Impact of Protected Areas on Forests in Madagascar

By

Theo Gimenez

Dr. Alexander Pfaff, Adviser

May 2012

Masters project submitted in partial fulfillment of the
requirements for the Master of Environmental Management degree in
the Nicholas School of the Environment of
Duke University

Abstract

Madagascar is known for having one of the most diverse ecosystems in the world. Unfortunately, the country's natural habitat has been declining for decades. One study estimated that only 7% of Madagascar's original vegetation is intact. In 1927, Madagascar implemented its first Protected Areas (PAs). Prior to 2003, approximately 3% of the country's land was protected. By 2008, this number more than tripled reaching 9.4% or just over 5.5 million hectares. Protected Areas have been used throughout the world to manage, conserve, and protect, geographic areas that would otherwise be subject to activities that could threaten their environmental and economic value. Nevertheless, few studies evaluate the effectiveness of PA networks with controls. The purpose of this study was to determine the impact Madagascar's PAs have had on deforestation from 1990 to 2005. I use matching to make 'apples to apples' comparisons with respect to geographic land characteristics. A control method such as matching is essential since PAs are often non-randomly distributed. Estimates produced from matching show a .244% lower annual deforestation rate from 2000 to 2005 (.303%) than from 1990 to 2000 (.547%) in Madagascar's PAs. Both rates are lower than naïve estimates and statistics reported in some previous studies. I also show that deforestation rates in Madagascar vary considerably with land characteristics regardless of the status of protection. This could have important implications for future forest conservation efforts since the impact of PAs vary with land characteristics as well.

Introduction

Madagascar is host to one of the most biologically diverse ecosystems in the world and is characterized by a variety of taxonomic groups endemic to the region. Despite global agreement and efforts to conserve the region, there have been high rates of deforestation throughout the country [1-4]. Over the past century, Madagascar has shown a strong commitment to the preservation of their natural resources primarily through the creation of Protected Areas (PAs). The PA network continued to grow steadily after the implementation of the country's first in 1927. Nevertheless, forested areas continued to decline. Unsurprisingly, the persistent decline is often attributed to corruption, failure of the government to effectively monitor the PAs and enforce national legislation, and to low-income, often rural communities use of forest resources [5, 6]. The management of the PA network remained largely unchanged until the collapse of the Soviet Union in the 1980s. When Madagascar adopted a democratic style government, they began receiving foreign aid, stimulating conservation and development [5]. In 1989 Madagascar ratified a National Environmental Action Plan (NEAP), which created the Association National pour la Gestion des Aires Protegees (ANGAP) to manage the country's PAs [7].

By 2003, 1.8 million hectares (ha), or about 3.1% of Madagascar's terrestrial surface area was protected. Malagasy president Marc Ravalomanana vowed to increase protection from 3.1% to 10% by 2012 at the 2003 International Union for Conservation of Nature World Parks Congress in Durban, South Africa. With this announcement came the creation of a new PA network, the System of Protected areas of Madagascar (Systeme des Aires Protegees de Madagascar, or SAPM). The SAPM would consist of some preexisting PAs and all additional PAs [5, 7].

A more exhaustive recount of Malagasy Forest Conservation policy is provided by Toillier et al. [8], Consiglio et al. [9], Miller and Porter Morgan [10], Harper et al. [2], and Raik [5]. The abbreviated history is designed to highlight the importance of PAs as the government's primary conservation policy tool. In fact, PAs are the most popular tool to conserve terrestrial resources worldwide, due in part to their high rate of success when properly implemented, monitored, and enforced. Nevertheless, few studies attempt to assess the effectiveness of PAs in Madagascar once they are put in place. This is surprising given the country's rapid augmentation of the network over the past 20 years, as well as the cost associated with their creation and management. The purpose of this study is to determine the impact protected areas have had on deforestation in Madagascar from 1990 to 2005. To my knowledge, this is the first study that goes beyond simply comparing deforestation rates inside and outside of Malagasy PAs and will attempt to eliminate the inherent bias associated with observational studies. I will use a statistical tool known as 'matching' to examine Malagasy PAs and compare the results to more conventional estimation methods.

