Universidade Federal Rural do Semi-Árido
Pró-Reitoria de Pesquisa e Pós-Graduação
https://periodicos.ufersa.edu.br/index.php/caatinga
ISSN 1983-2125 (online)

http://dx.doi.org/10.1590/1983-21252026v3915548rc

Scientific Article

Spatiotemporal variation of deforestation in the Piauí Caatinga under a panel and machine learning approach

Variação espaço-temporal do desmatamento na Caatinga piauiense sob abordagem de painel e aprendizado de máquina

Eduardo Ganassoli[1]* , Aldemir L. Junior[1] , Igor E. S. Machado[2] , Jader N. Cachoeira[3] , Olavo da C. Leite[4] , Marcos Giongo[1]

[1]Graduate Program in Forest and Environmental Sciences, Universidade Federal do Tocantins, Gurupi, TO, Brazil. [2]Graduate Program in Plant Production, Universidade Federal do Tocantins, Gurupi, TO, Brazil. [3]Center for Environmental Monitoring and Fire Management, Universidade Federal do Tocantins, Gurupi, TO, Brazil. [4]Secretaria da Educação do Estado do Tocantins, Gurupi, TO, Brazil.

ABSTRACT - The Piauí Caatinga has shown a marked rise in deforested area in recent years, but the factors associated with this dynamic remain poorly investigated under a spatiotemporal approach at the municipal scale. This study identified the factors associated with the spatiotemporal variation of deforestation in the biome across 140 municipalities of Piauí state between 2017 and 2024, integrating fixed-effects panel regression (FE), a spatial autoregressive model (SAR), and Random Forest (RF), based on deforestation data (PRODES/INPE), land use (MapBiomas), socioeconomic indicators (IBGE), rural credit (SICOR/BACEN), and protected areas (CNUC/MMA). The time series revealed a 105.3% increase in mean annual deforested area from 2021 onward, without statistically significant confirmation of either a trend (Mann-Kendall) or a change point (Pettitt) at the 5% level. Moran's index confirmed significant spatial dependence, intensified in the post-2021 period. The FE model identified agricultural land-use mosaics as the strongest correlate of intramunicipal temporal variation ($\beta = 1.032$; $p < 0.001$), a result potentially confounded by overlap with the PRODES response variable, whereas the SAR revealed significant spatial dependence ($Rho = 0.294$; $p < 0.001$) and associations of productive-pressure and infrastructure variables not captured by the FE. Random Forest, under municipality-grouped cross-validation, explained 50.98% of the variance in deforestation, identifying pasture, cattle herd size, distance to hydrography, and protected area coverage as additional structural predictors. The results evidence methodological complementarity among the three approaches and point to land-use conversion as the main correlate of deforestation in the biome, supporting the territorial prioritization of environmental command-and-control actions.

Keywords: Spatial autocorrelation. Fixed effects. MapBiomas. PRODES. Random Forest.

RESUMO - A Caatinga piauiense apresentou aumento expressivo do desmatamento nos últimos anos, mas seus fatores associados permanecem pouco investigados sob abordagem espaço-temporal na escala municipal. Este estudo identificou os fatores associados à variação espaço-temporal do desmatamento no bioma em 140 municípios do Piauí entre 2017 e 2024, integrando regressão em painel com efeitos fixos (EF), um modelo autorregressivo espacial (SAR) e Random Forest (RF), com dados de desmatamento (PRODES/INPE), uso do solo (MapBiomas), indicadores socioeconômicos (IBGE), crédito rural (SICOR/BACEN) e áreas protegidas (CNUC/MMA). A série temporal revelou aumento de 105,3% na área média anual desmatada a partir de 2021, sem significância estatística a 5% para tendência (Mann-Kendall) ou ponto de mudança (Pettitt). O índice de Moran confirmou dependência espacial significativa, intensificada após 2021. O modelo de EF identificou os mosaicos de uso agrícola do solo como o principal correlato da variação temporal intramunicipal ($\beta = 1,032$; $p < 0,001$), resultado potencialmente confundido pela sobreposição com a variável resposta, enquanto o SAR revelou dependência espacial significativa ($Rho = 0,294$; $p < 0,001$) e associações de pressão produtiva e infraestrutura não capturadas pelo EF. O Random Forest, sob validação cruzada agrupada por município, explicou 50,98% da variância do desmatamento, identificando pastagem, tamanho do rebanho bovino, distância à hidrografia e cobertura de áreas protegidas como preditores estruturais adicionais. Os resultados evidenciam complementaridade metodológica entre as três abordagens e apontam a conversão do uso do solo como principal correlato do desmatamento no bioma, subsidiando a priorização territorial de ações de comando e controle ambiental.

Palavras-chave: Autocorrelação espacial. Efeitos fixos. MapBiomas. PRODES. Random Forest.

Conflict of interest: The authors declare no conflict of interest related to the publication of this manuscript.

This work is licensed under a Creative Commons Attribution-CC-BY https://creativecommons.org/licenses/by/4.0/

Received for publication in: July 21, 2026.
Accepted in: September 29, 2026.

Editor in Chief: Aurélio Paes Barros Júnior
Section Editor: Antonio Policarpo Souza Carneiro

Data Availability: The data that support the findings of this study can be made available, upon reasonable request, from the corresponding author.

***Corresponding author**: <engf.edu@gmail.com>

INTRODUCTION

The Caatinga is the only exclusively Brazilian biome, occupying approximately 10% of the national territory, distributed across nine states of the Northeast region and northern Minas Gerais. Despite harboring high biological diversity and a significant degree of endemism, it remains among the least studied and most vulnerable biomes in the country, with only about 10% of its area covered by protected areas (FRANÇA ROCHA et al., 2024).

Native vegetation loss in the biome already reaches 42.6% of its original area, a process historically associated with extensive cattle ranching and agricultural expansion (REFATI et al., 2023), and more recently with the installation of wind and solar power facilities (BRASIL, 2024; TURKOVSKA et al., 2021; KLINGLER et al., 2024). In Piauí state, this process has intensified: while other Brazilian biomes recorded a marked reduction in deforestation in 2024, the Caatinga maintained an upward trajectory, with proportionally larger increases concentrated precisely in Piauí territory.

Recent studies have advanced the identification of deforestation drivers in Brazilian agricultural frontiers, highlighting the role of rural credit, livestock farming, and cropland expansion as direct and indirect determinants of native

vegetation removal (KUSCHNIG et al., 2021; MARCA; ALVIM, 2025).

