Global Journal of HUMAN-SOCIAL SCIENCE: E
Economics
Volume 25 Issue 1 Version 1.0 Year 2025
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals
Online ISSN: 2249-460X & Print ISSN: 0975-587X

Hybrid Model of Artificial Neural Networks and Principal Component Decomposition for Predicting Greenhouse Gas Emissions in the Brazilian Region of MATOPIBA

By Milena Monteiro Feitosa & E José de Jesus Sousa Lemos
Federal University of Ceará

Abstract- Greenhouse gas (GHG) emissions in agricultural production represent a global environmental challenge, and it is necessary to understand the factors that influence them to develop sustainable practices. The general objective of this research is to investigate some of the factors that probably influence GHG emissions and reductions in agricultural production in the MATOPIBA region of Brazil between 2006 and 2017. A hybrid methodology was used, and the first stage used linear models (decomposition into principal components) and non-linear models (artificial neural networks) to determine the relationships that should exist between the dependent variable (GHG emissions) and 11 variables. The data was obtained from the 2006 and 2017 Brazilian Agricultural Census, MapBiomas, SEEG, and NOAA. The results showed that of the 373 municipalities that make up MATOPIBA, only 100 did not see an increase in GHG emissions between 2006 and 2017. The principal component decomposition method reduced the 11 initial variables into 3 orthogonal and unobserved variables. In one of the unobserved variables, 4 of the five variables that are supposed to cause a reduction in GHG emissions were brought together. The 5 variables thought to have caused an increase in GHG emissions were condensed into 5.

Keywords: brazilian agriculture, EMBRAPA, change in land use, cerrado biome, evolution of GHG emissions.

GJHSS-E Classification: *JEL Code: Q54, Q15*

Strictly as per the compliance and regulations of:

Hybrid Model of Artificial Neural Networks and Principal Component Decomposition for Predicting Greenhouse Gas Emissions in the Brazilian Region of MATOPIBA

Milena Monteiro Feitosa [α] & E José de Jesus Sousa Lemos [σ]

Abstract- Greenhouse gas (GHG) emissions in agricultural production represent a global environmental challenge, and it is necessary to understand the factors that influence them to develop sustainable practices. The general objective of this research is to investigate some of the factors that probably influence GHG emissions and reductions in agricultural production in the MATOPIBA region of Brazil between 2006 and 2017. A hybrid methodology was used, and the first stage used linear models (decomposition into principal components) and non-linear models (artificial neural networks) to determine the relationships that should exist between the dependent variable (GHG emissions) and 11 variables. The data was obtained from the 2006 and 2017 Brazilian Agricultural Census, MapBiomas, SEEG, and NOAA. The results showed that of the 373 municipalities that make up MATOPIBA, only 100 did not see an increase in GHG emissions between 2006 and 2017. The principal component decomposition method reduced the 11 initial variables into 3 orthogonal and unobserved variables. In one of the unobserved variables, 4 of the five variables that are supposed to cause a reduction in GHG emissions were brought together. The 5 variables thought to have caused an increase in GHG emissions were condensed into 5.

Keywords: brazilian agriculture, EMBRAPA, change in land use, cerrado biome, evolution of GHG emissions.

I. Introduction

Brazilian agriculture is recognized worldwide for its prominence in the global market, especially in producing grains and foods such as soybeans, corn, cotton, orange juice, cocoa, coffee, sugar, and meat (FAOSTAT, 2020). In recent years, the expansion of soybean cultivation in Brazil has intensified, consolidating the country as one of the world's leading exporters. This growth is due to the advance of new agricultural frontiers such as the MATOPIBA region, which covers parts of the frontiers of the states of Maranhão, Tocantins, Piauí, and Bahia, located predominantly in the Cerrado biome (Santos & Naval, 2022).

Author α: PhD Student in Program of Development and Environment at Federal University of Ceará (UFC), Brazil.
e-mail: milenamonteirofeitosa@gmail.com
Author σ: Permanent Professor of the Graduate Programs in Agricultural Economics and Development and Environment at the Federal University of Ceará. Researcher and Productivity Fellow of the Brazilian National Council for Scientific and Technological Development (CNPq).
e-mail: lemos@ufc.br

Since the creation of the Brazilian Agricultural Research Corporation (EMBRAPA) in 1973, the knowledge generated by this institution has been fundamental in developing and adapting technologies for the tropical conditions of this type of production, especially in the Cerrado. This has boosted the development of the country's agricultural sector. These innovations have established Brazil as one of the largest global food producers and exporters (Nehring, 2016; Souza et al., 2020; EMBRAPA, 2024).

However, despite the economic growth driven by the agricultural sector, this is one of the activities responsible for greenhouse gas (GHG) emissions, due to the use of fossil fuelbased fertilizers, the burning of biomass, the high density of cattle per unit area, and the use of heavy agricultural machinery, which also uses this type of fuel in its energy matrix (Liu et al., 2017). According to the Food and Agriculture Organization of the United Nations (FAO), the agricultural sector may be responsible for up to 21% of the world's GHG emissions (FAO, 2016).

The increase in the greenhouse effect can be exacerbated by the rising levels of carbon dioxide (CO2) in the atmosphere (Myhre et al., 2013), which is attributed to various causes. Agricultural production can contribute to this process through the actions described in the previous paragraph. However, paradoxically, plants play a crucial role in reducing these emissions through the biochemical phenomenon known as photosynthesis. During photosynthesis, plants capture CO2, solar energy, water, and nutrients from the soil, transforming them into organic matter and releasing oxygen as a byproduct. Because of this phenomenon, CO2 is often known as the gas of life. It can therefore be inferred that deforestation may be one of the primary causes of the reduced capacity to capture CO2, which is ultimately released by various sources (Felício, 2014).

