Received: 5 April 2022 | Revised: 26 November 2022 | Accepted: 5 January 2023

DOI: 10.1111/cobi.14058

Conservation Biology

CONTRIBUTED PAPERS

Identifying socioeconomic and biophysical factors driving forest loss in protected areas

Kathryn A. Powlen[1] | **Jonathan Salerno**[1,2] | **Kelly W. Jones**[1] | **Michael C. Gavin**[1,3]

[1]Department of Human Dimensions of Natural Resources, Colorado State University, Fort Collins, Colorado, USA

[2]Graduate Degree Program in Ecology, Colorado State University, Fort Collins, Colorado, USA

[3]Department of Linguistic and Cultural Evolution, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

Correspondence
Kathryn A. Powlen, Department of Human Dimensions of Natural Resources, Colorado State University, 1001 Amy Van Dyken Way, Fort Collins, CO 80523, USA.
Email: kapowlen@colostate.edu

Article impact statement: Identifying drivers of deforestation in protected areas can inform management and help to preemptively protect key biodiversity areas.

Abstract

Protected areas (PAs) are a commonly used strategy to confront forest conversion and biodiversity loss. Although determining drivers of forest loss is central to conservation success, understanding of them is limited by conventional modeling assumptions. We used random forest regression to evaluate potential drivers of deforestation in PAs in Mexico, while accounting for nonlinear relationships and higher order interactions underlying deforestation processes. Socioeconomic drivers (e.g., road density, human population density) and underlying biophysical conditions (e.g., precipitation, distance to water, elevation, slope) were stronger predictors of forest loss than PA characteristics, such as age, type, and management effectiveness. Within PA characteristics, variables reflecting collaborative and equitable management and PA size were the strongest predictors of forest loss, albeit with less explanatory power than socioeconomic and biophysical variables. In contrast to previously used methods, which typically have been based on the assumption of linear relationships, we found that the associations between most predictors and forest loss are nonlinear. Our results can inform decisions on the allocation of PA resources by strengthening management in PAs with the highest risk of deforestation and help preemptively protect key biodiversity areas that may be vulnerable to deforestation in the future.

KEYWORDS
conservation, deforestation, machine learning, Mexico, random forest

Identificación de los factores biofísicos y socioeconómicos que impulsan la pérdida de bosques en las áreas protegidas

Resumen: Las áreas protegidas son una estrategia de uso común para hacer frente a la conversión forestal y la pérdida de biodiversidad. Aunque determinar los factores que impulsan la pérdida de bosques es fundamental para el éxito de la conservación, su comprensión se ve limitada por los supuestos de modelación convencionales. Utilizamos la regresión de bosques aleatorios para evaluar los posibles impulsores de la deforestación en las áreas protegidas de México, considerando las relaciones no lineales y las interacciones de orden superior que subyacen a los procesos de deforestación. Los impulsores socioeconómicos (densidad de carreteras, densidad de población humana) y las condiciones biofísicas subyacentes (precipitaciones, distancia al agua, elevación, pendiente) fueron predictores más fuertes de la pérdida de bosques que las características de las áreas protegidas, como la edad, el tipo y la efectividad de la gestión. Dentro de las características de las áreas protegidas, las variables que reflejan una gestión colaborativa y equitativa y el tamaño del área protegida fueron los predictores más potentes de la pérdida de bosques, aunque con menor poder explicativo que las variables socioeconómicas y biofísicas. A diferencia de los métodos utilizados anteriormente, que suelen basarse en el supuesto de relaciones lineales, observamos que las asociaciones entre la mayoría de los predictores y la pérdida de bosques no son lineales. Nuestros resultados pueden servir de base para la toma de decisiones sobre la asignación de los recursos para las áreas protegidas, reforzando la gestión en las zonas protegidas con mayor riesgo de deforestación y ayudando a proteger de forma preventiva zonas clave para la biodiversidad que pueden ser vulnerables a la deforestación en el futuro.

PALABRAS CLAVE
aprendizaje automático, bosque aleatorio, conservación, deforestación, México

【摘要】
建立保护区是应对森林土地利用变化和生物多样性丧失的常用策略。虽然确定森林丧失的驱动因素是有效保护的核心,但对它们的理解却受到传统模型假设的限制。本研究利用随机森林回归法评估了墨西哥保护区内森林砍伐的潜在驱动因素,同时考虑了森林砍伐过程中的非线性关系和高阶互作。我们发现,社会经济驱动因素 (如道路密度、人口密度) 和潜在的生物物理条件 (如降水、到水源地的距离、海拔、坡度) 比保护区特征 (如建立时长、类型和管理有效性) 更能预测森林丧失情况。在保护区特征中,反映合作和公平管理的变量以及保护区的大小是森林丧失情况最强的预测因素,尽管其解释力低于社会经济和生物物理变量。相较于之前使用的通常基于线性关系假设的方法,我们发现大多数预测因素与森林丧失之间的关联是非线性的。基于以上结果,我们建议通过加强对森林砍伐风险最高的保护区的管理,来优化保护区资源分配的决策,我们的发现还有助于预先保护未来可能容易遭受砍伐的生物多样性关键地区。**【翻译: 胡怡思; 审校: 聂永刚】**

关键词: 保护, 森林砍伐, 机器学习, 墨西哥, 随机森林

INTRODUCTION

Global forests harbor much of the world's terrestrial biodiversity, provide carbon sequestration and other critical ecosystem services, and support over a billion people's livelihoods (FAO & UNEP, 2020; Fedele et al., 2021). Nonetheless, forest cover continues to decline rapidly, with an estimated 10 million ha lost annually from 2015 to 2020 (FAO & UNEP, 2020). Land-use changes, including conversion for agriculture, urban expansion, and forest product demands, are primary causes of forest loss (Armenteras et al., 2017; Curtis et al., 2018). Protected areas (PAs) are one of the most common conservation tools used to counter forest loss; 18% of the world's remaining forest are under some form of protection (FAO & UNEP, 2020; UNEP-WCMC et al., 2020). However, forest conversion and biodiversity loss persist in PAs, albeit often less than in unprotected areas (Geldmann et al., 2019; Wolf et al., 2021; Yang et al., 2021). Understanding the factors driving forest cover change in PAs is critical to ensuring conservation success.

A number of PA characteristics (e.g., size, strictness), socioeconomic drivers (e.g., distance to roads, population density), and underlying biophysical factors (e.g., slope, elevation) affect PA outcomes (e.g., Barnes et al., 2017; Geist & Lambin, 2002). However, previous methods used to examine the impacts of these factors have been limited by the methodological assumptions they are based on (Yang et al., 2021). For example, oversimplification of nonlinear relationships has often resulted in biased PA evaluations and low explanatory power (Andam et al., 2008; Vaca et al., 2019). Moreover, predicted drivers of forest loss are often not independent (e.g., urban area proximity can lead to more roads and greater population density). Therefore, examining a suite of forest-loss drivers can often increase risk of multicollinearity, leading to unstable models and masked relationships, or require the omission of key predictor variables to avoid violating model assumptions. We used a machine learning technique—random forest regression—to overcome these limitations and advance understanding of drivers of forest cover change in PAs in Mexico, a megadiverse country.

In this study, we sought to assess hypothesized drivers of forest loss related to PA characteristics, socioeconomic drivers, and biophysical factors in influencing forest cover change inside PAs. High-resolution satellite imagery allows researchers to analyze patterns of environmental change globally (Pettorelli et al., 2014; Turner et al., 2003). We conducted our analysis at the country level, given its relevance for PA policy. Through a country-level analysis, we aimed to contribute unique empirical evidence on the importance of PA characteristics, beyond International Union for Conservation of Nature (IUCN) category, by integrating scores from a national-level management effectiveness evaluation, thus filling a critical gap previously identified in PA research (Macura et al., 2015).

