Revista Brasileira de Cartografia
ISSN 1808-0936 | https://doi.org/10.14393/revbrascartogr
Sociedade Brasileira de Cartografia, Geodésia, Fotogrametria e Sensoriamento Remoto

Spatial Variables and Land-Use Change Models: A Study on Conditioning Patterns of Natural Vegetation Suppression and Persistence

Variáveis Espaciais e Modelos de Mudanças de Uso do Solo: Estudo Sobre Padrões Condicionantes de Supressão e Persistência de Vegetação Natural

Macleidi Varnier [1], Eliseu José Weber [2]

[1] Graduate Program in Remote Sensing, Federal University of Rio Grande do Sul, Porto Alegre, Brazil, macleidi.varnier@ufrgs.br
ORCID: https://orcid.org/0000-0002-2255-7294
[2] Graduate Program in Remote Sensing, Federal University of Rio Grande do Sul, Porto Alegre, Brazil, eliseu.weber@ufrgs.br
ORCID: https://orcid.org/0000-0003-4873-0232

Recebido: 11.2024 | Aceito: 03.2025

Abstract: Land-use change models are formulated by identifying patterns of change and persistence. In modeling software, this step is usually performed by characterizing samples based on spatial variables. Despite the importance of this stage, the evaluation of the change and persistence patterns is often neglected by the scientific community. Thus, this study evaluated the conditioning factors of natural vegetation suppression and persistence in three study areas in different Brazilian biomes. The patterns were investigated for five different time periods, 1995 to 2000 (representing training) and 2000 to 2005, 2000 to 2010, 2000 to 2015 and 2000 to 2020 (representing extrapolation). The spatial variables used to identify the patterns were formulated to represent the environmental context of the training period (1995 to 2000). The method used to analyze the data was Violin Plot graphs. Among the modeling challenges investigated, the following stand out: 1) The ability of variables to explain changes; 2) The variation of change patterns across different time periods; and 3) The variation of change patterns across different study areas and within the same study area. Among the main findings, it was shown that: 1) within the set of analyzed variables, some had a greater ability to differentiate between vegetation suppression and persistence.; 2) the farther the extrapolation was from the training period, the lower the ability of the variables to differentiate the patterns; and 3) Vegetation suppression and persistence in different study areas were described by the variables in distinct ways. As possible recommendations, it is highlighted that modelers analyze patterns of change and persistence using statistical techniques.
Keywords: Predictive variables. Spatial modeling. Violin plot graphs.

Resumo: Modelos de mudanças de uso do solo são formulados com base na identificação de padrões associados às mudanças e persistências. Nos programas de modelagem, essa etapa geralmente é realizada por meio da caracterização de amostras com base em variáveis espaciais. Apesar da importância dessa fase, a avaliação dos padrões de mudança e persistência é frequentemente negligenciada pela comunidade científica. Assim, este estudo avaliou os fatores condicionantes da supressão e persistência da vegetação natural em três áreas de estudo situadas em diferentes biomas brasileiros. Os padrões foram investigados para cinco períodos distintos: 1995 a 2000 (representando o treinamento) e 2000 a 2005, 2000 a 2010, 2000 a 2015 e 2000 a 2020 (representando a previsão). As variáveis espaciais utilizadas para identificar os padrões foram formuladas para representar o contexto ambiental do período de treinamento (1995 a 2000). O método utilizado para analisar os dados foi a construção de gráficos Violin Plot. Entre os desafios de modelagem investigados, destaca-se: 1) A capacidade das variáveis em explicar mudanças; 2) A variação de padrões de mudanças em diferentes períodos de tempo; e 3) A variabilidade espacial dos padrões de mudança, observada entre e dentro de áreas de estudo. Dentre os principais achados, demonstrou-se que: 1) dentro do conjunto de variáveis analisadas, algumas possuíram maior capacidade de diferenciação entre a supressão e a persistência da vegetação; 2) quanto mais distante a previsão esteve do período de treinamento, menor foi a capacidade das variáveis em diferenciar os padrões; e 3) a supressão e a persistência da vegetação natural em diferentes áreas de estudo foram descritas pelas variáveis de formas distintas. Como possíveis recomendações, destaca-se que os modeladores analisem os padrões de mudança e persistência utilizando técnicas estatísticas.
Palavras-chave: Variáveis preditivas. Modelagem espacial. Gráficos de violino.

1 INTRODUCTION

Land-use change models are used to project expected changes in the future. These models are developed by identifying patterns between two time points, t_0 and t_1, and then extrapolating from t_1 to t_x (Pontius et al., 2004). To build these models, samples of change and persistence, spatial variables with explanatory capabilities, and either parametric, non-parametric or semi-parametric methods are utilized. For each sampled point, the information from the spatial variables is extracted based on the corresponding location. Next, a method is used to identify the characteristics that can differentiate the training sample set. The result of this process is a surface indicating the areas considered most and least susceptible to future changes (Soares-Filho et al., 2002; Sangermano et al., 2010; Lin et al., 2011).

Normally, the definition of the spatial variables used to train the models is based on researchers' knowledge of the processes of change in a study area. A wide range of variables can be found, especially geospatial data that varies continuously in space. The most commonly used variables include Euclidean distance to spatial objects such as roads, conservation units, and specific land-uses, as well as terrain attributes like elevation and slope (Soares-Filho et al., 2002; Van Vliet et al., 2016).

The identification of change and persistence patterns aims to find characteristics that conditionally explain these occurrences. In land-use change models, spatial variables are employed to identify attributes correlated with the samples. Using information such as proximity or distance to spatial objects, it is possible to characterize and differentiate the sample group (Soares-Filho et al., 2002; Sangermano et al., 2010). In commercial and open-source modeling software, pattern identification is automated, and its results are often accepted without further evaluation (Lin et al., 2011; Van Vliet et al., 2016).

However, knowing how the training process works, some difficulties can be listed: 1) The variables used in the training may have different capacities for predicting change (Adhikari & Southworth, 2012; Chavan et al., 2018); 2) The predictive variables used in the training may not be related to land-use changes (Adhikari & Southworth, 2012; Chavan et al., 2018); 3) The patterns of change identified in the training may not persist over the extrapolation period; 4) Different study areas have different patterns of change (Trigueiro et al., 2020); 5) within the same study area, patterns of change may vary (Trigueiro et al., 2020); and 6) Different land-use transitions have different patterns of occurrence (Kucsicsa et al., 2019).

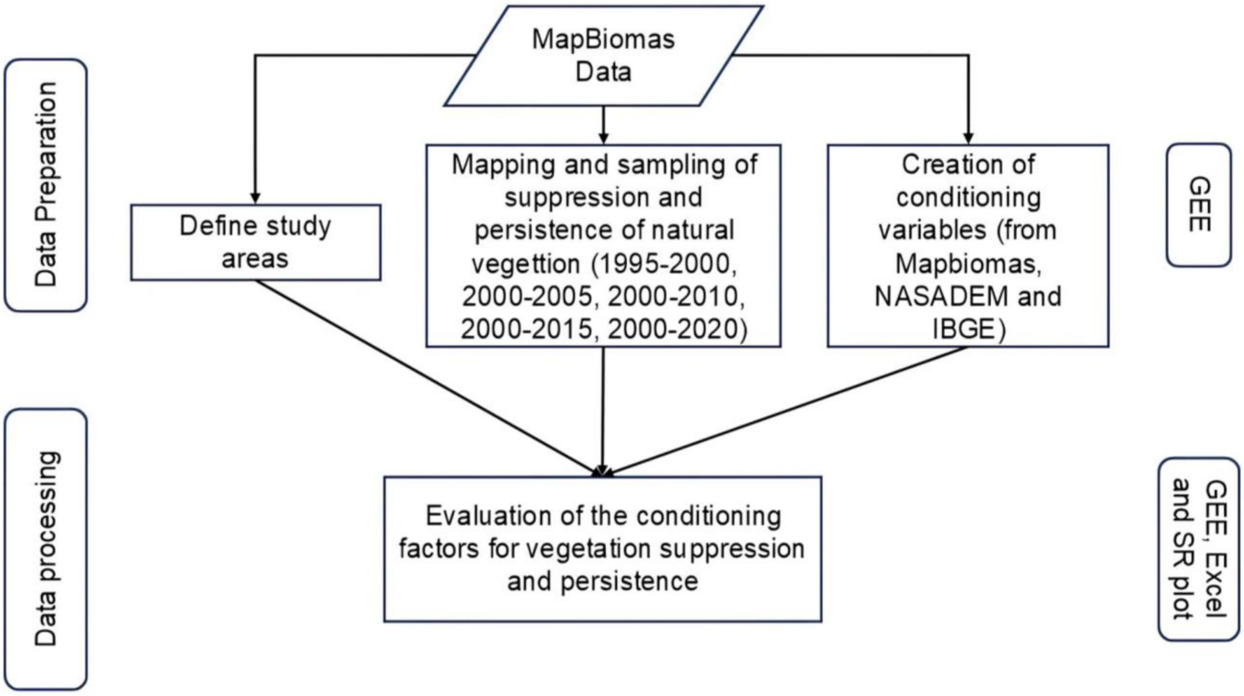
Given these complications, assessing the coherence of the identified patterns enables a better understanding of how the models work. One possible method of investigation is the evaluation of conditioning factors. This technique is based on the statistical representation of how the samples are described by the predictive variables. While this technique is widely used in studies like landslide assessment (Brito et al., 2017), initiatives to evaluate patterns of land-use change are rare (Trigueiro et al., 2020). Consequently, there is a gap in the scientific community's understanding of how models work. Therefore, most modeling is carried out using ready-made methods and the results are not critically assessed.

