolo, E. L. G.; Freitas, J. S. Using mesocosms to evaluate the impacts of pasture intensification and pasture-sugarcane conversion on tadpoles in Brazil. *Environ. Sci. Pollut. Res.* **2023**, *30* (8), 21010−21024.
(29) Mosier, A.; Wassmann, R.; Verchot, L.; King, J.; Palm, C. Methane and nitrogen oxide fluxes in tropical agricultural soils: sources, sinks and mechanisms. In *Tropical Agriculture in Transition—Opportunities for Mitigating Greenhouse Gas Emissions?*; Springer: Dordrecht, 2004, 11- 49. DOI: .
(30) Wang, C.; Amon, B.; Schulz, K.; Mehdi, B. Factors that influence nitrous oxide emissions from agricultural soils as well as their representation in simulation models: a review. *Agronomy* **2021**, *11* (4), 770.
(31) Freitas, I. B. F.; Duarte-Neto, P. J.; Sorigotto, L. R.; Yoshii, M. P. C.; de Palma Lopes, L. F.; de Almeida Pereira, M. M.; Girotto, L.; Athayde, D. B.; Goulart, B. V.; Montagner, C. C.; Schiesari, L. C.; Martinelli, L. A.; Espíndola, E. L. G. Effects of pasture intensification and sugarcane cultivation on non-target species: a realistic evaluation in pesticidecontaminated mesocosms. *Sci. Total Environ.* **2024**, *922*, No. 171425.
(32) Euclides, V. P. B.; Montagner, D. B.; Barbosa, R. A.; Nantes, N. N. Pasture and grazing management of Brachiaria brizantha (Hochst) Stapf and Panicum maximum Jacq. *Ceres Magazine* **2014**, *61*, 808−818.
(33) Dias-Filho, M. B. *Diagnóstico das pastagens no Brasil*; Embrapa Amazônia Oriental: Belém, 2014, *1*, 1-38.
(34) Cezário, A. S.; Ribeiro, K. G.; Santos, S. A.; de Campos Valadares Filho, S.; Pereira, O. G. Silages of Brachiaria brizantha cv. Marandu harvested at two regrowth ages: Microbial inoculant responses in silage fermentation, ruminant digestion and beef cattle performance. *Anim. Feed Sci. Technol.* **2015**, *208*, 33−43.
(35) Freitas, I. B. F.; Duarte-Neto, P. J.; de Palma Lopes, L. F.; Yoshii, M. P. C.; Girotto, L.; de Mello Gabriel, G. V.; Sorigotto, L. R.; do Carmo, J. B.; Montagner, C. C.; Schiesari, L. C.; Martinelli, L. A.; Espíndola, E. L. G. Soil management effects of extensive pastures, intensive pastures and sugarcane crops on the availability of metals and nutrients in freshwater: A realistic mesocosm approach. *Agric., Ecosyst. Environ.* **2023**, *350*, No. 108473.
(36) Davidson, E. A.; Savage, K. V. L. V.; Verchot, L. V.; Navarro, R. Minimizing artifacts and biases in chamber-based measurements of soil respiration. *Agricultural and Forest Meteorology* **2002**, *113* (1−4), 21−37.
(37) do Carmo, J. B.; de Sousa Neto, E. R.; Duarte-Neto, P. J.; Ometto, J. P. H. B.; Martinelli, L. A. Conversion of the coastal Atlantic forest to pasture: Consequences for the nitrogen cycle and soil greenhouse gas emissions. *Agric., Ecosyst. Environ.* **2012**, *148*, 37−43.
(38) Ramsay, J. O.; Silverman, B. W. *Functional data analysis*; Springer: Verlag, New York, 1997.
(39) Bande, M. F.; de la Fuente, M. O.; Galeano, P.; Nieto, A.; Garcia-Portugues, E. *fda.usc: Functional Data Analysis and Utilities for Statistics Computing*. R package, version 2.1.0, **2022**, *4*, 1, 28. .

(40) Ullah, S.; Finch, C. F. Applications of functional data analysis: A systematic review. *BMC Med. Res. Methodol.* **2013**, *13* (1), 43.