Identifying drivers of forest loss in PAs has implications for the design of future PAs and the management of existing ones. Understanding the factors contributing to increased deforestation risk can inform PA design and the strategic allocation of resources to vulnerable forested areas. Because we aimed to identify drivers of forest loss in PAs and not measure avoided forest loss due to protected status or determine PA effectiveness, we did not take a counterfactual approach. Thus, we measured associations between drivers of forest loss and PA outcomes without assessing causal relationships between PAs and forest loss. However, we aimed to improve the design of PA evaluations that apply a counterfactual approach to control for confounding factors by providing critical information for variable selection, which would otherwise lead to over- or underestimating PA impact (see Andam et al., 2008; Baylis et al., 2016).

DRIVERS OF PA OUTCOMES

An extensive body of literature examines potential drivers of forest loss and other conservation outcomes as determined by various analytical methods, including correlative and causal analyses (e.g., Aide et al., 2012; Barnes et al., 2017; Busch & Ferretti-Gallon, 2017; Geist & Lambin, 2002; Salafsky et al., 2008). We organized relevant drivers of PA outcomes into 3 categories—PA characteristics and socioeconomic and biophysical factors—to compare relative influence of drivers. PA characteristics include design and management factors. Socioeconomic drivers of forest loss serve as proxies for development pressure (e.g., human population density, population growth) and forest accessibility (e.g., proximity to roads), and biophysical drivers represent potential for economic productivity (e.g., slope, climate conditions).

Various PA characteristics influence PA effectiveness, including age, size, strictness, enforcement capacity, stakeholder engagement, and management resources (Barnes et al., 2017; Ghoddousi et al., 2022). However, the estimated degree and direction of influence of many characteristics vary across studies, potentially due to geographical differences, scale of analysis, and study design (e.g., control variables included, outcome variable examined). For example, a recent study showed that PAs with greater management effectiveness prevented more forest loss than those with less effective management across Mexico (Powlen et al., 2021). Similar patterns occur between more effective management and animal population trends in Southeast Asia (Graham et al., 2021) and global vertebrate abundance (Geldmann et al., 2018). However, the influence of management effectiveness on PA outcomes has been less clear when examining forest loss in Madagascar (Eklund et al., 2019) and fire occurrence and avoided conversion in the Amazon (e.g., Carranza et al., 2014; Nolte & Agrawal, 2013).

Young PAs are often more successful than old PAs because management resources (i.e., financial and staff capacity) increase in the first few years following establishment (Barnes et al., 2017). Research in Mexico has pointed to similar relationships between age and effectiveness. Blackman et al. (2015) found a negative correlation between avoided deforestation and age. However, young PAs may be established in response to recent threats and thus could appear less successful if threat levels are not controlled for (Geldmann et al., 2018; Kere et al., 2017).

Large PAs are expected to be more successful due to decreased edge effects from development pressures, encroachment, or concentration of ecological stresses along PA boundaries (Barnes et al., 2016). Research in Mexico shows that large PAs better conserve forest cover compared with small PAs on average (Blackman et al., 2015). However, larger areas can require more resources for monitoring and management, and were found to experience more forest loss than smaller PAs in a recent global assessment (Wolf et al., 2021).

Strictness of a PA refers to resource-use regulations inside PAs, ranging from strict no-access zones to multiuse areas that provide livelihood opportunities. Research has found strict PAs to be more successful at the global level (Jones et al., 2018) and in Mexico (e.g., Figueroa & Sanchez-Cordero, 2008; Pfaff et al., 2017). However, other studies have found multiuse PAs to be more successful than strict PAs in Mexico (e.g., Blackman, 2015; Sims & Alix-Garcia, 2017), as well as in other countries (e.g., Ferraro et al., 2013; Pfaff et al., 2014). Blackman (2015) found that although multiuse PAs have more heterogenous outcomes than strict PAs, those that provide livelihood opportunities result in less forest loss.

Socioeconomic factors, such as urbanization and infrastructure development, can increase forest accessibility and associated deforestation threats. Proximity to roads and higher road density can increase deforestation risk in PAs due to increased accessibility for natural resource extraction or development opportunities (Joppa & Pfaff, 2009; Kere et al., 2017; Laurance et al., 2009). Similarly, population growth, population density, and proximity to urban areas can increase biodiversity risk due to encroachment, demand for resources, and a large labor force available for extractive industries (Busch & Ferretti-Gallon, 2017). Alternatively, PAs providing tourism opportunities may incentivize compliance from nearby communities, reducing encroachment (Barnes et al., 2017). Increases in both accessibility and human populations may also promote more surveillance and enforcement efforts (Andam et al., 2008), which could decrease forest loss linked to noncompliant activities.

Land tenure, including land ownership and tenure security, has also been identified as an important factor that shapes forest conservation outcomes (e.g., Barnes et al., 2017; Blackman et al., 2015; Bonilla-Moheno et al., 2013; Robinson et al., 2014; Skutsch et al., 2014). Mexico's communal tenure system, *ejidos*, provides usufruct rights and requires *ejidatarios* to make land-use decisions collectively (Bray et al., 2008). Although some researchers have found that community forest management works as a conservation strategy in Mexico (e.g., Ellis & Porter-Bolland, 2008; Porter-Bolland et al., 2012), others have found no significant relationship between tenure and forest loss (e.g., Bray et al., 2008; Mas & Cuevas, 2013; Skutsch et al., 2014). Additionally, overlapping PAs may weaken or undermine existing tenure systems, leading to increased biodiversity risks (Geldmann et al., 2019).

Physical landscape characteristics, such as slope and elevation, may serve as proxies for suitability for other land uses, such as agriculture and livestock grazing. Lands with relatively low agricultural preparation costs may be more likely to be deforested (Busch & Ferretti-Gallon, 2017; Joppa & Pfaff, 2009; Vaca et al., 2019). Areas with favorable climate conditions are also expected to be more at risk of deforestation for development or agriculture (Aide et al., 2012; Busch & Ferretti-Gallon, 2017). Water proximity is linked to increased agricultural activities (Vaca et al., 2019) and increased accessibility via water transportation (Bax & Francesconi, 2018). Research has found biophysical variables—specifically temperature and elevation—to be stronger predictors of land-cover change than socioeconomic variables at the municipality level in Mexico (Bonilla-Moheno et al., 2012) and across Latin America (Aide et al., 2012).

METHODS

We used random forest regression to investigate the relationships between PA design and management characteristics, socioeconomic and biophysical factors, and change in forest cover in PAs in Mexico. We sought to identify the main drivers of, and their relative influence on, forest loss inside PAs. This is not equivalent to measuring PA impact or determining PA effectiveness, which would require the establishment of a counterfactual. We examined how random forest regression can help advance understanding of drivers of PA outcomes due to its capacity to examine a large number of predictor variables simultaneously, while adjusting for multicollinearity, higher order interactions, and nonlinear relationships (Breiman, 2001).

Study area

In 2015, an estimated 33% of Mexico was forested (regenerated and primary), largely concentrated in the southern states of Oaxaca, Chiapas, Campeche, and Quintana Roo (FAO, 2015). However, Mexico continues to experience high rates of deforestation, which have increased over the past decade (GFW, 2022). Mexico has established an extensive network of over 180 national PAs managed by the National Commission for Natural Protected Areas (CONANP) and more at the subnational level. We focused on CONANP-managed PAs, including national parks (35% of national PAs), biosphere reserves (23%), flora and fauna protection areas (21%), sanctuaries (10%), national monuments (6%), and natural resource protection areas (5%).

Our sample included a subset of Mexico's national PAs in forest regions that had management effectiveness scores from CONANP's *i-efectividad* evaluation, a standardized self-employed survey taken by PA staff yielding a management score for each PA based on 48 indicators representing 5 dimensions of management. The 5 dimensions are context and planning, administration and financial, use and benefits, governance and social participation, and management quality. Our unit of analysis was 1-km^2 grid cells, which covered the entirety of PAs with management effectiveness scores. Cells containing <75% forest cover in 2000 were removed. The final sample included 30,888 cells across 51 PAs.