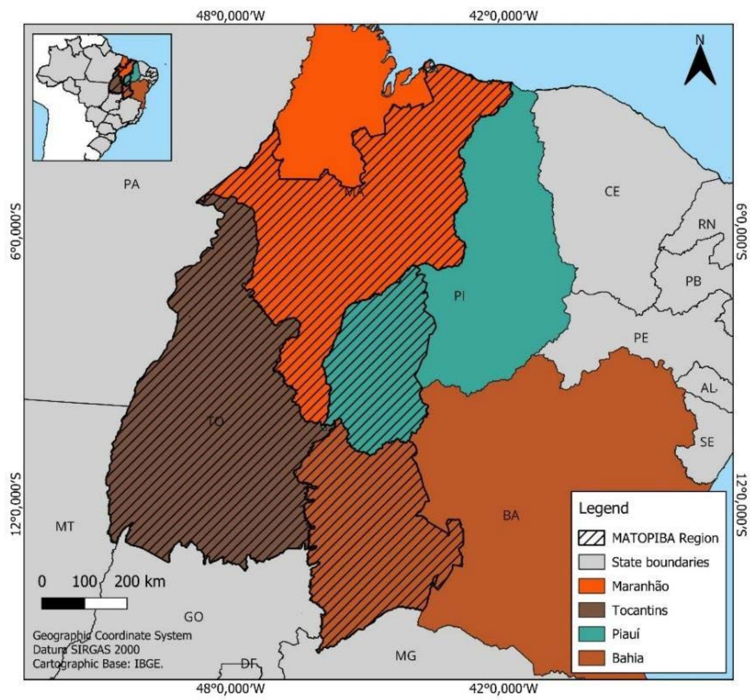
Thus, one of the primary sources of CO2 emissions is the burning of fossil fuels (Forster et al., 2007). Additionally, land use changes can alter the flow of carbon dioxide (CO2), methane (CH4), and nitrous oxide (N2O)- greenhouse gases that result from modifications to biogeochemical processes (Forster et al., 2007; Houghton et al., 2012; Kirschbaum et al., 2012; Kim & Kirschbaum, 2015).

Worldwide, it is estimated that approximately 420 million hectares of forest have been cleared since

1990. More than half (54%) of the world's forests are concentrated in just five countries: Russia, Brazil, Canada, the United States, and China. Meanwhile, agricultural land areas expanded by 6% between 2000 and 2021, contributing to a growth of 38% in permanent crops and 12% in temporary crops (FAO, 2020).

In 2021, Brazil had 66 million hectares of arable land, with a 20% increase in agricultural areas due to the expansion of temporary crops (FAO, 2023). It is, therefore, necessary to evaluate more rigorously how land is utilized as the main production factor in agriculture and its relationship with greenhouse gas emissions, given that it is directly related to changes in soil organic carbon stocks. This, in turn, is important for determining soil quality, natural fertility, agricultural productivity, and the fixation of atmospheric carbon dioxide (CO_2) (Kumar et al., 2022). In addition, the use of fire related to land use change can also reduce soil organic carbon stocks (Van der Werf et al., 2006; Van der Werf et al., 2010; Kim & Kirschbaum, 2015).

In recent decades, since its implementation in the 1980s, the MATOPIBA agricultural frontier, which includes parts of the states of Maranhão (135 of its 217 municipalities), Tocantins (all 139 municipalities), Piauí (33 of its 224 municipalities), and Bahia (30 of its 417 municipalities on the frontier) (Figure 1), has emerged as one of the main regions for the expansion of grain production, especially soybeans.



Source: Developed by the author.

Figure 1: Location Map of the MATOPIBA Region

According to data from IBGE's Municipal Agricultural Production (PAM), in 2023 the 337 municipalities of MATOPIBA recorded a 9.6% increase in cultivated area compared to the previous year. Consequently, production reached 18,943,144 tons, an increase of 11.2%, with an average yield of 3,581.94 kg/ha, representing a growth of 1.44% compared to 2022.

In addition, between 2013 and 2017, the GDP of agriculture in MATOPIBA reached R$17.1 billion, with an average annual growth rate of 7.7% (Souza, Magalhães & Castro, 2022). However, this economic growth has also had environmental impacts. Between 2010 and 2013, the MATOPIBA region was responsible for 45% of forest carbon emissions resulting from agricultural expansion in the Cerrado, with Maranhão making one of the largest contributions (14.42%) (Noojipady et al., 2017).

The conversion of native Cerrado areas into agricultural land intensifies deforestation (Rausch et al., 2019). In 2023, according to the Annual Report on Deforestation in Brazil, the MATOPIBA region was responsible for almost half (47%) of the loss of native vegetation in Brazil, with 858,952 hectares deforested, representing an increase of 59% over the previous year. Of the 50 municipalities with the most deforestation in the country, 33 are in the Cerrado, and the 10 with the largest deforested areas are in this biome. The state of Maranhão led the national ranking, with 331,225 hectares deforested, an increase of 95.1%. The state of Tocantins experienced a 177.9% increase in deforestation, with 230,253 hectares cleared, while

Bahia deforested 290,606 hectares, representing a 27.5% increase compared to 2022 (RAD, 2023).

This change in land use, resulting from the conversion of native areas into agricultural land and pastures, drives deforestation and promotes the frequent use of fire (Reddington et al., 2015; Spera et al., 2016). According to data from MapBiomas Fogo, between 1985 and 2023, around 199.1 million hectares were burned in Brazil, of which 44.6% (equivalent to 88.5 million hectares) were in the Cerrado biome, while the Amazon biome accounted for 19.6% (MapBiomas Project, 2023).

Although fire is a natural element used to clear areas where crops or pastures will be planted, it is becoming a threat due to its increasing frequency and the possibility of it getting out of control in the Cerrado due to agricultural expansion and deforestation. The Cerrado is responsible for 48% of the country's soybean production and approximately a quarter of this production area is located in MATOPIBA (Pitta et al., 2017; Soterroni et al., 2019; Silva et al., 2021; MapBiomas Project, 2023).

Given this scenario, studying the evolution of greenhouse gas (GHG) emissions and identifying the factors that likely contribute to the reduction or expansion of these emissions in the MATOPIBA region is essential for guiding public policies and investments that promote agriculture with lower or zero GHG emissions (Bezerra, 2022).

Based on the above, this research aims to answer the following questions: 1 - How many municipalities in MATOPIBA had an increase or decrease in GHG emissions between 2006 and 2017? 2 - Which variables, and in what proportions, probably influenced these emissions in this time interval?

To answer these questions, the general objective of this study is to investigate the factors that influence GHG emissions and reductions in agricultural production in the MATOPIBA region between 2006 and 2017. Specifically, the study seeks to: a - Ascertain the number of municipalities in the MATOPIBA region and, by state, identify which had an increase or decrease in GHG emissions between 2006 and 2017; b - Analyze the interaction between the variables tested in determining GHG emissions in this period; c - Evaluate how GHG emissions are influenced by the synergies between the indicators analyzed.

II. Material and Methods

a) Database and Construction of Indicators

The research uses secondary data extracted from the 2006 and 2017 Agricultural Censuses, MapBiomas, NOAA (National Oceanic and Atmospheric Administration), and SEEG (System of Estimates of Greenhouse Gas Emissions and Removals), from which information was obtained on the variables that are supposed to affect greenhouse gas emissions in the municipalities of the MATOPIBA region over these 11 years. The variables and data sources used are shown in Table 1.