However, most of the literature on deforestation determinants focuses on the Amazon and the Cerrado, and studies systematically analyzing the spatiotemporal structure of the phenomenon in the Caatinga using municipal panel data remain scarce (FRANÇA ROCHA et al., 2024). Furthermore, a substantial share of this literature relies on parametric regression models that, although interpretable, impose restrictions on capturing nonlinear relationships and structural predictors with low temporal variation, a limitation that machine learning techniques, such as Random Forest, have shown capacity to overcome in recent studies on Amazonian deforestation (DIAS et al., 2024).

Likewise, traditional deforestation models that ignore spatial dependence among neighboring municipalities are subject to estimation bias. This occurs because spillover effects, the propagation of the phenomenon beyond the administrative boundaries of the originating municipality, among adjacent units, can mask or distort the true magnitude of the identified determinants (KUSCHNIG et al., 2021; MARCA; ALVIM, 2025).

Given this scenario, the present study aimed to identify the factors associated with the spatiotemporal variation of deforestation in the Piauí Caatinga between 2017 and 2024, integrating municipal fixed-effects panel regression (FE), a spatial autoregressive model (SAR), and Random Forest (RF). This multi-methodological approach seeks not only to identify the determinants of temporal deforestation variation within each municipality, but also to examine the spatial dependence of the phenomenon among neighboring municipalities and to reveal structural predictors, such as pasture and cattle herd size, whose low intramunicipal temporal variation limits their detection by conventional fixed-effects models.

MATERIAL AND METHODS

The study was conducted in the portion of the Caatinga biome located within Piauí state, corresponding to approximately 40% of the state's area (Figure 1). The spatial unit of analysis was the municipality: of the 141 municipalities overlapping the biome, 140 comprised the data panel. Vector layers were reprojected to SIRGAS 2000/Brazil Polyconic (EPSG:5880).

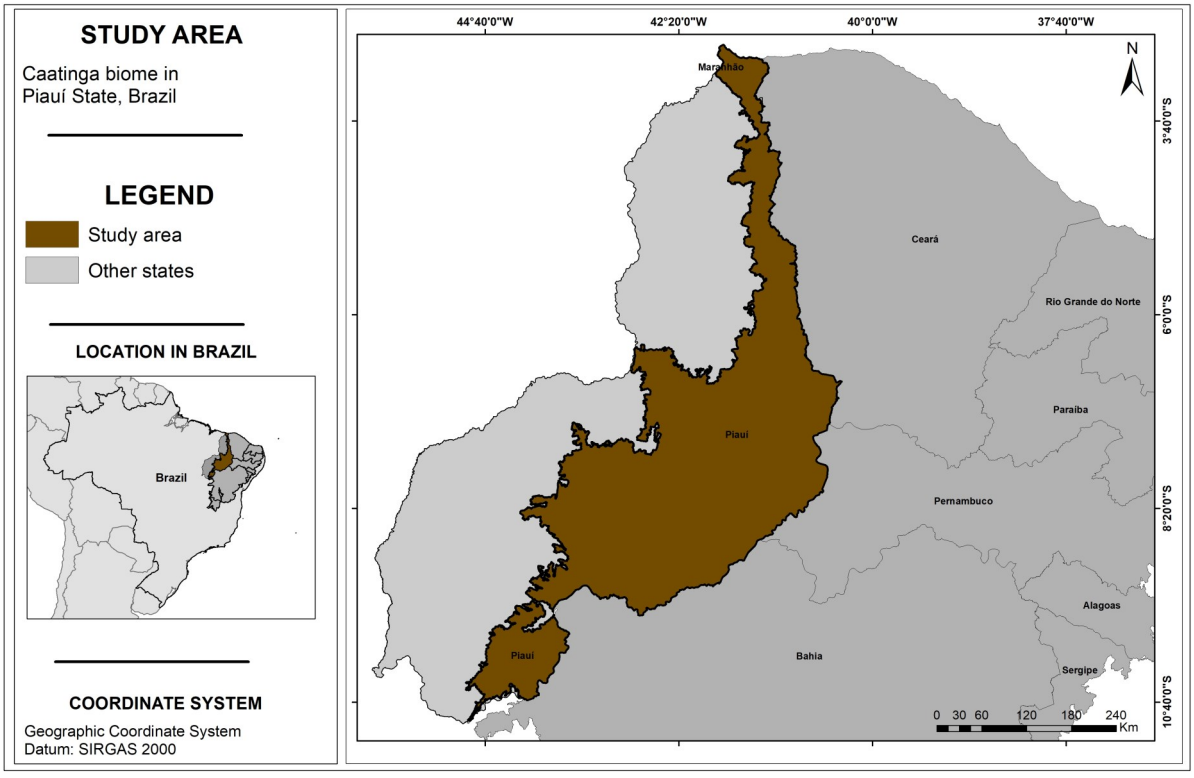


Figure 1. Location of the study area, Caatinga biome in Piauí state, Brazil. Source: IBGE (2026).

The dependent variable, annual deforested area (ha) per municipality within the Caatinga biome, was obtained from PRODES/INPE alerts (2017–2024), totaling 54,132 polygons and 211,746.90 ha, monitored by the Landsat 8, Landsat 9, Sentinel-2, and Sentinel-2B satellites. PRODES alerts were spatially aggregated by municipality and year, with a zero value assigned to municipality-years with no occurrence, producing a balanced panel of 140 municipalities between 2017 and 2024.

The explanatory variables were organized into four thematic groups, corresponding to one control set and three working hypotheses (Figure 2).