Dependent and independent variables were gathered from publicly available sources. Due to the 75% forest cover threshold in grid-cell sampling, we calculated relative rather than absolute forest loss from 2015 to 2019 by dividing forest loss by the baseline forest cover per cell based on Global Forest Watch data produced at 30-m resolution (Hansen et al., 2013). This relative measure of forest loss better accounts for total forest at risk of deforestation and avoids underestimating forest loss rates based on little forest cover. We calculated 2015 baseline forest cover by subtracting 2001–2014 forest loss from 2000 forest cover. Independent variables consisted of 9 PA characteristics, 9 socioeconomic predictors, and 5 biophysical predictors (Table 1), each specific to 2015–2019. We also included 3 categorical independent variables (individual PAs, state, and ecoregion) to control for differences not captured by the other predictors. Grid cells with missing data were removed. More details on included variables are in Appendices S1 and S2.

Analyses

Random forest is a machine learning approach that can classify a categorical variable or predict a dependent variable (i.e., regression) (Breiman, 2001). Random forest regression is an exploratory approach for analyses with many potential predictor variables with higher order interactions and nonlinear relationships (Breiman, 2001; Strobl et al., 2009). The method applies a bootstrap aggregating, or "bagging," approach to create multiple subsamples of a data set for training and testing a model (Strobl et al., 2009). With the training data, the algorithm applies a split-variable randomization process to develop a collection of uncorrelated decision trees and averages the prediction across all trees (Hastie et al., 2017). Averaging the predictions across diverse trees reduces the variance and bias and ultimately increases prediction performance. Using between 60% and 80% of the data is recommended for training; the remaining 20–40% is used for testing the model's predictive power (Breiman, 2001). We used a 70:30 training–test split ratio.

Random forest is often referred to as an "off-the-shelf" machine learning algorithm due to its high predictive capability with low hyperparameter tuning requirements (Hastie et al., 2017). Optional tuning parameters include the total numbers of trees (ntree) and the number of variables randomly sampled to split each tree node (mtry) (Breiman, 2001). Similar to Epstein et al. (2021), we used the caret package (Kuhn, 2008) in R 3.6.1 with RStudio 1.2.1335 to determine optimal tuning parameters, and a 10-fold cross validation method to assess model accuracy (Epstein et al., 2021).

We used the mean absolute error as a measure of model performance (Chai & Draxler, 2014; Willmott & Matsuura, 2005) and report total variance explained (R^2) and the importance values (IVs). Random forest regression calculates IVs as the increase in node purity, or the reduction in the sum of square errors from splitting with each specific variable (Hastie et al., 2017). The values were adjusted using the caret package to fall along a 0–100 scale for interpretability; 100 is the highest degree of influence. To test the sensitivity of our findings, we

TABLE 1 Summary statistics of variables included in the random forest model predicting forest loss in protected areas

Variable	Median	Mean	SD	Min	Max
Socioeconomic					
Distance to roads (km)	3.20	4.12	4.07	0.00	14.99
Distance to urban areas (km)	14.91	16.58	13.62	0.00	49.99
Ejido tenure (%)	0.00	39.55	46.85	0.00	100.00
Extreme poverty (%)	23.55	21.92	14.02	0.18	69.97
Moderate poverty (%)	48.27	45.19	8.62	4.23	72.25
PES (%)	100.00	65.32	46.73	0.00	100.00
Pop. change rate ($\triangle$pop/km^2)	−0.01	−0.49	3.86	−205.23	8.38
Pop. density (pop/km^2)	0.37	5.28	39.53	0.00	1959.03
Road density (0.00–5.00)	0.09	0.24	0.43	0.00	4.15
Biophysical					
Elevation (m)	600.00	906.52	853.16	1.00	3618.00
Distance to water (km)	1.07	2.06	2.73	0.00	14.99
Precipitation (mm)	1154.00	1240.35	570.07	281.00	3100.00
Slope (degree)	2.69	4.64	5.30	0.00	31.85
Temperature (°C)	22.00	19.85	5.07	6.00	27.00
Protected area characteristics					
Age (years)	32.00	35.27	16.41	6.00	86.00
Area (km^2)	3,312	3,806	2,379	20	7231.00
Management effectiveness	74.00	72.24	11.28	42.00	88.00
ME: context	71.00	69.22	12.73	41.00	95.00
ME: admin	56.00	56.37	7.60	17.00	92.00
ME: use	76.00	77.08	17.18	38.00	100.00
ME: governance	87.00	83.1	17.67	30.00	100.00
ME: management quality	81.00	77.91	15.93	38.00	90.00

Abbreviations: ME, management effectiveness; PES, payment for ecosystem services; pop., population.

also ran the model without individual PAs and ecoregions and compared model performance and IVs. To examine differences in deforestation predictors across forest types, we aggregated data into subsets based on simplified ecoregions from World Wildlife Fund (Olson et al., 2001) and ran separate models for moist forest (n = 15,087 cells), dry forest (n = 2094), pine-oak forest (n = 10,450), montane forest (n = 592), and mangrove forest (n = 1541) (Appendices S3 & S4). We tested for spatial autocorrelation with Moran's I and the residuals of each model (Negret et al., 2020) and ran a general linear model for comparison.

An IV is an estimate of the degree of influence of each variable in the prediction and does not reflect the direction of influence. Therefore, random forest analyses are often paired with partial dependence plots or accumulated local effect (ALE) plots to better depict the nature and direction of the relationship (Apley & Zhu, 2020; Hastie et al., 2017). We used ALE plots to present the conditional relationship between our predictors and forest loss due to the multicollinearity of many of our predictor variables (Apley & Zhu, 2020; Strobl et al., 2008). We use the iml package (Molnar et al., 2018) to produce ALE plots for the 9 variables with the greatest IVs.

RESULTS

Descriptive statistics

The final sample contained grid cells in 51 PAs located in 27 different states and represented 5 of Mexico's PA management categories, including biosphere reserves (74%), flora and fauna PAs (14%), national parks (6%), natural resource PAs (6%), and national monuments (0.001%). Ages of the PAs ranged from 6 to 86 years (mean = 35 years). The smallest area was 20 km^2 and the largest was 7232 km^2 (mean = 3806 km^2). The mean overall management effectiveness score was 72 (minimum 42, maximum 88). The highest scoring management dimensions were governance and social participation (mean = 83), management quality (mean = 78), and use and benefits (mean = 77).

The average coverage of ejido tenure in the sampled cells was 40%, and the average overlap with PES enrollment was 65% (Table 1). Population density was relatively low, and the population change rate was negative, on average. The mean estimate for the percent of the population in extreme poverty was 22%; around twice as many were living in moderate poverty (see Appendix S5 for subgroup descriptive statistics).

