# A k-Means-Based-Approach to Analyze the Emissions of GHG in the Municipalities of MATOPIBA Region, Brazil

Lucas Ferreira-Paiva, Attawan G. L. Suela, Elizabeth R. Alfaro-Espinoza, Nancy A. Cardona-Casas, Domingos S. M. Valente, and Rodolpho V. A. Neves, *Member, IEEE*

***Abstract*—Brazil is the sixth largest emitter of greenhouse gases (GHG), with land use change and agriculture being the main source of these emissions. The recent expansion of agriculture in Brazil has been occurring mainly in the MATOPIBA region, a territorial division inserted on the Cerrado. Clustering methods are useful for driving hyper-local GHG emission reduction strategies, but they have yet to be applied at the municipal level, nor from the emissions of various economic sectors. In order to contribute to the identification of the most critical areas in relation to GHG emissions for MATOPIBA, this study proposes an approach of municipal clustering according to the percentage contribution of Agriculture, Land Use Change, Energy and Waste sectors in total emissions. The clustering was performed with the k-means algorithm, using the elbow method and the silhouette score to define the number of clusters. In addition, statistical and geostatistical analyses were conducted to assess the consistency and spatial autocorrelation of the groups formed. The approach was able to generate six clusters with distinct characteristics, showing the heterogeneous profile of GHG emissions from MATOPIBA. At the same time, the clustering of similar municipalities can help in making decisions about the best pro-environmental measures to reduce/remove GHG to contain global warming.**

***Index Terms*—biodiversity, carbon offset pricing, climate change impact, kmeans, sustainability risk assessment, spatial autocorrelation.**

## I. Introduction

Brazil is globally important for food production and conservation of natural resources [1]–[3]. However, the country is currently the sixth largest emitter of greenhouse gases (GHG). In 2020 it emitted approximately 2.16 billion tons of equivalent $CO_2$ ($GtCO_2e$), being the Agriculture and Land Use Change sectors achieved the largest contribution and represented 66% of emissions when added [4].

Due to this high level of emissions, the Brazilian government created pro-environmental strategies to mitigate emissions by reducing deforestation, without harming agricultural production. The main actions are the Brazilian credit program for low-carbon agriculture (ABC+) [5] and Payment for Environmental Services (PES). These strategies are important to encourage the maintenance, recovery and improvement of Brazilian ecosystems. They also provide benefits such as genetic heritage and associated traditional knowledge preservation, reduction of deforestation and forest degradation, and climate regulation [6], [7].

One of the biomes that needs these actions is the Cerrado, which has been the focus of agricultural and industrial expansion over the last 60 years [8]. This biome has suffered an alarming loss of native forest [1], [9], [10]. In the last 40 years it has reduced native cover to less than half of its original size [3]. These losses are critical, given the geographic importance, ecological characteristics and biodiversity of this biome [11], [12]. The recent agricultural expansion in Cerrado occurred mainly in the north of the biome, a region named MATOPIBA, acronym for the states of Maranhão (MA), Tocantins (TO), Piauí (PI) and Bahia (BA), known as the last agricultural frontier of Brazil. MATOPIBA is a territorial division created through a technical cooperation agreement signed in 2014 between some government departments and federal agencies to designate the potential area for agricultural expansion [13].

MATOPIBA is being extensively explored for the cultivation of agricultural commodities and beef production [8]. Soybean stands out among the main crops in the region [14], but other crops such as corn, cotton and rice also play a significant role [3], [15]. The strong agricultural expansion, based on the suppression of the native forest of the Cerrado/MATOPIBA, caused the temperature of this region to increase by $1°C$, in 2021, being the biggest increase registered in Brazil [16]. This scenario highlights the low effectiveness of the implementation of the current GHG mitigation plans [17] and the need to strengthen actions that ensure compliance with the Paris agreement to contain global warming.

In this context, clustering methods have been used to contribute to strategies for monitoring and containing climate change. Among them, hierarchical clustering by Ward's linkage was used to identify similar rainfall patterns [18], [19] and to locate areas with homogeneous distribution of fire outbreaks

Lucas Ferreira-Paiva is with Programa de Pós-Graduação em Ciência da Computação, Departamento de Informática, Universidade Federal de Viçosa, Viçosa, Brazil e-mail:lucas.f.paiva@ufv.br

Attawan Guerino Locatel Suela is with Programa de Pós-Graduação em Economia Aplicada, Departamento de Economia Rural, Universidade Federal de Viçosa, Viçosa, Brazil e-mail:attawan.suela@ufv.br

Elizabeth R. Alfaro-Espinoza is with Programa de Pós-Graduação em Bioinformática, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil e-mail:elizaespinoza@ufmg.br

Nancy A. Cardona-Casas is with Programa de Pós-Graduação em Solos e Nutrição de Plantas, Departamento de Solos, Universidade Federal de Viçosa, Viçosa, Brazil e-mail:nancy.casas@ufv.br

Domingos S. M. Valente is with Departamento de Engenharia Agrícola, Universidade Federal de Viçosa, Viçosa, Minas Gerais, Brazil e-mail:valente@ufv.br

Rodolpho V. A. Neves is with Departamento de Engenharia Elétrica, Universidade Federal de Viçosa, Viçosa, Minas Gerais, Brazil e-mail:rodolpho.neves@ufv.br

[20]. Recently, this method was used to analyze Brazilian biomes from the total emissions of GHG and the distribution of fire outbreaks [21].

