Landsc Ecol
https://doi.org/10.1007/s10980-023-01694-y

RESEARCH ARTICLE

Landscape context importance for predicting forest transition success in central Panama

Giulia Bardino · Gianrico Di Fonzo · Kendra Walker · Marcello Vitale · Jefferson S. Hall

Received: 16 February 2023 / Accepted: 21 May 2023

Abstract
Context Naturally recovering secondary forests are frequently re-cleared before they can recover to pre-disturbance conditions. Identifying landscape factors associated with persistence success will help planning cost-efficient and effective forest restoration.
Objectives The ability of secondary forest to persist is an often undervalued requisite for long-term ecosystem restoration. Here we identify the landscape context for naturally regenerated forests to persist through time within central Panama.
Methods We developed a random forest classification (RFC) calibration method to identify areas with high (≥ 90%) and low (< 90%) likelihood of forest persistence success based on their spatial relation with nine landscape explanatory variables.
Results The RFC model discriminated between secondary forests areas that persisted and did not persisted with an error rate of 2%. By tuning, we obtained a precision of 0.94 (94%) in the validation test. The two most important explanatory variables involved in the persistence dynamic were elevation and distance to the nearest rural area. Naturally regenerated forests lasted longer in patches that were closer to both Gatun and Alajuela Lakes as to protected areas, but further from rural communities, roads, urban areas and in patches with higher elevation and steeper slopes.
Conclusion By tracking remote sensed, landscape context metrics of easy collection, we developed a prediction map of central Panama areas with high (≥ 90%) and low (> 90%) probability of natural forest regeneration and persistence success within the next 30 years. This map represents a basis for management decisions and future investigations for effective, long-term forest-landscape restoration.

Keywords Tropical forest-landscape · Landscape restoration · Landscape analysis · Machine learning · Panama Canal Watershed · Passive regeneration · Human-modified landscapes · Persistence · Forest transition

G. Bardino (✉) · M. Vitale
Department of Environmental Biology, Sapienza University of Rome, Rome, RM, Italy
e-mail: giulia.bardino@uniroma1.it

M. Vitale
e-mail: marcello.vitale@uniroma1.it

G. Di Fonzo
Department of Statistical Sciences, Sapienza University of Rome, Rome, RM, Italy
e-mail: gianrico.difonzo@uniroma1.it

K. Walker
University of California, Santa Barbara, CA 93106, USA
e-mail: kendrawalker@ucsb.edu

J. S. Hall
Smithsonian Institution Forest Global Earth Observatory (ForestGEO), Smithsonian Tropical Research Institute, Balboa, Ancón, Panama
e-mail: hallje@si.edu

Published online: 27 May 2023

Introduction

Protecting and restoring forests is essential to meeting the Paris Climate Agreement and the new Global Biodiversity Framework goals, conserving biodiversity, and addressing food security and livelihood needs (Chazdon et al. 2017a; UN SDG 2017; Lewis et al. 2019; CBD 2022). Many reforestation activities are being conducted worldwide, at both local and landscape scales. While local-based projects could help determine the best specific solution for the target area, landscape-scale projects require a broader perspective and planning for restoration activities (Mansourian and Vallauri 2014). The forest landscape restoration (FLR) is a holistic approach that considers the relationship between landscape structure (the spatial patterning of different land uses across space) and landscape functioning (ecological processes), and holds promises to achieve both ecological and social goals. It also provides long-term, multi-objective and large-scale means to implement international targets in field interventions (Turner 2005; Hall et al. 2011, 2015; Holl 2017; Stanturf and Mansourian 2020). Within the different forest recovery scenarios, natural recovery, or "passive restoration" sensu Holl and Aide (2011), can play a major role in large-scale landscape restoration in tropical regions (van Breugel et al. 2013a, b; Chazdon and Guariguata 2016; Chazdon and Uriarte 2016; Hall et al. 2022).

Significant progress in recovering ecosystem services found in mature tropical forests have been observed in as little as two decades of naturally recovering secondary forests. For example, Martin et al. (2013) found aboveground carbon pools can recover to xxxx in yyy time while Rozendaal et al. (2019) reported an 80% recovery of tree species diversity in 20 years across the Neotropics. Poorter et al. (2021) found significant recovery of functional traits within this same time period for a sub sample of these same sites. Working at Agua Salud in Panama on water related ecosystem services, Hassler et al. (2011) found a significant recovery in saturated hydraulic conductivity, a measure of soil water infiltration in 12 to 15 years of natural forest recovery in the top layer of the mineral soils. Birch et al. (2021) reported recovery of hydraulic flow paths at the catchment scale at Agua Salud up to 30 cm depth within a decade of forest recovery while Chavarria et al. (2021) report a significant shift in stream water bacterial communities and measures of water quality at the same site and time scale. This means that to achieve an effective restoration of forest landscape, forests not only need suitable conditions to grow, but also to persist in time (Sloan 2022), where persistence is defined as a time-dependent process where forests grow, mature and continue to be present in time, without being cut again. Persistence of naturally regenerating forests could translate into a "forest transition" event, which has been increasingly observed in the tropics since the 1990s (e.g., Rudel et al. 2002; Aide et al. 2013; Sloan 2022). Several studies have found that newly recovered forests are frequently re-cleared before they can recover to pre-disturbance conditions (Breugel et al. 2013a, b; Mbow et al. 2017; Meli et al. 2017; Brondizio et al. 2019; Reid et al. 2019). Detecting this dynamic and identifying the context where naturally recovered forest do not persist could provide support for a more effective forest landscape restoration planning (Smith et al. 2003; Fagan et al. 2013; Sloan 2016; Reid et al. 2017; Schwartz et al. 2017, 2020; Borda-Niño et al. 2020). To our knowledge, no studies have yet provided spatially explicit insights and/or landscape attributes related to the persistance of naturally recovering secondary forests in central Panama.

Central Panama is a dynamic and heterogeneous area with a diversity of ecosystems and habitats in a relatively small area, including tropical lowland and pre-montane moist forests, freshwater wetlands, and mangroves (Holdridge et al. 1967; Ibanez et al. 2002), where its provision and dependence on ecosystem services are well known. The region hosts one of the most important inland commercial waterways, the Panama Canal, which is important for international trade and a crucial economic driver for the country (Adamowicz et al. 2019). It contains two large artificial freshwater lakes—Gatun Lake and Alhajuela Lake—both created to ensure canal operations and freshwater supply for Panama City, the city of Colon, and numerous towns in between (Hall et al. 2015). The long-term maintenance and sustainability of central Panama biodiversity and water supply depends on the preservation of the tropical forest cover; the latter is accomplished in part through a system of 21 protected and/or restricted areas (see, e.g. Hall et al. 2022). Nevertheless, it has experienced rapid economic growth over the past decade, followed by intense land use/land-cover changes (Heckadon

Moreno at al. 1999; Condit et al. 2001; Ibanez et al. 2002; Walker et al. 2020) which led to the construction of roadways and the expansion of pastures. This has stimulated the decrease of forest cover capable of regulating water supply and maintaining connectivity between the protected areas (Rompre et al. 2008).

As people have migrated to cities in central Panama to take advantage of economic opportunities found there, significant areas that had been cleared for agriculture have regenerated (Wright and Samaniego 2008; Sloan 2015) and also reduced the rate of deforestation by over 80% in recent decades (Walker 2020; Hall et al. 2022). It is therefore an area where the recovery and the conservation of forests is key to support ecosystem services supply and where it is crucial to assess wether and how the natural expansion of secondary forests is allowed to establish into forest transition. Yet, we do not know how much of the naturally recovered secondary forests persisted longer than 20 years, much less the relationship between landscape attributes and the probability to persist.