(41) Ramsay, J. O.; Hooker, G.; Graves, S. Functional Data Analysis with R and MATLAB. *Use R* **2009**, 1−207.
(42) Jansen, M. Generalized cross validation in variable selection with and without shrinkage. *Journal of statistical planning and inference* **2015**, *159*, 90−104.
(43) Febrero-Bande, M.; De La Fuente, M. O. Statistical computing in functional data analysis: The R package fda. usc. *J. Stat. Software* **2012**, *51*, 1−28.
(44) Mas, A.; Pumo, B. Functional linear regression with derivatives. *Journal of Nonparametric Statistics* **2009**, *21* (1), 19−40.
(45) Ramsay, J. O.; Silverman, B. W. *Applied functional data analysis: methods and case studies*; Springer: New York, 2002, 17−40. .
(46) Fuss, R.; Hueppi, R. *gasfluxes: Greenhouse Gas Flux Calculation from Chamber Measurements*. R package, version 0.4−4, 2024, v0.7, 1−16. https://git-dmz.thuenen.de/fuss/gasfluxes.
(47) Tsai, M. Y.; Lin, C. C. New model-averaged estimators of concordance correlation coefficients: simulation and application to longitudinal overdispersed Poisson data. *Communications in Statistics-Simulation and Computation* **2023**, *52* (3), 961−979.
(48) Lin, L. I. K. The concordance correlation coefficient I'm Evaluate reproducibility. *Biometrics* **1989**, *45*, 255−268.
(49) Stevenson, M.; Sergeant, E.; Heuer, C.; Nunes, T.; Heuer, C.; Marshall, J.; Sanchez, J.; Thornton, R.; Reiczigel, J.; Robison-Cox, J.; Sebastiani, P.; Solymos, P.; Yoshida, K.; Jones, G.; Pirikahu, S.; Firestone, S.; Kyle, R.; Popp, J.; Jay, M.; Cheung, A.; Singanallur, N.; Szabo, A.; Rabiee, A. *epiR: Tools for the Analysis of Epidemiological Data*. R package, version 2.0.53, 2022. https://CRAN.R-project.org/package=epiR.
(50) Silva, F. S. R. Prediction of Cr, Cu, Hg, Ni, Pb, Zn in polluted mangrove soils in northeastern Brazil using near-infrared reflectance spectroscopy. 46 f. Dissertation (Postgraduate Program in Soil Science) - Federal Rural University of Pernambuco: Recife, **2022**.
(51) Steichen, T. J.; Cox, N. J. A note on the concordance correlation coefficient. *Stata Journal* **2002**, *2* (2), 183−189.
(52) Pauthenet, E.; Roquet, F.; Madec, G.; Sallée, J. B.; Nerini, D. The thermohaline modes of the global ocean. *Journal of Physical Oceanography* **2019**, *49* (10), 2535−2552.
(53) Ramsay, J. O., Graves, S., Hooker, G. *fda: Functional Data Analysis*. R package version 6.0.5, 2022. https://CRAN.Rproject.org/package=fda.
(54) Pauthenet, E., Nerini, D., Roquet, F. *Functional Data Analysis of hydrographic profiles(v1.0.0)*; Zenodo: **2020**. .
(55) Warmenhoven, J.; Cobley, S.; Draper, C.; Harrison, A.; Bargary, N.; Smith, R. Bivariate functional principal components analysis: considerations for use with multivariate movement signatures in sports biomechanics. *Sports Biomechanics* **2019**, *18* (1), 10−27.
(56) Warmenhoven, J.; Cobley, S.; Draper, C.; Harrison, A.; Bargary, N.; Smith, R. Considerations for the use of functional principal components analysis in sports biomechanics: examples from on-water rowing. *Sports Biomechanics* **2019**, *18* (3), 317−341.
(57) Jacques, J.; Preda, C. Functional data clustering: a survey. *Adv. Data Anal Classif* **2014**, *8*, 231−255.
(58) Shang, H. L. survey of functional principal component analysis. *AStA Advances in Statistical Analysis* **2014**, *98* (2), 121−142.