Although clustering by biome can help environmental agencies, further studies of clustering GHG emissions from economic sources at the municipal level in MATOPIBA can help in public policies for conservation and sustainability, as well as efforts to minimize GHG emissions. Thus, clustering allows to identify the municipalities in which the application of a certain sort of pro-environmental action could have the greatest effect on GHG reduction. This decreases monitoring expenses and maximizes economic resources for the implementation of these actions.

Detailed emission information by municipality is available on the The Greenhouse Gas Emission and Removal Estimating System (SEEG) platform for five major economic sectors: Agriculture, Land Use Change (LUC), Energy, Industrial Processes and Product Use (IPPU) and Waste [17]. Therefore, the information of this platform could help in gaining a comprehensive understanding of MATOPIBA emissions by identifying patterns and exceptions at municipal level.

Given the above, the objectives of this work are: (i) to propose an approach based on k-means to group municipalities according to their GHG emissions by sector; and (ii) to conduct a case study with the MATOPIBA region to assess the ability of the proposed approach to identify and to map similarities and differences in municipal emission profiles, driving hyper-local GHG emission reduction strategies.

## II. SEEG Platform

SEEG estimates annual emissions of GHG for the whole of Brazil according to the IPPU guidelines and the methodology of the Brazilian Inventories of Anthropogenic Emissions and Removals of GHG, prepared by the Ministry of Science, Technology and Innovation [22]. Estimated emissions are available for all GHG and also considering the equivalence of other gases in relation to $CO_2$. The equivalence is done with Global Warming Potential (GWP) and Global Temperature Change Potential (GTP). In addition, gross and net emissions are provided, and removals are accounted for in net emissions.

The SEEG platform [23] has five major sectors that are sources of emissions. These sectors are Agriculture, Energy, LUC, IPPU, they are treated with the same level of detail contained in the Intergovernmental Panel on Climate Change emission inventories. The platform presents disaggregation at national, state and municipal levels. Table I shows the main emission sources for each sector in Brazil in 2018 [17].

TABLE I

Main GHG emitting activities by sector in Brazil in 2018 [17], ranked from highest to lowest emitter.

| Sector | Main sources of direct emission in the sector |
|---|---|
| LUC | Deforestation, liming and forest waste burning. |
| Agriculture | Breeding of bovine herd (by enteric fermentation), application of synthetic fertilizers, management of animal waste, rice cultivation irrigated and residue burning (such as sugarcane straw). |
| Energy | Transport (divided into cargo and passengers), energy consumption in industry, fuel production and electricity generation. |
| IPPU | Steel industry (such as pig iron and steel) and cement production. |
| Waste | Disposal of urban solid waste in controlled landfills, sanitary landfills and dumps, and the treatment of industrial and domestic effluents and waste incineration. |

## III. Model Construction and Evaluation

The characterization of the emission profile of the MATOPIBA municipalities was carried out in four main stages: *(i)* the collection of emission data by municipality in the SEEG platform and georeferenced data from the Geodata BR project; *(ii)* the pre-processing, with imputation of missing values and transformations of the data; *(iii)* the clustering of municipalities with similar profiles; and *(iv)* the cluster analysis to identify existing profiles. The entire process was performed using Python and it was available on GitHub [24].

### A. Data Collection

Data for the Agriculture, LUC, Energy, and Waste sectors were collected from the SEEG platform, taking into account the emissions of the 337 municipalities that form the MATOPIBA. Emissions from IPPU sector were not used due to lack of data of this sector for MATOPIBA. The gross $CO_2$ equivalent emissions were collected using the GWP methodology for 2018, the last year available until the access date (October 9th, 2021). In addition, georeferenced data were collected for the groups spatial visualization, from the Geodata BR project [25].

### B. Pre-Processing

The first pre-processing step was imputation of missing values of nine municipalities without Energy emissions information. For this we use the method for multivariate feature imputation from the scikit-learn library [26]. Preliminary analysis identified a high variability of absolute values for municipal emissions, as well as several outliers, which made it difficult to form the groups even after normalizing the data. To mitigate this effect, the municipalities were clustered using contributions from each sector, and absolute emissions were analyzed. The emissions contribution from a sector $CE_S$ was defined as the absolute emission of the sector ($E_S$) divided by the absolute emission of the four sectors $E_t$. In this way, four new features were created according to (1).

$$CE_S = \frac{E_S}{E_t} \tag{1}$$

### C. Clustering

A k-means algorithm was used with scikit-learn to identify MATOPIBA emission profiles. The first step was the identification of potential existing groups using the elbow method with initial $k$ ($k_i$) equals to two and final $k$ ($k_f$) equals to 30 groups of municipalities. For each value of $k$, the Within-Cluster Sum of Square (WCSS) was calculated. Finally, the value of $k_c$ was defined as the elbow point, which results in the greatest distance between the ($WCSS[k]$) curve and a straight line drawn between the points $(k_i, WCSS[k_i])$ and $(k_f, WCSS[k_f])$. The distance calculation ($d[k]$) for each value of $k$ is presented in (2), where $y = WCSS[k]$, $y_i = WCSS[k_i]$ and $y_f = WCSS[k_f]$.

$$d[k] = \frac{|(y_i - y_f)k - (k_i - k_f)y + k_f y_i - y_f k_i|}{\sqrt{(y_i - y_f)^2 + (k_i - k_f)^2}} \quad (2)$$

Furthermore, the silhouette value was calculated, which weights the distances of a point from the center of its respective cluster with the center of neighboring clusters. The silhouette value was calculated for the three potential group values found with the elbow method ($k_c - 1$, $k_c$ and $k_c + 1$). Both methods were used because they are recommended to find the value of $k$ in small databases [2].