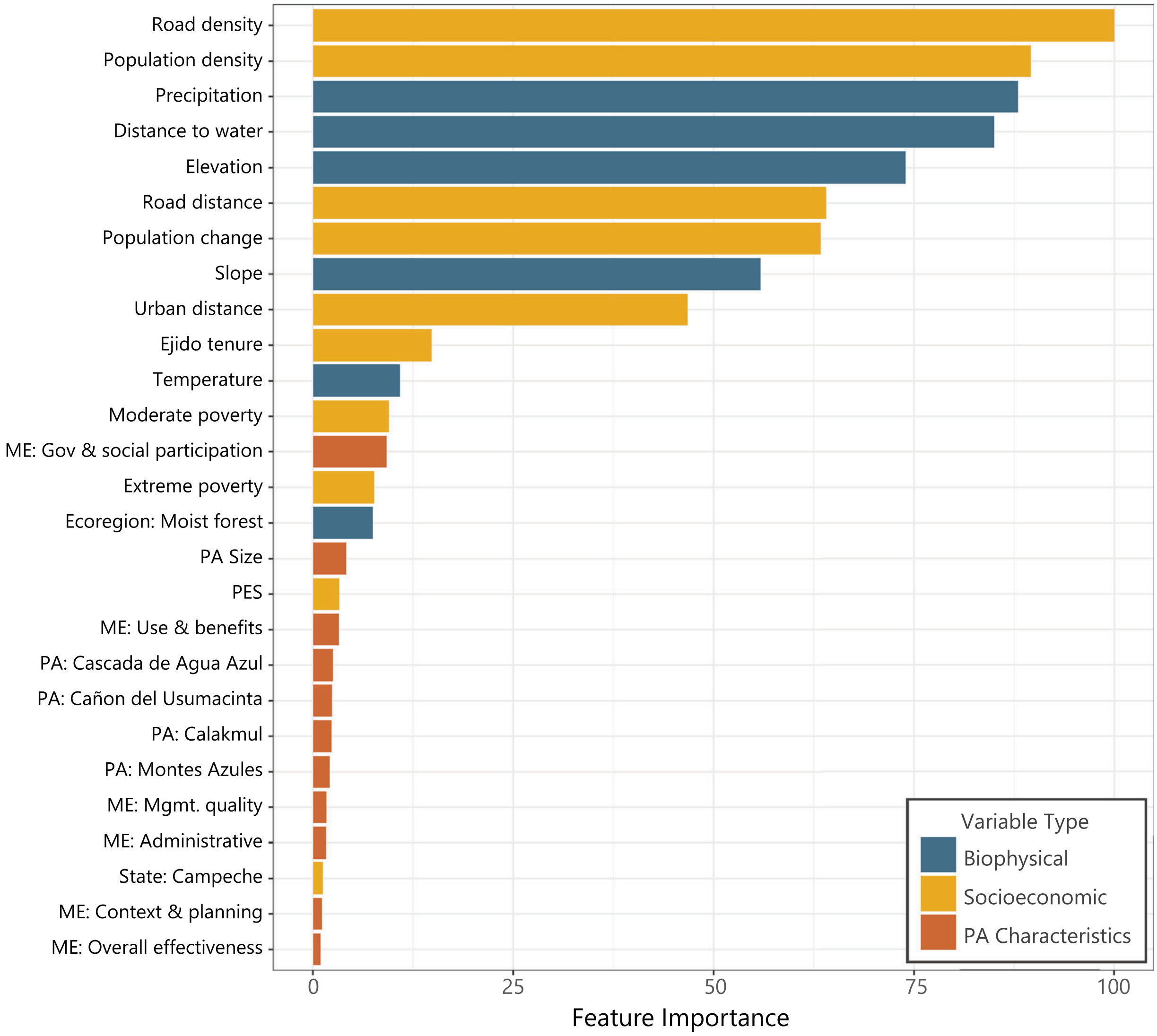

FIGURE 1 Variable importance (increase in node purity) from the random forest model for all variables associated with forest loss in protected areas with an importance value >1.0 (ME, management effectiveness component; PA, protected area). Importance values are an estimate of the level of influence and do not indicate direction of influence.

A Pearson correlation analysis showed strong relationships between multiple predictors, further emphasizing the need for modeling methods that adjust for multicollinearity when exploring a variety of forest-loss drivers. Correlated variables occurred within the same variable category (e.g., temperature and elevation [$r = -0.95$]) and across categories (e.g., extreme poverty and precipitation [$r = 0.76$]). A full list of Pearson correlation coefficients is in Appendix S6.

Random forest model

For the random forest model, $R^2 = 0.58$ (Appendix S7). Figure 1 shows variable importance, which reflects the degree of influence, for all variables with values >1.0 to increase figure interpretability. Two variables had IVs of <1.0, including PA age (0.79) and strictness (0.19). All biophysical and socioeconomic variables were >1.0 in at least 1 category (e.g., moist forest, state of Campeche). The top 12 variables of importance were socioeconomic or biophysical variables, and the top 9 had a substantially higher IV than all remaining variables (IV > 45).

Road density best predicted forest loss (IV = 100), followed by population density (IV = 90) and the biophysical variables of precipitation (IV = 88) and distance to water (IV = 85) (Figure 1). Other key socioeconomic variables included road distance (IV = 64) and population change (IV = 63). Percentage of the population in moderate poverty (IV = 9), extreme poverty (IV = 7), and enrollment in PES programs (IV = 3) had the least influence, although they were still above the 1.0 cutoff and higher than most PA characteristics. Additional biophysical variables of high importance were elevation (IV = 74) and slope (IV = 56); temperature (IV = 11) had the least influence of the biophysical variables.

The PA characteristic with the highest IV was the governance and social participation dimension of management effectiveness

(IV = 9). This reflects the degree to which all stakeholders' rights are respected and the level of involvement from neighboring resource users in management decision-making. The second most influential PA variable was PA size (IV = 4), followed by a second management dimension, use and benefits (IV = 3), which measures the fair distribution of benefits for all stakeholders. In addition to these variables, a number of specific PAs were strong predictors of forest loss, including Cascada de Agua Azul (IV = 2.5), Cañon del Usumacinta (IV = 2.4), Calakmul (IV = 2.3), and Montes Azules (IV = 2.1).

We found similar model performance and importance ranking results in our sensitivity tests, confirming that the categorical variables of PAs and ecoregion were not absorbing a substantial amount of variance (Appendices S8 & S9). We also found no evidence of spatial clustering in the residuals of the full model or the subgroup models; Moran's *I* index equaled ∼0.00 (Appendices S10–S15). In our separate ecoregion subgroup models, we found qualitatively similar patterns of variable importance. The 5 most important variables across all subgroup models were socioeconomic and biophysical factors (subgroup analysis results in Appendix S16).

Accumulated local effects

To better understand the direction of the relationships between the variables of highest importance and forest loss, we produced ALE plots for the 9 strongest predictors (Figure 2). None of the relationships were linear. We found higher road density increased the risk of forest loss. More specifically, forest loss increased steeply as road density increased, with some variation, then leveled off after a road density of 1.50 (out of 5.00) was reached. The risk of forest loss decreased as distance from roads and urban areas increased until about 8 and 20 km, respectively, then it gradually increased as PAs increased in remoteness. There was high uncertainty in the direction of influence of population density and population change on forest loss.

The risk of forest loss declined sharply within 1 km of water, but increased from around 1.25 km until approximately 7 km. The direction of the relationship between risk of forest loss and precipitation varied. Forest loss decreased at low levels of precipitation, began to increase around 1200 mm, and then dropped off again at very high levels of precipitation (∼ 2700 mm). Risk of forest declined as elevation increased to about 800 m, then it leveled off. A similar pattern was found with slope and forest loss; risk decreased until around a 7° slope and then began to increase, though with some uncertainty. The ALE plots for additional predictor variables are in Appendix S17.

DISCUSSION

Random forest regression has been used to identify predictors of marine PA performance (Di Franco et al., 2016; Gill et al., 2017), drivers of PA placement (Baldi et al., 2017), factors leading to social and conservation PA outcomes in a meta-analysis (Oldekop et al., 2016), and predictors of forest monitoring by forest-user groups (Epstein et al., 2021). However, prior to our present study, the method's application for understanding drivers of terrestrial PA outcomes remained relatively unexplored.

Overall, we found PA placement, including socioeconomic context and landscape characteristics, had substantial influence over forest loss. Specifically, most socioeconomic and biophysical variables were stronger drivers of PA outcomes than PA characteristics such as size, strictness, and management effectiveness (Figure 1). Within PA characteristics, management variables that reflected collaboration and equitable benefit sharing and PA size were the strongest predictors of forest loss, albeit with less explanatory power than socioeconomic and biophysical variables.

Our ALE plots demonstrated the complexity of relationships between various predictors and observed forest loss, suggesting that PA evaluations have frequently oversimplified these relationships by using modeling approaches based on the assumption of linearity of forest-loss drivers, including linear and generalized linear models. We built on a small number of studies in which random forest models were used to examine the complex relationship between drivers of forest loss and deforestation (e.g., Aide et al., 2012; Bax & Francesconi, 2018; Bonilla-Moheno et al., 2012) and contributed to a growing number of studies accounting for nonlinear drivers of PA performance (e.g., Shumba et al., 2021; Yang et al., 2021). Additionally, we adjusted for multicollinearity among predictor variables, using ALE plots (rather than partial dependence plots) to estimate the conditional relationships between predictor variables and forest loss.

Socioeconomic factors

Socioeconomic variables were strong predictors of forest loss across all models, including proxies for development pressures and forest accessibility (Figure 1). Existing research often highlights negative effects of linear infrastructure (e.g., roads) on forest conservation (e.g., Busch & Ferretti-Gallon, 2017; Laurance et al., 2009). Busch and Ferretti-Gallon (2017) found that presence of roads is consistently correlated with high forest loss across over 100 studies. Our results align with previous findings, identifying road density and distance as strong predictors of forest loss (Figure 1) and demonstrating the increased risk of forest loss in areas closer to roads and with higher road density (Figure 2a,g). However, we also found that very remote areas were also at increased risk of forest loss.