Thus, it is necessary to conduct research that investigates the difficulties exposed in order to better describe the functioning of the models. Considering this lack of understanding, this work aims to provide a basis for evaluating the behavior of predictive variables in land-use change models under different environmental conditions and time periods.

2 MATERIALS AND METHODS

The study was conducted in selected areas of three Brazilian biomes, using land-cover and land-use data from MapBiomas and predictor variables derived from MapBiomas, NASADEM (digital elevation model) and the Brazilian Institute of Geography and Statistics (IBGE). The methodology comprised two main stages, executed with the assistance of Google Earth Engine (GEE) (Gorelick et al., 2017), Excel (Microsoft Corporation, 2016), and SR plot (Tang et al., 2023). Figure 1 shows the flowchart of the work methodology. In the central part of the figure are the processes, on the left are the stages of the work, and on the right are the software used.

Figure 1 – Flowchart of the methodology.



Source: The authors, 2025.

2.1 Study area

The study areas were selected to represent different environmental contexts that experienced land-use changes and reductions in natural vegetation between 2000 and 2020. In three Brazilian biomes, areas that exhibited large land-use changes during this period were selected. The areas were chosen based on land-use maps from Collection 8 of the MapBiomas project (Souza et al., 2020) and the tiles of the respective 1:250,000 grid.

We first calculated the area of the natural vegetation classes for the years 2000 and 2020 in each grid cell of the map. Subsequently, we calculated the difference in the area of these vegetation formations between 2000 and 2020. Finally, we chose the grid cell with the largest absolute reduction in natural vegetation cover during the analyzed period in the Amazon, Cerrado, and Pampa biomes. The selection of areas in three different biomes sought to represent different environmental characteristics and pressures that may affect change.

The selected study areas encompass a wide variety of spaces and dynamics of land-use change. Figure 2 shows the location of the three study areas.

The Amazon biome study area is located in the state of Rondônia, in northern Brazil, within the longitudinal range of 64°30′ W to 63° W, and the latitudinal range of 9° S to 10° S. It is characterized by the predominance of dense, humid tropical forest, contrasting with extensive areas of pasture derived from recent deforestation. In the period from 2000 to 2020, there was a reduction of approximately 40% in the natural vegetation cover, with deforestation advancing from south to north. The suppression of the Amazon rainforest is primarily driven by timber harvesting and the expansion of livestock and agricultural production. Today, this process is regarded as a global environmental issue, attracting attention from various sectors of society (Souza et al., 2020).

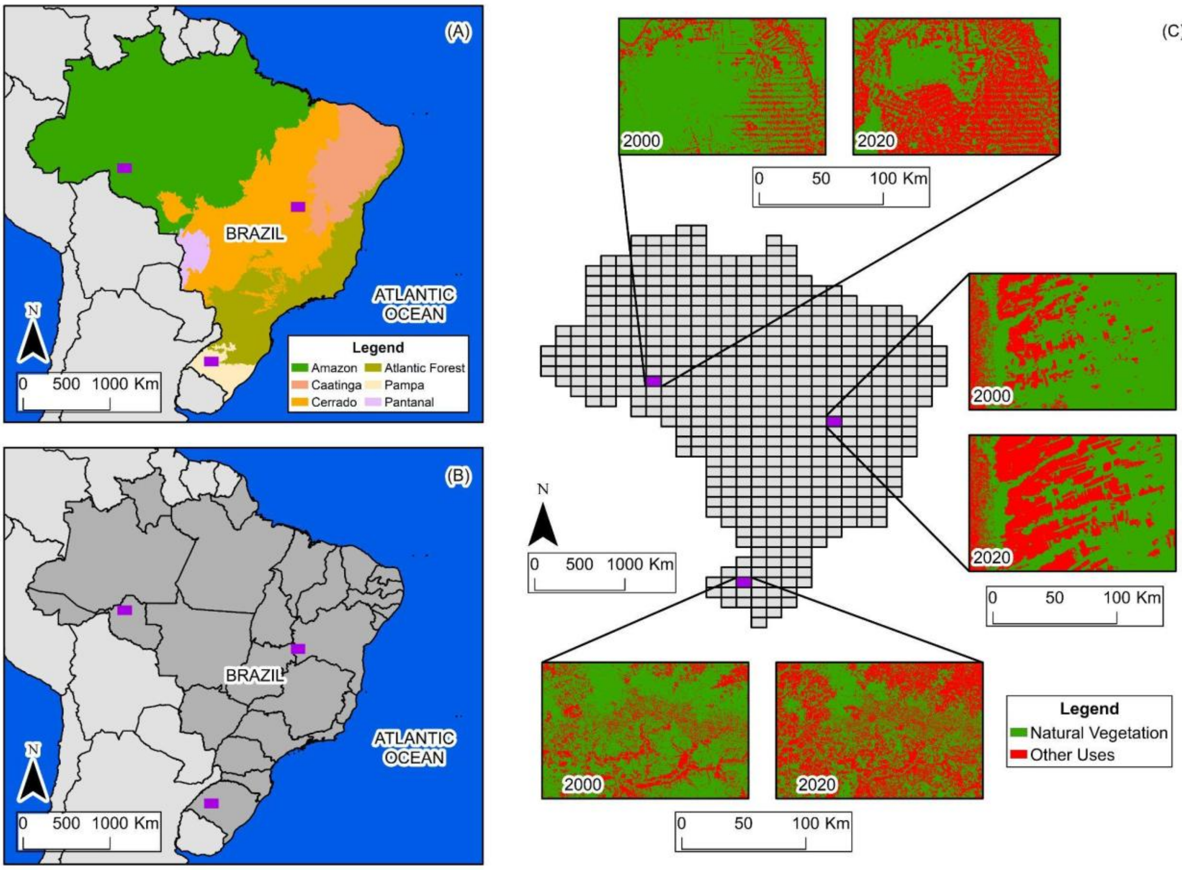
The Cerrado biome study area is located on the border of the states of Bahia and Goiás, in the central portion of Brazil, within the longitudinal range of 46°30′ W to 45° W, and the latitudinal range of 13° S to 14° S. It is characterized by mosaics of savannah, grass-land, agriculture, and pasture. In the period from 2000 to 2020, there was a reduction of approximately 28% in the natural vegetation cover, with suppression progressing from west to east. The reduction in natural vegetation in this part of Cerrado is motivated by the implementation of large-scale monocultures, a process that has received attention from the scientific community in order to understand its environmental impacts (Souza et al., 2020; Pontius et al., 2023).

The Pampa biome study area is located in the state of Rio Grande do Sul, in southern Brazil, within the longitudinal range of 55°30′ W to 54° W, and the latitudinal range of 29° S to 30° S. It is characterized by

 DOI: http://dx.doi.org/10.14393/rbcv77n0a-76156

grassland vegetation, pastures, agricultural uses, and forested areas. Between 2000 and 2020, natural vegetation decreased by approximately 29%, with suppression occurring without a defined spatial pattern. The conversion of natural vegetation in this part of Pampa is primarily driven by the establishment of agricultural crops, particularly the shift from grassland formations to soybean plantations. Between 2000 and 2020, soybean production expanded extensively in the biome, considerably altering the landscape. Another important process is the periodic alternation of land-use classes, with transitions from grassland vegetation to pasture and rice production, and vice versa (Souza et al., 2020).