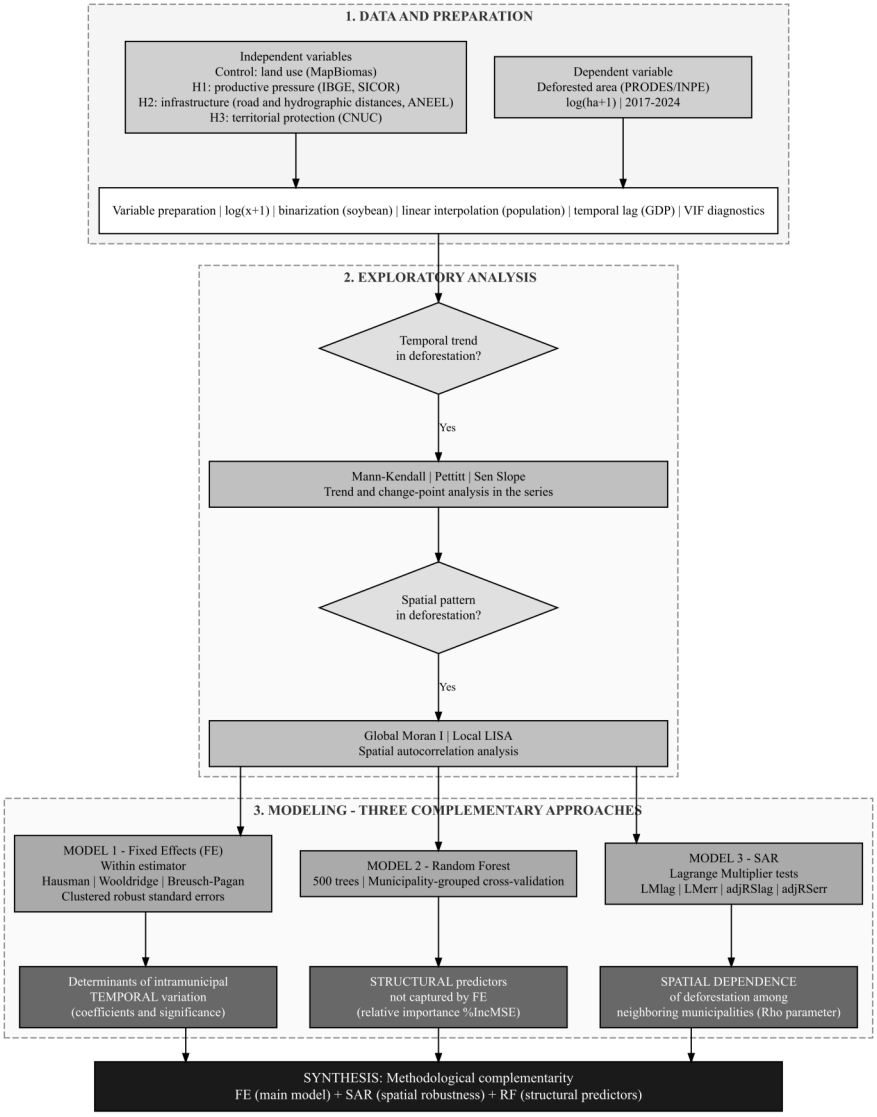


Figure 2. Methodological flowchart of the study: variable preparation, exploratory analysis (temporal and spatial), and modeling under three complementary approaches. FE = fixed effects; RF = Random Forest; SAR = spatial autoregressive model; %IncMSE = percentage increase in mean squared error; Rho = spatial lag parameter; LMlag/LMerr/adjRSlag/adjRSerr = Lagrange Multiplier tests (simple and robust) for selection between SAR and SEM.

Land use and land cover, treated as independent variables, were obtained from MapBiomas Collection 10 (30 m spatial resolution): native vegetation (sum of the forest, savanna, and grassland formation classes), pasture, land-use mosaic, and soybean, the last of which binarized due to zero-inflation in 91 to 94% of observations.

Agricultural production pressure (H1) was represented by cattle and goat herd size (IBGE – Municipal Livestock Survey), cropland, maize, bean, and soybean area (IBGE – Municipal Agricultural Production), total rural credit, PRONAF credit, and other program credit (SICOR/BACEN), and municipal GDP per capita, lagged by one period (t−1, IBGE).

Infrastructure and territorial access (H2) were measured by the mean distance of PRODES alerts to roads (IBGE, via the MapBiomas Infrastructure Module, Collection 7, MAPBIOMAS, 2022a; 2,734 segments) and to hydrography features (IBGE database; 917 features), both calculated per individual alert and aggregated by municipality and year, and by the number of wind and photovoltaic power plants (ANEEL, 2026).

Territorial protection (H3) was represented by protected area coverage, disaggregated into Strict Protection and Sustainable Use categories. Eight federal and state protected areas overlap with 47 municipalities in the study area, totaling 1,633,869.1 ha; as these variables showed no relevant temporal variation over the period, they were treated as static. Indigenous land coverage, restricted to a single, small unit within a single municipality, was excluded from the modeling due to insufficient spatial variation for statistical purposes.

Continuous variables were log(x+1)-transformed. Gaps in municipal population data (2022–2023) were filled by linear interpolation between 2021 and 2024. Multicollinearity diagnosis (VIF) resulted in the exclusion of total cropland area (VIF = 9.03) and pasture (r = 0.84 with land-use mosaic) from

the fixed-effects model; Random Forest, robust to collinearity, retained all 19 original variables.

This difference in the treatment of collinearity between the two models reflects distinct objectives: FE prioritizes the absence of multicollinearity to ensure coefficient interpretability for intramunicipal temporal variation, at the cost of excluding variables such as pasture; RF, robust to collinearity by design, retains this variable and reveals its relevance in the between-municipality comparison, additional evidence of the complementarity between the two approaches.

Two municipalities, Cristalândia do Piauí and Redenção do Gurguéia, recorded no PRODES alerts over the period, which precluded the calculation of the distance variables (H2) and resulted in their exclusion from the panel estimation. The fixed-effects model was thus estimated with 138 municipalities (n = 938; 2018–2024), after the additional loss of the year 2017 due to the GDP per capita lag.

The monotonic trend of the deforestation series was tested using the non-parametric Mann-Kendall method (MANN, 1945; KENDALL, 1975), and its magnitude was estimated using Sen's slope estimator (SEN, 1968; GILBERT, 1987), robust to outliers by using the median of the slopes between pairs of points in the series. A potential structural break was identified using Pettitt's test (PETTITT, 1979), which locates the change point that maximizes the statistic

$$U^* = max_t |U_t|$$

without assuming a parametric data distribution. All three tests were performed using the trend package in R software.

The spatial dependence of deforestation among municipalities was assessed using the global Moran's I index (MORAN, 1950) and the local LISA statistic (ANSELIN, 1995), separately for the pre-break (2017–2021) and post-break (2022–2024) subperiods, with a first-order queen contiguity matrix built from the municipal shapefile. The analyses were performed using the libpysal and esda packages in Python.

The panel model was estimated using municipal fixed effects (within estimator). The fixed-effects specification was selected based on the Hausman test ($p < 0.001$), which rejected the hypothesis that unobserved municipal characteristics, such as soil type, relief, or production history, are independent of the model's explanatory variables. Given this dependence, the random-effects estimator would be inconsistent, making fixed effects the methodologically appropriate choice.

Municipality-clustered robust standard errors were adopted due to heteroskedasticity detected by the Breusch-Pagan test ($p < 0.001$); serial autocorrelation, on the other hand, was not identified (Wooldridge test, $p = 0.885$). The model followed the specification:

$$log(desmat_ha+1)_{it} = \alpha_i + \beta X_{it} + \varepsilon_{it}$$

where deforestation in municipality i in year t is explained by three components: α_i, a fixed term that absorbs municipality-specific characteristics that do not vary over the analyzed period (relief, soil type, production history, among others); X_{it}, the set of explanatory variables that vary year by year for each municipality (rural credit, herd size, land use, among others); and ε_{it}, the error term, corresponding to the variation not explained by the model.

By isolating the fixed component of each municipality, the within estimator compares each unit against its own mean over time, eliminating omitted-variable bias arising from comparisons between structurally distinct municipalities. The model was estimated using the plm package (CROISSANT; MILLO, 2008).