We found high levels of uncertainty in the relationship between forest loss and human populations (Figure 2d,f), potentially reflecting complex and sometimes contradictory relationships between human population centers and forest cover change. For example, large populations near PAs can increase demand for resources and employment opportunities in extractive industries (Busch & Ferretti-Gallon, 2017). Alternatively, neighboring populations can increase surveillance for

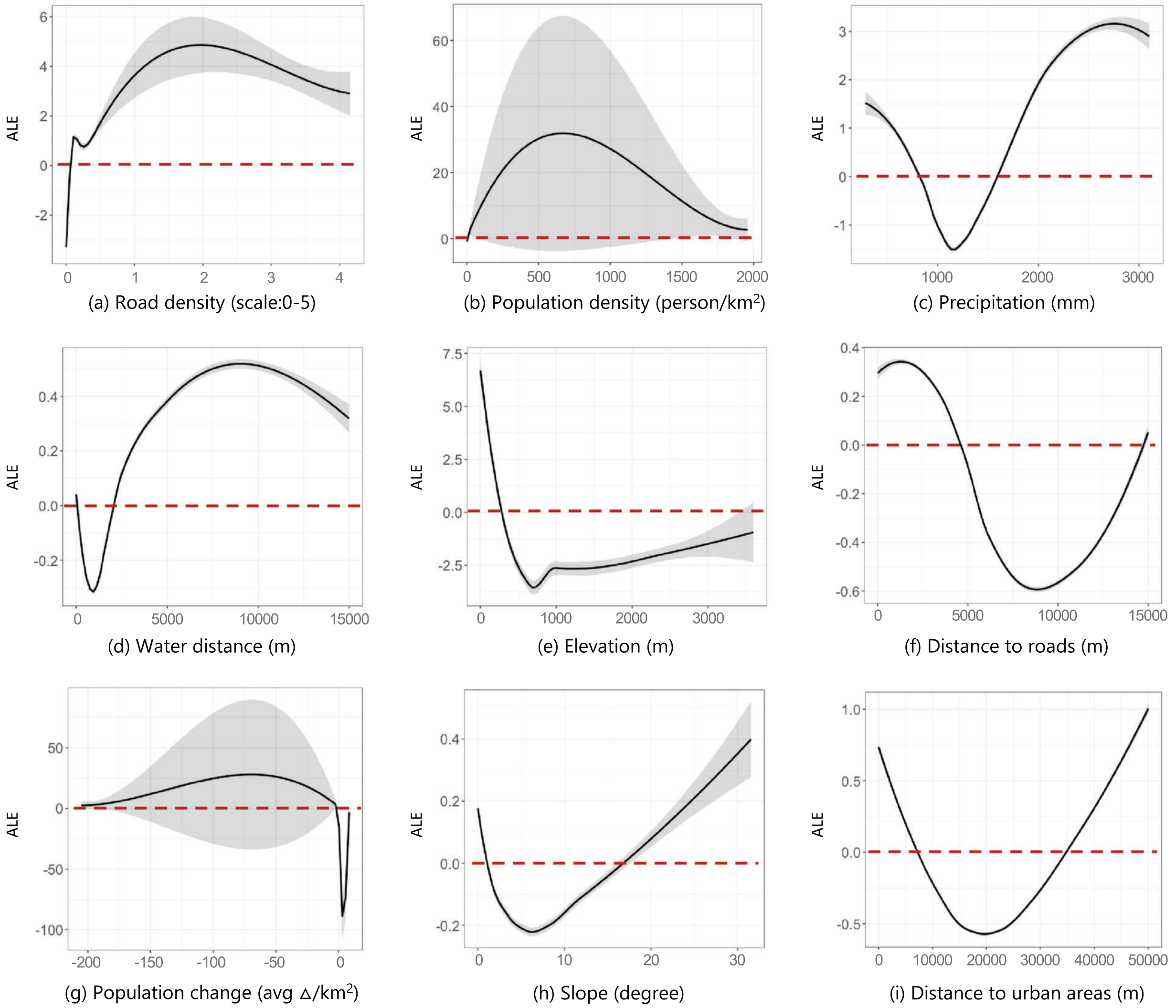

FIGURE 2 Relationships between predicted forest loss (accumulated local effect [ALE]) and 9 variables with the greatest effect (ALE values, change in predicted forest loss at a given value compared with the average prediction; gray shading, 95% confidence interval).

conservation efforts and be encouraged to engage in more environmentally sustainable livelihoods (Andam et al., 2008; Barnes et al., 2017; Solorzano & Fleischman, 2018). Solorzano and Fleischman (2018) found that tenure legacies, political inequality, and economic opportunities can also influence community support for conservation in PAs in Mexico and Guatemala. Future research should incorporate additional socioeconomic variables, such as tenure legacy and livelihood opportunities, to attempt to clarify this uncertainty.

Biophysical factors

We found that precipitation and distance to water strongly affected forest loss, consistent with other research conducted in Mexico (Bonilla-Moheno et al., 2012; Roy Chowdhury, 2006) and across Latin America (Aide et al., 2012). Bax and Francesconi (2018) found distance to rivers and precipitation to be important forest loss predictors in Peru; similar nonlinear patterns were found using partial dependence plots. Water access from rivers can increase forest accessibility, enabling human settlement and resource extraction (Bax & Francesconi, 2018). Although precipitation can influence agricultural productivity, Busch and Ferretti-Gallon (2017) found landscape wetness is negatively correlated with deforestation due to higher moisture decreasing agricultural suitability. Our ALE plots suggested a similar relationship: risk of forest loss increased as precipitation increased until a potential decline occurred after 2700 mm, though with less certainty. We also found high levels of predicted forest loss at very low precipitation rates, potentially reflecting fire risk. To test for this, we examined the relationship between burn scars and forest loss, but found little evidence of correlation (Appendices S18 & S19).

PAs are often established at high elevations and on steeper slopes, which are less suited for human use (Baldi et al., 2017; Joppa & Pfaff, 2009). We found that these variables affected PA outcomes, consistent with other research conducted in Mexico (e.g., Pfaff et al., 2017). Our ALE results were consistent with prior findings; risk of forest loss was greatest at low elevations and slopes, areas with potentially lower agricultural preparation costs and favorable livestock conditions. We also found substantially higher risk of forest loss on steep slopes, which could be linked to fire risk because fire containment is more challenging in steeper landscapes (Vaca et al., 2019). Future research should expand on these findings to determine directional thresholds in each of the nonlinear predictor variables.

PA characteristics

There has been a recent call to make PAs more just and equitable (e.g., Jonas et al., 2021; Zafra-Calvo & Geldmann, 2020; Zafra-Calvo et al., 2017); related indicators have been integrated into the most recent IUCN Green List standards. Community involvement in PA decision-making increases support for conservation and ultimately improves PA outcomes (Andrade & Rhodes, 2012; Solorzano & Fleischman, 2018). We found the most important PA management variables were governance and social participation, followed by use and benefits. In general, the relationships trended negative, with higher scores resulting in lower risks of forest loss, supporting calls for PAs to be more collaborative and equitable. Although the ALE plot of governance and social participation showed some evidence of nonlinearity, the results were skewed by 2 PAs with high governance scores and unusually high rates of forest loss (Montes Azules [score = 100] and Calakmul [score = 97]). Given that our results showed the main drivers of forest loss were proxies for human pressure, involvement of and equitable benefit sharing with stakeholders may be a key response to these threats.

The administrative and financial management dimension was also ranked in the importance plot, albeit lower than expected. Financial resources and human capacity are consistently highlighted as drivers of PA outcomes and are limited globally (e.g., Coad et al., 2019; Gill et al., 2017). Although the ALE plot of the administrative and financial management score showed a slight increase in forest loss risk at higher scores, we relationship was skewed by 1 PA with the highest score (Volcán Nevado de Colima [score = 95]), whereas all other administrative and financial scores were <75. Therefore, our findings support existing research highlighting the importance of sufficient resources for reducing forest loss risk and contribute evidence of the importance of specific management dimensions, such as collaborative decision-making and shared benefits.