Figure 2 – (A) Biome boundaries and study areas. (B) State boundaries and study areas. (C) Grid, study areas, and land use and land cover in 2000 and 2020.



Source: The authors, 2025.

2.2 Data used

All the geospatial data used in this study comes from open sources and is accessible for consultation. The land-use maps, used to map change and persistence and to formulate Euclidean distance variables, were derived from the MapBiomas project. MapBiomas is a collaborative effort involving researchers from various Brazilian institutions, including universities, NGOs, research institutes, and technology start-ups. One of its main products is the annual land-use and land-cover maps, available since 1985 for the entire Brazilian territory. These maps are generated using data from the Landsat program, offering a spatial resolution of 30 m. A key strength of this mapping is the robust accuracy assessment, allowing users to understand the disagreements in the classification for each land-use category (Souza et al., 2020; MapBiomas, 2023a; 2023b).

The elevation and slope data were derived from the NASADEM digital elevation model. NASADEM is a product resulting from the reprocessing of the Shuttle Radar Topographic Mission (SRTM) with a spatial resolution of 30 m. Compared to SRTM, NASADEM provides greater accuracy due to processing improvements and the incorporation of auxiliary data from the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) and the Ice, Cloud, and Land Elevation Satellite (ICESat)/Geoscience Laser Altimeter System (GLAS) (National Aeronautics and Space Administration, 2023).

The road network was derived from the Brazilian Institute of Geography and Statistics (IBGE), representing federal and state roads. The scale of this data is 1:250,000 and the features were delimited using images from the Sentinel-2, Planet and Maxar sensors. During the validation of the product, an average positional error of 125 m was considered acceptable. This error is attributed to the scale at which the mapping was conducted, as defined by the Cartographic Accuracy Standard of the Brazilian National Cartographic System (Instituto Brasileiro de Geografia e Estatística, 2024).

2.3 Data preparation

Data preparation was conducted on the GEE platform and involved three main stages: (1) reclassification of land-cover and land-use maps from the MapBiomas project; (2) mapping and sampling of natural vegetation suppression and persistence; and (3) preparation of predictive variables.

2.3.1 RECLASSIFICATION OF LAND-USE AND LAND-COVER MAPS

The land-use and land-cover maps from the MapBiomas project were reclassified in order to group the original classes into just two: natural vegetation and other uses. The land-use maps for the years 1995, 2000, 2005, 2010, 2015, and 2020 were selected for reclassification. These specific years and intervals were selected to ensure that the training period would match the duration of each extrapolation period, enabling us to assess how the variables developed during training behave across different extrapolation periods.

2.3.2 MAPPING AND SAMPLING OF NATURAL VEGETATION SUPPRESSION AND PERSISTENCE

The mapping of vegetation suppression and persistence was based on the reclassified land-use maps. Five different periods were mapped: 1995 to 2000, representing training; and 2000 to 2005, 2000 to 2010, 2000 to 2015 and 2000 to 2020, representing extrapolation.

The method used was based on comparing two land-use maps from different times. Equation 1 illustrates this process.

$$NVSP = Xt_0 - Xt_1 \tag{1}$$

where natural vegetation suppression and persistence ‘NVSP’ is obtained by comparing ‘$-$’ the state of the cells at time point t_0 ‘Xt_0’ with the state of the cells at time point t1 ‘Xt_1’. By applying this function to the entire study area, a mask is obtained indicating the areas of vegetation suppression and persistence for the period analyzed.

Based on this mapping, a random sampling of 10,000 points was carried out for each class (suppression and persistence of natural vegetation).

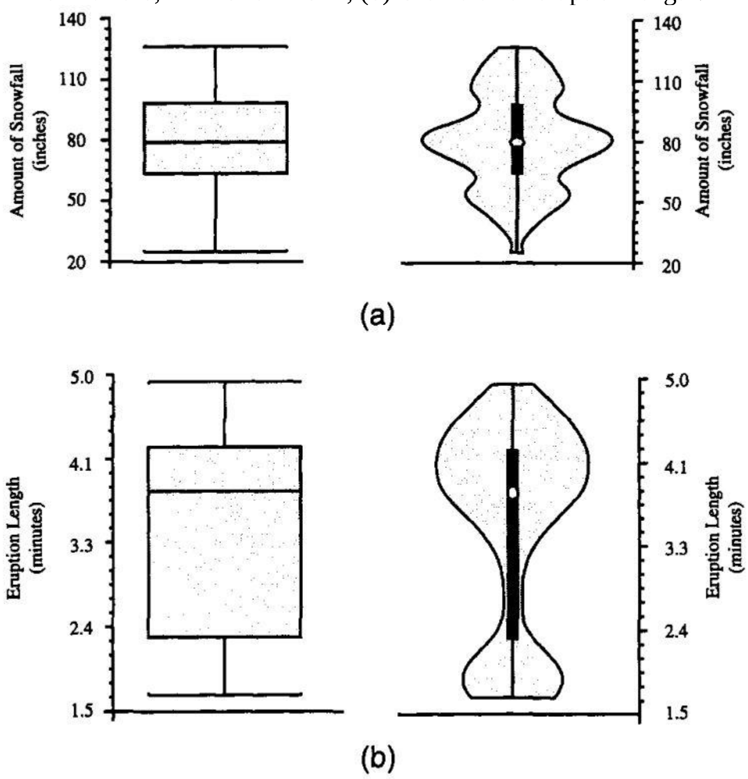
2.3.3 PREPARATION OF PREDICTIVE VARIABLES

The predictive variables for natural vegetation suppression were derived from MapBiomas, NASADEM, and IBGE data. From MapBiomas, we computed Euclidean distance from the water bodies in 1995, urban space in 1995, anthropogenic uses in 1995, and vegetation suppression between 1990 and 1995. From NASADEM we extracted elevation and slope data. And from the IBGE, we calculated Euclidean distance for the roads existing in 1995.

2.4 Evaluation of the conditioning factors for natural vegetation suppression and persistence

Based on the set of samples and the predictive variables, the attributes that explain the occurrence and non-occurrence of vegetation suppression were evaluated. For each suppression sample, the values contained in the predictive variables were extracted based on their x and y positions. This data was then organized in a table, and violin plots were generated to represent the distribution of values received by the samples.

Figure 3 – Example Comparing Box Plot (graphs on the left) and Violin Plot (graphs on the right). (a) Annual snowfall for Buffalo, NY 1910 – 1972; (b) Old Faithful eruption length.



Source: Hintze and Nelson, 1998.

Violin Plot graphs show the distribution of a set of samples in relation to their values. The ‘y’ axis shows the values of the variables, while the ‘x’ axis shows the kernel density for the set of samples. Kernel density is a way of distributing data in a continuous space. Each sample is distributed along a vertical axis based on its value. The different samples that have the same value are superimposed and added. The result is a curve that estimates the density of the samples for all values in relation to the sample population. Inside the violin graph is a box plot, displayed in black. The white circle represents the median. The space between the circle and the lower horizontal line represents the lower quartile (25% of the samples). And the space between the circle and the upper horizontal line represents the upper quartile (25% of the samples). The samples below the lower horizontal line and above the upper horizontal line represent 25% of the samples with the lowest and highest values, respectively (Turkey, 1977; Hintze & Nelson, 1998).

Equation 2 presents the method used to calculate the Kernel density.

$$d(x/h) = \frac{\sum_{i=1}^{n} \delta_i}{nh} \qquad (2)$$

The Kernel density ‘d(x|h)’ of a point ‘x’ is defined as the fraction of values per unit of measure that lie within an interval centered at ‘x’. Where ‘n’ is the number of samples, ‘h’ is the interval width, and ‘δi’ is one when the ith data value is in the interval [x−h/2x+h/2] and zero otherwise (Hintze & Nelson, 1998). For this study, the ‘h’ value was set at 2% of the total sample size.

Violin plots are considered an evolution of the Box Plot, as they allow for a more detailed analysis of sample distribution. Additionally, the Violin Plot technique encompasses all the traditional statistics of the Box Plot, providing a detailed description of the phenomenon under investigation (Hintze & Nelson, 1998).

For the application used in this work, analyzing Violin Plot graphs makes it possible to identify and differentiate the potential of variables to describe the occurrence and non-occurrence of natural vegetation suppression. It is expected that, for a variable with high descriptive potential, samples of vegetation suppression and persistence will be described differently.