The within estimator absorbs, by design, any variable with low intramunicipal temporal variation, even if structurally relevant to explaining differences in deforestation among municipalities. To identify these structural predictors, a Random Forest model was fitted (500 trees, a number sufficient to stabilize variable importance estimates without excessive computational cost; mtry = 4; k = 5 cross-validation), with the importance of each variable measured by the percentage increase in mean squared error after permutation (%IncMSE), using the randomForest package (LIAW; WIENER, 2002). The model's explanatory power is presented in the Results section.

The detection of spatial dependence in the residuals of the fixed-effects model evidenced a violation of the observation-independence assumption, requiring a spatial econometrics approach. The application of robust Lagrange Multiplier tests indicated a theoretical and statistical preference for the spatial autoregressive model (SAR) over the spatial error model (SEM), since only the spatial lag term reached statistical significance at the 5% level (adjRSlag = 5.91, $p = 0.015$; adjRSerr = 3.52, $p = 0.061$).

To capture these long-term determinants, the SAR model was estimated by maximum likelihood, structured on the cross-section of municipal temporal means. Territorial connectivity was modeled using a queen contiguity matrix, implemented through the spatialreg package (BIVAND; PIRAS, 2021). The autoregressive coefficient Rho thus measures the magnitude of spillover effects of deforestation among neighboring municipalities.

RESULTS AND DISCUSSION

Between 2017 and 2024, annual deforested area in the Piauí Caatinga ranged from 16,230.22 ha (2021, series minimum) to 52,517.68 ha (2024, series maximum), with a mean of 26,468.36 ha year^{-1} and a standard deviation of 13,180.57 ha. The coefficient of variation (49.8%) indicates high interannual dispersion, reflecting the level shift around the most likely change point identified by Pettitt's test, not statistically significant at the 5% level (see below).

The Mann-Kendall test did not identify a statistically significant trend ($S = 1{,}856$; $\tau = 0.571$; $p = 0.064$). This lack of formal significance may partly reflect the small number of annual observations (n = 8), a condition known to limit the statistical power of non-parametric tests, although the absence of a genuine monotonic trend cannot be ruled out. The magnitude of the trend, estimated by Sen's method, was +3,886.34 ha year^{-1} (95% CI: −195.90 to 8,117.52 ha year^{-1}), consistent with the positive, though not statistically significant, direction of the Mann-Kendall test (Figure 3).

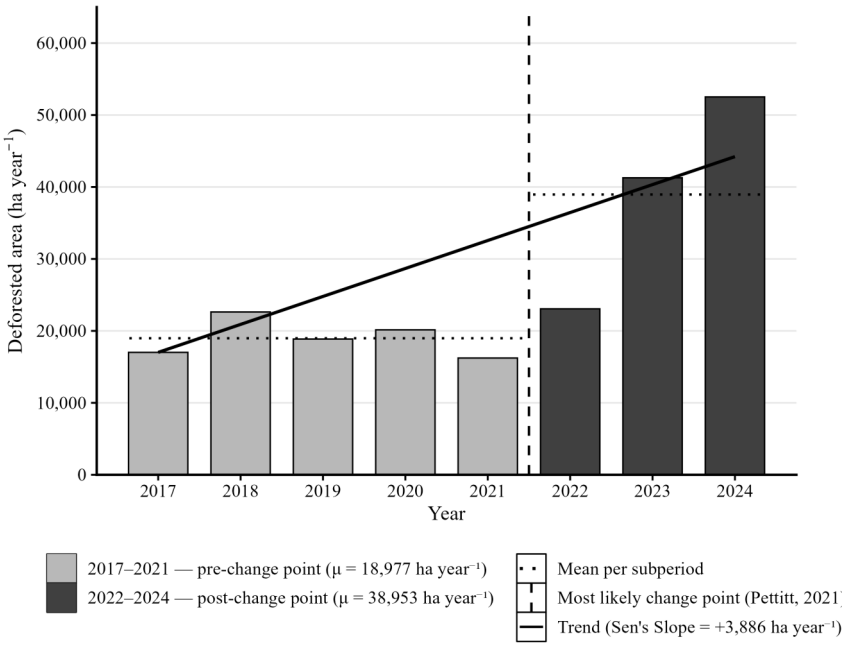


Figure 3. Annual deforested area in the Caatinga biome of Piauí state, Brazil (2017–2024). Light gray bars = period preceding the most likely change point (2017–2021); dark gray bars = subsequent period (2022–2024). Dotted line = mean deforested area per subperiod. Vertical dashed line = most likely change point identified by Pettitt's test (2021; U = 15; p = 0.192, not significant at the 5% level). Solid line = temporal trend estimated by Sen's slope (+3,886 ha year⁻¹; p = 0.064). Deforestation data source: PRODES/INPE.

This growth pattern is consistent with the broader dynamics of native vegetation conversion in the biome. França Rocha et al. (2024), when analyzing three decades of land use and land cover in the Caatinga (1985–2019) using MapBiomas data, identified continuous conversion of native formations into pasture and agricultural land-use mosaics as the main vector of landscape transformation in the biome, a pattern with which the increase observed in the present study is consistent, updated for the Piauí subset in the most recent period. National monitoring data further report an 88.9% increase in Caatinga deforestation between 2020 and 2021 (MAPBIOMAS, 2022b), one of the largest proportional increases among Brazilian biomes that year, although not directly comparable to the results of this study due to differences in scale and reference period.

Pettitt's test did not detect a statistically significant break (U = 15; p = 0.192). Nonetheless, a marked increase in the annual mean was observed: from 18,977.34 ha year^{-1} (2017–2021) to 38,953.40 ha year^{-1} (2022–2024), a +105.3% increase. This level shift, even without formal statistical confirmation by the break test, was substantially relevant enough to methodologically justify the adoption of the fixed-effects panel approach, capable of isolating intramunicipal temporal variation from time-invariant structural characteristics, a control unattainable through a single cross-sectional analysis.

Spatial dependence of deforestation was confirmed in both subperiods, with intensified spatial autocorrelation between the pre-break (I = 0.1193; p = 0.020) and post-break (I = 0.1802; p = 0.007) periods. This increase is consistent with the spatial dependence parameter estimated by the SAR model (Rho = 0.294; p < 0.001) and suggests that the intensification of deforestation after 2021 did not occur spatially at random, but was accompanied by greater territorial clustering of the phenomenon.