In this paper, we aim to identify the landscape context for naturally recovered forest to persist in time within central Panama. We studied the forest areas that naturally recovered in the two decades between 1990 and 2010 and persisted until 2020 using the random forest classification (RFC) method and relying on the tuning of important RFC parameters, to identify areas with higher likelihood of persistence success.

Materials and methods

Study area

Our study area is identical to that of Hall et al. (2022) and represents 23% of the country's land area (Fig. 1). The study area includes the Panama Canal Watershed (PCW), an area of critical relevance for biodiversity conservation and economic development (Ibanez et al. 2002; Adamowicz et al. 2019). Previous projections of deforestation and carbon storage have found that the unrealistic scenario of halting deforestation and allowing all available forested land within the study region to be protected, to grow, and to persist, central Panama could sequester between 56.0% of its national goal. In contrast, were deforestation rates to revert to those before the year 2000, it would have devastating consequences releasing an additional 73 million Mg CO2e by 2050, making it extremely difficult for Panama to achieve its planned land-based carbon sequestration objective (Hall et al. 2022).

Data processing and quantification of persistence

A unique 30-year time-series data set (1990–2020) of central Panama vegetation cover compiled by Walker (2020) was used to identify areas where natural regeneration occurred spontaneously and persisted between 1990 and 2020 (Fig. 2a–c). Vegetation cover layers were based on epochal composite of dry-season Landsat 4, 5, 7, 8 images with less than 70% total cloud cover and at least 10% clear pixels images (30 m). All images were downloaded as Landsat Surface Reflectance Products through the EROS science processing architecture (ESPA) interface with standard pre-processing for L1T data to correct for positional and atmospheric conditions and hand-classified in five macro-categories and 23 sub-categories (Table 1). Sub-categories provide additional information regarding habitat, disturbance and land-use, allowing for a better understanding of land-cover change dynamics. Unlike previous land-cover maps of Panama, which applied dichotomous or categorical definitions of forests, these maps are based on an ordinal scale to define vegetation, distinguishing different age classes of secondary forests: low vegetation (1–2 yrs), medium vegetation (3–5 yrs), high vegetation (6–20 yrs) and forest (≥ 20 yrs), and therefore capturing small and ephemeral clearings. Such gradient reflects the natural process of regeneration and acknowledges the fuzzy boundaries between forest and non-forest classes. The Walker forest cover and forest-cover change maps were designed to achieve an adequate spatial and temporal resolution; before these maps became available Panama lacked a consistent dataset by which to assess baselines for land-cover changes.

The regeneration and persistence dynamics were observed working at macro-categories level (NOVEG, LOWVEG, MEDVEG, HIGHVEG and FOREST), while the pre-selection of eligible patches also considered sub-categories. We adopted a patch-based approach, where a patch is defined as a minimum 0.5 ha width, homogeneous area that differs from its surroundings (Forman and Godron 1981; Farina 2000). The patches classified as HIGHVEG

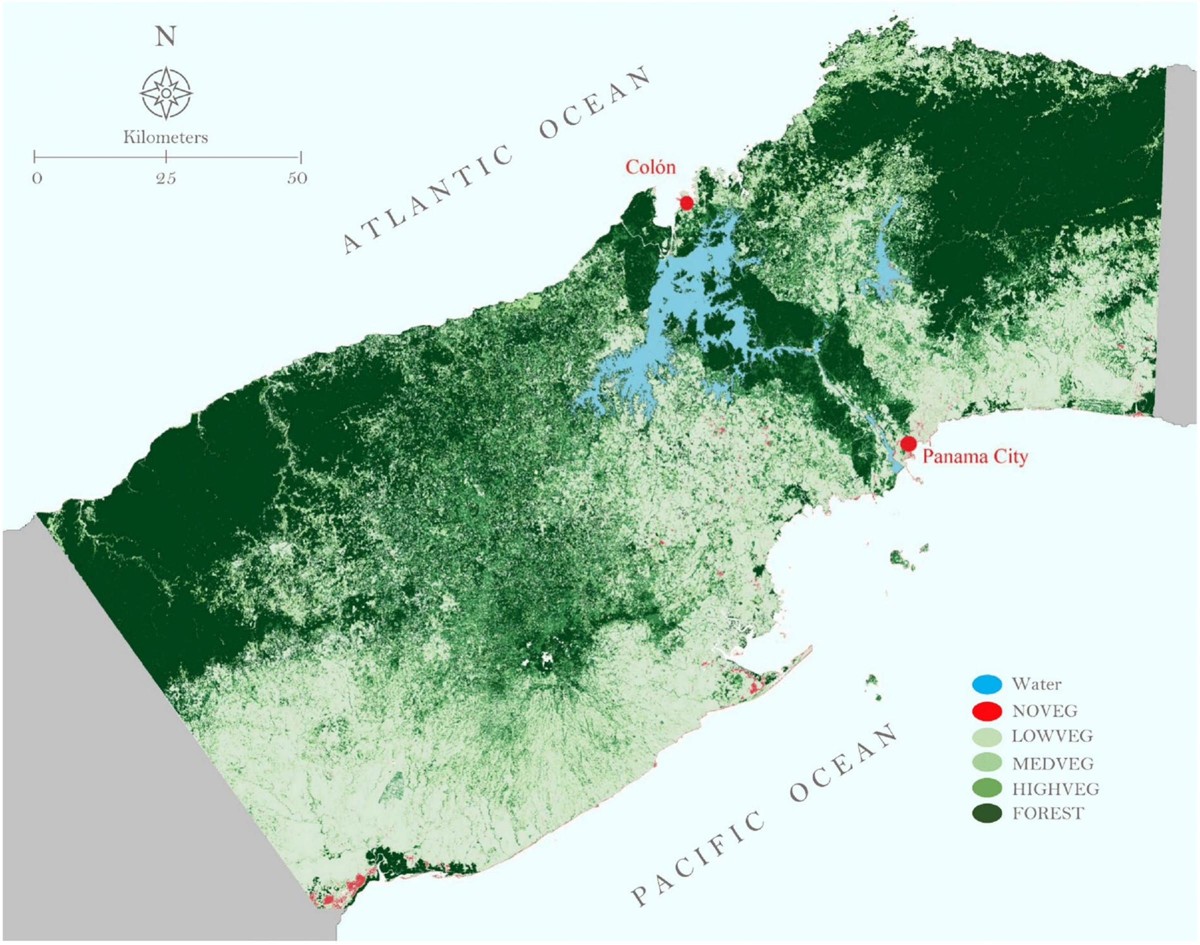


Fig. 1 Study area (Panama Canal Watershed)

in 1990 were excluded from the dataset to limit the fuzziness and highlight the transition from non-forest to forest. The patches classified as NOVEG, as LOWVEG sub-categories *low crops rainfed* (18) and *low crops irrigated* (28), and as MEDVEG sub-category *high crops* (48) were not considered in the analysis as our goal was to assess the forest-landscape natural regeneration potential in a multi-functional landscape, without altering the presence of urban/agricultural land-uses. Finally, we excluded the patches classified as FOREST sub-category *evergreen plantations* (68) to focus on natural regeneration only. We designated patches with 0 = *natural recovery BUT NO persistence* or 1 = *natural recovery AND persistence* histories as follows: Patches 0 were identified as ≥ 0.5 ha patches classified as LOWVEG or MEDVEG in 1990, as FOREST in 2010 and again as LOWVEG or MEDVEG in 2020 land-cover map; patches 1 were identified as ≥ 0.5 ha patches classified as LOWVEG or MEDVEG in 1990 land cover map and as FOREST in both subsequent 2010 and 2020 land cover maps, for at least 50% of their total area. Our sample was composed by 1,234 patches ranging from 0.5 to 35,978 ha, representing a total area of 150,381 ha.