Matching has been used in a variety of program evaluation studies where generating data would be is infeasible (e.g., a comprehensive global analysis) or unethical (e.g., studying the effects of smoking). The former best describes this analysis since it would be impractical for a government to conduct a randomized experiment with PAs in order to measure their impact. Consequently, data that already exists, termed observational data, must be used. If a PA is located far from cities, matching can help determine whether lower deforestation rates are a result of protection or geographic location by making ‘apples to apples’ comparisons. I will begin by elaborating on the challenges of evaluating PAs and provide a brief literature review of studies addressing the effectiveness of PAs. Then I will proceed to the methods, results, and discussion.

Section 1: Background

1.1 Challenges

The objective of evaluating PAs is the same as it would be if it were a randomized experiment; to answer the question of what would have happened had the land never been protected (the ‘counterfactual’). This becomes a difficult task because people choose where PAs will be located. Consequently, locations that receive protection tend to be characteristically different than those that remain unprotected. Were it not for this, it would suffice to measure and compare deforestation occurring in protected and unprotected land, or possibly a more sophisticated method such as econometric modeling with regression analysis.

Intuitively, determining the ways in which protected and unprotected land differ is central to matching observations on their similarities to make apples to apples comparisons. Many of these features have been teased out of observational data through various forms of statistical analysis including those used in this report. Joppa and Pfaff [11] found that PAs were more likely to be located farther from urban areas and roads, and at higher elevations and steeper slopes. Collectively, these types of characteristics are often appropriately labeled ‘observable’ or ‘measured’ covariates. However, we don’t pretend to have the ability to describe all of the ways PAs may be different. Many differences are either not measured and thus can’t be included in comparisons, or not observed. These are referred to as ‘unobservable’ covariates. Both unobservable and observable covariates can be said to perfectly describe PAs.

While matching on observable covariates is achievable, it is impossible to determine whether the matched pair is also similar with respect to unobservable covariates. Unobservable covariates could be abstract and unknown, or could be an unmeasured concrete characteristic such as distance from roads. If distance from roads cannot be calculated, it will undoubtedly be difficult to ensure that the each individual in the match is similar in this respect. This is the root of the aforementioned inherent bias plaguing this genre of analysis. Eliminating, or more likely reducing, this distortion is essential to procuring meaningful results from an evaluation of PAs.

1.2 Literature Review

That majority of match based studies to date have analyzed the impact of PAs globally or focused on methods designed to reduce bias [11-14]. There have been a variety of studies utilizing methods similar to matching described in Joppa and Pfaff [13] including Vogt et al. [15], Oliveira et al. [16], and Mas [17]. Vogt et al. [15] conducted a study in Uganda reporting that protected land experienced lower rates of deforestation even when soil quality was comparable to unprotected land. Oliveira et al. [16] revealed a connection between roads and deforestation in Peru. Through their analysis they found that 75% of deforestation was occurring within a 20 km radius of roads. Importantly, the authors found that unprotected land near road networks was roughly four times more likely to lose forest than protected land. Mas [17] incorporated buffers to estimate deforestation rates in the Calakmul Biosphere Reserve in Mexico. Mas found PAs to be much less effective than had been reported previously after controlling for soil type, elevation, slope, distance from settlements, and distance from roads.

Andam et al. [12]'s analysis of Costa Rican PAs used methods most similar to those applied in this report. Like Madagascar, Costa Rica is also one of the most ecologically diverse regions of the world. Andam studied 150 PAs from 1960 to 1997. They controlled for land productivity, and distances from forest edges, roads, and cities. They concluded that deforestation rates were 11% lower in PAs. This is a shocking deviation from the 44% estimated by comparing deforestation in PAs to all unprotected land. Even when Andam controlled for spillover effects by using a 10km buffer, the estimate only fell to 38%.