Table 1: Variables that, by Hypothesis, Affect Greenhouse Gas (GHG) Emissions Positively (+) or Negatively (-) between 2006 and 2017 in MATOPIBA in this Research

Variables	Hypothesis of the relation between Yi and Xij	Definition	Sources
Y_i	GHG Emissions	(GHG_{2017}/GHG_{2006}) Emissions	Greenhouse Gas Emissions - SEEG (OC, 2022).
X_{i1}	(-)	(Municipal average annual rainfall2017) / (Municipal average annual rainfall2006)	NOAA (2022)
X_{i2}	(-)	(Vegetation cover) [(Crop areas + forest areas) / (total establishment area 2017)] / [(Crop areas + forest areas)/ (total establishment area 2006)]	Agricultural Censuses of 2006 and 2017 / IBGE
X_{i3}	(-)	[(Agricultural production value2017) / (Harvested agricultural area2017)] / [(Agricultural production value2006) / (Harvested agricultural area2006)]	Agricultural Censuses of 2006 and 2017 / IBGE
X_{i4}	(-)	[(Livestock production value2017) / (Pasture area2017)]/ [(Livestock production value2006) / (Pasture area2006)]	Agricultural Censuses of 2006 and 2017 / IBGE
X_{i5}	(-)	[(Recovered areas2017) / (Deforested areas2017)]/ [(Recovered areas2006) / (Deforested areas2006)]	MapBiomas

Global Journal of Human-Social Science (E) XXV Issue I Version I Year 2025

X_{i6}	(+)	[(Cattle quantity2017) / (pasture areas 2017)] / [(Cattle quantity2006) / (pasture areas 2006)]	Agricultural Censuses of 2006 and 2017 / IBGE
X_{i7}	(+)	[(Total tractors and machinery2017) / (total establishment area2017)] / [(Total tractors and machinery2006) / (total establishment area2006)]	Agricultural Censuses of 2006 and 2017 / IBGE
X_{i8}	(+)	[(Expenditure on agricultural pesticides2017) / (total area of municipal establishments 2017)] / [(Expenditure on agricultural pesticides2006) / (total area of municipal establishments 2006)]	Agricultural Censuses of 2006 and 2017 / IBGE
X_{i9}	(+)	[(Industrial sector GDP2017) / (Total municipal GDP2017)] / [(Industrial sector GDP2006) / (Total municipal GDP2006)]	Agricultural Censuses of 2006 and 2017 / IBGE
X_{i10}	(+)	CV rainfall 2017/ CV rainfall 2006	NOAA (2022)
X_{i11}	(+)	(Burn scars areas 2017 / Burn scars areas 2006)	MapBiomas

Source: Compiled based on data from SEEG, MapBiomas, Agricultural Census (2006 and 2017), IBGE, and NOAA (2022).

The methodological approaches adopted to achieve the objectives of this research begin with the development of the indicators used. To assess changes in greenhouse gas (GHG) emissions between 2006 and 2017, in addition to the impact of the variables presented in Table 1 on these emissions, the indicators are constructed as follows: the relationship between the values observed in 2017 (final year) and those in 2006 (initial year) is estimated for both GHG emissions and the explanatory variables.

This makes it possible to identify whether each variable increased or decreased over the period analyzed. In municipalities where the ratio between GEE2017 and GEE2006 is greater than 1, there has been an increase in emissions; if it is less than 1, there has been a reduction. The same process is applied to the explanatory or independent variables (Table 1).

b) Methodology Adopted to Assess the First Research Objective

To achieve the first objective of the research, we estimated the total number of municipalities where the ratios of GHG emissions in 2017 were higher than those observed in 2006. In these instances, the ratios are represented as Yi2017/Yi2006. Additionally, we measured the relationships between the variables believed to have influenced GHG emissions between 2006 and 2017, denoted as Xij2017/Xij2006.

c) Methodology for Achieving the Second Objective

To estimate the synergy between the variables that are thought to have influenced GHG emissions, Factor Analysis (FA) was used, using the principal component decomposition technique.

Before using the principal component decomposition model, it was decided to transform all the variables into indices. The indices range from 1 to 100. In the case of the dependent variable, the ratio of GHG emissions between 2006 and 2017, the following procedure was adopted. The municipalities were ranked in descending order by the ratio of GHG emissions between 2006 and 2017. Therefore, the higher the value of this ratio, the higher the GHG emissions between 2006 and 2017. For this reason, the highest emissions value was assigned the index=100. The other values were adjusted proportionally using a simple, straightforward rule of three. Thus, in the municipality where the GHG emissions index = 100, there was the highest emission of this gas between 2006 and 2017. In those municipalities where the GHG index is close to 1, this means that there was the greatest reduction in these emissions.

About the 11 independent variables used to cause GHG emissions, the following criteria were adopted. All 5 variables whose hypothesis in this study establishes that they should cause a reduction in GHG emissions between 2006 and 2017 (GHG2017/2006 $\leq$1) were ranked in ascending order. The lowest value (worst case) is assigned an index of 100. The remaining values are adjusted proportionally using a simple inverse rule of three. These variables are marked with a (-) sign in Table 1 indicating that, by hypothesis, they cause a reduction in GHG emissions.

The other 6 independent variables which, by hypothesis, should cause an increase in GHG emissions (GHG2017/2006) were ranked in descending order. The highest value of these variables (worst case) was assigned an index of 100. The other values are adjusted proportionally using a simple, direct rule of three. These variables are marked with a (+) sign in Table 1 indicating that, by hypothesis, they cause an increase in GHG emissions.

i. Summary of the Factor Analysis Model as it Applies to the Study

A summary of the factor analysis method applied in this study is presented below. In general, the factor analysis model can be expressed as follows:

$$X = af + e \tag{1}$$

In equation (1), $X = (X_1, X_2, ..., X_p)^T$ is a transposed vector of observable variables, while $f = (f_1, f_2, ... , f_r)^T$ represents a transposed vector consisting of r latent (r < p) factors that are not directly observable. The matrix a of coefficients has dimension (p x r) of fixed coefficients, known as factor loadings; and $e = (e_1, e_2, ..., e_r)^T$ is a transposed vector of random terms. It is generally assumed that E(e) = E(f) = 0.

At the outset, the estimated factor loadings may not be definitive; however, the factor analysis method enables the rotation of this initial structure for enhanced interpretation. In this study, varimax orthogonal rotation was used, which has the advantage of making the factors independent (Dillon & Goldstein, 1984; Johnson & Wichern, 1988; Basilevsky, 1994; Fávero et al., 2017).

To construct the index, the factor scores are estimated after the orthogonal rotation of the initial structure. The factor score positions each observation in the space of common factors. Thus, for each factor fi, the i-th factor score that can be extracted is defined by Fi, and can be expressed as:

$$F_i = b_1X_{i1} + b_2X_{i2} + .. + b_pX_{ip};\ i = 1, 2,.., n; j = 1,2 .., p \tag{2}$$

Where $b_1, b_2, ..., b_p$ are regression coefficients; $X_{i1}, X_{i2}, ..., X_{ip}$ are "p" observable variables.