### D. Consistency Assessment of the Groups

To assess the consistency of the identified groups, we tested the existence of significant differences between the means of groups emissions for each of the sectors. As the groups are constituted by different municipalities, it was assumed that they are independent samples. Thus, the t-student test was used to compare normal distributions and non-different variances. The U Mann-Whitney test was used for distributions that did not match the criterion of normality or homoscedasticity.

The null hypothesis was tested that the emissions of a given sector of a group were equal to the emissions of the same sector of another group. If $p < 0.05$, the null hypothesis was rejected and the emissions were assumed to be different. In this analysis, we seek to assess if the used approach to identify the groups allowed us to find groups that have significant differences. All statistical analyses were performed using scipy [27]

### E. Interpretation of Groups

With the groups defined, we aim to understand the distinct characteristics of each group, which might be neglected when considering MATOPIBA as a homogeneous region. First, the average contributions of each sector to each group's emissions were observed. This allowed us to identify the predominance of one or more sectors inside the groups of municipalities. Furthermore, we observed the spatial distribution of the groups using GeoPandas [28].

In addition, we detect spatial autocorrelations among MATOPIBA municipalities based on their emissions contribution per sector using the univariate Moran index. We used the Global Moran I to calculate the degree of global spatial autocorrelation for each emission source in MATOPIBA. For local autocorrelation, we calculated the Local Indicators of Spatial Association (LISA) to determine similarity and correlation among the municipalities and nearby municipalities. All autocorrelation analyses were performed using PySAL [29].

Once the normality of the data was confirmed, Pearson correlations between emissions from different sectors and the average absolute emissions of each group of municipalities were evaluated. With this information, we sought to find priority groups for the application of emission-reduction measures, as well as to highlight probable heterogeneity between the emission profiles of the municipalities in MATOPIBA.

## IV. MATOPIBA'S CASE STUDY

In this section, we presented the results of the proposed approach with MATOPIBA emission data. Therefore, we presented the discovered groups, their statistical comparison and spatial distribution and autocorrelation.

### A. Consistency of Identified Groups

The performance of the elbow method shown that the ideal group number is six ($k = 6$), as illustrated in Fig. 1a. In addition, silhouette coefficients were evaluated for $k = 5$ and $k = 7$. The silhouette coefficients ($CS$) reinforced the assumption of the existence of six groups, since this division had the highest coefficient ($CS = 0.424$), as shown in Fig. 1b. The negative coefficients presented for some municipalities in G1, G4 and G6 groups (Fig. 1b) indicate that these municipalities may belong to other groups. The low occurrence of municipalities with negative $CS$ indicates that the formed groups are consistent.

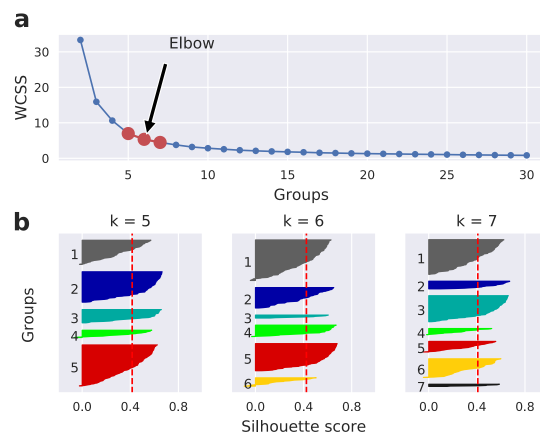



Fig. 1. Selection of the number of groups using the elbow method (a) and the silhouette score (b).

The group heterogeneity is verified in Fig. 2, which shows the means and standard deviations for each group by sector. The density estimation of the emissions from each sector to each group is represented in Fig. 2a. It is observed that the groups have greater variation in relation to Agriculture, LUC and Energy than Waste.

Sixty tests were performed to compare all groups in pairs for each sector, with only eight tests did not show a significant difference, as shown in Fig 2b. It is clear that all groups differ statistically from the others for at least two emission sectors. Therefore, it was possible to prove the ability of the proposed model to identify groups with different emission profiles in MATOPIBA.