Four individual PAs were ranked as strongly influencing forest loss outcomes, all with higher-than-expected levels of forest loss, including 2 biosphere reserves and 2 flora and fauna PAs in 3 states in southern Mexico. Three of these PAs were located either on (Cañón del Usumacinta, Calakmul) or close to (Montes Azules) the border with Guatemala. Further empirical research is required to explore the specific characteristics leading to this outcome.

Limitations

By accounting for nonlinearity, our random forest models had much larger explanatory power compared with a linear model (Appendix S20). However, the model's R^2 value showed that some predictors were missing from the analysis. Data availability can be a limiting factor when conducting spatial analyses with secondary data, particularly given the requirement of close temporal alignment of independent variables with observed loss, often resulting in omission of potential predictors. Our list of predictors is not comprehensive and others may be influential, such as tenure security (Robinson et al., 2014), overlap with Indigenous lands (see Garnett et al., 2018), and political trends predating PA establishment (Solorzano & Fleischman, 2018). Additionally, a marginalization index from 2010 showed that human marginalization affected forest loss in a study conducted at the municipality level in Mexico (Bonilla-Moheno et al., 2012) and livestock density was negatively correlated with forest conservation in another (Mas & Cuevas, 2015). However, to our knowledge, these data were not available for our period of focus.

Additional limitations to spatial analyses are differing scales of key variables and nested variables (e.g., roads are in PAs within states). Our results, particularly the relative order of importance, may be influenced by the scale of the independent variables. For example, poverty rates at the municipal level may have limited ability to explain variation in forest loss occurring at smaller scales. However, our sample across 51 PAs included appreciable variation in even coarser resolution predictors, and random forest models are particularly well suited to tease apart relationships in nested data. Nonetheless, analyses such as ours should ideally be paired with qualitative work to inform appropriate management actions.

We do not claim that PAs are ineffective due to forest loss experienced, and we did not try to determine PA impact on forest conservation outcomes. Our objective was to identify factors influencing forest loss patterns, similar to Aide et al. (2012), Bax and Francesconi (2018), and Bonilla-Moheno et al. (2012). Moreover, we do not claim that the factors we identified are similar or different between PAs and forest outside of PAs, which would require more empirical research. Because we sought only to examine drivers of forest loss inside PAs and did not examine the impact of PA establishment on forest loss, establishing a control group was not appropriate. Instead, we demonstrated how random forest models can be used to increase understanding of drivers of forest loss in PAs and encourage the use of our results on drivers to inform counterfactual research aimed at testing the impact of PAs, which requires establishing reliable control groups through an understanding of both the factors that influence outcomes and placement of PAs.

Broader implications

Our findings highlight the influence of placement characteristics in driving forest loss in PAs, including proximity to populations, road access, waterways, and physical landscape characteristics. This has important implications for PA managers, as well as researchers. By combining results of our ALE plots, landscape-level patterns can be identified to determine areas with the highest deforestation risk (e.g., low elevation with low slopes, close to urban areas and roads, areas with cooler temperatures and greater precipitation) and resources can be appropriately allocated to PAs that may face the greatest threat based on these characteristics. Additionally, identifying the important design and management characteristics that can counter this pressure will be key to future conservation success.

Researchers have called for more careful PA effectiveness evaluations after finding evidence of PA placement biases (see Baldi et al., 2017; Joppa & Pfaff, 2009). In response, these biases are accounted for by including external factors as independent variables or evaluation approaches that estimate a counterfactual (i.e., what would have happened if the area was not protected) (e.g., Baylis et al., 2016; Sarkki et al., 2015; Schleicher et al., 2019). Our findings reinforce the need for these approaches given that many variables driving outcomes are external to PA design. We also presented the nonlinear relationships between many drivers of forest loss and deforestation and highlight the need to account for this complexity. We demonstrated the utility of random forest as an exploratory method, which can be used to help inform matching or other counterfactual design measurements, to avoid omitted variable biases and, ultimately, design more accurate evaluations. Finally, our results highlight a number of future research opportunities, including more investigation into thresholds determining the direction of influence of predictors.

ACKNOWLEDGMENTS

We thank S. Haider for her assistance in the early development and execution of the random forest analysis code. We also thank R. Boone and J. Solomon for their valuable comments on earlier drafts.

ORCID

Kathryn A. Powlen https://orcid.org/0000-0002-9685-0063
Jonathan Salerno https://orcid.org/0000-0001-9402-6479
Kelly W. Jones https://orcid.org/0000-0001-9664-7615
Michael C. Gavin https://orcid.org/0000-0002-2169-4668

REFERENCES

Aide, T. M., Clark, M. L., Grau, R., López-Carr, D., Levy, M. A., Redo, D., Bonilla-Moheno, M., Riner, G., Andrade-Núñez, M. J., & Muñiz, M. (2012). Deforestation and reforestation of Latin America and the Caribbean (2001–2010). *Biotropica*, *45*, 262–271.

Andam, K. S., Ferraro, P. J., Pfaff, A., Sanchez-Azofeifa, G. A., & Robalino, J. A. (2008). Measuring the effectiveness of protected area networks in reducing deforestation. *Proceedings of the National Academy of Sciences of the United States of America*, *105*, 15–20.

Andrade, G. S. M., & Rhodes, J. R. (2012). Protected areas and local communities: An inevitable partnership toward successful conservation strategies? *Ecology and Society*, *17*, 14. https://doi.org/10.5751/ES-05216-170-414

Apley, D. W., & Zhu, J. (2020). Visualizing the effects of predictor variables in black box supervised learning models. *Journal of the Royal Statistical Society B*, *82*, 1059–1086.

Armenteras, D., Espelta, J. M., Rodríguez, N., & Retana, J. (2017). Deforestation dynamics and drivers in different forest types in Latin America: Three decades of studies (1980 –2010). *Global Environmental Change*, *46*, 139–147.

Baldi, G., Texeira, M., Martin, O. A., Grau, H. R., & Jobbágy, E. G. (2017). Opportunities drive the global distribution of protected areas. *PeerJ*, *5*, e2989. https://doi.org/10.7717/peerj.2989

Barnes, M. D., Craigie, I. D., Harrison, L. B., Geldmann, J., Collen, B., Whitmee, S., Balmford, A., Burgess, N. D., Brooks, T., Hockings, M., & Woodley, S. (2016). Wildlife population trends in protected areas predicted by national socio–economic metrics and body size. *Nature Communications*, *7*, 12747.

Barnes, M. D., Craigie, I. D., Dudley, N., & Hockings, M. (2017). Understanding local-scale drivers of biodiversity outcomes in terrestrial protected areas. *Annals of the New York Academy of Sciences*, *1399*, 42–60.

Bax, V., & Francesconi, W. (2018). Environmental predictors of forest change: An analysis of natural predisposition to deforestation in the tropical Andes region, Peru. *Applied Geography*, *91*, 99–110.

Baylis, K., Honey-Roses, J., Borner, J., Corbera, E., Ezzine-de-Blas, D., Ferraro, P. J., Lapeyre, R., Persson, U. M., Pfaff, A., & Wunder, S. (2016). Mainstreaming impact evaluation in nature conservation. *Conservation Letters*, *9*, 58–64.

Blackman, A. (2015). Strict versus mixed-use protected areas: Guatemala's Maya Biosphere Reserve. *Ecological Economics*, *112*, 14–24.

Blackman, A., Pfaff, A., & Robalino, J. (2015). Paper park performance: Mexico's natural protected areas in the 1990s. *Global Environmental Change*, *31*, 50–61.

Bonilla-Moheno, M., Aide, T. M., & Clark, M. L. (2012). The influence of socioeconomic, environmental, and demographic factors on municipality-scale land-cover change in Mexico. *Regional Environmental Change*, *12*, 543–557.

Bonilla-Moheno, M., Redo, D. J., Aide, T. M., Clark, M. L., & Grau, H. R. (2013). Vegetation change and land tenure in Mexico: A country-wide analysis. *Land Use Policy*, *30*, 355–364.