Because studies tend to use data with varying degrees of spatial and temporal resolution, as well as differ in the particular methodology employed, it can be difficult to compare their results directly to my own. However, with the rapidly growing body of literature revealing gross miscalculations in the effectiveness of PAs, it is inevitable that more sophisticated studies will be demanded by global lending and aid organizations. If good data is available, protected area networks can be evaluated in relatively short periods of time, and can cost next to nothing when considering how great the true cost could be of forgoing the analysis.

Section 2: Methods

In randomized experiments, unobserved covariates are balanced by random assignment to treatment, being protected in this case [18, 19]. Equation 1 describes the method that could be used for estimating the effect of being protected, if it were randomly assigned, and not intentionally selected. This effect is referred to as the Average Treatment Effect on the Treated (ATT), because it looks only at the protected group ('treatment') and unprotected individuals that are similar to the protected group ('control'). This is often the starting point for most studies using observational data but only provides a naïve estimate of the treatment effect (Table 2). The ATT can be calculates by,

$$\boldsymbol{\tau}|(\mathbf{T} = 1) = \mathbf{E}(\mathbf{Y}_{\mathbf{i1}}|\mathbf{T}_{\mathbf{i}} = 1) - \mathbf{E}(\mathbf{Y}_{\mathbf{i0}}|\mathbf{T}_{\mathbf{i}} = 1), \qquad (1)$$

where $\boldsymbol{\tau}$ is the ATT, $\mathbf{T}$ is the treatment assignment equal to 1 for protected and 0 for unprotected, and $\mathbf{Y}_{\mathbf{i1}}$ is the expected deforestation rate ('response') for a protected observation that is protected, and $\mathbf{Y}_{\mathbf{i0}}$ is the expected response of a protected observation had it not been protected. The last term represents the counterfactual, or the individual matching attempts to simulate. Since the assumption that assignment to treatment is independent of potential outcomes Y_{i1} and Y_{i0} rarely holds in observational studies, Equation 1 will tend to overestimate or underestimate the true treatment effect as exemplified by Andam et al. [12] and confirmed empirically [13].

Matching on the other hand relies on the Conditional Independence Assumption (CIA) as the crux of its ability to have a causal interpretation. In this way matching is similar to Ordinary Least Squares (OLS) regression. That is to say, conditional on X (one or more observable characteristics such as slope or distance from urban areas), selection to treatment is independent of potential outcomes: $[Y_0, Y_1 \perp\!\!\!\perp T]\,|X$ [20, 21]. This would imply that if protected and unprotected observations are matched on several observable covariates, they would only differ by the fact that one happens to be protected. Theoretically, any statistical analysis attempting to determine which individual is treated would be as accurate as flipping a coin. If the covariate(s)

X achieve this outcome, treatment is said to be strongly ignorable. If treatment assignment is in fact strongly ignorable conditional on X, the ATT can be calculated using observational data by solving

$$\boldsymbol{\tau|(T = 1)=E(Y_{ij} |X_i , T_i = 1) - E(Y_{ij} |X_i , T_i = 0).} \qquad \mathbf{(2)}$$

Note that Equation 2 is identical to Equation 1 except that treatment is now conditional on $\mathbf{X_i}$. In practice, the Mahalanobis distance matrix is often used to identify the treatment and control representing the pair separated by the smallest multivariate distance [22-24].