Explanatory variables

While climatic and edaphic factors affect plant germination, growth, and survival, leading variations in the regeneration process across regions, landscape context determines where regeneration is permitted to occur and persist, playing a central role in driving successional pathways within human-modified

Fig. 2 Panama Canal Watershed Land-Cover Maps: **a** 1990; **b** 2010; **c** 2020

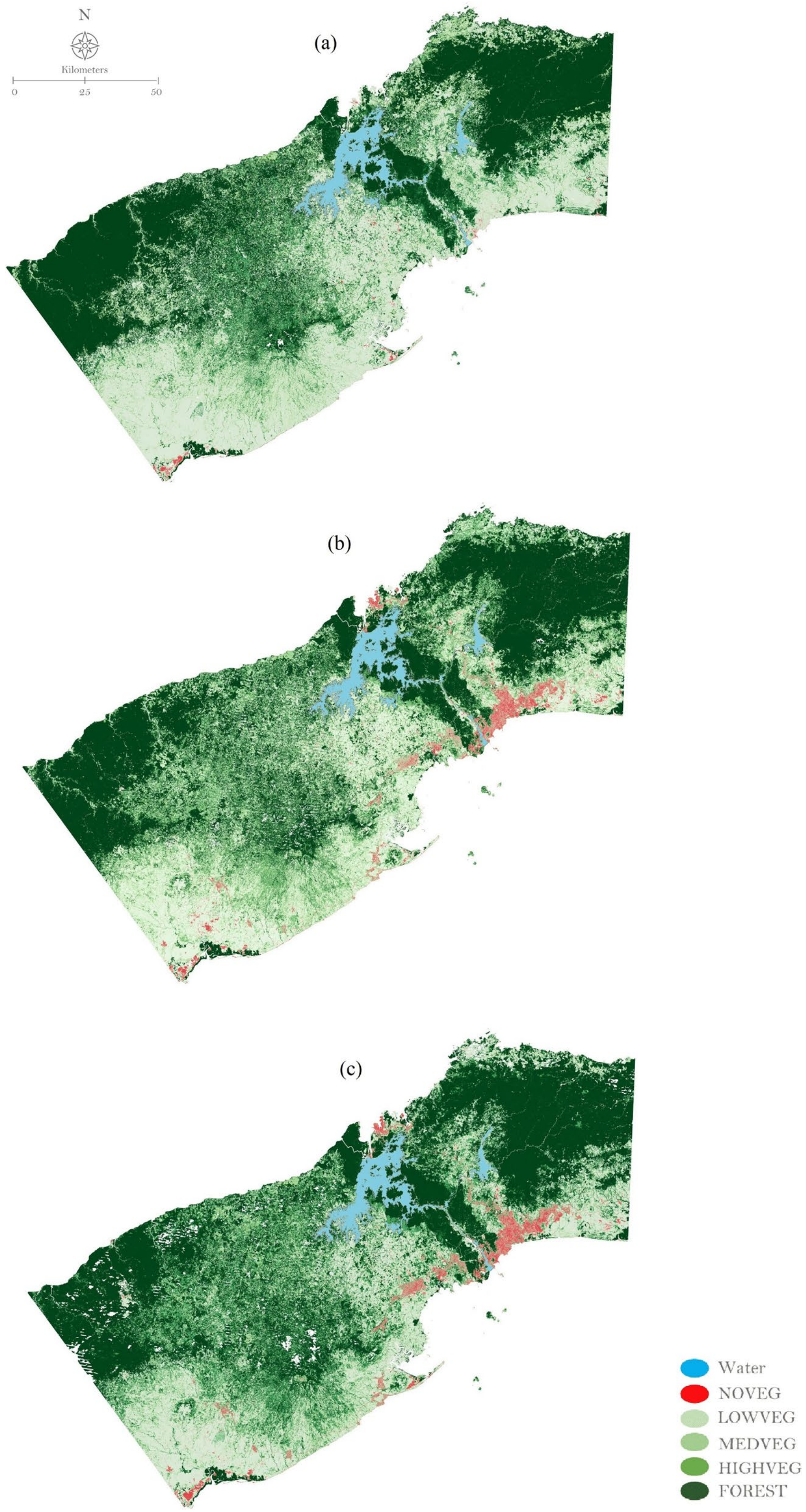


Table 1 Macro and Sub-categories of land-cover maps

Macro-categories	Summary	Sub-categories: vegetation		Sub-categories: wet detail		Sub-categories: anthropological detail	
NOVEG	Water/Bare/Beach/Built			**− 7**	Water	**1**	Built/Bare
				7	Water and bare		
LOWVEG	Low growth (1–2 yrs)/Pasture/Agriculture/Low wetland	**10**	Low veg, seasonally dry	**27**	Low wetland	**18**	Low crops rainfed
		20	Low veg			**28**	Low crops irrigated
MEDVEG	Medium growth (3–5 yrs)/Shrubby crops/Medium gallery vegetation	**30**	Thin gallery/edge veg	**37**	Low to Med gallery	**48**	High crops
		40	Med veg	**47**	Med gallery		
HIGHVEG	Young high vegetation (6–10 yrs)/Immature forest (11–20 yrs)	**50**	High veg	**57**	High gallery	**58**	Deciduous plantations
FOREST	Fragmented/disturbed forest/Forest within 50 m of edge/Secondary forest, cleared > 20 yrs prior/Riparian trees (wet or heavily shaded)/Forested wetland/Undisturbed upland forest	**60**	Disturbed forest	**67**	Mature gallery	**68**	Evergreen plantations
		70	Forest wetland	**77**	Mixed water & mangrove		
		80	Mature upland forest	**87**	Gallery in mature forest		

landscapes, such as tropical ones (Jakovac et al. 2021). Nine remote-sensed, easily collectable explanatory variables linked to anthropic use and access of the area were measured for each of the 1234 patches included in the dataset (Table 2): *mean patch elevation* (*meanElev*); *mean patch slope* (*meanSlope*); *minimum distance from the patch centroid to the nearest road* (*RoadsDista*); *minimum distance from the patch centroid to the nearest river* (*Hydro123D*); *minimum distance from the patch centroid to the nearest urban area* (*UrbanDista*), defined by the National Institute of Statistics and Census of Panama as the populated places that concentrates 1500 or more inhabitants; *minimum distance from the centroid to the nearest rural area* (*RuralDist*), defined by the National Institute of Statistics and Census of Panama as the populated places that concentrates > 1500 inhabitants; *minimum distance from the patch centroid to either Gatun or Alajuela lake* (*LLDistDi*); *minimum distance from the patch centroid to the nearest protected area* (*PADistanc*); *mean distance between the centroid and the patch's edge* (*distpole*). The selection of explanatory variables was based on a previous study on deforestation scenarios and future carbon dynamics conducted in the PCW (Hall et al. 2022). The variable *distpole*, which was not used in the previous study, was included to consider the patch nature of our samples and their variability in terms of extent..