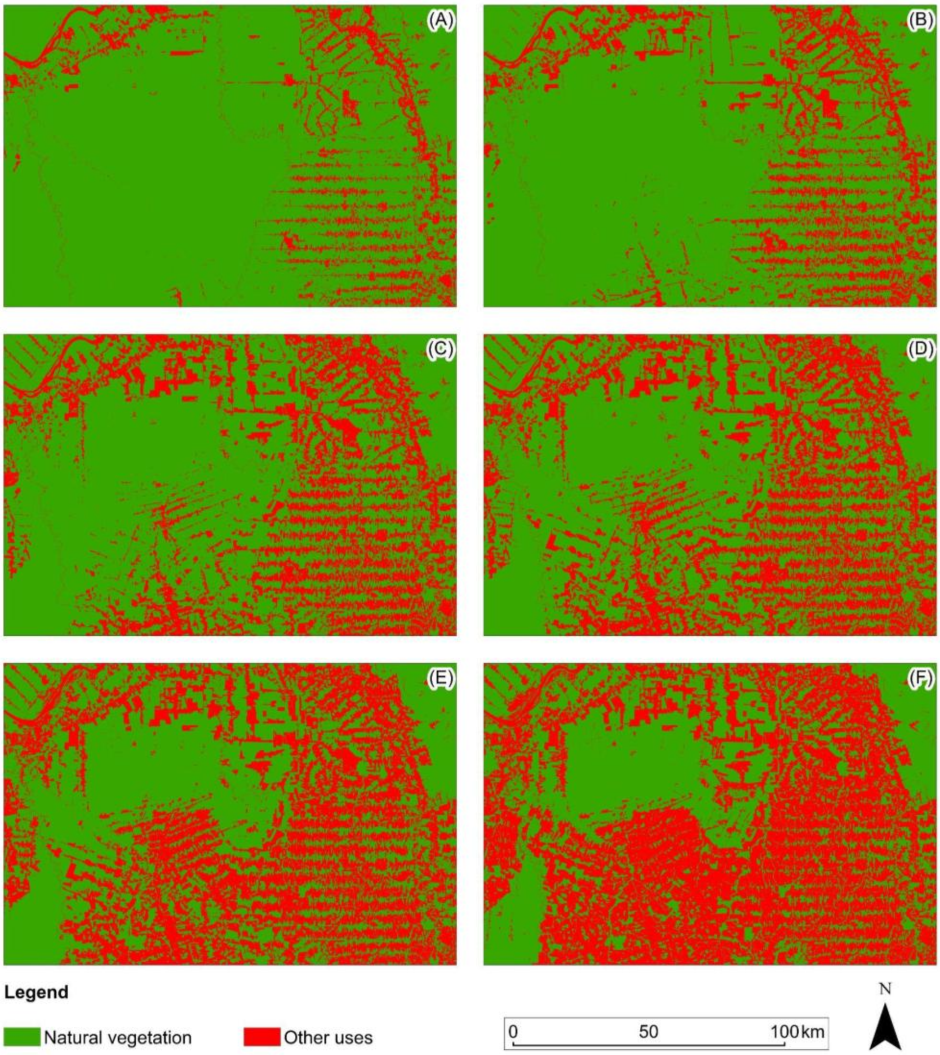
 DOI: http://dx.doi.org/10.14393/rbcv77n0a-76156

3 RESULTS

3.1 Amazon biome

Figure 4 shows the temporal evolution of land-use and land-cover for the study area in the Amazon biome based on MapBiomas data. In 1995, the natural vegetation class covered 89.6% of the study area. In 2000, this percentage had decreased to 82.3%; 73.02% in 2005; 66.8% in 2010; 59.7% in 2015; and 49.3% in 2020.

Figure 4 – Temporal evolution of land-use and land-cover in the Amazon Biome study area. (A) 1995. (B) 2000. (C) 2005. (D) 2010. (E) 2015. (F) 2020.



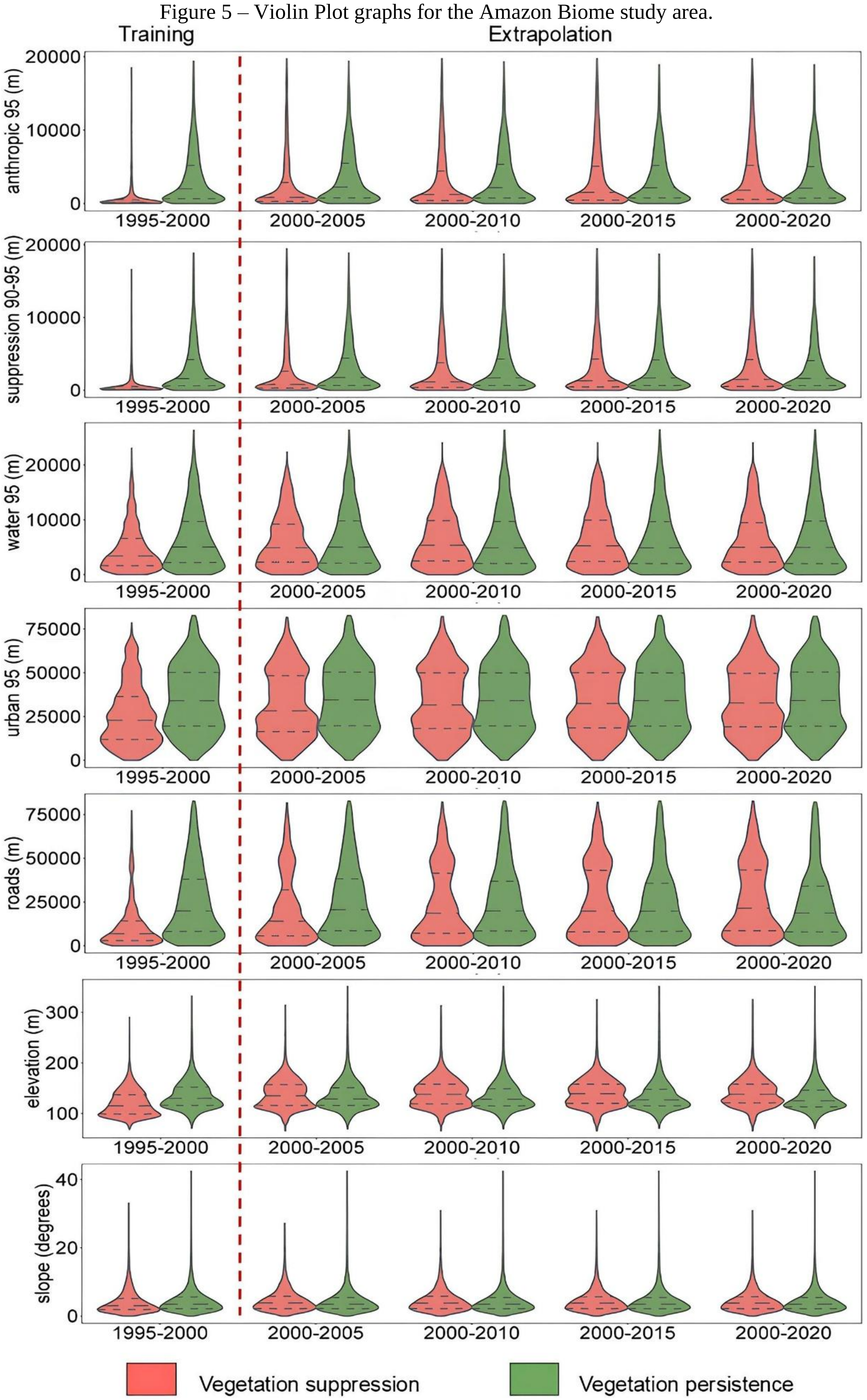
Source: The authors, 2025.

Figure 5 shows the results found for the study area in the Amazon biome. For all variables, the ability to differentiate between suppression and persistence of natural vegetation was greater during the training period compared to the extrapolation period. It is also observed that as the extrapolation periods advance, the ability of the variables to discriminate between suppression and persistence of natural vegetation decreases.

When analyzing the data for the variables of Euclidean distance for anthropic uses in 1995 (anthropic 95) and Euclidean distance for suppressions of natural vegetation between 1990 and 1995 (suppression 90-95), it is evident that vegetation suppression is concentrated in the lowest values across all periods. This indicates that the suppression of natural vegetation predominantly occurred near areas of anthropic uses in 1995 and areas of vegetation suppression between 1990 and 1995.

 DOI: http://dx.doi.org/10.14393/rbcv77n0a-76156

Figure 5 – Violin Plot graphs for the Amazon Biome study area.