(59) Wang, J. L.; Chiou, J. M.; Müller, H. G. Functional data analysis. *Annual Review of Statistics and its application* **2016**, *3*, 257−295.
(60) Hair, J. F.; Black, W. C.; Babin, B. J.; Anderson, R. E.; Tatham, R. L *Multivariate data analysis*; 5nd ed., Bookman: Porto Alegre, 2005, 593 p.
(61) Muniswamaiah, M.; Agerwala, T.; Tappert, C. C. Applications of Binary Similarity and Distance Measures. arXiv preprint *arXiv*, **2023**, 2307, 00411.
(62) Chang, C. C.; Lin, C. J. LIBSVM: a library for supporting vector machines. *ACM transactions on intelligent systems and technology (TIST)* **2011**, *2* (3), 1−27.
(63) Salazar, J. J.; Garland, L.; Ochoa, J.; Pyrcz, M. J. Fair train-test split in machine learning: Mitigating spatial autocorrelation for improved prediction accuracy. *J. Pet. Sci. Eng.* **2022**, *209*, No. 109885.
(64) Géron, A. *Hands on: machine learning with Scikit-Learn. Keras & TensorFlow: Concepts, tools and techniques for building intelligent systems*; 2nd ed., Alta Books: Rio de Janeiro, 2021, 640 p.
(65) R Development Core Team *A: A Language and Environment for Statistical Computing*. Vienna, Austria, 2019. Available at: http://www.Rproject.org.
(66) Bi, J.; Kuesten, C. Using Functional Data Analysis (FDA) Methodology and the R Package, "fda" for Sensory Time-Intensity Evaluation. *Journal of Sensory Studies* **2013**, *28* (6), 474−482.
(67) Silva, J. P.; Lasso, A.; Lubberding, H. J.; Pena, M. R.; Gijzen, H. J. Biases in greenhouse gases static chambers measurements in stabilization ponds: Comparison of flux estimation using linear and non-linear models. *Atmos. Environ.* **2015**, *109*, 130−138.
(68) Wang, N.; Zhu, X.; Zuo, Y.; Liu, J.; Yuan, F.; Guo, Z.; Zhang, L.; Sun, Y.; Gong, C.; Guo, D.; Song, C.; Xu, X. Microbial mechanisms for methane source-to-sink transition after wetland conversion to cropland. *Geoderma* **2023**, *429*, No. 116229.
(69) Sun, Y.; Genton, M. G. Functional boxplots. *Journal of Computational and Graphical Statistics* **2011**, *20* (2), 316−334.
(70) Debouk, H.; Altimir, N.; Sebastià, M. T. Maximizing the information obtained from chamber-based greenhouse gas exchange measurements in remote areas. *MethodsX* **2018**, *5*, 973−983.
(71) 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* **2007**, *12* (4), 292−305.
(72) Matthias, A. D.; Blackmer, A. M.; Bremner, J. M. A. Simple chamber technique for field measurement of emissions of nitrous oxide from soils. *American Society of Agronomy, Crop Science Society of America, and Soil Science Society of America* **1980**, *9* (2), 251−256.
(73) Bouwman, A. F.; Boumans, L. J. M.; Batjes, N. H. Emissions of N2O and NO from fertilized fields: Summary of available measurement data. *Global Biogeochem. Cycles* **2002**, *16* (4), 6-1−6-13.
(74) Rochette, P.; Eriksen-Hamel, N. S. Chamber Measurements of Soil Nitrous Oxide Flux: Are Absolute Values Reliable. *Soil Science Society of America Journal* **2008**, *72* (2), 331.
(75) Collier, S. M.; Ruark, M. D.; Oates, L. G.; Jokela, W. E.; Dell, C. J. Measurement of greenhouse gas flux from agricultural soils using static chambers. *J. Visualized Exp.* **2014**, *90*, 52110.
(76) Zanatta, J. A.; Alves, B. J. R.; Bayer, C.; Tomazi, M.; Fernandes, A. H. B. M.; Costa, F. D. S.; Carvalho, A. D. *Protocol for measuring soil greenhouse gas fluxes*; 1st ed., Embrapa Florestas: **2014**, ISSN 1980−3958.