The local analysis (LISA) identified 8 municipalities as High-High (HH) in the pre-break period and 9 in the post-break period, with 7 municipalities, Colônia do Gurguéia, Manoel Emídio, Queimada Nova, Caridade do Piauí, Curimatá, Júlio Borges, and Acauã, remaining as persistent hotspots of elevated deforestation in both periods. Paulistana and Betânia do Piauí emerged as new HH clusters only in the post-break period, indicating a possible expansion of the deforestation frontier into municipalities previously not identified as critical. Curral Novo do Piauí, in turn, no longer appeared as an HH cluster in the most recent period.

Rural credit directed to municipalities in the Piauí Caatinga grew markedly between 2017 and 2024, rising from R$387.9 million to R$1,558.6 million, a 302% increase over the period (Figure 4a). PRONAF credit, specifically, more than tripled over the same interval (from R$237.9 million to R$835.7 million), with a more pronounced acceleration from 2022 onward.

The increase in rural credit disbursements coincides with the rise in deforestation observed in the time series and is consistent with the hypothesis that the effect of agricultural credit on deforestation depends on credit modality and biome. The study by Faria et al. (2025) on the relationship between rural credit policy and deforestation, which examined 5,354 Brazilian municipalities between 2013 and 2019, showed that investment-oriented credit is negatively associated with deforestation, whereas subsidized credit with poorly controlled allocation is positively associated with it, particularly in the Cerrado and the Pantanal. This distinction reinforces the need for caution in treating total rural credit as a variable of only marginal significance in the fixed-effects model, since the aggregated measure (total credit) does not allow discrimination between modalities with potentially opposing effects.

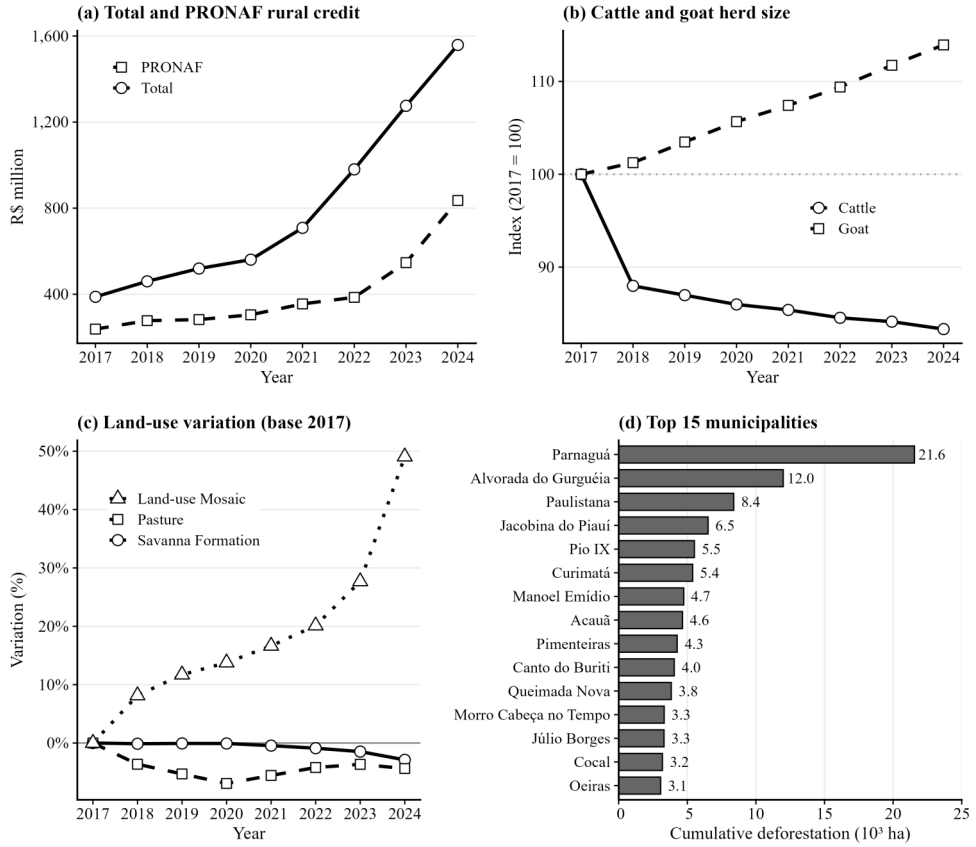


Figure 4. Variables associated with productive pressure and land use in the Piauí Caatinga (2017–2024). (a) Evolution of total rural credit and PRONAF credit (R$ million). (b) Evolution of cattle and goat herd size, as an index number (2017 = 100). (c) Percentage change in area of the main land use and land cover classes (MapBiomas) relative to 2017. (d) The 15 municipalities with the highest cumulative deforestation over the period (10^3 ha). Data sources: SICOR/BACEN (a); IBGE-PPM (b); MapBiomas Collection 10 (c); PRODES/INPE (d).

Cattle herd size declined continuously over the period, decreasing by 16.7% between 2017 and 2024 (index 100 → 83.3), whereas goat herd size grew steadily, with a 13.9% increase over the same interval (index 100 → 113.9) (Figure 4b). This pattern suggests livestock substitution, possibly associated with the greater hardiness of goats under semi-arid edaphoclimatic conditions; this interpretation, however, was not directly tested in this study and should be treated with caution.

Among land use and land cover classes (Figure 4c), the land-use mosaic showed the largest relative expansion, increasing by 49.1% relative to 2017, with a marked acceleration from 2022 onward. Pasture area followed a non-linear trajectory, with an initial decline (−6.9% in 2020) followed by a partial recovery (−4.4% in 2024). Savanna formation remained relatively stable, with a marginal negative variation throughout the period (−2.9% in 2024).

This combination, strongly expanding land-use mosaic alongside a non-linear pasture trajectory, partially diverges from the national pattern described by Caballero et al. (2023), who identified a peak in Brazilian pasture area in 2006, followed by gradual replacement with soybean and other crops. In the Piauí Caatinga, pasture does not show this definitive replacement with cropland, but rather a reconfiguration into a land-use mosaic, a category combining multiple low-technology agricultural activities, suggesting that the conversion frontier in the biome follows a trajectory distinct from that observed in more consolidated agricultural regions, such as the Cerrado.

Cumulative deforestation over the period was concentrated in a small number of municipalities (Figure 4d): Parnaguá led with 21,600 ha, followed by Alvorada do Gurguéia (12,000 ha) and Paulistana (8,400 ha). The 15 municipalities with the highest cumulative deforestation accounted for 89,700 ha, 42.4% of the total deforested area in the biome over the period (211,746.90 ha), confirming the spatial concentration of deforestation described above.

The SAR model revealed statistical significance for several variables that had not reached significance in the fixed-effects model, evidencing that the predominantly spatial variation among municipalities (SAR) and the temporal variation within each municipality (FE) capture distinct mechanisms of association with deforestation (Table 1).