Source: The authors, 2025.

For the variables of Euclidean distance to roads (roads), Euclidean distance to urban areas in 1995 (urban 95), and Euclidean distance to water bodies in 1995 (water 95), the suppression of natural vegetation was distributed across a wide range of values in all periods analyzed. This wide distribution makes it difficult to identify a clear pattern that explains the occurrence of vegetation suppressions in a simple and effective way. There is also a similarity between the way vegetation suppression and persistence were described, indicating a low discriminative ability of these variables.

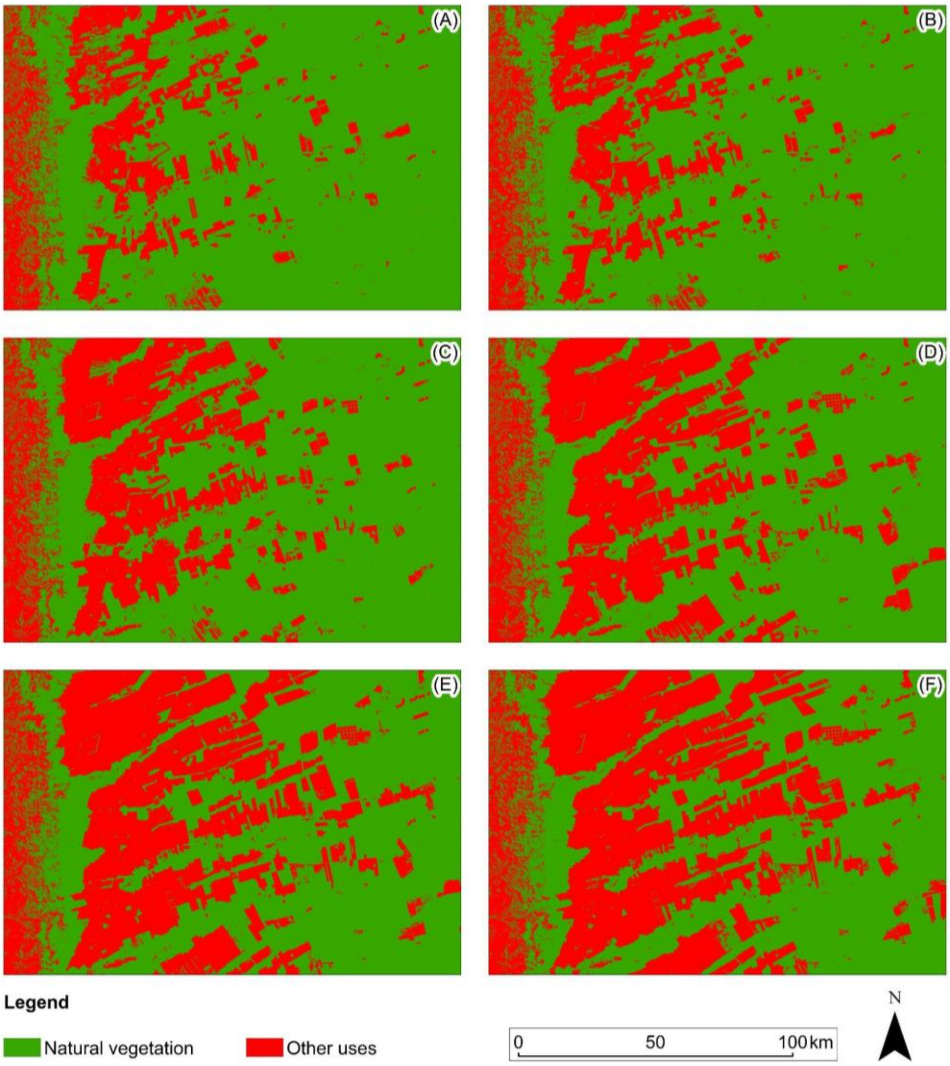
And for the topographical variables of elevation and slope, vegetation suppression during the analyzed periods was primarily concentrated in areas with low altitudes and gentle slopes. However, knowing that the study area in the Amazon biome is predominantly low and flat terrain, the results obtained using these variables do not reveal efficient vegetation suppression patterns.

For this study area, the variables with the greatest ability to differentiate between suppression and persistence of natural vegetation were Euclidean distance for anthropic uses in 1995 (anthropic 95) and Euclidean distance for suppressions of natural vegetation between 1990 and 1995 (suppression 90-95).

3.2 Cerrado biome

Figure 6 shows the temporal evolution of land-use and land-cover for the study area in the Cerrado biome based on MapBiomas data. In 1995, the natural vegetation class covered 79.1% of the study area. In 2000, this percentage had decreased to 76.2%; 69.2% in 2005; 62.8% in 2010; 58.5% in 2015; and 55.1% in 2020.

Figure 6 – Temporal evolution of land-use and land-cover in the Cerrado Biome study area. (A) 1995. (B) 2000. (C) 2005. (D) 2010. (E) 2015. (F) 2020.



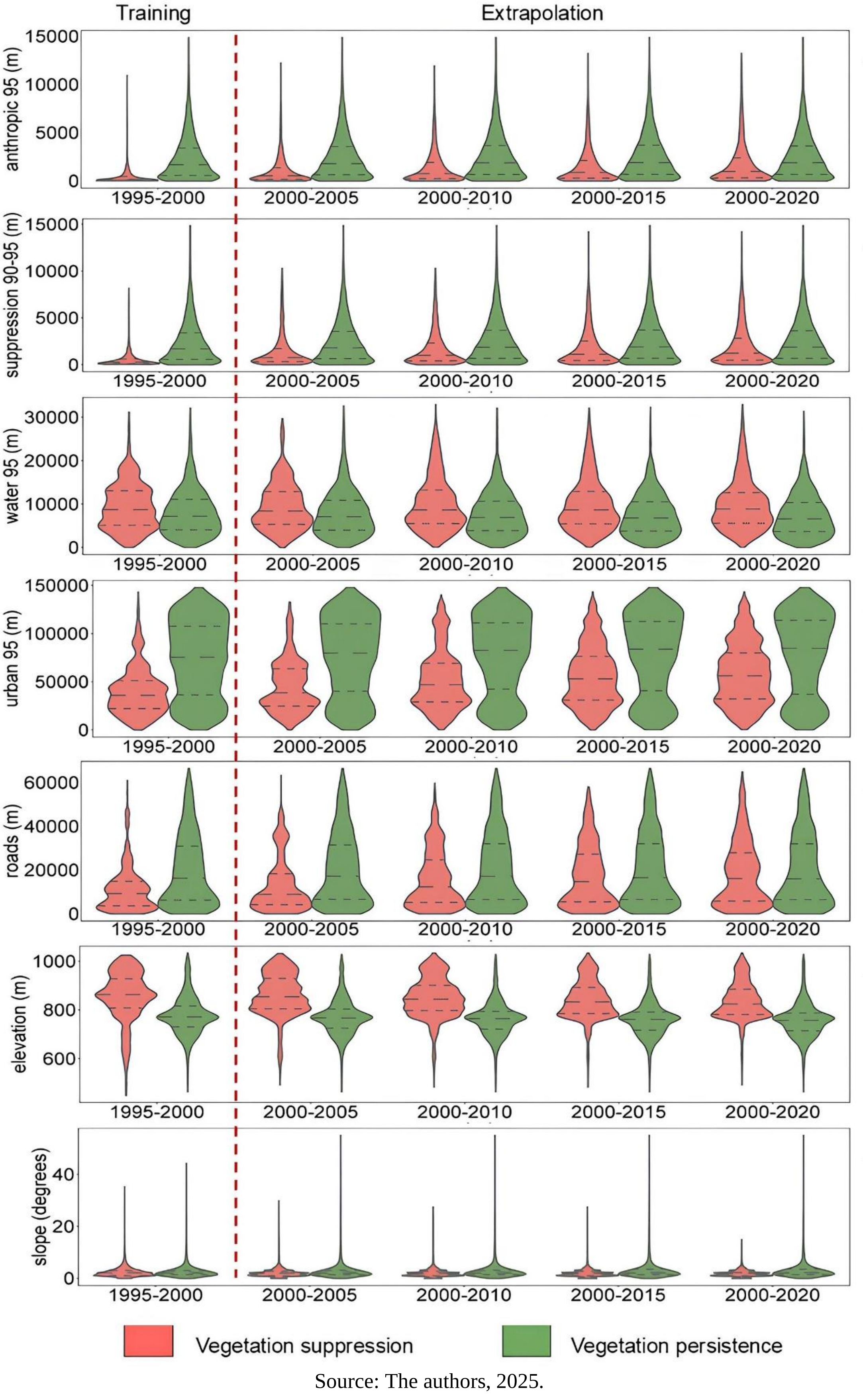
Source: The authors, 2025.