(77) Lakshani, M. M. T.; Deepagoda, T. K. K. C.; Clough, T. J.; Jayarathne, J. R. R. N.; Thomas, S.; Balaine, N.; Elberling, B.; Smits, K. Effects of Soil Moisture on Simulated Methane Flow Under Varying Levels of Soil Compaction. In *12th International Conference on Structural Engineering and Construction Management: Proceedings of the ICSECM*; Springer Nature Singapore: Singapore, **2021**; pp 271−282.
(78) Vasconcelos, A. L. S.; Pires, I. C.; Ferrão, G. E.; Neto, M. S. Agriculture and greenhouse gases - case studies in Brazil. In *Trópica Magazine: Agricultural and Biological Sciences*; **2019**, 10.2.
(79) Siqueira-Neto, M. Estoque de carbono e nitrogênio do solo com diferentes usos no Cerrado em Rio Verde (GO). Tese (Doutorado em Energia Nuclear na Agricultura) - Centro de Energia Nuclear na Agricultura, University of São Paulo: Piracicaba, **2006**. DOI:. Acesso em: 2023−12−05.
(80) Frazão, L. A. Conversion of the Cerrado into pasture and agricultural systems: effects on the dynamics of soil organic matter. Dissertation (Master's in Chemistry in Agriculture and the Environment) − Center for Nuclear Energy in Agriculture: University of São Paulo, Piracicaba, 2007. doi:10.11606/D.64.2007.tde-18092007-113334. Accessed on: 2023−12−05.
(81) Carvalho, J. L. N. Carbon dynamics and greenhouse gas flow in croplivestock integration systems in the Amazon and Cerrado. Thesis (Doctorate in Soils and Plant Nutrition) - Escola Superior de

Agricultura Luiz de Queiroz, University of São Paulo: Piracicaba, 2010. DOI: , Accessed on: 2023−12−05.
(82) Siqueira-Neto, M.; Popin, G. V.; Piccolo, M. C.; Corbeels, M.; Scopel, E.; Camargo, P. B.; Bernoux, M. Impacts of land use and cropland management on soil organic matter and greenhouse gas emissions in the Brazilian Cerrado. *European Journal of Soil Science* **2021**, *72* (3), 1431−1446.
(83) Oertel, C.; Matschullat, J.; Zurba, K.; Zimmermann, F.; Erasmi, S. Greenhouse gas emissions from soils—A review. *Geochemistry* **2016**, *76* (3), 327−352.
(84) Tubiello, F. N.; Rosenzweig, C.; Conchedda, G.; Karl, K.; Gütschow, J.; Xueyao, P.; Obli-Laryea, G.; Wanner, N.; Qiu, S. Y.; Barros, J. D.; Flammini, A.; Mencos-Contreras, E.; Souza, L.; Quadrelli, R.; et al. Greenhouse gas emissions from food systems: building the evidence base. *Environ. Res. Lett.* **2021**, *16* (6), No. 065007.
(85) Funk, R.; Pascual, U.; Joosten, H.; Duffy, C.; Pan GenXing, P. G.; Scala, N. L.; Gottschalk, P.; Banwart, S. A.; Batjes, N.; ZuCong, C.; Six, J.; Noellemeyer, E. From potential to implementation: an innovation framework to realize the benefits of soil carbon. In *Soil carbon: Science, management and policy for multiple benefits*, CABI: Wallingford UK, 2015; pp 47−59.
(86) Da Costa, L. M.; de Araújo Santos, G. A.; de Mendon, G. C.; Morais Filho, L. F. F.; de Meneses, K. C.; de Souza Rolim, G.; La Scala, N. Jr. Spatiotemporal variability of atmospheric CO2 concentration and controlling factors over sugarcane cultivation areas in southern Brazil. *Environ., Dev. Sustainability* **2022**, *24* (4), 5694−5717.
