nature sustainability

Article

https://doi.org/10.1038/s41893-022-01018-z

Forest conservation in Indigenous territories and protected areas in the Brazilian Amazon

In the format provided by the authors and unedited

Table of Contents

Driving factors for the changes of annual forest area loss rates before and after ITs and PAs establishment

1. Input data for the Generalized Linear Model (GLM)

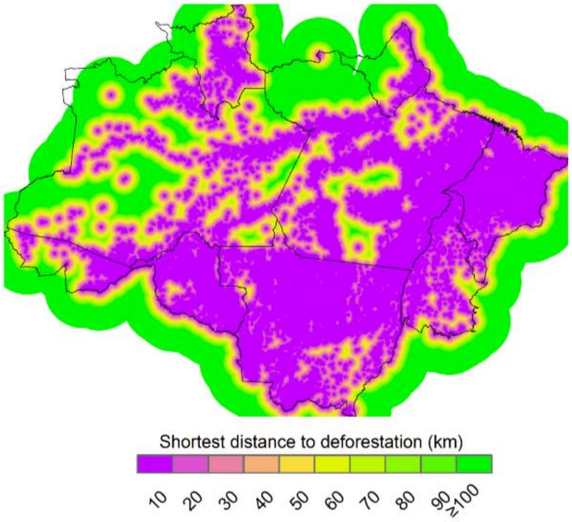
We used the Generalized Linear Model (GLM) to investigate the driving factors for the changes of annual forest area loss rates before and after ITs and PAs establishment. We organized a table that lists individual ITs and PAs with various attributes, including IT or PA name, governance or protected area types (IT, nPAs, sPAs), management objectives (sustainable use, strict protection), state, year of establishment, forest area and forest area fraction in the establishment year, average forest area loss rates before their establishment, changes in annual forest area loss rates before and after their establishment, total human population before their establishment, changes in human population between 2020 and the establishment year, and the average shortest distance to deforestation. This table has categorical variables (e.g., protected area types, management objectives, state), continuous variables (e.g., forest area in the establishment year, changes in population between 2020 and the establishment year). The forest area, forest area fraction, and forest area loss rates are calculated from the forest cover data generated in this study.

1.1. Gridded population density

We collected the Gridded Population of the World (Version 4.11, https://sedac.ciesin.columbia.edu/data/set/gpw-v4-population-density-rev11) at 1-km spatial resolution in 2000, 2005, 2010, 2015, and 2020. We used the population density data close to the establishment year of each ITs and PAs as the population density before ITs and PAs establishment. We used the population density in 2020 as the population density after each ITs and PAs establishment. We then calculated the total population before and after each ITs and PAs establishment by multiplying the total area in each ITs and PAs. We calculated the population changes before and after the ITs and PAs establishment.

1.2. The average shortest distance to deforestation

We calculated the average shortest distance of each ITs and PAs to deforestation boundary in the BLA. First, we calculated the maximum spatial extent of deforestation (gross forest area loss) from 2002 to 2021 and then converted it from GeoTIFF format to shapefile format. We calculated the area for each deforestation polygons. Second, to avoid the impacts of the sparse and random deforestation, we only chose deforestation polygons with more than or equal to the area of nine 500-m MODIS pixels. Third, we applied the Multiple Ring Buffer tool in the ArcGIS (version 10.1) to generate buffers inside and outside (Dissolve_Option = "ALL") deforestation polygons and each buffer had a distance of 10 km. We made sure the buffer rings covering all the ITs and PAs in the BLA. Fourth, we converted the buffer rings (and deforestation shapefile) into a GeoTIFF raster file at the 500-m spatial resolution same as MODIS. Fifth, we overlaid each ITs and PAs with the shortest distance to deforestation raster map (Supplementary Figure 1) and calculated their average shortest distance values to deforestation.



Supplementary Figure 1. The shortest distance to deforestation in the BLA.

1.3. Dummy variables for categorical variables

We converted the categorical variables into dummy variables, including protected area types (IT, nPAs, sPAs), management objectives (sustainable use, strict protection), and state (Supplementary Tables 1-3).