This method results in statistically efficient estimates of the treatment effect when continuous variables are limited to one. Additional continuous covariates will result in increasingly biased estimates [19]. The most common solution to mitigate the presence of multiple continuous covariates is to employ a propensity score [20]. The theoretical intuition behind a propensity score is to define similarity by the probability that an individual will be assigned to treatment, rather than on the specific characteristics that influence treatment assignment (vector X) as before. The objective is exactly the same as in covariate matching except now matches are a function of the probability of being treated, which as stated previously is ultimately how the quality of the match is tested [25]. The CIA now becomes $[Y_0, Y_1]\perp\perp T|\ e(X)$, where e(X) is the propensity score. Typically, e(X) is calculated using the linear predicted values of a probit function or logistical regression as was used in this analysis. The following equation:

$$\boldsymbol{\tau\ |(T = 1) = E[E(Yi\ |e(Xi\ ), Ti = 1)-E(Yi\ |e(Xi\ ), Ti = 0)|Ti = 1],} \quad \mathbf{(3)}$$

is equivalent to Equation 2, and is used when conditioning on **e(Xi)**, the propensity score, as opposed to X. The Mahalanobis distance and propensity score can be used alone, or in a variety of combinations. In this study I match on covariates alone, covariates with the propensity score, as well as the propensity score alone. For each model, I evaluate covariate balance to evaluate the effectiveness of the matching method.

When multiple continuous variables are incorporated into the matching process, either alone, or in combination with a propensity score, bias adjustments can be used to reduce the associated bias. In this analysis, the bias-corrected matching estimator is obtained by using post-match regression functions to account for the differences in covariates between matched pairs [26-28]. Recently, nonparametric approaches to bias adjustment have been proposed because they would not depend on the correct model specification or particular statistical assumptions necessary for OLS to yield meaningful results [29, 30]. Jasjeet [31] recommends Genetic Matching as an effective way to achieve covariate balance using nonparametric methods to adjust for bias. This method was not included in this study because of computational limitations for large datasets.

2.1 Data

Land Cover – Response Variable

Land-cover data for 1990, 2000, and 2005 were originally produced by Conservation International (CI) using Thematic Mapper (TM) and Landsat Enhanced Thematic Mapper Plus (EMT+) data. The CI compiled the three raster images into one multi-date image with a 28.5m resolution [2, 3]. The land-cover data used for this analysis was identical to the CI data, aside from minor preprocessing by Kremen et al. [3]. Pixels that were obstructed by clouds during imaging for one or more of the three study dates were excluded from analysis. The number of forested pixels for 1990, 2000, and 2005 are 124,252,074 (1,009 ha), 114,507,467 (930 ha), and 111,536,234 (906 ha) respectively (Table 1).

Observable Covariates

Elevation data was provided by the NASA Shuttle Radar Topographic Mission (SRTM), and has a 900m resolution. I used ArcGIS 10.0 to derive slope in degrees from horizontal from the Digital Elevation Model (DEM). Stream and other inland water body (i.e., rivers, canals, and lakes) data came from the 1992 publication of the Digital Chart of the World (DCW). Urban area data came from the 1997 publication of the DCW. All of the DCW data had a scale of 1:1,000,000. Villages and road network data were acquired from The National Geographic & Hydrographic Institute of Madagascar (L'institut Geographique Et Hydrographique De Madagascar; www.ftm.mg) and had a scale of 1:500,000. Population density was produced by the Center for International Earth Science Information Network (CIESIN) at Columbia University as part of the 2000 Global Gridded Population Database. Density values were reported as the number of people per kilometer. Administrative Boundaries were extracted from the global administrative area database GADM, Version 1. Though the scale of the Administrative Boundaries is unknown and may vary by country, it is sufficient (roughly 30 Arc seconds) for the purpose of this analysis. The six original administrative provinces, Antsiranana, Antananarivo, Mahajanga, Toamasina, Fianarantsoa, and Toliara, were used for this report because the effective time frame of this report predates the establishment of other administrative regions (effective because there is a one year overlap) and the ultimate dissolution of the original six provinces (Figure 1).