(87) Abreu, N. L.; da Cruz Ribeiro, E. S.; de Sousa, C. E. S.; Moraes, L. M.; de Sousa, J. V. C.; de Abreu Faria, L.; Ruggieri, A. C.; da Silva Cardoso, A.; Faturi, C.; do Rêgo, A. C.; da Silva, T. C. Mudanças de uso da terra e emissão de gases de efeito estufa: uma explanação sobre os principais drivers de emissão. *Braz. Anim. Sci.* **2024**, *25*, No. 77646E.
(88) Bento, C. B.; Brandani, C. B.; Filoso, S.; Martinelli, L. A.; do Carmo, J. B. Effects of extensive-to-intensive pasture conversion on soil nitrogen availability and CO_2 and N_2O fluxes in a Brazilian oxisol. *Agric., Ecosyst. Environ.* **2021**, *321*, No. 107633.
(89) Bento, C. B.; Filoso, S.; Pitombo, L. M.; Cantarella, H.; Rossetto, R.; Martinelli, L. A.; do Carmo, J. B. Impacts of sugarcane agriculture expansion over low-intensity cattle ranch pasture in Brazil on greenhouse gases. *Journal of environmental management* **2018**, *206*, 980−988.
(90) Martins, M. R.; Jantalia, C. P.; Polidoro, J. C.; Batista, J. N.; Alves, B. J.; Boddey, R. M.; Urquiaga, S. Nitrous oxide and ammonia emissions from N fertilization of maize crop under no-till in a Cerrado soil. *Soil and Tillage Research* **2015**, *151*, 75−81.
(91) Lopes, I. M.; Pinheiro, E. F. M.; Lima, E.; Ceddia, M. B.; Campos, D. V.; Alves, B. J. R. *Emissões de N2O em Solos sob Cultivo de Cana-de-açúcar no bioma Mata Atlântica: Efeito dos Sistemas de Colheita e da Adubação com Vinhça*; Revista Virtual de Química: Niterói, **2017**, *9*, (5) 1930-1943.
(92) Meurer, K. H. E.; Franko, U.; Stange, C. F.; Rosa, J. D.; Madari, B. E.; Jungkunst, H. F. Direct nitrous oxide (N2O) fluxes from soils under different land use in Brazil—a critical review. *Environ. Res. Lett.* **2016**, *11* (2), No. 023001.
(93) Canadell, J. G.; Monteiro, P. M.; Costa, M. H.; Da Cunha, L. C.; Cox, P. M.; Eliseev, A. V.; Henson, S.; Ishii, M.; Jaccard, S.; Koven, C.et al.*Global carbon and other biogeochemical cycles and feedbacks*. IPCC AR6 WGI, final government distribution, chapter-5, **2021**; pp -58.
(94) Schmutz, A.; Jacques, J.; Bouveyron, C.; Cheze, L.; Martin, P. Clustering multivariate functional data in group-specific functional subspaces. *Computational Statistics* **2020**, *35* (3), 1101−1131.
(95) Jacques, J.; Preda, C. Model-based clustering for multivariate functional data. *Computational Statistics & Data Analysis* **2014**, *71*, 92−106.
(96) Besen, R. M.; Ribeiro, R. H.; Monteiro, A. N. T. R.; Iwasaki, G. S.; Piva, J. T. Soil conservation practices and greenhouse gases emissions in Brazil. *Sci. Agropecu.* **2018**, *9.3*, 429−439.
(97) da Silva, D. A. P.; Campos, M. C. C.; Mantovanelli, B. C.; dos Santos, L. A. C.; Soares, M. D. R.; da Cunha, J. M. Spatial variability of CO2 emission, temperature and soil humidity in a pasture area in the Amazon region, Brazil. *J. Agroveterinary Sci.* **2019**, *18* (1), 119−126.
(98) Bouwman, A. F.; Boumans, L. J. M.; Batjes, N. H. Modeling global annual N2O and NO emissions from fertilized fields. *Global Biogeochem. Cycles* **2002**, *16* (4), 28-1−28-9.
(99) Zheng, J.; Sakata, T.; Sukartiningsih; Fujii, K. Deciphering nitrous oxide emissions from tropical soils of different land uses. *Sci. Total Environ.* **2023**, *862*, No. 160916.