Figure 7 shows the results found for the study area in the Cerrado biome. For all variables, the ability to differentiate between suppression and persistence of natural vegetation was greater during the training period compared to the extrapolation period. It is also observed that as the extrapolation periods increase, the

 DOI: http://dx.doi.org/10.14393/rbcv77n0a-76156

ability of the variables to discriminate between suppression and persistence of natural vegetation decreases.

Figure 7 – Violin Plot graphs for the Cerrado Biome study area.



Source: The authors, 2025.

When analyzing the data for the variables of Euclidean distance to anthropic uses in 1995 (anthropic 95) and Euclidean distance to vegetation suppressions between 1990 and 1995 (suppression 90-95), it is evident that vegetation suppression is concentrated in the lowest values across all periods. This indicates that the suppression of natural vegetation primarily occurred near areas of anthropogenic uses in 1995 and areas of vegetation suppression between 1990 and 1995.

For the variables of Euclidean distance to roads (roads), Euclidean distance to urban areas in 1995 (urban 95), and Euclidean distance to water bodies in 1995 (water 95), the suppression of natural vegetation was distributed across a wide range of values in all the periods analyzed. Despite this, it is possible to see significant differences between the patterns of suppression and persistence of natural vegetation. In this way, these variables help to differentiate the two groups.

For topographical variables, in general, vegetation suppression occurred in higher areas than vegetation persistence during all periods. For slope, vegetation suppression and persistence were concentrated in areas with low slopes, not allowing easy differentiation.

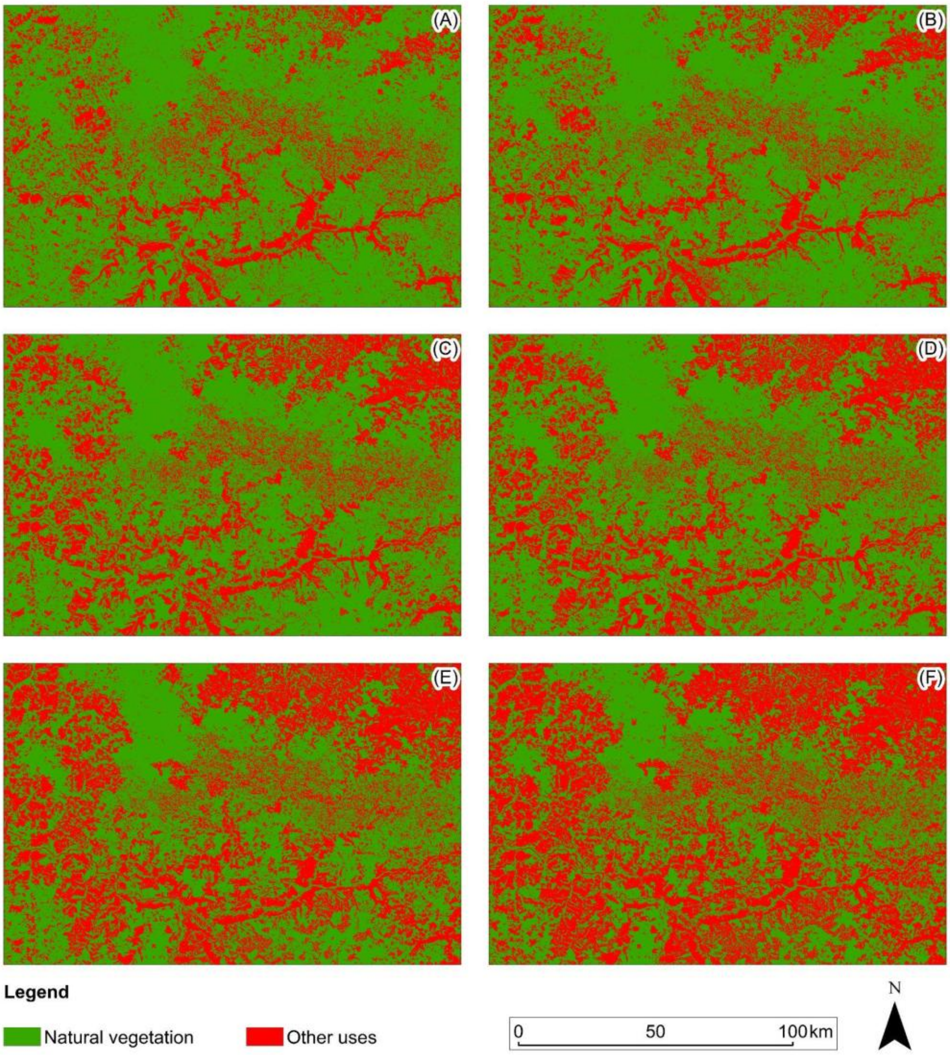
For this study area, the variables with the greatest ability to differentiate between suppression and persistence of natural vegetation were Euclidean distance for anthropic uses in 1995 (anthropic 95), Euclidean distance for suppressions of natural vegetation between 1990 and 1995 (suppression 90-95), Euclidean distance for urban in 1995 (urban 95) and elevation.

3.3 Pampa biome

Figure 8 shows the temporal evolution of land-use and land-cover for the study area in the Pampa biome based on MapBiomas data. In 1995, the natural vegetation class covered 75.3% of the study area. In 2000, this percentage had decreased to 73.5%; 68.1% in 2005; 65.6% in 2010; 59.2% in 2015; and 52.2% in 2020.

 DOI: http://dx.doi.org/10.14393/rbcv77n0a-76156

Figure 8 – Temporal evolution of land-use and land-cover in the Pampa Biome study area. (A) 1995. (B) 2000. (C) 2005. (D) 2010. (E) 2015. (F) 2020.



Source: The authors, 2025.

Figure 9 shows the results found for the study area in the Pampa biome. For all variables, the ability to differentiate between suppression and persistence of natural vegetation was greater during the training period compared to the extrapolation period. It is also observed that as the extrapolation periods increase, the ability of the variables to discriminate between suppression and persistence of natural vegetation decreases.

Analyzing the data for the variables of Euclidean distance for anthropic uses in 1995 (anthropic 95), Euclidean distance for vegetation suppressions between 1990 and 1995 (suppression 90-95) and Euclidean distance for water bodies in 1995 (water 95), it is evident that vegetation suppression is concentrated in the lowest values across all periods. Nevertheless, it is noted that, except for the variable anthropic 95, the others did not show significant differences between suppression and persistence of natural vegetation.

For the variables of Euclidean distance to roads (roads) and Euclidean distance to urban areas in 1995 (urban 95), the suppression of natural vegetation was distributed across a wide range of values in all the periods analyzed. In addition, there were no considerable differences between suppression and persistence of natural vegetation. In this way, it is impossible to identify a pattern that explains the occurrence of vegetation suppressions in a simple and effective way, indicating that there is no relationship between these variables and the suppressions.