### B. Groups Presentation

As shown in Table II, in 2018, the main responsible for MATOPIBA emissions was LUC sector with 54.25% of emissions, mainly due to the replacement of native vegetation with monoculture plantations such as soybeans, corn and cotton and livestock [30], [31]. G1, the group with the most municipalities, has an emission profile similar to MATOPIBA

TABLE II
COMPOSITION OF GROUPS BY THE PROPORTION OF EMISSIONS FROM SECTOR AND THE PROPORTION OF MUNICIPALITIES BY STATE.

| Groups | Emissions(%) | | | | Municipalities | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Agr | LUC | Ener | Was | MA (%) | TO (%) | PI (%) | BA (%) | Total |
| MATOPIBA | **35.48** | **54.25** | 7.53 | 2.74 | **40.06** | **41.25** | 9.79 | 8.90 | 337 |
| G1 | **30.85** | **62.38** | 4.44 | 2.33 | **45.08** | **40.16** | 9.02 | 5.74 | **122** |
| G2 | **51.07** | **42.48** | 4.34 | 2.11 | **30.65** | **61.29** | 6.45 | 1.61 | 62 |
| G3 | 14.70 | 17.42 | **60.65** | 7.23 | **40.00** | **50.00** | 0.00 | 10.00 | 10 |
| G4 | **75.18** | 15.63 | 5.66 | 3.53 | 8.57 | **42.86** | 11.43 | **37.14** | 35 |
| G5 | 16.90 | **78.57** | 3.11 | 1.42 | **54.88** | 24.39 | 14.63 | 6.10 | 82 |
| G6 | **33.16** | **33.66** | **25.70** | 7.47 | **34.62** | **46.15** | 7.69 | 11.54 | 26 |

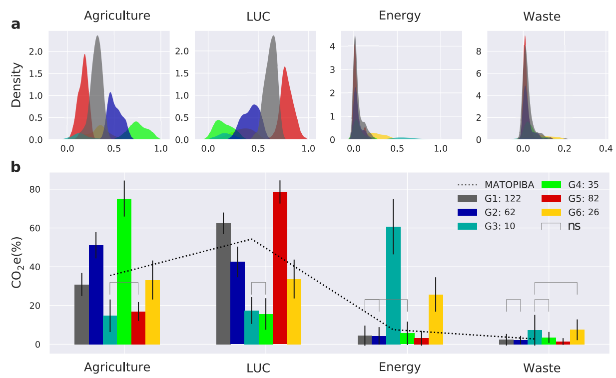



Fig. 2. Comparison of the emissions by sectors and groups. (a) Density estimation comparison of the emissions percentage from each economic sector to each group. (b) Average proportion comparison of emissions, number of municipalities allocated to each group and identification of sectors without significant difference (ns).

as a whole. In terms of emission contributions, the difference between them is less than 10%.

In Fig. 2 we observe proximity between G1 and G5. However, G5 stands out for the great contribution of LUC emissions (78.57%, Table II). In contrast to G5, the G4 group stands out for agricultural emissions (75.18%, Table II). Following G4 is the G2 group, which has a higher percentage of GHG emissions related to agriculture (51.07%, Table II), followed by emissions from LUC (42.48%, Table II). G6 has the best sector balance, with Agriculture, LUC and Energy each contributing over 25%. Finally, G3 stands out for being the only group with the largest contribution of emissions due to Energy, with 60.65% of emissions, as shown in Table II.

### *C. Group's Spatial Distribution and Autocorrelation*

In Table II we observe that all groups have more than 79% of cities belonging to Tocantins or Maranhão. This is expected since these states represent 81% of MATOPIBA's cities. Being G5 the highest number of cities in Maranhão state (54.88%). The only exception is the G4 group, with a strong representation of cities in Bahia (37.14%) and a low representation of cities in Maranhão.

The heterogeneity of MATOPIBA is highlighted in Fig 3a, as the groups have municipalities distributed over the region, especially for G2. The municipalities of G6 and G2 are dispersed as well, but are absent from much of the region. On the other hand, proximity between municipalities within the same group can be observed, forming subgroups. G5 subgroups can be found in Piauí and Maranhão. G1 is another example, with a subgroup that goes from the south to the southeast of Tocantins and extends to Bahia, Maranhão and Piauí. Finally, there is a G4 subgroup in Bahia.

In Fig 3b, we visualize the global and local univariate spatial autocorrelation for each sector, based on the Moran's I Index. For Agriculture, low-emission municipalities can be observed clustered with low-emission municipalities. Furthermore, various agglomerations of municipalities with a large percentage of emissions by Agriculture may be seen. When compared with Figure 3a, we identified that this high-high correlation corresponds with agglomerations of municipalities in G4 and G2, whose emissions are predominantly from Agriculture.

The high-high connections for LUC are mostly observed in regions with agglomeration of municipalities in the G1 and G5 groups. We also see the appearance of High-High correlations in areas where Agriculture had Low-Low correlations. Local autocorrelations related with Waste do not correlate with the existence of municipalities of specific groups, and they are mostly of the Low-Low quadrant, which is expected, since all groups have low emissions by Waste. Finally, few regions show significant autocorrelation in the Energy sector.

This spatial autocorrelations from the emissions proportion of Agriculture and LUC can facilitate the optimization of the economic resources necessary for the implementation of pro-environmental measures.