Data analysis

Random forest classification

We evaluated the associations between secondary forest persistence success and landscape-context predictors using the Machine-Learning (ML) based method of random forest classification (RFC) (Breiman 2001). The advantage of ML over traditional statistical techniques is the ability to model highly dimensional and non-linear data—such as ecological data (Knudby et al. 2010; Thessen 2016)—with complex interactions. There are several types of tasks that ML techniques can perform. The task we were interested in was the classification, the process of predicting the class of given data point (Kotsiantis 2007). RFC is the most widespread classifier ML technique, especially when working with land use/land cover change

Table 2 Independent/explanatory variables

Variables	Dataset name	Description	Source
Slope	*meanSlope*	Mean slope in degrees	Digital elevation model of the Republic of Panamá, generated by NASA SRTM program, 30 m
Elevation	*meanElevation*	Mean elevation in meters	Digital elevation model of the Republic of Panamá, generated by NASA SRTM program, 30 m
Distance to rivers	*Hydro123D*	Euclidean distance to nearest river (first, second and third order)	Hydrology,Instituto Geográfico Nacional Tommy Guardia, 1:50,000
Distance to roads	*RoadsDista*	Euclidean distance to nearest road	Roads, open street map, 1:50,000
Distance to protected areas	*PADistanc*	Euclidean distance to nearest protected area	ANAM (National Environmental Authority, now MiAmbiente), 2006
Distance to Gatun Lake or Alajuela Lake	*LLDist*	Euclidean distance to nearest Lake (Gatun Lake or Alajuela Lake)	Lakes and Lagunes, Instituto Geográfico Nacional Tommy Guardia, 1:50,000
Distance to urban areas	*UrbanDista*	Euclidean distance to nearest urban area	Urban Areas, Instituto Geográfico Nacional Tommy Guardia, 1:25,000
Distance to rural areas	*RuralDista*	Euclidean distance to nearest rural area	Rural Areas, Instituto Geográfico Nacional Tommy Guardia, 1:25,000
Distance centroid	*distpole*	Mean distance between centroid and patch edge	Fragstats 4.2

(LULC) (Vitale et al. 2014; Lowe and Kulkarni 2015; Pelletier et al. 2016; Nguyen et al. 2018; Talukdar et al. 2020). RFC is a supervised algorithm based on decision trees and improved bagging and bootstrap techniques that utilizes several classifiers to work together to identify the class label for each observation (Yang et al. 2010). In addition, RFC provides variable importance measures, which can be used to identify most relevant features and/or their effect (Archer and Kimes 2008). However, like most classifiers RFC faces problems when learning from an extremely unbalanced training data set (Zakariah 2014). As it is constructed to minimize the overall error rate, it will tend to focus more on the prediction accuracy of the majority class, which often results in poor accuracy for the minority class. However, there are several techniques that can be used with RFC to solve class imbalance problems (Chen 2004; Kuhn and Johnson 2013; More and Rana 2017). We solved the class unbalance problems to the training set, and we validated the model on the unbalanced test set to assess the performance of the model with real frequencies. The study used the *randomForest* package running in Rstudio (CRAN) (Liaw and Wiener 2002).

Training dataset balancing The dataset was composed of 0 = 182 and 1 = 1052 units. We divided the dataset into training and test sets following the rule of 70–30%, as it is demonstrated to represent the ratio with the best performance in ML models (Nguyen et al. 2021). The balance problem was solved for the training set through up-sampling technique (*upSample* function in the *caret* package running in Rstudio (CRAN)). Up-sampling is an oversampling technique that randomly selects and duplicates observations in the minority class to balance the dataset. We obtained a balanced training set with a total of 1474 samples. The optimal number of decision trees (n = 1000) and number of explanatory variables randomly sampled as candidates at each split (n = 3) were evaluated based on the lowest out of bag (OOB) error. The RFC is trained through bagging technique, where each new decision tree is fit from a subsample of the training units. The OOB error is the average classification error of the remaining training units not included in the subsample. That allows to fit and also validate the RFC while being trained (Breiman 1996).

Tuning The tuning consists of the identification of a set of optimal RFC hyperparameters' values to maximize the model's performance and produce better outputs. Hyperparameters are parameters whose values control the learning process of the model and can be tuned to increase some specific metrics like accuracy or precision. In our case, the RFC model was validated on an unbalanced test set to assess the real frequencies of 0 and 1 found in the PCW, within the past 30 years. Since RFC is designed to minimize the overall error rate, with an unbalanced dataset it will tend to focus more on the prediction accuracy of the majority class, which often results in poor accuracy for the minority class. For this reason, accuracy was not a suitable measure for our model's performance. We instead focused on the tuning of *precision* for 1, and *recall* for 0 (Table 3a, b). Here *precision* is defined as the ability of a classifier not to label "as TRUE" an instance that is actually "FALSE". For each class, it is defined as the ratio of *real trues* (RT) to the sum of *real trues* and *false trues* (FT), Eq. (1):

$$\text{Precision} = \frac{\text{RT}}{\text{RT} + \text{FT}} \tag{1}$$

Thus, the *precision* measures how good the model is at predicting a specific category. The *recall* (also known as *sensitivity*) is the ability of a classifier to find all the RT instances for the specified class. It is defined as the ratio of RT to the sum of RT and *false false* (FF), Eq. (2):

$$\text{Recall} = \frac{\text{RT}}{\text{RT} + \text{FF}} \tag{2}$$

Recall is a measure of how many times the model was able to detect a specific category. To maximize the *precision* of 1, we needed to improve the *Recall* of 0. In other words, we needed to reduce as much as possible the misclassification errors (FT) of 0. A straightforward approach to improve the performance of a classifier that predicts probabilities with unbalanced dataset is to tune the threshold used to map probabilities to class labels. ML algorithms predicts the probability or scoring of class membership by using a threshold. All values equal to or greater than the selected threshold are mapped to one class and all the other values are mapped to another class. We therefore tested several thresholds until identifying 0.90 as the optimal one to improve the *precision* of 1 and *recall* of 0, as described above. Scores above 0.90 were classified as 1, those below as 0.

Tuning one of these quantities nevertheless entails a trade-off: as we increase the *precision* of a class, we decrease the *recall* for the same class, and viceversa. In our case, this trade-off leads to low *sensitivity* for class 1 (several 1 RT were lost), and consequently low *precision* for class 0 (a high percentage of the total patches that were predicted as 0, were actually 1).

Relative importance of independent variables and partial dependencies To evaluate the specific behavior of each of the explanatory variables in the persistence dynamic we did not rely on the thresholds derived by the tree graph, rather on the relative importance of the independent variables and their partial dependencies. The underlying RFC algorithms are trained by randomly selected subsets of data, which makes thresholds potentially different for each of the "n" trees produced by the RFC, and therefore approximate for the purpose of our work. The relative importance of independent variables indicates their predictive power. It can be used to sort variable from most to least predictive, allowing to have more insights on the problem and to perform feature selection when there are too many input variables. The order of importance of the variables was measured through mean decrease

Table 3 Confusion matrix parameters for class 1 and 0

Confusion matrix 1	Real		
Predicted	0	1	Parameters
0	RF	FF	Precision: RT/RT + FT
1	FT	RT	Recall: RT/RT + FF
Confusion matrix 0	Real		
Predicted	0	1	Parameters
0	RT	FT	Precision: RT/RT + FT
1	FF	RF	Recall: RT/RT + FF

RF real false, *NRF* non-real false, *RT* real true, *NRT* non-real true in Gini (MDG). MDG is a measure of how important a variable is for estimating the class across all of the trees that make up the forest (Grömping 2009). A higher MDG indicates higher variable importance. The most important variables to the model will have the largest MDG. Conversely, the least important variable will have the smallest MDG values. In addition, partial dependencies plots contributed to clarifying which predictors have positive and negative effects on the persistence success. The RFC model allows graphical examination of partial dependencies of the model on each predictor. Partial dependencies plots show how the variables marginally affect the prediction based on the RFC model. Assuming all other variables fixed at the center, the values of a given predictor that increases along the y-axes indicate a positive effect on the prediction, while the values that decreases along the y-axes indicate a negative effect on the prediction.