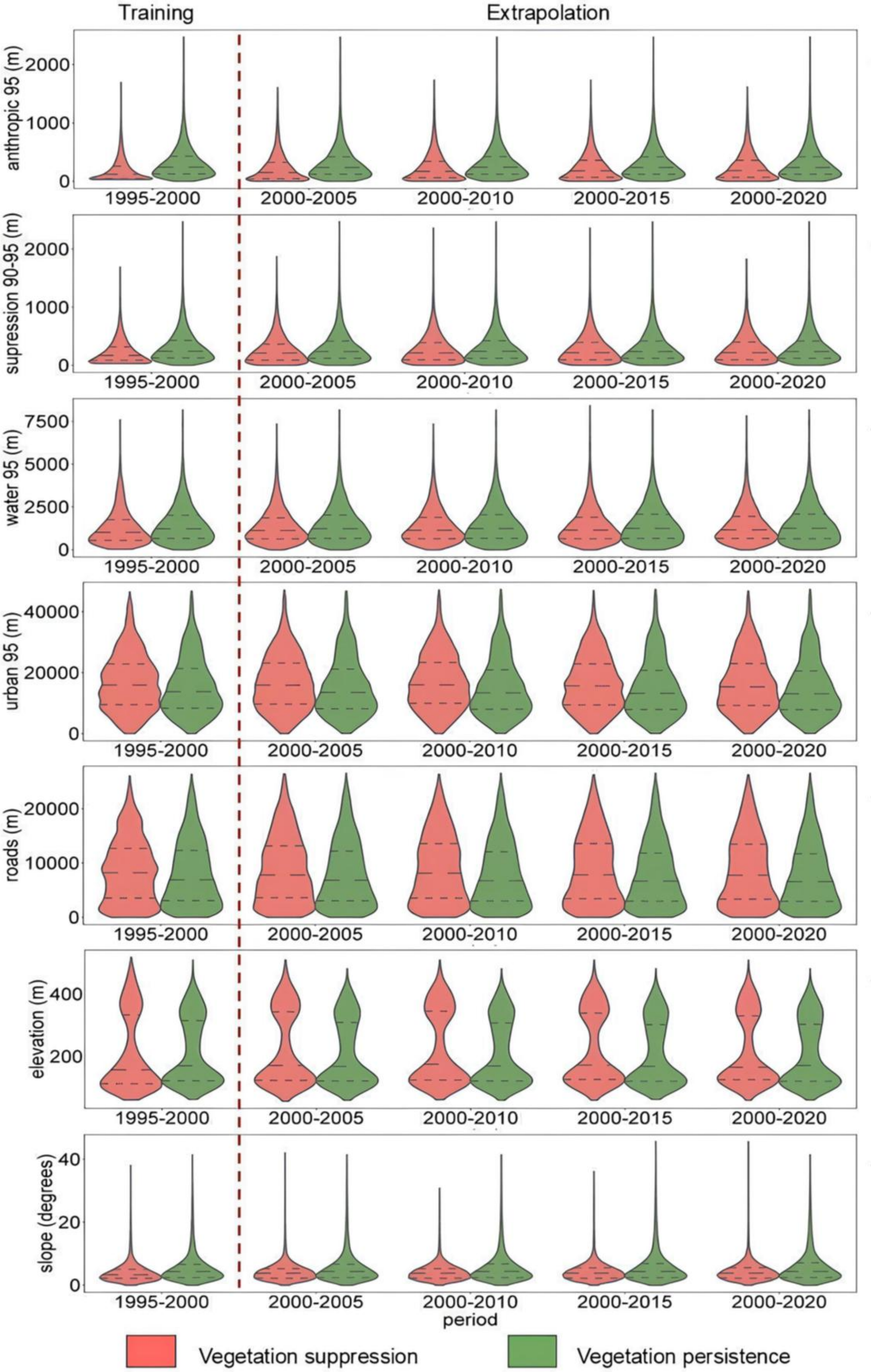
As for the topographical variables, it can be observed that during the analyzed periods, vegetation suppression was concentrated in two altitude groups, around 130 m and 350 m. These results reflect the occurrence of suppression in two distinct relief units: the Paraná basin plateau and the Central Depression. Thus, it is not possible to define an effective suppression pattern using these data, but only to demonstrate the spatial variability of suppressions in the study area. Regarding slope, there are no expressive differences

 DOI: http://dx.doi.org/10.14393/rbcv77n0a-76156

between the suppression and persistence of natural vegetation.

For this study area, the variable with the greatest ability to differentiate between suppression and persistence of natural vegetation were Euclidean distance for anthropic uses in 1995 (anthropic 95).

Figure 9 – Violin Plot graphs for the Pampa Biome study area.



Source: The authors, 2025.

 DOI: http://dx.doi.org/10.14393/rbcv77n0a-76156

4 DISCUSSION

The results presented in this article make it possible to provide answers to the difficulties outlined in the introduction:

1) "The variables used in the training may have different capacities for predicting change". When examining the data obtained for all the study areas (Figures 3.5, 3.7, and 3.9), it is evident that in all situations some variables were superior to the others. These results corroborate with data found in other studies, leading to the conclusion that among various variables used to train the models, some usually have a greater ability to describe the changes (Voight et al., 2019. Kucsicsa et al., 2019).

2) "The predictive variables used in the training may not be related to land-use changes". When analyzing the histograms of some variables (Figures 3.5, 3.7, and 3.9), it is remarkable the absence of relation with vegetation suppressions. One example of this issue can be demonstrated by the elevation variable in the Pampa biome (Figure 9). When analyzing the histograms for the different periods, it is observed that the suppression samples are allocated into two well-defined groups: from 80 to 190 m and from 280 to 450 m. This result highlighted the occurrence of vegetation suppressions in different relief compartments, the Sul Rio Grandense plateau and the Central Depression. Thus, the data range observed in the histograms is primarily justified by the geomorphological conditions of the study area. This illustrates that the patterns identified from predictive variables may not necessarily express the conditions that determine changes, but simply represent the diversity of characteristics in a study area. One possibility to circumvent this problem is to train the models for regions with similar environmental characteristics, facilitating the identification of valid patterns. If the goal is to model land-use changes over large areas, segmenting and training homogeneous sub-regions could be a viable alternative (Kucsicsa et al., 2019).

3) "The patterns of change identified in the training may not persist over the extrapolation period". Observing the graphs and median values in Figures 3.5, 3.7 and 3.9, it is noticed that changes are represented in a variable way in different periods. This variation can be elucidated by factors inherent to the modeling process, three main reasons can be listed: i) it is known that the mapping and extrapolation of changes in the training and extrapolation periods are based on land-use maps from different years. For training, changes are mapped based on the initial year t0, and for extrapolation, the base year considered is t1. Therefore, it is expected that using maps from different time periods will lead to the mapping and extrapolation of distinct change areas. This is justified because maps from different times express distinct scenarios and have different classification disagreements (Pontius et al., 2017). ii) The predictive variables are developed for a time point immediately before the training period. Thus, the patterns indicated by the variables tend to be more relevant for the training period than for extrapolation. It is expected that over time the chances of changes in patterns that explain land-use changes will increase. Furthermore, natural vegetation suppression often occurs in patches, which reduce the predictive ability of variables based on Euclidean distance as time moves further from the period of their formulation. And iii), as time periods advance, the amount of mapped/projected changes tends to increase (as identified in all study areas). Thus, it is natural to expect that further away from the training period, more changes will be allocated in areas considered less susceptible during the training.

4) "Different study areas have different patterns of change" and 5) "within the same study area, patterns of change may vary". When comparing the patterns of suppression of natural vegetation identified by the same variable in the different study areas (Figures 3.5, 3.7, and 3.9), different behaviors can be observed. This demonstrates that the predictive characteristics of changes are particular to each space due specific environmental conditions. This question can be illustrated with the predictive variable of distance to anthropogenic uses in 1995. For the study area in the Amazon biome, the median vegetation suppression for the analyzed periods ranged from 22,861 to 32,791 m. In the Cerrado biome, it was between 35,933 and 56,177 m, while for the Pampa biome, it ranged from 15,264 to 16,037 m.

Considering this data, the patterns of suppression found in each study area differ from the others, being representative of suppressions only in their specific context. Therefore, it is difficult to generate land-use change models for large areas. The variability of the characteristics that condition the changes, combined with the environmental diversity of the different locations, restricts the identification of valid patterns. In a similar assessment, Trigueiro et al. (2020) highlighted the existence of spatial variability in the predictors of land-use change in the Brazilian Cerrado biome. For the different sub-regions of the biome, it was demonstrated the relationship of vegetation suppression with different variables. In view of this, it is important to highlight that when developing a model, the statistical method used will adjust the patterns of change considering the values found in all the samples employed (Soares-Filho, 2002; Sangermano, 2010; Lin et al., 2011). As a result, modeling is unable to represent the spatial variability of all the factors of change.

6) “Different land-use transitions have different patterns of occurrence”. The type of change to be modeled is crucial for pattern identification. In this study, a one-way transition is evaluated. However, it is common to project multiple changes simultaneously (Voight et al, 2019; Kucsicsa et al, 2019). Considering that these models use the same set of variables to identify the patterns of all transitions, modeling multiple changes at the same time increases the chance of errors. This happens because there is no way to control which variables will be used in the training of specific changes, and data that have no relation to the changes may be used. Similarly, the evaluation of predictive factors of specific changes is hindered, requiring additional methodological steps and generating a large amount of data. As a result, the analysis of identified change patterns is commonly ignored.

Moreover, the variables used to identify patterns represent only a portion of the factors driving environmental change, failing to capture the full complexity of human and natural systems. Consequently, model results provide a simplified approximation of reality and are subject to inaccuracies. Furthermore, predictive variables used in model training often contain errors, such as disagreements in land-use and land-cover classifications, and errors in location, delineation, and referencing of geographic features. When generating Euclidean distance surfaces based on such data, the resulting change patterns will inevitably contain imperfections.