### *D. Emissions Profile*

The dispersion plot of emission proportions in relation to Agriculture, LUC and Energy is shown in Fig. 4. A negative correlation ($r2 = -0.76$) exists between the emission contributions of Agriculture and LUC. Furthermore, when the municipalities of G3 are excluded, the correlation increases ($r2 = -0.88$). This association shows that the cities that generated the most GHG by LUC also emitted the least by Agriculture, indicating that these MATOPIBA municipalities are at different stages of agricultural expansion [7], [12]. The G5 municipalities would be on the initial stage, with a high level of native forest suppression and lower agriculture. On the other hand, Agriculture would already be consolidated in the

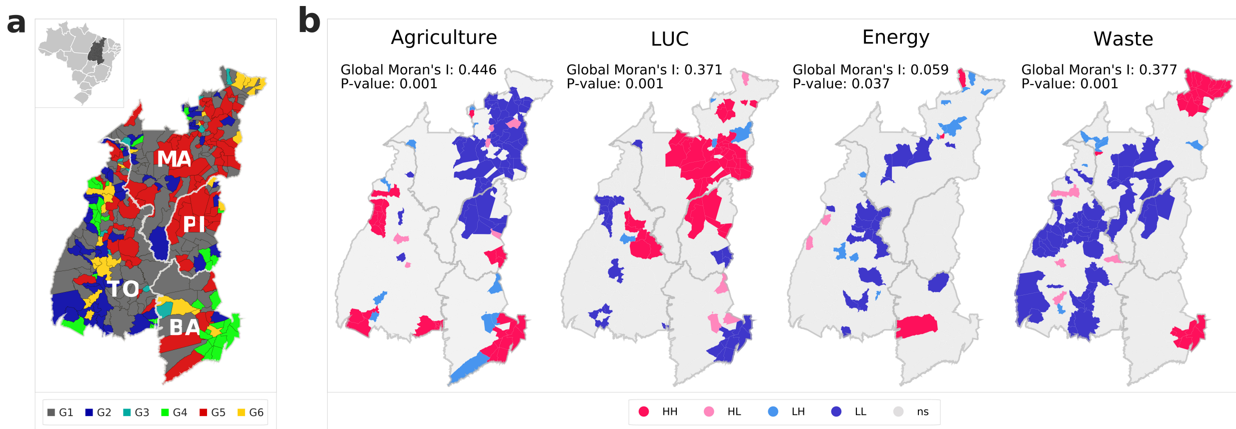



Fig. 3. Spatial distribution and autocorrelation. (a) Spatial distribution of the created groups. (b) Moran's I significant map for GHG emissions from each economic sector, the colored regions are defined by the LISA quadrants: HH (High-High), LL (Low-Low), HL (High-Low) and LH (Low-High).

G4 municipalities, and the majority of the native vegetation would already be devastated [1], [32]. Thus, groups G1, G2 and G6 would be in intermediate stages. However, the validation of this observation lacks verification of this change in the emission profile over time.

It is crucial to note that the groups G1 and G5 occupy the majority of MATOPIBA and emit the most per LUC. Deforestation, the principal emitting activity of LUC (Table1) is temporal, because the forest resources are limited. Consequently, if effective efforts to reduce deforestation are not implemented, emissions per LUC will fall in the future due to a lack of forest to cut.

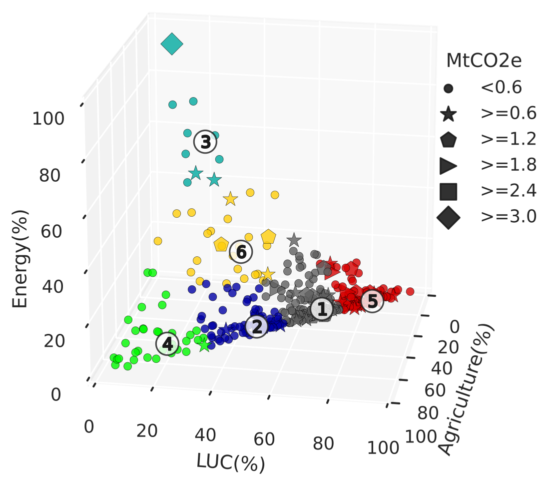



Fig. 4. Dispersion of absolute emissions for each group. The symbols in the graph represent the absolute value of total emissions of each municipality.

Furthermore, after excluding G3 and G6, which had a high contribution from the Energy sector, the groups that most emitted GHG were those belonging to the groups with the largest contribution of emissions due to LUC, as shown in Fig. 5. Indicating that the reduction of LUC is essential for the reduction of emissions from MATOPIBA.

Another sector with high GHG emissions is Energy, mainly represented by G3 and G5. GHG emissions from the Brazilian

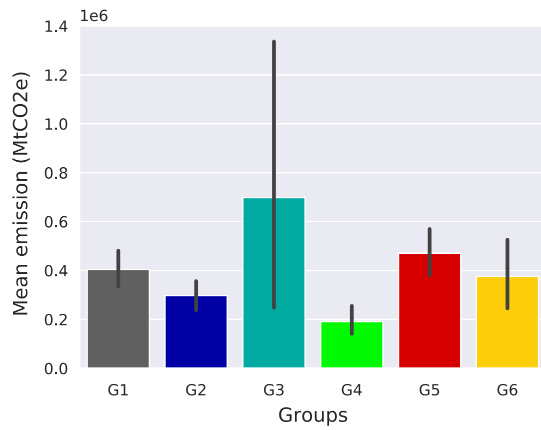



Fig. 5. The averages and standard deviations of the total emissions for each group.