Bray, D. B., Duran, E., Ramos, V. H., Mas, J., Velazquez, A., McNab, R. B., Barry, D., & Radachowsky, J. (2008). Tropical deforestation, community forests, and protected areas in the Maya Forest. *Ecology and Society*, *13*, 56. https://doi.org/10.5751/ES-02593-130256

Breiman, L. (2001). Random forests. *Machine Learning*, *45*, 5–32.

Busch, J., & Ferretti-Gallon, K. (2017). What drives deforestation and what stops it? A meta-analysis. *Review of Environmental Economics and Policy*, *11*, 3–23.

Carranza, T., Manica, A., Kapos, V., & Balmford, A. (2014). Mismatches between conservation outcomes and management evaluation in protected areas: A case study in the Brazilian Cerrado. *Biological Conservation*, *173*, 10–16.

Chai, T., & Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature. *Geoscience Model Development*, *7*, 1247–1250.

Coad, L., Watson, J. E. M., Geldmann, J., Burgess, N. D., Leverington, F., Hockings, M., Knights, K., & Di Marco, M. (2019). Widespread shortfalls in protected area resourcing undermine efforts to conserve biodiversity. *Frontiers in Ecology and the Environment*, *259*, 259–264.

Curtis, P. G., Slay, C. M., Harris, N. L., Tyukavina, A., & Hansen, M. C. (2018). Classifying drivers of global forest loss. *Science*, *361*, 1108–1111.

Di Franco, A., Thiriet, P., Di Carlo, G., Dimitriadis, C., Francour, P., Gutiérrez, N. L., Jeudy de Grissac, A., Koutsoubas, D., Milazzo, M., del Mar Otero, M., Piante, C., Plass-Johnson, J., Sainz-Trapaga, S., Santarossa, L., Tudela, S., & Guidetti, P. (2016). Five key attributes can increase marine protected areas performance for small-scale fisheries management. *Scientific Reports*, *6*, 38135.

Eklund, J., Coad, L., Geldmann, J., & Cabeza, M. (2019). What constitutes a useful measure of protected area effectiveness? A case study of management inputs and protected area impacts in Madagascar. *Conservation Science and Practice*, *1*, 1–12.

Ellis, E. A., & Porter-Bolland, L. (2008). Is community-based forest management more effective than protected areas? A comparison of land use/land cover change in two neighboring study areas of the Central Yucatan Peninsula, Mexico. *Forest Ecology and Management*, *256*, 1971–1983.

Epstein, G., Gurney, G., Chawla, S., Anderies, J. M., Baggio, J., Unnikrishnan, H., Tomas, S. V., & Cumming, G. S. (2021). Drivers of compliance monitoring in forest commons. *Nature Sustainability*, *4*, 450–456.

Fedele, G., Donatti, C. I., Bornacelly, I., & Hole, D. G. (2021). Nature-dependent people: Mapping human direct use of nature for basic needs across the tropics. *Global Environmental Change*, *71*, 102368.

Ferraro, P. J., Hanauer, M. M., Miteva, D. A., Canavire-Bacarreza, G. J., Pattanayak, S. K., & Sims, K. R. E. (2013). More strictly protected areas are not necessarily more protective: Evidence from Bolivia, Costa Rica, Indonesia, and Thailand. *Environmental Research Letters*, *8*, 025011. https://doi.org/10.1088/1748-9326/8/2/025011

Figueroa, F., & Sanchez-Cordero, V. (2008). Effectiveness of natural protected areas to prevent land use and land cover change in Mexico. *Biodiversity and Conservation*, *17*, 3223–3240.

Food and Agriculture Organization (FAO). (2015). *Forest cover in Mexico*. www.globalforestwatch.org

Food and Agriculture Organization (FAO) & United Nations Environment Programme (UNEP). (2020). *The state of the world's forests: Forests, biodiversity, and people*. Food and Agriculture Organization.

Garnett, S. T., Burgess, N. D., Fa, J. E., Fernández-Llamazares, Á., Molnár, Z., Robinson, C. J., Watson, J. E. M., Zander, K. K., Austin, B., Brondizio, E. S., Collier, N. F., Duncan, T., Ellis, E., Geyle, H., Jackson, M. V., Jonas, H., Malmer, P., McGowan, B., Sivongxay, A., & Leiper, I. (2018). A spatial overview of the global importance of Indigenous lands for conservation. *Nature Sustainability*, *1*, 369–374.

Geist, H. J., & Lambin, E. F. (2002). Proximate causes and underlying driving forces of tropical deforestation. *Bioscience*, *52*, 143–150.

Geldmann, J., Coad, L., Barnes, M. D., Craigie, I. D., Woodley, S., Balmford, A., Brooks, T. M., Hockings, M., Knights, K., Mascia, M. B., McRae, L., & Burgess, N. D. (2018). A global analysis of management capacity and ecological outcomes in terrestrial protected areas. *Conservation Letters*, *11*, 1–10.

Geldmann, J., Manica, A., Burgess, N. D., Coad, L., & Balmford, A. (2019). A global-level assessment of the effectiveness of protected areas at resisting anthropogenic pressures. *Proceedings of the National Academy of Sciences of the United States of America*, *116*, 23209–23215.

Ghoddousi, A., Loos, J., & Kuemmerle, T. (2022). An outcome-oriented, social–ecological framework for assessing protected area effectiveness. *Bioscience*, *72*, 201–212.

Gill, D. A., Mascia, M. B., Ahmadia, G. N., Glew, L., Lester, S. E., Barnes, M., Craigie, I., Darling, E. S., Free, C. M., Geldmann, J., Holst, S., Jensen, O. P., White, A. T., Basurto, X., Coad, L., Gates, R. D., Guannel, G., Mumby, P. J., Thomas, H., … Fox, H. E. (2017). Capacity shortfalls hinder the performance of marine protected areas globally. *Nature*, *543*, 665–669.

Global Forest Watch (GFW). (2022). Tree cover loss in Mexico. www.globalforestwatch.org

Graham, V., Geldmann, J., Adams, V. M., Grech, A., Deinet, S., & Chang, H.-C. (2021). Management resourcing and government transparency are key drivers of biodiversity outcomes in Southeast Asian protected areas. *Biological Conservation*, *253*, 108875. https://doi.org/10.1016/j.biocon.2020.108875

Hansen, M. C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S. A., Tyukavina, A., Thau, D., Stehman, S. V., Goetz, S. K., Loveland, T. R., Kommareddy, A., Egorov, A., Chini, L., Justice, C. O., & Townshend, J. (2013). High-resolution global maps of 21st-century forest cover change. *Science*, *342*(6160), 850–853.

Hastie, T., Tibshirani, R., & Friedman, J. (2017). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer.

Jonas, H. D., Ahmadia, G. N., Bingham, H. C., Briggs, J., Butchart, S. H. M., Cariño, J., Chassot, O., Chaudhary, S., Darlling, E., de Gemmis, A., & Dudley, N. (2021). Equitable and effective area-based conservation: Towards the conserved area paradigm. *Parks*, *27*, 71–84. https://doi.org/10.2305/IUCN.CH.2021.PARKS-27-1HJ.en

Jones, V. O., Fuller, R. A., Allan, J. R., Maxwell, S. L., Negret, P. J., & Watson, J. E. M. (2018). One-third of global protected land is under intense human pressure. *Science*, *360*, 788–791.

Joppa, L. N., & Pfaff, A. (2009). High and far: Biases in the location of protected areas. *PLoS ONE*, *4*, 1–6.

Kere, E. N., Choumert, J., Motel, P. C., Combes, J. L., Santoni, O., & Schwartz, S. (2017). Addressing contextual and location biases in the assessment of protected areas effectiveness on deforestation in the Brazilian Amazônia. *Ecological Economics*, *136*, 148–158.

Kuhn, M. (2008). Building predictive models in R using the caret package. *Journal of Statistical Software*, *28*, 1–26. https://doi.org/10.18637/jss.v028.i05

Laurance, W. F., Goosem, M., & Laurance, S. G. W. (2009). Impacts of roads and linear clearings on tropical forests. *Trends in Ecology & Evolution*, *24*, 659–669.