Results

Secondary forest persistence analysis

From a total of ~ 62,203,260 ha of LOWVEG and MEDVEG occurring in 1990 in the study area, ~ 150,380 ha (0, 24%) were classified as FOREST (natural regeneration) in 2010. Within this forest subset, approximately ~ 36,400 ha (24%) were cleared again by 2020, while ~ 113.900 ha (76%) persisted.

Random forest classification performance analysis

As shown in Table 4, in the training phase the RFC model discriminated between the two classes (0–1) with an OOB error rate of about 2% and a class specific error of 0.001 (0.1%) for the 0, and 0.046 (4.6%) for the 1. By tuning, *Precision* of 1 and *Recall* of 0 reached, respectively, 0.94 (94%) and 0.83 (83%) in the validation test (Table 5).

Table 4 Random forest classification training performance

Number of decision trees (*ntree*)	1000		
Number of variables tried at each split (*mtry*)	3		
OOB estimate of error rate	2,37%		
Confusion matrix	Real		Class error
Predicted	0	1	
0	736	1	**0.001356852**
1	34	703	**0.046132972**

Relative importance of independent variables and partial dependencies plots

Variables are sorted and displayed in the Variable Importance plot (Fig. 3). The variables with more predictive power were *meanElev* and *RuralDist,* followed by *LLDistDi*, *UrbanDist*, *meanSlope*, *PADistanc*, *RoadsDista*, *distpole* and *Hydro123D*. Figure 4a–i shows the variables-specific partial dependencies plots, which represent how each variable marginally affected the persistence of secondary forests. The variables *meanElev*, *meanSlope*, *RoadsDista*, *UrbanDist* and *RuralDist* appeared to be directly proportional to the probability of 1, while *distpole*, *PADistanc* and *LLDistDi* were inversely proportional to the probability of 1.

Prediction map of secondary forests persistence

Based on the RFC model, we developed a prediction map of central Panama areas that have high (≥ 90%) and low (< 90%) probability of secondary forests persistence success within the next 20 years (Fig. 5). From a total of 286,593 ha of LOWVEG and MEDVEG present in 2020 within our study area, only 16% (~ 48,179 ha) have a probability ≥ 0.90% to recover and persist until, at least, 2050.

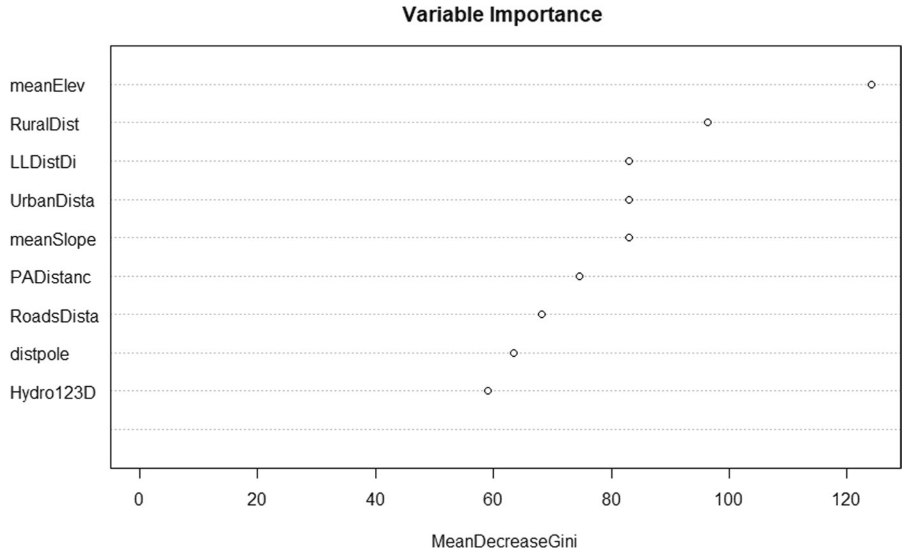
Discussion

Natural regeneration following land abandonment has been globally recorded starting from 1950 and is seen as an important opportunity for carbon sequestration and habitat restoration (Holl and Aide 2011; Chazdon

Table 5 Random forest classification test performance

Accuracy	0.5122			
95% CI	(0.4599, 0.5643)			
Mcnemar's test P-value	< 2e − 16			
Confusion matrix	Real		Precision	Recall
Predicted	0	1		
0	45	171	**0.46**	**0.83**
1	9	144	**0.94**	**0.20**

Fig. 3 Variables order of importance based on mean decrease of Gini index



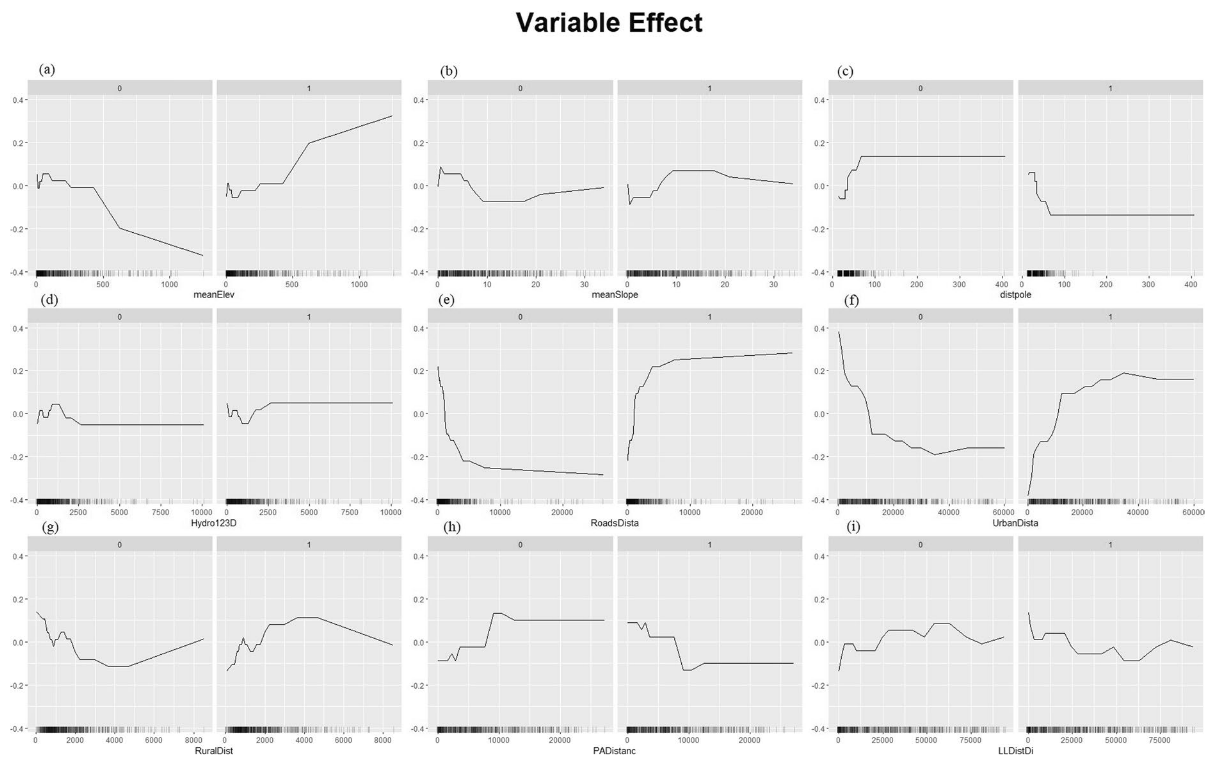


Fig. 4 Partial dependence plots for the marginal effects of the predictors on the random forest classification model: estimated probability of no persistence/persistence (0–1) versus the observed values of each predictor in the combined data: **a** elevation (*meanElev*); **b** slope (*meanSlope*); **c** area and shape (*distpole*); **d** nearest river (*Hydro123D*); **e** nearest road (*RoadsDista*); **f** nearest urban area (*UrbanDista*); **g** nearest rural area (*RuralDist*); **h** nearest protected area (*PADistanc*); **i** nearest lake (Panama Canal/Alajuela lake (*LLDistDi*)