The positive spatial lag coefficient (Rho = 0.294; $p < 0.001$) demonstrates that deforestation exhibits strong spatial dependence among neighboring municipalities. This finding is consistent with Kuschnig et al. (2021) and Santos et al. (2021), who identified municipal spillovers and novel spatial predictors as recent determinants of deforestation in different regions of Brazil, with conversion pressure propagating from neighboring municipalities regardless of each unit's local characteristics.

Table 1. Coefficients estimated by the fixed-effects (FE) and spatial autoregressive (SAR) models for the factors associated with deforestation in the Piauí Caatinga (2018–2024).

Variable	Hypothesis	β (FE)	SE (FE)	Sig. (FE)	β (SAR)	SE (SAR)	Sig. (SAR)
Native vegetation (ln)	Control	-3.688	1.895	.	0.379	0.077	***
Land-use mosaic (ln)	Control	1.032	0.241	***	0.414	0.062	***
Goat herd (ln)	H1	1.278	1.461	n.s.	-0.193	0.038	***
Cattle herd (ln)	H1	0.876	1.688	n.s.	0.211	0.096	*
Soybean presence (binary)	H1	-0.263	0.186	n.s.	-0.058	0.250	n.s.
Total rural credit (ln)	H1	0.153	0.080	.	0.115	0.095	n.s.
GDP per capita t-1 (ln)	H1	0.136	0.152	n.s.	-0.040	0.190	n.s.
Maize area (ln)	H1	0.127	0.078	n.s.	0.255	0.074	***
Bean area (ln)	H1	-0.020	0.112	n.s.	0.026	0.067	n.s.
Distance to hydrography (ln)	H2	0.119	0.107	n.s.	-0.215	0.048	***
Renewable power plants (n)	H2	-0.020	0.044	n.s.	0.040	0.022	.
Distance to roads (ln)	H2	0.018	0.089	n.s.	0.207	0.067	**
Strict Protection area (ln)	H3	—	—	—	-0.033	0.017	.
Sustainable Use area (ln)	H3	—	—	—	0.034	0.011	**
Rho — Spatial dependence	—	—	—	—	0.294	0.064	***

Note: FE: within estimator with municipality-clustered robust standard errors (n = 938; 138 municipalities; 2018–2024). SAR: spatial autoregressive model by maximum likelihood, cross-section of municipal means (n = 138); queen contiguity matrix; AIC = 229.1. β: coefficient; SE: standard error; Rho: spatial lag parameter. Sig.: *** p < 0.001; ** p < 0.01; * p < 0.05; . p < 0.10; n.s. = not significant.

Among land use and land cover variables, both native vegetation and the land-use mosaic maintained high significance in the SAR model (β = 0.379, p < 0.001; and β = 0.414, p < 0.001, respectively). Notably, however, native vegetation showed a sign reversal between the two models: negative and marginal in FE (β = −3.688; p < 0.10) and positive and significant in SAR (β = 0.379; p < 0.001).

This divergence directly illustrates the methodological complementarity sought in this study: FE captures the short-term effect, within each municipality, whereby the reduction of remaining native vegetation accompanies deforestation's own advance; SAR, by comparing means across municipalities, reveals that those with larger native vegetation stocks also exhibit greater absolute deforestation, reflecting the availability of area still subject to conversion, rather than a direct causal protective relationship.

Among agricultural production pressure variables, maize area (β = 0.255; p < 0.001), cattle herd (β = 0.211; p < 0.05), and goat herd (β = −0.193; p < 0.001) showed significant associations with deforestation in the SAR model, capturing predominantly structural, between-municipality variation, with no corresponding significance in the FE model, which captures intramunicipal temporal variation; the goat herd coefficient, notably, showed an opposite sign to the non-significant FE coefficient (β = 1.278). Total rural credit, soybean presence, lagged GDP per capita, and bean area remained non-significant in both models.

Among infrastructure and access variables, distance to roads (β = 0.207; p < 0.01) and distance to hydrography (β = −0.215; p < 0.001) showed significant associations with deforestation in the SAR model, with no equivalent correspondence in FE; the number of renewable power plants showed only marginal significance (β = 0.040; p < 0.10).

The positive sign of distance to roads in the SAR model is consistent with Trigueiro, Nabout, and Tessarolo (2020), who found a spatially variable effect of this variable in the Brazilian Cerrado, with a positive sign predominating precisely in the Northeast and North regions, a result the authors attribute to the greater relevance of rural credit, rather than road proximity per se, as the driving force of deforestation in these regions.

The disaggregation of territorial protection by category revealed opposing effects between the two governance regimes, both estimable only in the SAR model, as they were omitted from FE due to time invariance. Strict Protection area was negatively associated with deforestation, with marginal significance (β = −0.033; p < 0.10), while Sustainable Use area was positively associated, with statistical significance (β = 0.034; p < 0.01).

This result diverges from the national pattern described by Gonçalves-Souza et al. (2021), who found no significant difference in effectiveness between Strict Protection and Sustainable Use across Brazilian biomes as a whole. The disaggregated specification also showed superior fit compared to the aggregated protected-area specification (AIC = 229.1 versus 237.4), confirming that the aggregated coefficient masked opposing effects between the two protection categories.

The Random Forest model (500 trees; mtry = 4; n = 938) explained 65.45% of the variance in deforestation under random cross-validation (R^2 = 0.6545). Because the panel structure implies multiple observations of the same municipality over time, this validation may overestimate the model's predictive capacity due to dependence between years within the same geographic unit (PLOTON et al., 2020). Under municipality-grouped cross-validation (grouped k-fold, k = 5), in which all years of a given municipality remain in the same fold, explanatory power dropped to R^2 = 0.5098, a more conservative estimate of the model's generalization capacity to unobserved municipalities.

This difference relative to the fixed-effects model (R^2 within = 0.231) is expected and does not indicate the superiority of one method over the other. RF captures non-linear relationships and structural effects among municipalities, with predominantly cross-sectional variation between units, whereas FE isolates only the temporal variation within each municipality, being more conservative by methodological design (Table 2). This pattern of greater explanatory power of RF relative to linear approaches has already been documented by Zanella et al. (2017) for the Atlantic Forest in Minas Gerais, where RF substantially outperformed stepwise multiple regression in explaining deforestation and forest fragmentation metrics.

Table 2. Predictive importance of variables in Random Forest (%IncMSE) and comparison with estimates from the municipal fixed-effects model (FE). Dependent variable: log(deforested area + 1). Period: 2018–2024.