Although Fi is not directly observable, it can be estimated using existing factor analysis techniques, using the matrix of observable variables X. Thus, equation (2) can be rewritten in a more compact form using matrix notation:

$$F_{(n\ x\ q)} = X_{(n\ x\ p)}.B_{(p\ x\ q)} \tag{3}$$

In equations (2) and (3), the factor scores are influenced by the magnitude and units of the X variables. To avoid this problem, the X variables are normalized, resulting in:

$$Z_{ij} = [(X_i - m_{xi}) / s_{xi}]; \tag{4}$$

Where m_{xi} is the mean of X_i, and S_{xi} is its standard deviation. Thus, equation (4) can be modified to:

$$F_{(n\ x\ q)} = Z_{(n\ x\ p)}.b_{(p\ x\ q)} \tag{5}$$

In equation (5), the vector "b" replaces "B", since the variables are already normalized. Pre-multiplying both sides of the equation by $(1/n)Z^T$, where n is the number of observations and ZT is the transposed matrix of Z, gives us:

$$(1/n)Z^TF = (1/n)Z^TZb. \tag{6}$$

The expression $(1/n)\ Z^TZ$ corresponds to the correlation matrix of the X variables, called R. The matrix $(1/n)\ Z^TF$ represents the correlation between the factor scores and the factors themselves, called L. The equation is redefined as:

$$L = R.b \tag{7}$$

If R is a non-singular matrix, verified by the Bartlett test, the analysis can proceed. Thus, the hypothesis that the matrix of correlations between the variables is not an identity matrix must be rejected, with at least a 5% error level (Fávero, 2017).

If R is non-singular, multiply both sides of equation (7) by the inverse matrix of R (R^1), resulting in:

$$b = R^- .L. \tag{8}$$

For the estimated model to be statistically valid, it is essential to conduct the Kaiser-MeyerOlkin (KMO) test, which should yield a value greater than 0.5. Additionally, the total variance explained by the orthogonal factors must exceed 50% (Hair et al., 2005; Maroco, 2003; Fávero, 2017).

After determining the "b" vector (as shown in equation 8), the compositions of each estimated factor are identified based on the magnitudes of the factor loadings. The factors, which are a reduction of the original variables (k < n), can be redefined and renamed according to the magnitudes of the factor loadings that each variable presents in each component factor.

The principal component decomposition process permits the generation of factor scores. These factor scores, represented by FE, are normalized variables with a mean of zero and a standard deviation of one. Positive and negative values gravitate around this zero mean FE. These factor scores can be converted into partial indices associated with each municipality (Ii) using equation (9). These partial indices can be supplemented, depending on the variables grouped in the composition of each factor score that generated it.

$$I_i = (FE_i - FE_{MN})/(FE_{MX} - FE_{MN}) \tag{9}$$

In equation (9), FEi is the i-th normalized factor score, FEMN is the minimum value of the factor score, and FEMX is the maximum value. In this manner, the Ii indices will range between zero and one, and they will be utilized in this study to identify the expected results for the second objective.

The relationship between GHG emissions and the partial indices (I1; I2, ..., Ip) can be described by the following equation:

$$GHG_i = f(I_1; I_2, ..., I_p) \tag{10}$$

This equation summarizes how the indices derived from the factor scores explain the variation in GHG emissions in the different municipalities.

ii. *Methodology Adopted to Achieve the Third Objective*

To achieve the third objective of this research, the Artificial Neural Networks (ANN) model was used to investigate how GHG emissions are influenced by the synergy of partial indices (I_1, I_2, ..., I_p). ANNs are part of computational artificial intelligence. One of the main areas of application of ANNs is in the prediction of multivariate statistical data that is both nonlinear and non-parametric (Sharda & Patil, 1992; Lee et al., 2017).

Zhang et al (1998) report that one of the procedures of computational artificial intelligence normally used to predict time series is the training of ANNs, based on the architecture and learning of the human brain. In this way, according to Zhang et al (1998), ANNs work like the human brain, seeking to recognize regularities and patterns in data, being able to learn from experience, and make generalizations based on previously accumulated knowledge. ANNs are non-linear models, unlike traditional forecasting models such as Box & Jenkins (1976) and Pankratz (1983), which assume that the series studied are generated by linear processes.

When designing an ANN model, we can envision it as a network of artificial 'neurons' organized into layers. The variables used to predict (inputs) a dependent variable (output) form the lower layer, while the predicted variables form the upper layer. The ANN model also allows for the possibility of intermediate layers, generally known as hidden layers (Sharda & Patil, 1992).

Designed to represent how the human brain processes information, ANNs are computer algorithms that add knowledge by detecting patterns and correlations and can be trained through experience. They consist of hundreds of artificial neurons (or nodes) interconnected in hierarchical layers. Each neuron has a specific output function and the connection between each two nodes has a weight, constituting its artificial neural network memory. It is through these weights that the power of neural computations is reflected, i.e. the degree of influence that one cell exerts on another.

Built to simulate the biological function of a neuron, each node has weighted inputs, a transfer function, and an output. Feedforward neural networks linearly transmit information, from the input layer to the output layer, and are among the most popular types used in various applications (Figure 1) (Agatonovic-Kustrin; Beresford, 2000; Gómez, Fernández & Peñuela, 2021).

In this research, the process begins with data entry, in which the explanatory variables correspond to the partial indices generated by factor analysis, and the dependent variable is represented by GHG emissions (Yi). The data was randomly divided into two sets: 70% was used to train the model and 30% was reserved for the test set (Liu &Cocea, 2017; Dao et al., 2020). The output of a neuron can be written mathematically:

$$\mathrm{Yi} = \mathrm{f(n)} \tag{12}$$

Where n is the weighted sum of the input signals plus an adjustment term (bias), defined as: n =

$$\mathrm{n} = \sum_{i=1}^{p}(w_j \,.\, X_j) + b \tag{13}$$

Where Xi_1, Xi_2, ..., Xi_p are the neuron's input signals (partial indices generated by factor analysis); w_1, w_2, ..., w_p are the weights associated with each input, determining the importance of each signal in the process; "b" is the bias term, used to adjust the flexibility of the model; and f (*) is the activation function, responsible for the non-linearity of the model, enabling the network to learn complex relationships between the data.

The model's performance was assessed using quantitative metrics, including the Root Mean Square Error (RMSE), the Mean Absolute Error (MAE), and the Mean Absolute Percentage Error (MAPE). The lower the estimated values for these measurements, the better the adjustments. The RMSE is calculated by the root mean square difference between the predicted and observed values, shown in Equation 14. It provides an overview of the model's accuracy. The lower the RMSE, the more accurate the model. The Mean Absolute Error (MAE), as measured by equation (15), is also a metric used for evaluating models. Finally, the Mean Absolute Percentage Error (MAPE), as measured by equation (16), expresses errors as a percentage, making it easier to interpret the observed values (Pham et al., 2018; Elsaraiti, 2024). Using several metrics is advantageous for obtaining a broader view of the model's performance from different perspectives (Tripathy & Prusty, 2021).