Energy sector were mostly represented by freight transport in 2018 [17]. This scenario can be represent G5, which seems to be composed of urbanized municipalities such as 'Barreiras' on Bahia, 'Timon' on Mato Grosso, 'Teresina', capital of Piauí, and 'Palmas', capital of Tocantins and the city with the largest population of MATOPIBA in 2018 according to SEEG data.

On the other hand, the high contribution of Energy emissions in the G3 group can be better explained through the concentration of thermoelectric plants existing in some municipalities of the group, such as 'Miranda do Norte' and 'Santo Antônio dos Lopes' of Maranhão state. In 'Santo Antônio dos Lopes' there is a thermoelectric complex that places it among the Brazilian municipalities with the highest emission of GHG per energy generation [33]. The high standard deviation of G3, as shown in Fig 5, is mainly due to this municipality, with more than 90% of GHG emissions related to Energy, while the other municipalities had emissions between 45% and 73%.

Therefore, measures to reduce GHG are necessary for the municipalities of G1 and G5, with a focus on reducing deforestation and for G3 with a focus on making improvements in the energy matrix and implementing practices that contribute to the direct removal of GHG, such as use microalgae in the biofixation of $CO_2$ in thermoelectric power plants [34], [35]. It should be noted that mitigation measures for GHG are necessary in all groups and the clusterization proposed in this study facilitates targeting the sectors to be prioritized.

This study contributes to decision making based on the magnitude of emissions and the proximity of areas. However, other factors have to be considered, such as social, economic, institutional, environmental, and technological aspects [36] for proposing pro-environmental measures. The impacts of mitigation measures are context specific, so it is important to consider each mitigation strategy on a case-by-case basis.

## V. Conclusion

In this work, an approach based on k-means was presented to group municipalities according to their GHG emissions by sector. When applied to the emissions data from the MATOPIBA municipalities in 2018, the approach identified six groups with different emission profiles. The main sectors that differentiated the groups were LUC, Agriculture and Energy. The wide variation in the contribution to total emissions by these sectors showed that the MATOPIBA municipalities are in different stages of deforestation and agricultural exploitation. In addition, one of the groups identified is composed of municipalities with the largest contribution of emissions due to the Energy sector.

Through spatial autocorrelation analysis, the effects of spatial differentiation and agglomeration of municipalities by the emissions of each economic sector were evidenced. Therefore, while exposing the complexity of MATOPIBA, the proposed approach makes it possible to group similar municipalities and highlight important spatial autocorrelations. This grouping can facilitate the targeting of GHG reduction policies such as the ABC+ plan and the PES, helping Brazil to comply with global warming containment treaties such as the Paris agreement.

The study has two main limitations: (i) emissions were not divided by economic sub-sectors, requiring data complementation to gain more conclusive evidence; and (ii) data from 2018 were collected, which may not completely reflect the current state of the MATOPIBA municipalities. These constraints can be reduced in future works that take into account the subsectors of emissions as well as the timeline of emissions of the municipalities in order to carry out the grouping. It is worth noting that the proposed approach can be used to assess the profile of GHG emissions from other regions of the country due to its capability and practicability.

## Acknowledgment

The authors thank the Coordination for the Improvement of Higher Education Personnel (CAPES) - Financing code 001 and Fellowship 88887.517813/2020-00 for providing financial support for the development of this work.

## References

[1] G. Sparovek, V. Guidotti, L. F. G. Pinto, G. Berndes, A. Barretto, and F. Cerignoni, "Asymmetries of cattle and crop productivity and efficiency during Brazil's agricultural expansion from 1975 to 2006," *Elementa: Science of the Anthropocene*, vol. 6, 1 2018.

[2] Y. l. P. d. Waroux, R. D. Garrett, J. Graesser, C. Nolte, C. White, and E. F. Lambin, "The restructuring of south american soy and beef production and trade under changing environmental regulations," *World Development*, vol. 121, pp. 188–202, 9 2019.

[3] S. P. Polizel, R. M. d. S. P. Vieira, J. Pompeu, Y. d. C. Ferreira, E. R. d. Sousa-Neto, A. A. Barbosa, and J. P. H. B. Ometto, "Analysing the dynamics of land use in the context of current conservation policies and land tenure in the cerrado – MATOPIBA region (Brazil)," *Land Use Policy*, vol. 109, p. 105713, 10 2021.

[4] R. F. Potenza, G. d. O. Quintana, A. M. Cardoso, D. S. Tsai, M. d. S. Cremer, F. B. e. Silva, K. Carvalho, I. Coluna, J. Shimbo, C. Silva, E. Souza, B. Zimbres, A. Alencar, C. Angelo, and T. Azevedo, "Análise das emissões brasileiras de e suas implicações para as metas climáticas do brasil 1970 – 2020," tech. rep., Sistema de Estimativas de Emissões e Remoções de Gases de Efeito Estufa (SEEG), 2021.