5 CONCLUSIONS

The results obtained in this study highlight important points about the use of predictive variables in the representation of land-use changes: (1) In a group of variables used to represent vegetation suppression, each variable showed a different degree of representativeness. (2) The same set of variables used to represent changes in different locations showed specific behavior in each area. (3) When the same set of variables was applied to characterize vegetation suppression over different time intervals, each period displayed distinct behavior. (4) Some predictive variables showed no relationship with vegetation suppression. And (5), assessing the ability of predictive variables to represent land-use changes provides valuable insights into how the models work.

Considering the main conclusions listed above, it is recommended that when defining predictive variables for training land-use change models, researchers conduct a statistical assessment of the variables' ability to represent the investigated changes. This approach makes it possible to analyze predictive variables from a robust database, reducing the complexity of the models and improving understanding of how they work. Furthermore, it would be beneficial to compare the assessment of conditioning factors with the importance of variables in model training.

References

Adhikari, S., & Southworth, J. (2012). Simulating Forest Cover Changes of Bannerghatta National Park Based on a CA-Markov Model: A Remote Sensing Approach. *Remote Sensing*, 4(10), 3215-3243. https://doi.org/10.3390/rs4103215.

Brito, M. M., Weber, E. J., Krigger, V. S., & Leitzke, F. P. (2017). Análise dos fatores condicionantes de movimentos de massa no município de porto alegre a partir de registros históricos. *Revista Brasileira de Cartografia*, 68(9), 1853-1872. http://dx.doi.org/10.14393/rbcv68n9-44450.

Chavan, S. B., Reddy, C. S., Rao, S. S., & Rao, K. K. (2018). Assessing and Predicting Decadal Forest Cover Changes and Forest Fragmentation in Kinnerasani Wildlife Sanctuary, Telangana, India. *Journal of The Indian Society of Remote Sensing*, 46(5), 729-735. http://dx.doi.org/10.1007/s12524-017-0739-x.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, 202. http://dx.doi.org/10.1016/j.rse.2017.06.031.

Hintze, J., & Nelson, R. (1998). Violin Plots: a box plot-density trace synergism. *The American Statistician*, 52(2), 181-184. http://dx.doi.org/10.1080/00031305.1998.10480559.

Instituto Brasileiro de Geografia e Estatística. (2024). *Base Cartográfica Contínua*. [Map]. https://geoftp.ibge.gov.br/cartas_e_mapas/bases_cartograficas_continuas/bc250/versao2023/shapefile/.

Kucsicsa, G., Popovici, E.-A., Bălteanu, D., Dumitraşcu, M., Grigorescu, I., & Mitrică, B. (2019). Assessing the Potential Future Forest-Cover Change in Romania, Predicted Using a Scenario-Based Modelling. *Environmental Modeling & Assessment*, 25(4), 471-491. http://dx.doi.org/10.1007/s10666-019-09686-6.

Lin, Y. P., Chu, H. J., Wu, C. F., & Verburg, P. H. (2011). Predictive ability of logistic regression, auto-logistic regression and neural network models in empirical land-use change modeling – a case study. *International Journal of Geographical Information Scienc*e, 25(1), 65-87. http://dx.doi.org/10.1080/13658811003752332.

MapBiomas. (2023a). *O Projeto*. https://mapbiomas.org/o-projeto.

MapBiomas. (2023b). *Análise de acurácia*. https://brasil.mapbiomas.org/analise-de-acuracia/.

Microsoft Corporation. (2016). *Microsoft Excel* [Software] (Versão 16.0).

National Aeronautics and Space Administration. (2023) *NASADEM: Creating a New NASA Digital Elevation Model and Associated Products*. https://www.earthdata.nasa.gov/esds/competitive-programs/measures/nasadem.

Pontius, R. G., Huffaker, D., & Denman, K. (2004). Useful techniques of validation for spatially explicit land-change models. *Ecological Modelling*, 179(4), 445-461. http://dx.doi.org/10.1016/j.ecolmodel.2004.05.010.

Pontius, R. G., Krithivasan, R., Sauls, L., Yan, Y., & Zhang, Y. (2017). Methods to summarize change among land categories across time intervals. *Journal Of Land Use Science*, 12(4), 218-230. http://dx.doi.org/10.1080/1747423x.2017.1338768.

Pontius, R. G., Bilintoh, T., Oliveira, G., & Shimbo, J. (2023). *Trajectories of losses and gains of soybean cultivation during multiple time intervals in western Bahia, Brazil*. In: Proc. Space Week Nordeste 2023, Fortaleza, Ceará.

Sangermano, F., Eastman, J. R., & Zhu, H. (2010). Similarity Weighted Instance-based Learning for the Generation of Transition Potentials in Land Use Change Modeling. *Transactions In Gis*, 14(5), 569-580. http://dx.doi.org/10.1111/j.1467-9671.2010.01226.x.

Soares-Filho, B. S., Cerqueira, G. C., & Pennachin, C. L. (2002). Dinamica—a stochastic cellular automata model designed to simulate the landscape dynamics in an Amazonian colonization frontier. *Ecological Modelling*, 154(3), 217-235. http://dx.doi.org/10.1016/s0304-3800(02)00059-5.

Souza Jr., C. M., Shimbo, J. Z., Rosa, M. R., Parente, L. L., Alencar, A. A., Rudorff, B. F. T., Hasenack, H., Matsumoto, M., Ferreira, L. G., Souza-Filho, P. W. M. de, Oliveira, S. W. de, Rocha, W. F., Fonseca, A. V., Marques, C. B., Diniz, C. G., Costa, D., Monteiro, D., Rosa, E. R., Vélez-Martin, E., ... Azevedo, T. (2020). Reconstructing Three Decades of Land Use and Land Cover Changes in Brazilian Biomes with Landsat Archive and Earth Engine. *Remote Sensing*, [S.L.], v. 12, n. 17, p. 2735. http://dx.doi.org/10.3390/rs12172735.

Tang, D., Chen, M., Huang, X., Zhang, G., Zeng, L., Zhang, G., Wu, S., & Wang, Y. (2023). SRplot: a free online platform for data visualization and graphing. *Plos One*, 18(11). http://dx.doi.org/10.1371/journal.pone.0294236.

Trigueiro, W. R., Nabout, J. C., & Tessarolo, G. (2020). Uncovering the spatial variability of recent deforestation drivers in the Brazilian Cerrado. *Journal of Environmental Management*, 275, 111243. http://dx.doi.org/10.1016/j.jenvman.2020.111243.

Turkey, J. (1977). *Exploratory Data Analysis*. Addison Wesley.

Van Vliet, J., Bregt, A. K., Brown, D. G., Van Delden, H., Heckbert, S., & Verburg, P. H. (2016). A review of current calibration and validation practices in land-change modeling. *Environmental Modelling & Software*, 82, 174-182. http://dx.doi.org/10.1016/j.envsoft.2016.04.017.

Voight, C., Hernandez-Aguilar, K., Garcia, C., & Gutierrez, S. (2019). Predictive Modeling of Future Forest Cover Change Patterns in Southern Belize. *Remote Sensing*, 11(7), 823. http://dx.doi.org/10.3390/rs11070823.

Supplementary materials

The computational codes used in this article can be accessed through the link: https://github.com/macleidivarnier/Spatial-Variables-and-Land-Use-Change-Models-A-Study-on-Conditioning-Patterns

Acknowledgments

The authors are grateful to professors Guilherme Garcia de Oliveira, Heinrich Hasenack, and Laurindo Antonio Guasselli; the two anonymous reviewers; and the editor Rogério Galante Negri for their comments that contributed to the improvement of the article.

Authors Contribution

Conceptualization, M.V. and E.J.W.; methodology, M.V. and E.J.W.; software, M.V.; validation, M.V.; formal analysis, M.V.; investigation, M.V.; data curation, M.V.; writing—original draft preparation, M.V.; writing—review and editing, M.V. and E.J.W.; visualization, M.V.; supervision, E.J.W. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest

The authors declare no conflicts of interest.

Biography of the main author

Macleidi Varnier is a PhD student in the Graduate Program in Remote Sensing at the Federal University of Rio Grande do Sul (UFRGS). Varnier does research in Remote Sensing, Geographic Information Systems (GIS), Land Use Changes, Spatial Modeling, and Accuracy Assessment.