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\, v_{\mathrm{observed}} - v_{predicted}\,)^2} \tag{14}$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n} |\, v_{observed} - v_{predicted}\, | \tag{15}$$

$$\mathrm{MAPE} = \frac{1}{n}\sum_{i=1}^{n} \left|\, \frac{v_{observed} - v_{predicted}}{v_{predicted}} \,\right| * 100 \tag{16}$$

III. Results and Discussion

To enhance the clarity of the presentation and discussion of the results, they have been organized according to the timeline of the research objectives.

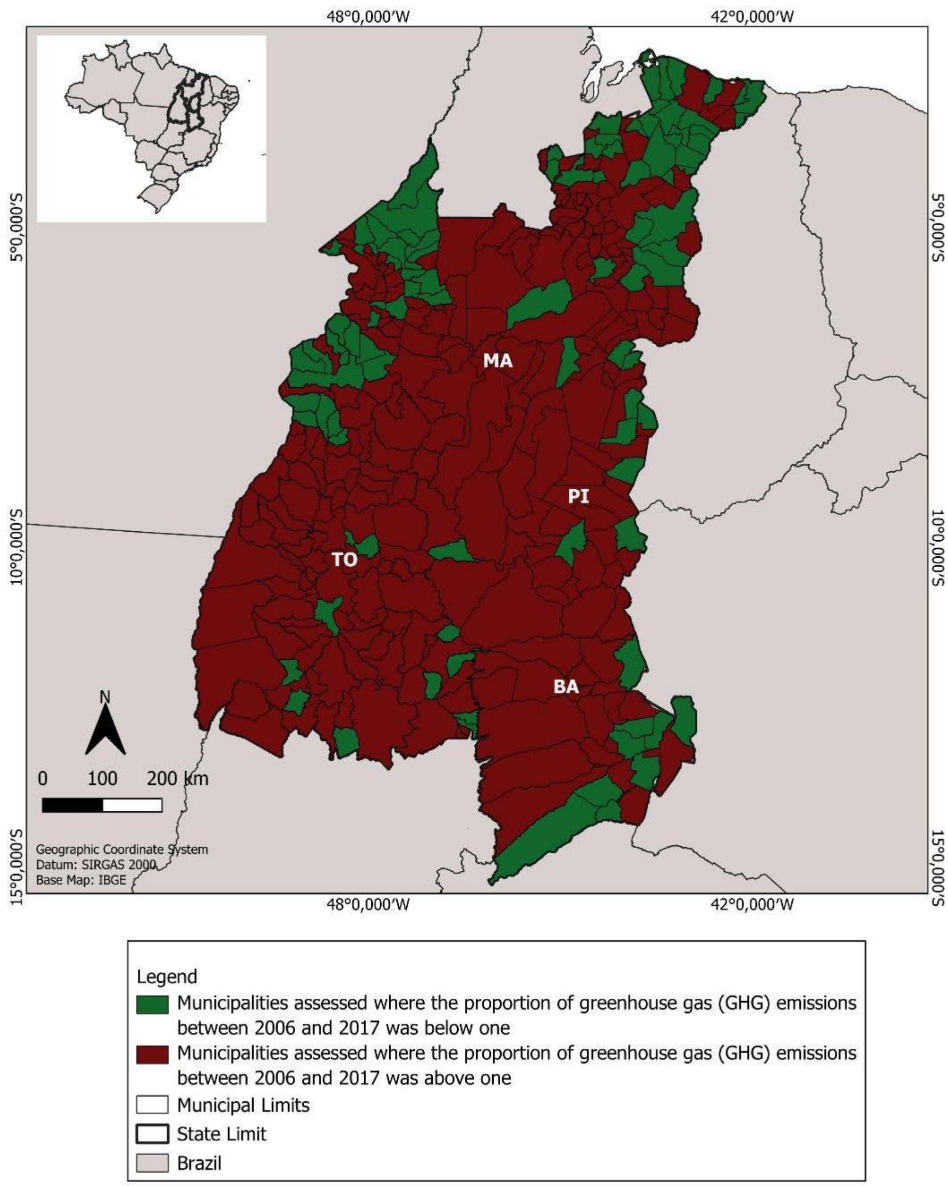
a) *Results Found for the First Objective*

Table 3 and Figure 2 show the absolute and relative frequencies of the MATOPIBA municipalities for states that showed an increase or decrease in GHG emissions between 2006 and 2017.

Table 3: Absolute and Relative Frequencies of GHG Missions in MATOPIBA Municipalities from 2006 to 2017

States	Absolute frequencies of municipalities where the ratio of GHG emissions between 2017 and 2006 was less than 1 (GHG < 1)	Relative Frequency (%)	Absolute frequencies of municipalities where the ratio of GHG emissions between 2017 and 2006 was greater than 1 (GHG >1)	Relative Frequency (%)
Maranhão	50	37	85	63
Tocantins	33	23,7	106	76,3
Piauí	8	24,2	25	75,8
Bahia	9	30	21	70
Total	100	100	237	100

Source: Based on the survey results.



Source: Based on the survey result.

Figure 2: 2017/2006 GHG Ratio for the MATOPIBA Region

There was a predominance of municipalities with an increase in GHG emissions during the analyzed period. Of the 337 municipalities in the MATOPIBA region, 237 (approximately 70.3%) experienced an increase in emissions, while 100 (29.7%) reported a reduction. Among the municipalities in the state of Maranhão, 50 out of a total of 135 (37%) experienced a reduction in emissions, while 85 (63%) reported an increase. In Tocantins, of the 139 municipalities, 33 (23.7%) registered a decrease in emissions, while 106 (76.3%) showed an increase. In Piauí, out of 33 municipalities, only 8 (24.2%) reduced emissions, while 25 (75.8%) experienced an increase. In the state of Bahia, out of a total of 30 municipalities in MATOPIBA, 9 (30%) reduced emissions, while 21 (70%) showed an increase (Table 3).

b) Results Obtained for the Second Objective

Table 4 presents the results obtained from the principal component decomposition procedure of the factor analysis (FA) conducted in this research.