[5] P. e. A. Ministério da Agricultura, *Plano setorial para adaptação à mudança do clima e baixa emissão de carbono na agropecuária com vistas ao desenvolvimento sustentável (2020-2030): visão estratégica para um novo ciclo*. 2021.

[6] F. L. M. Freitas, *Brazilian public protection regulations and the preservation of ecosystem services and biodiversity*. PhD thesis, 2019.

[7] R. Loyola, C. Rezende, and B. Ribeiro, "Áreas prioritárias para conservação e restauração no MATOPIBA," tech. rep., Conservação Internacional Brasil, 2021.

[8] G. R. Lopes, M. G. B. Lima, and T. N. P. d. Reis, "Maldevelopment revisited: Inclusiveness and social impacts of soy expansion over Brazil's cerrado in MATOPIBA," *World Development*, vol. 139, p. 105316, 3 2021.

[9] B. Rudorff and J. Risso, "Análise geoespacial da dinâmica das culturas anuais no bioma cerrado: 2000 a 2014," tech. rep., Agrosatélite Geotecnologia Aplicada, 2015.

[10] J. Bonanomi, F. R. Tortato, R. d. S. R. Gomes, J. M. Penha, A. S. Bueno, and C. A. Peres, "Protecting forests at the expense of native grasslands: Land-use policy encourages open-habitat loss in the Brazilian cerrado biome," *Perspectives in Ecology and Conservation*, vol. 17, pp. 26–31, 1 2019.

[11] D. Ezzine-de-Blas, E. Corbera, and R. Lapeyre, "Payments for environmental services and motivation crowding: Towards a conceptual framework," *Ecological Economics*, vol. 156, pp. 434–443, 2 2019.

[12] K. Costa, J. Brandão, A. Carneiro, J. Monti, and H. v. d. Vliet, "Potencial regional de expansão da soja no MATOPIBA," tech. rep., Solidaridad Brasil, 2021.

[13] E. d. Miranda, "MATOPIBA, delimitação, caracterização, desafios e oportunidades para o desenvolvimento," tech. rep., Empresa Brasileira de Pesquisa Agropecuária, 2015.

[14] C. N. de Abastecimento, *Acompanhamento da safra brasileira de grãos*. 2020.

[15] B. Rudorff and J. Risso, "Análise geoespacial da soja no bioma cerrado: Dinâmica da expansão. aptidão agrícola da soja. sistema de avaliação para compensação financeira: 2001 a 2019," tech. rep., Agrosatélite Geotecnologia Aplicada, 2020.

[16] Intergovernmental Panel on Climate Change, "Climate change 2021: Impacts, adaptation and vulnerability. summary for policymakers," 2022.

[17] C. Angelo and C. Rittl, "Análise das emissões brasileiras de gases de efeito estufa e suas implicações para as metas climáticas do brasil 1970 – 2018," tech. rep., Sistema de Estimativas de Emissões e Remoções de Gases de Efeito Estufa (SEEG), 2019.

[18] G. B. Lyra, J. F. Oliveira-Júnior, and M. Zeri, "Cluster analysis applied to the spatial and temporal variability of monthly rainfall in alagoas state, northeast of brazil," vol. 34, pp. 3546–3558, 2 2014.

[19] P. E. Teodoro, J. F. de Oliveira-Júnior, E. R. da Cunha, C. C. G. Correa, F. E. Torres, V. M. Bacani, G. Gois, and L. P. Ribeiro, "Cluster analysis applied to the spatial and temporal variability of monthly rainfall in mato grosso do sul state, brazil," vol. 128, pp. 197–209, 10 2015.

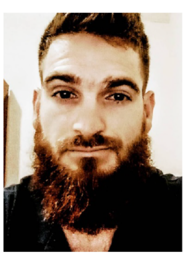
[20] R. H. Caúla, J. F. Oliveira-Júnior, G. B. Lyra, R. C. Delgado, and P. F. L. H. Filho, "Overview of fire foci causes and locations in brazil based on meteorological satellite data from 1998 to 2011," vol. 74, pp. 1497–1508, 2 2015.

[21] C. A. da Silva Junior, P. E. Teodoro, R. C. Delgado, L. P. R. Teodoro, M. Lima, A. de Andréa Pantaleão, F. H. R. Baio, G. B. de Azevedo, G. T. de Oliveira Sousa Azevedo, G. F. Capristo-Silva, D. Arvor, and C. U. Facco, "Persistent fire foci in all biomes undermine the paris agreement in brazil," vol. 10, 10 2020.

[22] T. R. De Azevedo, C. Costa Junior, A. Brandão Junior, M. d. S. Cremer, M. Piatto, D. S. Tsai, P. Barreto, H. Martins, M. Sales, T. Galuchi, *et al.*, "Seeg initiative estimates of brazilian greenhouse gas emissions from 1970 to 2015," *Scientific data*, vol. 5, no. 1, pp. 1–43, 2018.