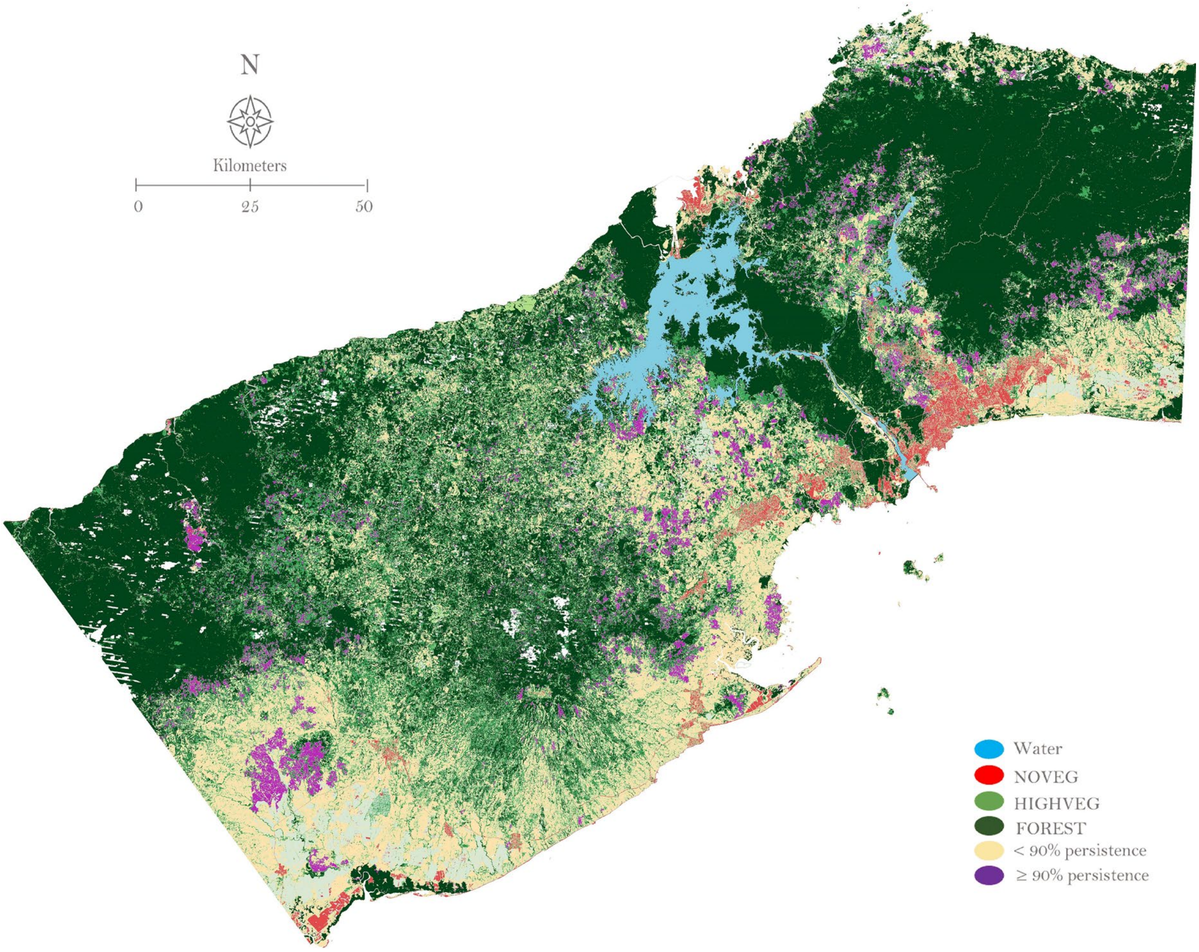


Fig. 5 Map of secondary forest persistence potential within 2020–2050

et al. 2017b; Hall et al. 2022). Yet, those benefits largely depend on the persistence of the forests, which has been found to be largely transient across the globe, with an average residence time of ~ 14.22 years (Crawford et al. 2022). Walker (2020) found significant secondary forest regeneration in central Panama. Our analysis, however, reveals that the secondary forest expansion recorded in the past 30 years did not necessarily persist. Indeed, 24% of the naturally regenerating secondary forest found in 2010 was lost by 2020. As pointed out by several authors, the potential ephemeral nature of secondary forests could result in a limited role for these forests in mitigating climate change and ecosystem service recovery. Piffer et al. (2022) found that the re-clearance of native secondary forests in the Atlantic Forests of Brazil greatly limited carbon sequestration, where second-growth forests could have sequestered over three times more carbon than the actual estimated carbon sequestration. Van Breugel et al. (2013a, b) noted that the proportion of species that could sustain reproductive populations (effective diversity) in a dynamic patchwork of young secondary forests found in central Panama depended on how old the secondary forests became and how many tree species were able to reach reproductive size within that time frame. By investigating the role of secondary forests in providing habitat for many vertebrates within tropical region, Acevedo-Charry et al. (2019) recorded slow recovery of species composition (at least 40 years) and strongly argued for the role of secondary forest persistence for the conservation of forest specialist's species as well as for source populations to recover in secondary forest sites. Identifying conditions that encourage the persistence of naturally regenerated forests is therefore crucial in assessing the real potential for ecosystem restoration within dynamic landscapes as those found in the tropics, as well as for informing evaluation of the specific supporting actions needed to reach the UN goals and to ensure a sustained ecosystem services provision. With a precision of 94%, we predicted areas with high (≥ 90%) and low (> 90%) probability of natural forest regeneration and persistence success within the next 30 years in central Panama (Fig. 5). These results demonstrate that a set of remotely sensed, easily collected landscape context variables could be used to predict persistence success of secondary forests. The two variables with more predictive power were found to be the elevation and the distance to rural areas**.** Naturally regenerated forests lasted longer in patches that were at higher elevation, further from rural communities, roads and urban areas, and on steeper slopes. This findings underlie the contribution of remoteness in achieving forest transition, mirroring what found in other areas within the tropics (Aide et al. 2013, 2019; Molin et al. 2017; Calaboni et al. 2018; Reid et al. 2019). Ashton et al (2001) provided a framework for tropical forest restoration in Sri Lanka that is adaptable throughout the tropics (see, e.g., Griscom and Ashton 2011 for an example involving dry forest in Panama). They point out that the ability for forests to recover depends upon the level of degradation and the disturbance regime in terms of both chronicity and acuteness. More recently, Chazdon et al. (2021) commented on ecosystems recovery following damage or destruction. People can participate in enabling natural mechanisms of recovery of a targeted ecosystem by interventions that both facilitate the inherent capacity for natural recovery and inhibit drivers of ecosystem degradation. Our approach contributes to restoration planning by identifying areas where forests have higher and/or lower potential to persist in time, setting the scene for the identification of the stakeholders involved and the potential drivers of ecosystem degradation, allowing informed prioritization and planning of appropriate actions for the support of forest natural regeneration and its persistence in time. In addition, our approach allows for a more detailed future exploration of the social and economic drivers underlying the ephemerality of secondary forests within central Panama and represents a basis for future investigations and management decisions for successful, long-term forest-landscape restoration.