Table 4: Results Found Showing the Decomposition of the 11 Original Variables into 3 Main Components

		Rotated components Matrix			Component Score Coefficient Matrix		
Variables		1	2	3	1	2	3
Index of rainfall	X_{i1}	0.000	0.009	**1.000**	-0.001	-0.003	**0.500**
Index of vegetation cover	X_{i2}	-0.296	**0.856**	-0.006	0.036	**0.281**	-0.008
Index of agricultural production value	X_{i3}	-0.225	**0.896**	0.013	0.061	**0.306**	0.001
Index of relative livestock value	X_{i4}	-0.035	**0.788**	0.008	0.098	**0.291**	-0.001
Index of rlative recovered areas	X_{i5}	-0.264	**0.941**	0.009	0.056	**0.317**	-0.001
Index or relative cattle quantity	X_{i6}	**0.927**	-0.260	-0.009	**0.219**	0.029	-0.004
Index of_relative_machinefy	X_{i7}	**0.947**	-0.170	-0.024	**0.237**	0.066	-0.013
Index of relative expenditure in pesticides	X_{i8}	**0.865**	-0.258	0.013	**0.202**	0.021	0.007
Index of_relative_industrial GNP	X_{i9}	**0.927**	-0.083	-0.001	**0.243**	0.096	-0.002
Index of relative_cv rainfall	X_{i10}	0.000	0.009	**1.000**	-0.001	-0.003	**0.500**
Index of relative Burn scars areas	X_{i11}	**0.937**	-0.199	0.020	0.230	0.053	0.010
Kaiser-Meyer-Olkin Measure of Sampling Adequacy:							0.738
Bartlett's Test of Sphericity							
Approx. Chi-Square							8186,727
Degrees of Freedom							55
Significance level.							0.000
Total Variance Explained (%)							88.232
Variance Explained by Component 1 (%)							40.464
Variance Explained by Component 2 (%)							29.575
Variance Explained by Component 3 (%)							18.193

Sources: Results found in the search

Observations: Extraction Method: Principal Component Analysis; Rotation Method: Varimax with Kaiser Normalization. Component Scores. Rotation converged in 4 iterations.

The evidence presented in Table 4 indicates that the adjustment obtained using the principal component decomposition (PCD) method was statistically significant. The Bartlett test, which yielded a high level of significance ($p = 0.00$), demonstrated that the correlation matrix of the independent variables is not an identity matrix. The estimated statistic for the Kaiser-Meyer-Olkin Measure of Sampling Adequacy (KMO) test was 0.738, and the total variance explained by the adjusted model was approximately 88.232¨%. The variances explained by each estimated component after Varimax orthogonal rotation were 40.464%, 19.373%, and 18.137%, respectively, for components 1, 2, and 3. These results indicate that the greatest synergy captured by the DCP procedure was in component 1, which measures the ratios of: the number of cattle per hectare; the number of tractors and machinery; spending on pesticides; and industrial GDP about the total GDP of the municipalities; and the evolution of burn scars in the period studied. All these variables captured in this component, as assumed in this research, must have positively affected greenhouse gas emissions between 2006 and 2017. This synergy, as we have seen, is responsible for explaining 40.464% of the total explanatory capacity of the model generated (Table 4).

From the results shown in Table 4, it can also be seen that associated with the second component generated in the research, whose variance explains approximately 29.575% of the total explained variance, are four of the five variables that are supposed to cause a reduction in greenhouse gas emissions: the vegetation cover index; the index that measures the productive potential of crops; the index that measures the productive capacity of animal husbandry and the index that measures the recovery of degraded areas.

For the third component, the greatest synergies were between the variables rainfall index, which is supposed to reduce GHG emissions, and the index measuring rainfall instability, as measured by the coefficients of variation, which is supposed to increase GHG emissions. These two variables account for 18.193% of the total explained variance. Based on these results, the matrix was generated, which is made up of 3 factor scores that capture these synergies (Table 4).

c) Results Obtained for the Third Objective

As shown above, the third objective of this research sought to assess how GHG emissions are influenced by the synergies between the indicators analyzed. To do this, the GHG emissions ratios between 2006 and 2017 were transformed into indices. In the previous step, when using FA, it was assumed that the relationships between the variables were linear. Thus, the three estimated factor scores are linearly independent. In this step, it is assumed that the relationship between the GHG emissions index between 2006 and 2017 and the independent variables transformed into factor scores is non-linear. The artificial neural network (ANN) model is used to perform the test. The results are shown in Table 5 and Figure 3.

Table 5: Estimation results using ANN, with the GHG emissions index as the dependent variable and the FS1, FS2 and FS3 indices as independent variables

$GHG_{2017/2006}$ emissions ≤ 1		$GHG_{2017/2006}$ emissions > 1	
Relative error in the Training period (%)	1.011	Relative error in the Training period (%)	1.002
Relative error in the Testing period (%)	0.908	Relative error in the Testing period (%)	0.997
Number of Units in Hidden Layer including 1 bias Hiden Layer	3	Number of Units in Hidden Layer including 1 bias Hiden Layer	4
Independent Variable Importance		Independent Variable Importance	
Variabels	Importance (weights)	Variabels	Importance (weights)
FS1	0.276	FS1	0.486
FS2	0.496	FS2	0.391
FS3	0.228	FS3	0.123
Acuracy tests		Acuracy tests	
RMSE	3.12	RMSE	10.49
MAE	2.52	MAE	6.58
MAPE	13.11	MAPE	21.29

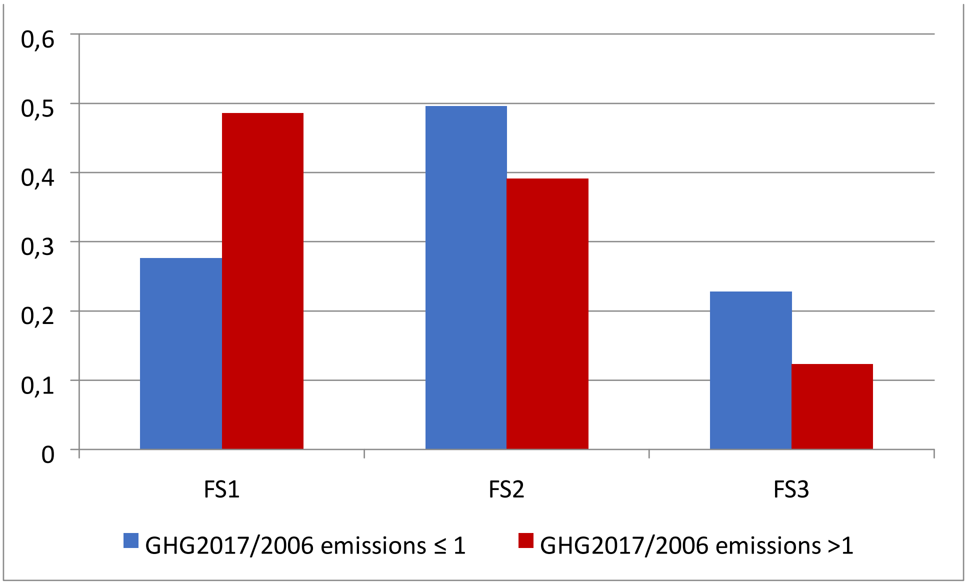
Sources: Results found in the search

The results presented in Table 5 and Figure 3 show that, in the 100 municipalities where there was a reduction in GHG emissions between 2006 and 2017, the score of the factor that brings together the four variables (FS2) that hypothetically impact this reduction had the highest weight (0.496), followed by the score of FS2, bringing together the variables that, by hypothesis, contributed to the increase in GHG emissions between 2006 and 2017, with a weight of 0.276. Meanwhile, FS3, which brings together the variables associated with rainfall and instability, weighted 0.228.