Supplementary Table 1. Dummy variables for governance types (ITs, nPAs, and sPAs)

Types of PAs and ITs	Values		
Variables	National PAs	ITs	State PAs
Dum_Type_1	1	0	0
Dum_Type_2	0	1	0

Supplementary Table 2. Dummy variables for management categories of ITs and PAs

Management categories	Values	
Variable	Strict protection	Sustainable use
Dum_Management	1	0

Supplementary Table 3. Dummy variables for states where ITs and PAs are located

States	Values								
Variables	Amapa	Acre	Roraima	Maranhao	Tocantins	Mato Grosso	Rondonia	Amazonas	Para
Dum_State_ID_1	1	0	0	0	0	0	0	0	0
Dum_State_ID_2	0	1	0	0	0	0	0	0	0
Dum_State_ID_3	0	0	1	0	0	0	0	0	0
Dum_State_ID_4	0	0	0	1	0	0	0	0	0
Dum_State_ID_5	0	0	0	0	1	0	0	0	0
Dum_State_ID_6	0	0	0	0	0	1	0	0	0
Dum_State_ID_7	0	0	0	0	0	0	1	0	0
Dum_State_ID_8	0	0	0	0	0	0	0	1	0

2. GLM models

We built three GLM models to estimate the change (difference) in annual forest area loss rates before and after the establishment of each ITs and PAs in SPSS (version 28.0). We used Linear model, included the predictor variables as main effects, included intercept in the model, and estimated the scale factor using the maximum likelihood method. For the GLM models, the predictor variables include protected area types (IT, nPAs, sPAs), management objectives (sustainable use, strict protection), state, forest area and forest area fraction in the establishment year, average forest area loss rates before the establishment, total population before the establishment, changes in population between 2020 and the establishment year, and the average shortest distance to deforestation.

First, we built a GLM model (Supplementary Table 4) to understand the driving factors for the difference in annual forest area loss rates before and after the establishment for all the 270 ITs/PAs established after 2002 (Two ITs/PAs were excluded as they don't have population data). The GLM model predicted values and our observed values had a significant relationship (P-value < 0.01) but weak exploratory power ($R^2 = 0.16$). The management objectives and total population are significant variables for the difference in annual forest area loss rates before and after the establishment. However, this model did not meet the normality (P-value < 0.001 of Kolmogorov-Smirnov test and Shapiro-Wilk test) and homoscedasticity assumptions (Supplementary Table 5, Supplementary Figure 2a). The variance inflation factor (VIF) values of factors are all below 3, suggesting no multicollinearity.

Second, we built a GLM model to understand the driving factors for those ITs/PAs with increased forest area loss rates after the establishment (a total of 97 ITs and PAs). We applied a data transformation (Log10) to convert the difference in annual forest area loss rates before and after the establishment. Then we built a GLM model using the same settings (Supplementary Table 4). Forest area in the establishment year, forest area loss rate before the establishment, change in population between 2020 and the establishment year, and the average shortest distance to deforestation showed significant impacts on the forest area loss rate changes. The states of Tocantins, Mato Grosso, and Amazonas had significant differences in forest area loss rate changes after the establishment. However, the protected area types and management objectives of ITs/PAs were not significant variables. The GLM predicted values and our observed values had a significant relationship ($R^2 = 0.49$, $P < 0.01$). Based on the Kolmogorov-Smirnov test ($P = 0.20 > 0.05$) and Shapiro-Wilk test ($P = 0.22 > 0.05$) for the residual errors, this model met the normality assumption (Supplementary Table 5). The two-dimension scatter plot between the regression standardized predicted value and regression standardized residual showed that this GLM model met the homoscedasticity assumption (Supplementary Figure 2b). The VIF values of forest area loss rate before the establishment and change in population between 2020 and the establishment year are 6.8 (< 10) and 6.0 (< 10), and the VIF values of the other factors are below 3, suggesting no multicollinearity.

Third, we built a GLM model to understand the driving factors for those ITs/PAs with decreased forest area loss rates after the establishment (a total of 140 ITs and PAs). We applied a data transformation (Log10) to convert the difference in forest area loss rates before versus after the establishment. Then we built the GLM model using the same settings (Supplementary Table 4). We find that the nPAs had significant impact on the forest area loss rate changes after establishment, compared to sPAs and ITs. Forest area loss rate before the establishment also had significant impact on the forest area loss rate changes after establishment. The GLM predicted values and our observed values had a significant relationship ($R^2 = 0.52$, $P < 0.01$). Based on the Kolmogorov-Smirnov test ($P = 0.20 > 0.05$) and Shapiro-Wilk test ($P = 0.08 > 0.05$) for the residual errors, this GLM model met the normality assumption (Supplementary Table 5). The two-dimension scatter plot between the regression standardized predicted value and regression standardized residual showed that this model met the homoscedasticity assumption (Supplementary Figure 2c). The VIF values of factors are all below 3, suggesting no multicollinearity.