Acknowledgements We thank two anonymous reviewers for comments on earlier versions of this manuscript. This work is a contribution of the Agua Salud Project, a collaboration between the Smithsonian Tropical Research Institute (STRI), the Panama Canal Authority (ACP) and the Ministry of the Environment of Panama (MiAmbiente) and the Smart Reforestation® program. Agua Salud is part of the Smithsonian Institution Forest Global Earth Observatory (ForestGEO). We thank Prof. Giovanna Jona Lasinio from Sapienza University, Department of Statistical Sciences who provided insight and expertise that greatly assisted the research. We thank Adriana Tapia, Mario Bailon, and Miguel Nunez for help in ground truthing. We are also immensely grateful to Martina Perez, Danilo Lombardi (Sapienza University, Department of Environmental Biology) and Massimo Pulejo (New York University, Department of Politics) for comments that greatly improved the manuscript. Any errors are our own and should not tarnish the reputations of these esteemed persons.

Author contributions GB conceived of the presented idea; developed the theory; conceived, planned and conducted the data collection; wrote the manuscript. KW contributed with land cover maps. GB and GDF verified the analytical methods and performed the statistical analysis. MV contributed to writing correction. JSH contributed to improving the manuscript, supervised the findings of this work and encouraged GB to investigate on passive regeneration.

Funding Open access funding provided by Università degli Studi di Roma La Sapienza within the CRUI-CARE Agreement. The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Declarations

Competing interest The authors have no relevant financial or non-financial interests to disclose. No authors have any conflict of interest associated with funders or any aspect of data collection.

Open research Walker, K. L. 2021. Panama Vegetation Time Series Maps. Figshare data repository. https://doi.org/10.6084/m9.figshare.14120603. Available at: https://figshare.com/articles/dataset/Panama_Vegetation_Time_Series_Maps/14120603 with supplementary 2020 dataset at: https://figshare.com/articles/dataset/PVCTS2020v2_part_tif/22046813. Shapefiles of Panama topography, roads, populated places, rivers and lakes can be found at STRI GIS Data Portal: https://stridata-si.opendata.arcgis.com/

Research involving in human and participants This study did not collect human subject data, nor did it include the study of animals.

References

Acevedo-Charry O, Mitchell Aide T (2019) Recovery of amphibian, reptile, bird and mammal diversity during secondary forest succession in the tropics. Oikos 128(8):1065–1078

Adamowicz W, Calderon-Etter L, Entem A et al (2019) Assessing ecological infrastructure investments. Proc Natl Acad Sci 116:5254–5261

Aide TM, Clark ML, Grau HR et al (2013) Deforestation and Reforestation of Latin America and the Caribbean (2001–2010). Biotropica 45:262–271

Aide TM, Grau HR, Graesser J et al (2019) Woody vegetation dynamics in the tropical and subtropical Andes from 2001 to 2014: satellite image interpretation and expert validation. Glob Change Biol 25:2112–2126

Archer KJ, Kimes RV (2008) Empirical characterization of random forest variable importance measures. Comput Stat Data Anal 52:2249–2260

Ashton PMS, Gunatilleke CVS, Singhakumara BMP, Gunatilleke IAUN (2001) Restoration pathways for rain forest in southwest Sri Lanka: a review of concepts and models. For Ecol Manage 154:409–430

Birch AL, Stallard RF, Barnard HR (2021) Precipitation characteristics and land cover control wet season runoff source and rainfall partitioning in three humid tropical catchments in central Panama. Water Resour Res 57(2):e2020WR028058

Borda-Niño M, Meli P, Brancalion PH (2020) Drivers of tropical forest cover increase: a systematic review. Land Degrad Dev 31:1366–1379

Breiman L (1996) Out-of-bag estimation

Breiman L (2001) Random forests. Mach Learn 45:5–32

Brondizio ES, Settele J, Díaz S, Ngo HT (2019) Global assessment report on biodiversity and ecosystem services of the Intergovernmental science-policy platform on biodiversity and ecosystem services

Calaboni A, Tambosi L, Igari A et al (2018) The forest transition in São Paulo, Brazil: historical patterns and potential drivers. Ecol Soc. https://doi.org/10.5751/ES-10270-230407

CBD (2022) A new global framework for managing nature through 2030: first detailed draft agreement debuts.

Chavarria KA, Saltonstall K, Vinda J et al (2021) Land use influences stream bacterial communities in lowland tropical watersheds. Sci Rep 11:21752

Chazdon RL, Guariguata MR (2016) Natural regeneration as a tool for large-scale forest restoration in the tropics: prospects and challenges. Biotropica 48:716–730

Chazdon RL, Uriarte M (2016) Natural regeneration in the context of large-scale forest and landscape restoration in the tropics. Biotropica 48:709–715

Chazdon RL, Brancalion PH, Lamb D et al (2017a) A policy-driven knowledge agenda for global forest and landscape restoration. Conserv Lett 10:125–132

Chazdon RL, et al. (2017b) Partnering with nature: the case for natural regeneration in forest and landscape restoration. *FERI policy brief, Montreal Canada. Montreal, Canada: forest ecosystem restoration initiative*.

Chazdon RL et al (2021) The intervention continuum in restoration ecology: rethinking the active–passive dichotomy. Restor Ecol. https://doi.org/10.1111/rec.13535

Chen C (2004) Using random forest to learn imbalanced data.

Condit R, Robinson WD, Ibáñez R et al (2001) The status of the Panama Canal Watershed and its biodiversity at the beginning of the 21st century: long-term ecological studies reveal a diverse flora and fauna near the Panama Canal, harbored within a corridor of forest stretching from the Caribbean to the Pacific, but deforestation, land degradation, erosion, and overhunting remain threats. Bioscience 51:389–398

Crawford CL et al (2022) Rural land abandonment is too ephemeral to provide major benefits for biodiversity and climate. Sci Adv 8(21):eabm8999

Fagan ME, DeFries RS, Sesnie SE et al (2013) Land cover dynamics following a deforestation ban in northern Costa Rica. Environ Res Lett 8:034017

Farina A (2000) Landscape ecology in action. Springer, Netherlands, Dordrecht

Forman RTT, Godron M (1981) Patches and structural components for a landscape ecology. Bioscience 31:733–740

Griscom HP, Ashton MS (2011) Restoration of dry tropical forests in Central America: a review of pattern and process. For Ecol Manage 261:1564–1579

Grömping U (2009) Variable importance assessment in regression: linear regression versus random forest. Am Stat 63:308–319

Hall JS, Ashton MS, Garen EJ, Jose S (2011) The ecology and ecosystem services of native trees: implications for reforestation and land restoration in Mesoamerica. For Ecol Manage 261:1553–1557

Hall J, Moss D, Stallard R et al (2015) Managing watersheds for ecosystem services in the steepland neotropics. IDB, Washington

Hall JS, Plisinski JS, Mladinich SK et al (2022) Deforestation scenarios show the importance of secondary forest for meeting Panama's carbon goals. Landsc Ecol 37:673–694

Hassler SK, Zimmermann B, van Breugel M et al (2011) Recovery of saturated hydraulic conductivity under secondary succession on former pasture in the humid tropics. For Ecol Manage 261:1634–1642