Among the six variables transformed into indices, which were hypothesized to contribute to the increase in GHG emissions between 2006 and 2017, five were grouped into score factor 1 (SF1), with an estimated weight of 0.486. Factor score 2 (FS2) had a weight of 0.391 in this definition, while factor score 3 (FS3) had a weight of 0.123 (see Table 5 and Figure 3).

These results prove the hypotheses of this research, indicating that the variables that were supposed to cause a reduction and increase in GHG emissions between 2006 and 2017 in Matopiba. It can also be seen that the prediction errors in both the training and testing phases were quite low. In addition, the RMSE, MAE, and MAPE tests also showed very low values, thus confirming the accuracy of the adjustments.

Global Journal of Human-Social Science (E) XXV Issue I Version I Year 2025



Sources: Results found in the search

Figure 3: Relative importance of each factor score in Matopiba municipalities with GHG emissions ≤ and GHG emissions > 1

As shown in the evidence presented in Table 5, to estimate the weights associated with each of the independent (explanatory) variables in the group of 100 MATOPIBA municipalities where the GHG ratio ≤ 1, the ANN model used 3 Units in Hidden Layer, including 1 Hiden Layer bias. To estimate the importance of the explanatory variables in municipalities where the GHG ratio > 1, the model used 4 Units in the Hidden Layer, including 1 Hiden Layer bias. In municipalities where GHG emissions are less than or equal to 1, the highest weight was estimated for variable FS2 (0.480), followed by FS3 (0.286), and the lowest weight was associated with the independent variable FS1 (0.234). These results highlight the significance of all the original variables synthesized in FS2, as they are expected to contribute to reducing greenhouse gas emissions. Therefore, the research hypothesis is confirmed.

About the estimated weights for the variables constructed linearly for the municipalities in which the GHG ratios > 1, it can be seen that, as expected, the greatest weighting was given to the FS1 combination (0.629), which synergistically brings together practically all the original variables that are assumed in this research to have contributed to the increase in greenhouse gas emissions between 2006 and 2017. It can be seen that the rainfall ratios, as well as the rainfall instability ratios measured by the coefficients of variation, both contributed to a reduction and an increase in GHG emissions.

IV. Conclusions

From the results of the evidence extracted from the research, it was shown that of the 327 municipalities that are part of the MATOPIBA agricultural frontier region, 100 experienced a reduction in GHG emissions between 2006 and 2017, the period in which this research was carried out. In the remaining 237 municipalities, GHG emissions increased over the same period.

We tested 5 variables that were assumed to have led to a reduction in GHG emissions in the period under investigation and 6 variables that were assumed to have led to an increase in emissions. The methodological procedures used in this research are unprecedented in this type of study, in that it uses two models in sequence. This study employed a model that assesses linear relationships through factor analysis using the principal component decomposition technique. In this process, the 11 observed variables were reduced to three unobserved variables, referred to as factor scores. These factor scores are orthogonal and linearly independent.

In the second methodological stage, the relationships were estimated between the dependent variable that measures GHG emissions between 2006 and 2017 and the three-factor scores in which the 11 original independent variables were synthesized by synergy. At this stage, it was assumed that the relationships were non-linear. Therefore, artificial neural networks (ANN) were employed to evaluate the weights that each of the three factor scores contributes to explaining the dependent variable.

It was observed that the assumptions made for this research were confirmed. Of the five variables hypothesized to contribute to a reduction in GHG emissions, four were grouped together in one of the factor scores and had the highest weighting in explaining greenhouse gas emissions in the MATOPIBA

Global Journal of Human-Social Science (E) XXV Issue I Version I Year 2025

municipalities that experienced a decline in GHG emissions between 2006 and 2017.

On the other hand, of the 6 variables that were tested and assumed to have led to an increase in greenhouse gas emissions between the two periods, 5 were brought together in synergy in another factor score (unobserved variable generated by the linear procedure) and had the highest weighting in the municipalities that had an increase in GHG emissions between 2006 and 2017.

Thus, the results of this research can indicate the paths that should be followed in agricultural practices on this agricultural frontier. Furthermore, they can guide future studies by identifying which variables may contribute to the emission or reduction of GHG emissions.

The overall conclusion of this research is that the two questions motivating its execution were answered, and the proposed objectives were achieved. The municipalities in the MATOPIBA region that increased and decreased GHG emissions between 2006 and 2017 were identified, along with the variables and their respective weightings that influenced these changes in emissions.