Macura, B., Secco, L., & Pullin, A. S. (2015). What evidence exists on the impact of governance type on the conservation effectiveness of forest protected areas? Knowledge base and evidence gaps. *Environmental Evidence*, *4*, 1–29.

Mas, J., & Cuevas, G. (2013). *Análisis de los patrones de deforestación en México con regresiones ponderadas geográficamente*. Anais XVI Simpósio Brasileiro de Sensoriamento Remoto- SBSR, Foz do Iguaçu, PR: INPE, pp. 6292–6299.

Mas, J. F., & Cuevas, G. (2015). Identifying local deforestation patterns using geographically weighted regression models. In C. Grueau & R. J. Gustavo (Eds.), *Geographical information systems theory, applications and management* (pp. 36–49). Springer.

Molnar, C., Bischl, B., & Casalicchio, G. (2018). iml: An R package for interpretable machine learning. *Journal of Open Source Software*, *3*(26), 786. https://doi.org/10.21105/joss.00786

Negret, P. J., Di Marco, M., Sonter, L. J., Rhodes, J., Possingham, H. P., & Maron, M. (2020). Effects of spatial autocorrelation and sampling design on estimates of protected area effectiveness. *Conservation Biology*, *34*, 1452–1462.

Nolte, C., & Agrawal, A. (2013). Linking management effectiveness indicators to observed effects of protected areas on fire occurrence in the Amazon rainforest. *Conservation Biology*, *27*, 155–165.

Oldekop, J. A., Holmes, G., Harris, W. E., & Evans, K. L. (2016). A global assessment of the social and conservation outcomes of protected areas. *Conservation Biology*, *30*, 133–141.

Olson, D. M., Dinerstein, E., Wikramanayake, E. D., Burgess, N. D., Powell, G. V. N., Underwood, E. C., D'amico, J. A., Itoua, I., Strand, H. E., Morrison, J. C., Loucks, C. J., Allnutt, T. F., Ricketts, T. H., Kura, Y., Lamoreux, J. F., Wettengel, W. W., Hedao, P., & Kassem, K. R. (2001). Terrestrial ecoregions of the world: A new map of life on Earth. *Bioscience*, *51*, 933–938.

Pettorelli, N., Safi, K., & Turner, W. (2014). Satellite remote sensing, biodiversity research and conservation of the future. *Philosophical Transactions Royal Society B: Biological Sciences*, *369*, 20130190. https://doi.org/10.1098/rstb.2013.0190

Pfaff, A., Robalino, J., Lima, E., Sandoval, C., & Herrera, L. D. (2014). Governance, location and avoided deforestation from protected areas: Greater restrictions can have lower impact, due to differences in location. *World Development*, *55*, 7–20.

Pfaff, A., Santiago-Ávila, F., & Joppa, L. (2017). Evolving protected-area impacts in Mexico: Political shifts as suggested by impact evaluations. *Forests*, *8*, 1–14.

Porter-Bolland, L., Ellis, E. A., Guariguata, M. R., Ruiz-Mallén, I., Negrete-Yankelevich, S., & Reyes-García, V. (2012). Community managed forests and forest protected areas: An assessment of their conservation effectiveness across the tropics. *Forest Ecology and Management*, *268*, 6–17.

Powlen, K. A., Gavin, M. C., & Jones, K. W. (2021). Management effectiveness positively influences forest conservation outcomes in protected areas. *Biological Conservation*, *260*, 109192.

Robinson, B. E., Holland, M. B., & Naughton-Treves, L. (2014). Does secure land tenure save forests? A meta-analysis of the relationship between land tenure and tropical deforestation. *Global Environmental Change*, *29*, 281–293.

Roy Chowdhury, R. (2006). Landscape change in the Calakmul Biosphere Reserve, Mexico: Modeling the driving forces of smallholder deforestation in land parcels. *Applied Geography*, *26*, 129–152.

Salafsky, N., Salzer, D., Stattersfield, A. J., Hilton-Taylor, C., Neugarten, R., Butchart, S. H. M., Collen, B., Cox, N., Master, L. L., O'connor, S., & Wilkie, D. (2008). A standard lexicon for biodiversity conservation: Unified classifications of threats and actions. *Conservation Biology*, *22*, 897–911.

Sarkki, S., Rantala, L., & Karjalainen, T. P. (2015). Local people and protected areas: Identifying problems, potential solutions and further research questions. *International Journal of Environment and Sustainable Development*, *14*, 299–314.

Schleicher, J., Eklund, J., Barnes, M. D., Geldmann, J., Oldekop, J. A., & Jones, J. P. G. (2019). Statistical matching for conservation science. *Conservation Biology*, *34*, 538–549.

Shumba, T., De Vos, A., Biggs, R., Esler, K. J., & Clements, H. S. (2021). The influence of biophysical and socio-economic factors on the effectiveness of private land conservation areas in preventing natural land cover loss across South Africa. *Global Ecology and Conservation*, *28*, e01670.

Sims, K. R. E., & Alix-Garcia, J. M. (2017). Parks versus PES: Evaluating direct and incentive-based land conservation in Mexico. *Journal of Environmental Economics and Management*, *86*, 8–28.

Skutsch, M., Mas, J. F., Bocco, G., Bee, B., Cuevas, G., & Gao, Y. (2014). Deforestation and land tenure in Mexico: A response to Bonilla-Moheno et al. *Land Use Policy*, *39*, 390–396.

Solorzano, C. R., & Fleischman, F. (2018). Institutional legacies explain the comparative efficacy of protected areas: Evidence from the Calakmul and Maya Biosphere Reserves of Mexico and Guatemala. *Global Environmental Change*, *50*, 278–288.

Strobl, C., Boulesteix, A., Kneib, T., Augustin, T., & Zeileis, A. (2008). Conditional variable importance for random forests. *BMC Bioinformatics*, *9*, 307.

Strobl, C., Malley, J., & Tutz, G. (2009). An introduction to recursive partitioning: Rationale, application and characteristics of classification and regression trees, bagging and random forest. *Psychology Methods*, *14*, 323–348.

Turner, W., Spector, S., Gardiner, N., Fladeland, M., Sterling, E., & Steininger, M. (2003). Remote sensing for biodiversity science and conservation. *Trends in Ecology & Evolution*, *18*, 306–314.

UNEP-WCMC, IUCN, & NGS. (2020). *Protected Planet Live Report 2020*. https://livereport.protectedplanet.net/

Vaca, R. A., Golicher, D. J., Rodiles-Hernández, R., Castillo-Santiago, Á., Bejarano, M., & Navarrete-Gutiérrez, D. A. (2019). Drivers of deforestation in the basin of the Usumacinta River: Inference on process from pattern analysis using generalised additive models. *PLoS ONE*, *14*, 1–21.

Willmott, C. J., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research*, *30*, 79–82.

Wolf, C., Levi, T., Ripple, W. J., Zárrate-Charry, D. A., & Betts, M. G. (2021). A forest loss report card for the world's protected areas. *Nature Ecology & Evolution*, *5*, 520–529. https://doi.org/10.1038/s41559-021-01389-0

Yang, H., Viña, A., Winkler, J. A., Chung, M. G., Huang, Q., Dou, Y., McShea, W. J., Songer, M., Zhang, J., & Liu, J. (2021). A global assessment of the impact of individual protected areas on preventing forest loss. *Science of the Total Environment*, *777*, 145995. https://doi.org/10.1016/j.scitotenv.2021.145995

Zafra-Calvo, N., & Geldmann, J. (2020). Protected areas to deliver biodiversity need management effectiveness and equity. *Global Ecology and Conservation*, *22*, e01026.

Zafra-Calvo, N., Pascual, U., Brockington, D., Coolsaet, B., Cortes-Vazquez, J. A., Gross-Camp, N., Palomo, I., & Burgess, N. D. (2017). Towards an indicator system to assess equitable management in protected areas. *Biological Conservation*, *211*, 134–141.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Powlen, K. A., Salerno, J., Jones, K. W., & Gavin, M. C. (2023). Identifying socioeconomic and biophysical factors driving forest loss in protected areas. *Conservation Biology*, *37,* e14058. https://doi.org/10.1111/cobi.14058