[23] http://plataforma.seeg.eco.br/cities/statistics
[24] https://github.com/lucas-fpaiva/matopiba-emissions
[25] https://github.com/tbrugz/geodata br
[26] https://scikit learn.org/stable
[27] https://docs.scipy.org/doc/scipy/reference/stats.html
[28] https://geopandas.org/en/stable/
[29] https://pysal.org/
[30] A. Suela, M. S. Nazareth, and D. A. d. Cunha, "Efeitos ambientais da implementação do plano ABC no MATOPIBA: uma abordagem por insumo-produto," *Revista Brasileira de Estudos Regionais e Urbanos*, vol. 14, pp. 629–656, 10 2020.
[31] A. G. L. Suela, E. A. d. Santos, G. L. Suela, and I. M. Trotter, "Um modelo de resposta a choques no crescimento das emissões de gases de efeito estufa," in *Anais do 59º Congresso da Sociedade Brasileira de Economia, Administração e Sociologia Rural (SOBER) e 6º Encontro Brasileiro de Pesquisadores em Cooperativismo (EBPC)*, 2021.
[32] B. B. N. Strassburg, T. Brooks, R. Feltran-Barbieri, A. Iribarrem, R. Crouzeilles, R. Loyola, A. E. Latawiec, F. J. B. O. Filho, C. A. d. M. Scaramuzza, F. R. Scarano, B. Soares-Filho, and A. Balmford, "Moment of truth for the cerrado hotspot," *Nature Ecology and Evolution*, vol. 1, 3 2017.
[33] Instituto de Energia e Meio Ambiente, "Municípios da amazônia dominam emissões de carbono," 2021.
[34] M. G. de Morais and J. A. V. Costa, "Isolation and selection of microalgae from coal fired thermoelectric power plant for biofixation of carbon dioxide," *Energy Conversion and Management*, vol. 48, no. 7, pp. 2169–2173, 2007.
[35] E. M. Valdovinos-García, J. Barajas-Fernández, M. Olán-Acosta, M. A. Petriz-Prieto, A. Guzmán-López, and M. G. Bravo-Sánchez, "Techno-economic study of co2 capture of a thermoelectric plant using microalgae (chlorella vulgaris) for production of feedstock for bioenergy," *Energies*, vol. 13, no. 2, 2020.
[36] M. Bustamante, C. Robledo-Abad, R. Harper, C. Mbow, N. H. Ravindranat, F. Sperling, H. Haberl, A. de Siqueira Pinto, and P. Smith, "Co-benefits, trade-offs, barriers and policies for greenhouse gas mitigation in the agriculture, forestry and other land use (afolu) sector," *Global Change Biology*, vol. 20, no. 10, pp. 3270–3290, 2014.

**Lucas Ferreira-Paiva** graduated in 2020 with a Bachelor's degree in Electrical Engineering from the Universidade Federal de Viçosa (UFV) in Viçosa, Brazil. He is now pursuing a MSc in Computer Science at UFV and a Specialization in Data Science for Aeronautical Industry at Universidade Federal de Pernambuco in Recife, Brazil. His research interests are Computational Intelligence and Signal Processing.



**Attawan G. L. Suela** received a Bachelor's degree in Agribusiness from the Universidade Federal de Viçosa in 2017, and a Master's degree in Applied Economics from UFV in 2019. He is currently a PhD candidate in Applied Economics at UFV. His research interests include research in agribusiness, environment, economics and modeling methods for data analysis.

**Elizabeth R. Alfaro-Espinoza** graduated from the Universidad Nacional de Trujillo in Trujillo, Peru, with a degree in Microbiology and Parasitology. She is a PhD student in Bioinformatics at the Federal University of Minas Gerais in Belo Horizonte, Brazil. Her research interests include structural bioinformatics and software development for managing biological data.

**Nancy A. Cardona-Casas** earned a Bachelor's degree in Agronomic Engineering from the Universidad Nacional de Colombia in Medellín, Colombia, and a Master's degree in Soils and Plant Nutrition from the Universidade Federal de Viçosa in Brazil, where she is currently completing her PhD. She has an experience in agronomy, with a focus on entomology and agroecology, and has worked mostly in family farming, soil management, conservation, and fruit producing.

**Domingos S. M. Valente** received a Dsc. in Agricultural Engineering at Universidade Federal de Viçosa, Brazil, in 2010, and a Sabbatical (post-doc) at the University of Illinois (Urbana-Champaign). Currently, he is an Associate Professor at UFV, in the Department of Agricultural Engineering, working in the areas of Agricultural Mechanization, Digital Farming, Machine Learning, Remote Sensing, and GIS.

**Rodolpho V. A. Neves** received a Bachelor's degree in Electrical Engineering from the Universidade Federal de Viçosa, Brazil, in 2011, and the M.Sc. and D.Sc. in Electrical Engineering from the University of São Paulo, São Carlos, Brazil, in 2013 and 2018, respectively. From 2015 to 2016, he was a Visiting Researcher at Aalborg University, Denmark. He is currently an Adjunct Professor in the Department of Electrical Engineering at UFV. His research interests include intelligent control of dynamic systems and management of micro power grids.

## ACRONYMS

| | |
|---|---|
| GTP | Global Temperature Change Potential |
| GWP | Global Warming Potential |
| GHG | greenhouse gases |
| LISA | Local Indicators of Spatial Association |
| LUC | Land Use Change |
| PES | Payment for Environmental Services |
| IPPU | Industrial Processes and Product Use |
| SEEG | The Greenhouse Gas Emission and Removal Estimating System |