Heckadon Moreno S, Ibañez DR, Condit R (eds) (1999) La Cuenca del Canal: deforestación, contaminación y urbanización: proyecto de monitoreo de la cuenca del Canal de Panama (PMCC): sumario ejecutivo del informe final. Instituto Smithsonian de Investigaciones Tropicales, Balboa

Holdridge LR (1967) Life zone ecology. Life Zone Ecol. 1967:10

Holl KD (2017) Research directions in tropical forest restoration. Ann Mo Bot Gard 102:237–250
Holl KD, Mitchell Aide T (2011) When and where to actively restore ecosystems? For Ecol Manag 261(10):1558–1563
Ibanez R, Condit R, Angehr G et al (2002) An ecosystem report on the Panama Canal: monitoring the status of the forest communities and the watershed. Environ Monit Assess 80:65–95
Jakovac CC, Junqueira AB, Crouzeilles R et al (2021) The role of land-use history in driving successional pathways and its implications for the restoration of tropical forests. Biol Rev Camb Philos Soc 96:1114–1134
Knudby A, Brenning A, LeDrew E (2010) New approaches to modelling fish–habitat relationships. Ecol Model 221:503–511
Kotsiantis SB (2007) Supervised machine learning: a review of classification techniques. Emerg Artif Intell Appl Comput Eng 160(1):3–24
Kuhn M, Johnson K (2013) Applied predictive modeling. Springer, New York
Lewis SL, Wheeler CE, Mitchard ET, Koch A (2019) Restoring natural forests is the best way to remove atmospheric carbon. Nature 568:25–28
Liaw A, Wiener M (2002) Classification and regression by random forest. R News 2:18
Lowe B, Kulkarni A (2015) Multispectral image analysis using random forest. Int J Soft Comput 6:1–14
Mansourian S, Vallauri D (2014) Restoring forest landscapes: important lessons learnt. Environ Manage 53:241–251
Martin PA, Newton AC, Bullock JM (2013) Carbon pools recover more quickly than plant biodiversity in tropical secondary forests. Proc Royal Soc b Biol Sci 280:20132236
Mbow HOP, Reisinger A, Canadell J, O'Brien P (2017) Special Report on climate change, desertification, land degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystems (SR2). IPCC, Geneva, p 650
Meli P, Holl KD, Benayas JMR et al (2017) A global review of past land use, climate, and active vs passive restoration effects on forest recovery. PLoS ONE 12:e0171368
Molin P, Gergel S, Soares-Filho B, Ferraz S (2017) Spatial determinants of Atlantic forest loss and recovery in Brazil. Landsc Ecol 32:857
More AS, Rana DP (2017) Review of random forest classification techniques to resolve data imbalance. 2017 1st international conference on intelligent systems and information management (ICISIM). IEEE, pp 72–78
HTTNguyen TMDoan VRadeloff 2018 Applying random forest classification to map land use/land cover using landsat 8 oli Int Archiv Photogramm Remote Sens Spatial Info Sci 10.5194/isprs-archives-XLII-3-W4-363-2018 Nguyen HTT, Doan TM, Radeloff V (2018) Applying random forest classification to map land use/land cover using landsat 8 oli. Int Archiv Photogramm Remote Sens Spatial Info Sci. https://doi.org/10.5194/isprs-archives-XLII-3-W4-363-2018
Nguyen Q, Ly HB, Lanh H et al (2021) Influence of data splitting on performance of machine learning models in prediction of shear strength of soil. Math Probl Eng. https://doi.org/10.1155/2021/4832864
Pelletier C, Valero S, Inglada J et al (2016) Assessing the robustness of random forests to map land cover with high resolution satellite image time series over large areas. Remote Sens Environ 187:156–168
Piffer PR, Calaboni A, Rosa MR et al (2022) Ephemeral forest regeneration limits carbon sequestration potential in the Brazilian Atlantic Forest. Glob Change Biol 28:630–643
Poorter L, Rozendaal DMA, Bongers F et al (2021) Functional recovery of secondary tropical forests. Proc Natl Acad Sci USA 118:e2003405118
Reid JL, Wilson SJ, Bloomfield GS et al (2017) How long do restored ecosystems persist? 1. Ann Mo Bot Gard 102:258–265
Reid JL, Fagan ME, Lucas J et al (2019) The ephemerality of secondary forests in southern Costa Rica. Conserv Lett 12:e12607
Rompre G, Robinson DW, Desrochers A (2008) Causes of habitat loss in a neotropical landscape: the Panamá Canal corridor. Landsc Urban Plan 87:129
Rozendaal DMA, Bongers F, Aide TM et al (2019) Biodiversity recovery of Neotropical secondary forests. Sci Adv 5:eaau3114
Rudel TK, Bates D, Machinguiashi R (2002) A tropical forest transition? Agricultural change, out-migration, and secondary forests in the Ecuadorian amazon. Ann Assoc Am Geogr 92:87–102
Schwartz NB, Uriarte M, DeFries R et al (2017) Land-use dynamics influence estimates of carbon sequestration potential in tropical second-growth forest. Environ Res Lett 12:074023
Schwartz NB, Aide TM, Graesser J et al (2020) Reversals of reforestation across Latin America limit climate mitigation potential of tropical forests. Front for Global Change 3:85
Sloan S (2015) The development-driven forest transition and its utility for REDD+. Ecol Econ 116:1–11
Sloan S (2016) Tropical forest gain and interactions amongst agents of forest change. Forests 7:55
Sloan S (2022) Reforestation reversals and forest transitions. Land Use Policy 112:105800
Smith J, Ferreira S, van de Kop P et al (2003) The persistence of secondary forests on colonist farms in the BrazilianAmazon. Agrofor Syst 58:125
Stanturf JA, Mansourian S (2020) Forest landscape restoration: state of play. Royal Soc Open Sci 7:201218
Talukdar S, Singha P, Mahato S et al (2020) Land-use land-cover classification by machine learning classifiers for satellite observations—a review. Remote Sens 12:1135
Thessen A (2016) Adoption of machine learning techniques in ecology and earth science. One Ecosyst 1:e8621
Turner MG (2005) Landscape ecology: what is the state of the science? Annu Rev Ecol Evol Syst 36:319–344
United Nations Sustainable Development Goals (2017). http://www.un.org/sustainabledevelopment/sustainable-development-goals/. Accessed 17 Nov 2017
van Breugel M, Hall JS, Craven D et al (2013a) Succession of ephemeral secondary forests and their limited role for the conservation of floristic diversity in a human-modified tropical landscape. PLoS ONE 8:e82433
Van Breugel M, Hall JS, Craven D, Bailon M, Hernandez A, Abbene M, Van Breugel P (2013b) Succession of

ephemeral secondary forests and their limited role for the conservation of floristic diversity in a human-modified tropical landscape. PLoSONE 8:e82433
Vitale M, Proietti C, Cionni I et al (2014) Random forests analysis: a useful tool for defining the relative importance of environmental conditions on crown defoliation. Water Air Soil Pollut. https://doi.org/10.1007/s11270-014-1992-z
Walker K (2020) Capturing ephemeral forest dynamics with hybrid time-series and composite mapping in the Republic of Panama. Int J Appl Earth Obs Geoinf 87:102029
Wright S, Samaniego M (2008) Historical, demographic, and economic correlates of land-use change in the Republic of Panama. Ecol Soc. https://doi.org/10.5751/ES-02459-130217
Yang P, Yang YH, Zhou BB, Zomaya AY (2010) A review of ensemble methods in bioinformatics. CBIO 5:296–308
Zakariah M (2014) Classification of large datasets using random forest algorithm in various applications. Survey 4:10