3. Spatial autocorrelation test

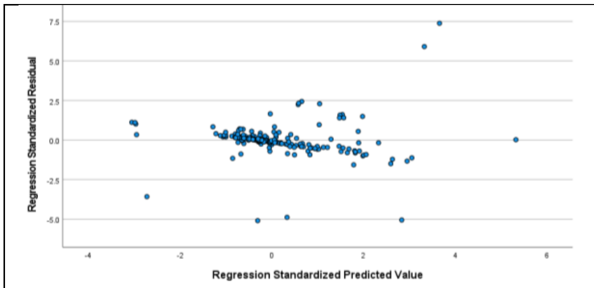
We analyzed the spatial autocorrelation of the model residuals using the Global Moran's I statistic tool in the ArcGIS, which measures spatial autocorrelation based on feature locations and attribute values. The Global Moran's I index has a value range from -1 to 1. -1 is perfect clustering of dissimilar values (perfect dispersion), 0 is no autocorrelation (perfect randomness), and 1 indicates perfect clustering of similar values. The Global Moran's I analysis showed that the residuals of the GLM model for all the ITs/PAs had an index of 0.009 (close to 0), a Z-score value of 0.31, and a P-value of 0.76 (> 0.05), suggesting no spatial autocorrelation. The Global Moran's I analysis showed that the residuals of the GLM model for the ITs/PAs with less deforestation had an index of 0.092, a Z-score value of 1.65, and a P-value of 0.10 (> 0.05), suggesting no spatial autocorrelation. The Global Moran's I analysis showed that the residuals of the GLM model for the ITs/PAs with more deforestation had an index of -0.036, a Z-score value of -0.39, and a P-value of 0.70 (> 0.05), suggesting no spatial autocorrelation.

Supplementary Table 4. The significance and variance inflation factor (VIF) for GLM models.

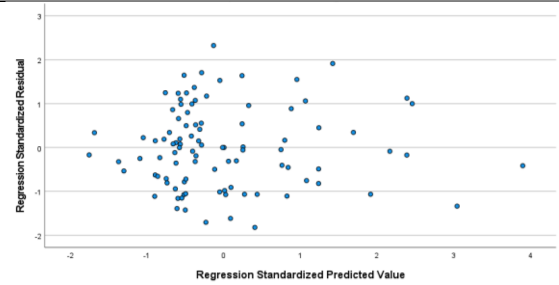
Variables	All ITs and PAs		ITs and PAs with more forest area loss rates		ITs and PAs with less forest area loss rates	
	Sig	VIF	Sig	VIF	Sig	VIF
Intercept	0.167		< 0.001		< 0.001	
Dum_Type_1	0.617	1.8	0.518	1.9	0.048	1.9
Dum_Type_2	0.278	1.7	0.175	1.9	0.791	1.9
Dum_Management	0.017	1.3	0.698	1.3	0.538	1.6
Dum_State_ID_1	0.843	1.2		1.3	0.429	1.3
Dum_State_ID_2	0.563	1.3	0.859	1.1	0.096	1.2
Dum_State_ID_3	<0.001	1.2	0.948	1.1	0.104	1.4
Dum_State_ID_4	0.254	1.3	0.843	2.2	0.243	1.5
Dum_State_ID_5	0.066	1.3	0.01	1.3	0.057	
Dum_State_ID_6	0.287	1.6	0.002	1.6	0.151	1.8
Dum_State_ID_7	0.988	1.4	0.105	1.8	0.592	1.4
Dum_State_ID_8	0.356	1.9	0.037	2.4	0.237	1.9
Forest area in the establishment year	0.999	1.9	0.049	2.6	0.32	2.6
Forest area fraction in the establishment year	0.652	1.9	0.050	6.8	0.007	1.8
Forest area loss before the establishment	0.061	2.3	< 0.001	1.6	0.732	1.6
Total population before the establishment	< 0.001	1.8	0.890	6	0.949	2.4
Change in population between 2020 and the establishment year	0.164	2.8	0.004	2	0.96	2.5
Average shortest distance to deforestation	0.915	1.9	0.004		0.766	2.4

Supplementary Table 5. Normality test for GLM models

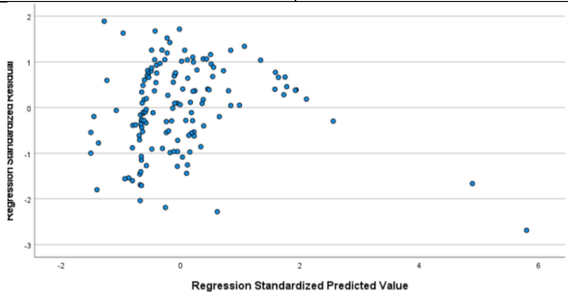
Normality test	Kolmogorov-Smirnov	Shapiro-Wilk
All PAs and ITs	< 0.001	< 0.001
PAs and ITs with more forest area loss rates	0.2	0.219
PAs and ITs with less forest area loss rates	0.2	0.077



a. ITs/PAs



b. ITs/PAs with increased forest area loss rates



c. ITs/PAs with decreased forest area loss rates

Supplementary Figure 2. The two-dimension scatter plots between the regression standardized predicted values and regression standardized residual.