References Références Referencias

1. Agatonovic-Kustrin, S., & Beresford, R. (2000). Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research. Journal of Pharmaceutical and Biomedical Analysis, 22 (5), 717–727.
2. Bezerra, F. N. R. (2022). Evaluation of low carbon emission and climate-smart agriculture in Brazil (Doctoral thesis, Federal University of Ceará, Center for Agricultural Sciences, Graduate Program in Rural Economics). Fortaleza.
3. Box, G. E. P., & Jenkins, G. M. (1976). Time series analysis: Forecasting and control (Revised ed.). San Francisco: Holden-Day.
4. Dao, D. V., et al. (2020). A sensitivity and robustness analysis of GPR and ANN for high-performance concrete compressive strength prediction using a Monte Carlo simulation. Sustainability, 12 (3).
5. Elsaraiti, M. (2024). Short-term power consumption forecasting using neural networks with first- and second-order differencing. Academia Green Energy, 1 (3).
6. Empresa Brasileira de Pesquisa Agropecuária (Embrapa). (n.d.). Sobre a Embrapa. Retrieved October 17, 2024, from https://www.embrapa.br/en/sobre-a-embrapa
7. Felício, R. A. (2014). Mudançasclimáticas e aquecimento global – Nova formatação e paradigma para o pensamentocontemporâneo? Ciência e Natura, 36 (3), 257–266.
8. Food and Agriculture Organization of the United Nations (FAO). (2010). "Climatesmart" agriculture: Policies, practices and financing for food security, adaptation and mitigation. Rome, Italy.
9. Food and Agriculture Organization of the United Nations (FAO). (2013). Climate smart agriculture sourcebook. Rome, Italy: Food and Agriculture Organization of the United Nations.
10. Food and Agriculture Organization of the United Nations (FAO). (2016). The state of food and agriculture. Food and Agriculture Organization of the United Nations. Retrieved from http://www.fao.org/3/a-i6030e.pdf
11. Food and Agriculture Organization of the United Nations (FAO). (2020). Global forest resources assessment 2020 – Key findings. Rome. Retrieved from https://doi.org/10.4060/ca8753en
12. Food and Agriculture Organization of the United Nations (FAO). (2023). Land use statistics and indicators 2000–2021: Global, regional and country trends (FAOSTAT
13. Analytical Briefs Series No. 71). Rome. Retrieved from https://doi.org/10.4060/cc6907en
14. Food and Agriculture Organization of the United Nations (FAO) Database. (n.d.). Retrieved September 5, 2024, from http://www.fao.org/faostat/en/#data
15. Forster, P., et al. (2007). Changes in atmospheric constituents and in radiative forcing. In S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K. B. Averyt, M. Tignor, &
16. H. L. Miller (Eds.), Climate change 2007: The physical science basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press.
17. Miranda, I. N. G., Jaramillo, J. M. F., & Peñuela, G. A. (2021). Hybrid multivariate statistical and neural network model to predict greenhouse gas emissions. Arabian Journal for Science and Engineering, 46 (10).
18. Houghton, R. A., et al. (2012). Emissões de carbono do uso da terra e mudançanacobertura da terra. Biogeosciences, 9, 5125–5142.
19. Kim, D.-G., & Kirschbaum, M. U. F. (2015). The effect of land-use change on the net exchange rates of greenhouse gases: A compilation of estimates. Agriculture, Ecosystems & Environment, 208, 114–126.
20. Kirschbaum, M. U. F., et al. (2012). Comprehensive evaluation of the climate-change implications of shifting land use between forest and grassland: New Zealand as a case study. Agriculture, Ecosystems & Environment, 150, 123–138.
21. Kumar, S., et al. (2022). Remote sensing for agriculture and resource management. In M. K. Jhariya, R. S. Meena, A. Banerjee, & S. N. Meena (Eds.), Natural resources conservation and advances for sustainability (pp. 91–135). Elsevier.

22. Lee, J. G., et al. (2017). Deep learning in medical imaging: General overview. Korean Journal of Radiology, 18 (4), 570–584.
23. Liu, H., & Cocea, M. (2017). Semi-random partitioning of data into training and test sets in granular computing context. Granular Computing, 2, 357–386.
24. Liu, X., Zhang, S., & Bae, J. (2017). The nexus of renewable energy-agriculture environment in BRICS. Applied Energy, 204, 489–496.
25. Myhre, G., et al. (2014). Anthropogenic and natural radiative forcing. In Climate change 2013 - The physical science basis (pp. 659–740). Cambridge University Press.
26. Nehring, R. (2016). Yield of dreams: Marching west and the politics of scientific knowledge in the Brazilian Agricultural Research Corporation (Embrapa). Geoforum, 77, 206–217.
27. Noojipady, P., et al. (2017). Forest carbon emissions from cropland expansion in the Brazilian Cerrado biome. Environmental Research Letters, 12 (2).
28. Pankratz, A. (1983). Forecasting with univariate Box-Jenkins models. New York: Wiley & Sons, Inc.
29. Pham, B. T., et al. (2018). Prediction of shear strength of soft soil using machine learning methods. Catena, 166, 181–191.
30. ProjetoMapBiomas. (2024). Mapeamento das áreasqueimadas no Brasil entre 1985 a 2023 - Coleção 3. Retrieved October 28, 2024, from https://plataforma.brasil.mapbiomas.org/hotsite-fogo
31. Pitta, F. T., & Vega, G. C. (2017). Impacts of agribusiness expansion in the MATOPIBA region: Communities and the environment. Rio de Janeiro: Action Aid Brasil.
32. MapBiomas. (2024). RAD 2023: RelatórioAnual do Desmatamento no Brasil 2023. São Paulo: MapBiomas. Retrieved October 28, 2024, from http://alerta.mapbiomas.org
33. Rausch, L., et al. (2019). Soy expansion in Brazil's Cerrado. Conservation Letters, 12.
34. Reddington, C. L., et al. (2015). Air quality and human health improvements from reductions in deforestation-related fire in Brazil. Nature Geoscience, 8 (10), 768–771.
35. Santos, J. F. S., & Naval, L. P. (2022). Soy water footprint and socioeconomic development: An analysis in the new agricultural expansion areas of the Brazilian cerrado (Brazilian savanna). Environmental Development, 42, 100670.
36. Sharda, R., & Patil, R. B. (1992). Connectionist approach to time series prediction: An empirical test. Journal of Intelligent Manufacturing, 3, 317–323.
37. Silva, P. S., et al. (2021). Putting fire on the map of Brazilian savanna ecoregions. Journal of Environmental Management, 296.
38. Sims, B., & Kienzle, J. (2017). Sustainable agricultural mechanization for smallholders: What is it and how can we implement it? Agriculture, 7 (6).
39. Soterroni, A. C., et al. (2019). Expanding the soy moratorium to Brazil's Cerrado. Science Advances, 5.
40. Souza, G. S., Gomes, E., Alves, E. R. A., & Gasques, J. G. (2020). Technological progress in Brazilian agriculture. Socio-Economic Planning Sciences, 72, 100879.
41. Souza, D. T. de, Magalhães, L. A., & Castro, G. S. A. (2022). Uma avaliação do impacto do crédito rural e do mercado de trabalho à agropecuária do Matopiba. In Anais do 50º Encontro Nacional de Economia (pp. 1–21). Fortaleza: ANPEC.
42. Spera, S. A., Galford, G. L., Coe, M. T., Macedo, M. N., & Mustard, J. F. (2016). Land-use change affects water recycling in Brazil's last agricultural frontier. Global Change Biology, 22(10), 3405–3413.
43. Tripathy, D. S., & Prusty, B. R. (2021). Forecasting of renewable generation for applications in smart grid power systems. In A. Tomar & R. Kandari (Eds.), Advances in smart grid power system (pp. 265–298). London: Academic Press.
44. Van der Werf, G. R., et al. (2010). Global fire emissions and the contribution of deforestation, savanna, forest, agricultural, and peat fires (1997–2009). Atmospheric Chemistry and Physics, 10, 11707–11735.
45. Van der Werf, G. R., et al. (2006). Interannual variability in global biomass burning emissions from 1997 to 2004. Atmospheric Chemistry and Physics, 6, 3423–3441.
46. Zhang, G., Patuwo, B. E., & Hu, M. Y. (1998). Forecasting with artificial neural networks: The state of the art. International Journal of Forecasting, 14, 35–62.