RF Rank	Variable	Hypothesis	%IncMSE (RF)	β (FE)	Robust SE	Sig.	Convergence
1	Land-use mosaic (ln)	Control	34.367	1.032	0.241	***	Convergent
2	Native vegetation (ln)	Control	28.417	-3.688	1.895	.	Convergent
3	Pasture (ln)	Control	25.726	—	—	—	RF only
4	Total rural credit (ln)	H1	25.69	0.153	0.080	.	Convergent
5	Cattle herd (ln)	H1	25.21	0.876	1.688	n.s.	RF > FE*
6	Distance to hydrography (ln)	H2	24.352	0.119	0.107	n.s.	RF > FE*
7	Bean area (ln)	H1	24.247	-0.020	0.112	n.s.	RF > FE*
8	PRONAF credit (ln)	H1	24.192	—	—	—	RF only
9	Goat herd (ln)	H1	24.031	1.278	1.461	n.s.	RF > FE*
10	Total planted area (ln)	H1	21.438	—	—	—	RF only
11	Maize area (ln)	H1	21.127	0.127	0.078	n.s.	Low importance
12	GDP per capita t-1 (ln)	H1	18.895	0.136	0.152	n.s.	Low importance
13	Total protected area (ln)	H3	18.069	—	—	—	RF > FE†
14	Distance to roads (ln)	H2	17.667	0.018	0.089	n.s.	Low importance
15	Other credit programs (ln)	H1	16.978	—	—	—	RF only
16	Sustainable Use area (ln)	H3	14.751	—	—	—	RF only
17	Renewable power plants (n)	H2	9.235	-0.020	0.044	n.s.	Low importance
18	Strict Protection area (ln)	H3	7.679	—	—	—	RF only
19	Soybean area (ha)	H1	5.516	—	—	—	RF only

Notes: %IncMSE = percentage increase in mean squared error after variable permutation (Random Forest, 500 trees, k=5 cross-validation, R^2=0.6545 under random CV and R^2=0.5098 under municipality-grouped CV). β = coefficient estimated by the within estimator with municipality-clustered robust standard errors. Sig.: *** p<0.001; ** p<0.01; * p<0.05; p<0.10; n.s. not significant; — variable not included in the FE model. Convergent: both methods identify the variable as relevant. RF > FE: high predictive importance in RF, but non-significant coefficient in FE, indicating a structural determinant with low intragroup temporal variation. RF > FE†: variable included in the FE model but dropped by the within estimator due to time invariance (protected area is nearly constant across years); RF captures the structural effect that FE cannot estimate. RF only: variable not included in the FE model.

The three variables with the highest predictive importance in RF, land-use mosaic (%IncMSE = 34.37), native vegetation (%IncMSE = 28.42), and pasture (% IncMSE = 25.73), belong to the land use and land cover dimension, reinforcing that landscape conversion, rather than isolated socioeconomic factors, is the most robust correlate of deforestation in the region. Of these, land-use mosaic and native vegetation were classified as convergent, identified as relevant by both RF and FE ($\beta = 1.032$, $p < 0.001$; and $\beta = -3.688$, $p < 0.10$, respectively), lending additional robustness to this finding through two methods with distinct assumptions.

It should be noted that MapBiomas land use and land cover classes, including land-use mosaic and pasture, are assigned annually through image interpretation, such that a recently deforested pixel tends to be reclassified into one of these categories within the same year. Part of the association found between these variables and deforestation may thus reflect operational overlap between predictor and response, rather than exclusively a determinative relationship.

Four variables, cattle herd, distance to hydrography, bean area, and goat herd, showed high predictive importance in RF (%IncMSE between 24.0 and 25.2), yet non-significant coefficients in FE. This divergence suggests that these variables function as structural predictors of deforestation, relevant in the between-municipality comparison, but with insufficient temporal variation within each municipality to be captured by the within estimator.

The same pattern explains protected area coverage, included in the FE model but dropped by the estimator due to time invariance. Nonetheless, this variable showed moderate importance in RF (%IncMSE = 18.07; rank 13), indicating that RF was able to capture a structural effect inaccessible to FE by design.

Seven variables were identified only by RF, as they

were not included in the FE model: pasture, PRONAF credit, total planted area, other credit programs, Sustainable Use protected area, Strict Protection area, and soybean area (in hectares, not binary). Among these, pasture stood out with the third-highest overall predictive importance (%IncMSE = 25.73), pointing to a land-use dimension not fully captured in the original FE specification.

Maize, lagged GDP per capita, distance to roads, and renewable power plants showed low predictive importance in both methods (%IncMSE between 9.24 and 21.13; non-significant FE coefficients), suggesting a secondary role for these factors in explaining deforestation in the Piauí Caatinga within the set of predictors evaluated.

The partial convergence between RF and linear models, relevant for some variables and divergent for others, is consistent with what Dias et al. (2024) observed for the Brazilian Amazon, whose Random Forest model with a 1999–2020 panel explicitly highlighted the limitations of linear approaches in capturing structural predictors of deforestation, such as permanent cropland area and public spending.

Recent literature on deforestation determinants in Brazilian biomes has explored these three approaches either in isolation or combined in pairs: Zanella et al. (2017) and Dias et al. (2024) compared Random Forest to linear models in the Atlantic Forest and the Amazon, respectively; Kuschnig et al. (2021) applied spatial modeling to deforestation in Mato Grosso; and Espindola et al. (2021) combined linear and spatial regression in Piauí state itself. The simultaneous integration of all three approaches remains little explored for Brazilian tropical dry forests, positioning the joint application of FE, SAR, and RF adopted in this study as a relevant methodological contribution to understanding deforestation determinants in these biomes.

A methodological limitation of this study concerns the construction of the distance-to-roads and distance-to-hydrography variables, calculated as the weighted mean distance of the PRODES deforestation alerts themselves to the nearest infrastructure element, which makes them dependent on the prior existence of alerts rather than an independent measure of territorial accessibility. This dependence also explains the exclusion of two municipalities and the loss of specific observations in another eight, for which no alert was recorded in certain years. Future studies should reconstruct this variable using municipal metrics independent of deforestation occurrence, such as road density, road network length per km², or the proportion of territory at different distances from infrastructure, allowing for a more robust causal interpretation of territorial accessibility on deforestation.

CONCLUSION

Deforestation in the Piauí Caatinga showed a 105.3% increase in mean annual deforested area from 2021 onward, without statistically significant confirmation of either a trend (Mann-Kendall) or a change point (Pettitt) at the 5% level. This pattern accompanied the expansion of rural credit and goat herd size, the decline of cattle herd size, and the conversion of native vegetation and pasture into agricultural land-use mosaics, with spatial concentration of cumulative deforestation in a small number of municipalities.

The fixed-effects model identified the land-use mosaic as the main correlate of intramunicipal temporal variation in deforestation ($\beta = 1.032$; $p < 0.001$), a result that should be interpreted with caution given its operational overlap with the PRODES response variable. Total rural credit and remaining native vegetation showed only marginal significance.

The spatial autoregressive model identified significant spatial dependence among neighboring municipalities and significant associations for cattle herd, goat herd, maize area, distance to roads, and distance to hydrography, variables that were non-significant in the fixed-effects model.

Random Forest explained 65.45% of the variance in deforestation under random cross-validation and 50.98% under municipality-grouped cross-validation, identifying pasture, cattle herd, distance to hydrography, and protected area coverage as additional structural predictors.

The local spatial autocorrelation analysis (LISA) identified an intensification of spatial clustering of deforestation between the pre- and post-break subperiods, with seven municipalities persistently classified as High-High in both periods.

The disaggregation of territorial protection by category revealed opposing effects between governance regimes: Strict Protection area was negatively associated with deforestation, while Sustainable Use area was positively associated, a result that diverges from the national pattern of similar effectiveness across categories and points to a dynamic specific to the Piauí Caatinga. Quantifying deforestation by protected-area category remains a complementary research agenda.

These results support the territorial prioritization of environmental command-and-control actions in the Piauí Caatinga, directing attention to municipalities with the highest cumulative deforestation and to the spatial dependence identified among neighboring units.

REFERENCES

ANEEL - Agência Nacional de Energia Elétrica. **Sistema de Informações Georreferenciadas do Setor Elétrico - SIGEL**. Available at: <https://sigel.aneel.gov.br/>. Access on: Feb. 18, 2026.

ANSELIN, L. Local indicators of spatial association - LISA. **Geographical Analysis**, 27: 93-115, 1995.

BIVAND, R.; PIRAS, G. **spatialreg**: spatial regression analysis. R package version 1.2-1, 2021. Available at: <https://CRAN.R-project.org/package=spatialreg>. Access on: Feb. 1, 2026.

BRASIL. Ministério do Meio Ambiente e Mudança do Clima. **Plano de Ação para Prevenção e Controle do Desmatamento e das Queimadas na Caatinga (PPCaatinga)**. Brasília, DF: MMA, 2024. Available at: <https://dados.mma.gov.br/dataset/plano-de-acao-para-a-prevencao-e-controle-do-desmatamento-e-das-queimadas-na-caatinga-2024-2027>. Access on: Mar. 1, 2026.

CABALLERO, C. B. et al. Transformation of Brazil's biomes: the dynamics and fate of agriculture and pasture expansion into native vegetation. **Science of the Total Environment**, 896: 166323, 2023.

CROISSANT, Y.; MILLO, G. Panel data econometrics in R: the plm package. **Journal of Statistical Software**, 27: 1-43, 2008.

DIAS, F. et al. Predicting the pulse of the Amazon: machine learning insights into deforestation dynamics. **Journal of Environmental Management**, 362: 121359, 2024.

ESPINDOLA, G. M. et al. Cropland expansion as a driver of land-use change: the case of Cerrado-Caatinga transition zone in Brazil. **Environment, Development and Sustainability**, 23: 17146-17160, 2021.

FARIA, W. R. et al. The relationship between rural credit policy and deforestation: evidence from Brazil. **Environmental Economics and Policy Studies**, 27: 231-263, 2025.

FRANÇA ROCHA, W. J. S. et al. Towards uncovering three decades of LULC in the Brazilian drylands: Caatinga biome dynamics (1985–2019). **Land**, 13: 1250, 2024.

GILBERT, R. O. **Statistical methods for environmental pollution monitoring**. New York: Van Nostrand Reinhold, 1987. 320 p.

GONÇALVES-SOUZA, D. et al. The role of protected areas in maintaining natural vegetation in Brazil. **Science Advances**, 7: eabh2932, 2021.

IBGE - Instituto Brasileiro de Geografia e Estatística. **Downloads: geociências**. Available at: <https://www.ibge.gov.br/geociencias/downloads-geociencias.html>. Access on: Feb. 1, 2026.

KENDALL, M. G. **Rank correlation methods**. 4. ed. London: Charles Griffin, 1975. 202 p.

KLINGLER, M. et al. Large-scale green grabbing for wind and solar photovoltaic development in Brazil. **Nature Sustainability**, 7: 747-757, 2024.

KUSCHNIG, N. et al. Spatial spillover effects from agriculture drive deforestation in Mato Grosso, Brazil. **Scientific Reports**, 11: 21804, 2021.

LIAW, A.; WIENER, M. Classification and regression by randomForest. **R News**, 2: 18-22, 2002.

MANN, H. B. Nonparametric tests against trend. **Econometrica**, 13: 245-259, 1945.

MAPBIOMAS. **Módulo de Infraestrutura**: Coleção 7**.** São Paulo: MapBiomas, 2022a. Available at: <https://brasil.mapbiomas.org/downloads/>. Access on: Feb. 1, 2026.

MAPBIOMAS. **Relatório Anual do Desmatamento no Brasil**: RAD 2021. São Paulo: MapBiomas, 2022b. 126 p. Available at: <https://alerta.mapbiomas.org/wp-content/uploads/sites/17/2024/03/RAD2021_Completo_FINAL_Rev1.pdf>. Access on: Jul. 18, 2026.

MARCA, L.; ALVIM, A. M. Deforestation in the Brazilian Amazon: an analysis of spatial interactions between economic and demographic factors. **Environment, Development and Sustainability**, s/v.: 1-31, 2025.

MORAN, P. A. P. Notes on continuous stochastic phenomena. **Biometrika**, 37: 17-23, 1950.

PETTITT, A. N. A non-parametric approach to the change-point problem. **Journal of the Royal Statistical Society: Series C (Applied Statistics)**, 28: 126-135, 1979.

PLOTON, P. et al. Spatial validation reveals poor predictive performance of large-scale ecological mapping models. **Nature Communications**, 11: 4540, 2020.

REFATI, D. C. et al. Influence of drought and anthropogenic pressures on land use and land cover change in the Brazilian semiarid region. **Journal of South American Earth Sciences**, 126: 104362, 2023.

SANTOS, A. M. et al. Deforestation drivers in the Brazilian Amazon: assessing new spatial predictors. **Journal of Environmental Management**, 294: 113020, 2021.

SEN, P. K. Estimates of the regression coefficient based on Kendall's tau. **Journal of the American Statistical Association**, 63: 1379-1389, 1968.

TRIGUEIRO, W. R.; NABOUT, J. C.; TESSAROLO, G. Uncovering the spatial variability of recent deforestation drivers in the Brazilian Cerrado. **Journal of Environmental Management**, 275: 111243, 2020.

TURKOVSKA, O. et al. Land-use impacts of Brazilian wind power expansion. **Environmental Research Letters**, 16: 024010, 2021.

ZANELLA, L. et al. How well does random forest analysis model deforestation and forest fragmentation in the Brazilian Atlantic forest? **Environmental and Ecological Statistics**, 24: 529-549, 2017.