Ecoregion data were classified by the World Wildlife Fund and was an updated version of Olson et al. [32] that was published in 2004. The Madagascar Vegetation Map (Forest Classification data) was produced for the Madagascar Vegetation Mapping Project (http://www.vegmad.org). The data was published in 2007, though the timeframe of the grid is 2001 and has a resolution of 30m [33].

Protected Area Network

Protected areas in this analysis was a subset of the data used in Kremen et al. [3] (Figure 2). Implementation of the PAs used in this analysis predates the “Durban Vision” and the creation of the SAPM. However, by 2005 PA prioritization, implementation, and management responsibilities shifted from ANGAP to the SAPM [7]. While the PAs in Madagascar represent different designations by the International Union for Conservation of Nature (IUCN), these designations were not part of the dataset and therefore were not considered in this analysis.

2.2 Data Processing

Spatial manipulation of raw data was achieved by using ArcGIS 10 and Python 3.2. All data was projected to the Universal Transverse Mercator (UTM) geographic coordinate system using Zone 28 South based on the 1984 World Geodetic System (WGS84) datum. The ‘Path Distance’ tool was used in conjunction with the elevation raster grid to derive distance grids for primary and secondary roads, inland water, villages, and urban areas. These derived grids had the same resolution as elevation, 900m. Two land-cover grids were derived from the raw data. The first consisted of two classifications, pixels that remained forested from 1990 to 2000 and pixels that were deforested from 1990 to 2000. The second consisted of two classifications as well, pixels that remained forested from 2000 to 2005 and pixels that were deforested from 2000 to 2005. Both of these grids were then converted to generalized polygon coverages and classifications were dissolved. I then used 'Create Random Points' to create 100,000 random points within each of the new coverages. I used 'Extract Values to Points' to overlay the points with the raster grids, and 'Spatial Join' to overlay the points with the vector data. X-coordinate and Y-coordinate fields were added prior to exportation.

Observations were considered protected if they were located inside a PA that was created in or before 1990 and 2000 for the 1990-2000 (referred to as Period 1) and 2000-2005 (referred to as Period 2) datasets respectively. Some of the observations had invalid or missing data due to some points being located outside of the raster grid's spatial domain. This is because raster data is formed with pixels while vector data consists of points, lines and polygons so there will inevitably be some error. The effect was minimal and the observations were removed. Table 2 shows the average values of the covariates in each time frame for protected and unprotected observations, as well as for observations that were added to the PA network by Period 2.

2.3 Statistical Analysis

Statistical analysis and all further manipulations of data were performed with R 2.13.1 [34]. The naïve estimate of deforestation in protected and unprotected areas found in Table 2 were calculated with an unpaired Welch Two Sample t-test. Tables 3A and 3B illustrate OLS models with different degrees of covariate specification. For administrative boundaries, forest classification, and ecoregions, dummy variables were used because they were discrete variables.

For the purpose of matching treated and control observations, the 'Matching Package' was used [28]. In order to find the most appropriate match for each treated observation, I used Mahalanobis distance matching, propensity score matching, and a combination of the two methods. The combined approach has the advantage of minimizing propensity score distances as well as multi-variate distances orthogonal to the propensity score between treated and control observations [28, 35]. Despite having a large dataset, I matched with replacement allowing multiple treatment observations to potentially be matched to the same control observation. Because the Match function matches treatments in the order of the dataset, allowing replacement reduces bias. I also allowed for ties permitting one treatment to be matched to multiple controls. Ultimately, the matched dataset is weighted to account for multiple matches. Typically, when ties are not randomly broken, the variance in the outcome variable will be less likely to be underestimated, reducing bias associated with the treatment effect [28]. I adjusted two factors to manipulate whether or not multiple matches were considered tied, calipers and distance tolerance. Bias adjustments were calculated for each model and incorporated into the estimated treatment effect.

Covariate balance achieved after matching is summarized in Tables 4A and 4B. For each covariate, a paired Welch Two Sample t-test was used to compare matched observations. Other tests that were performed to assess balance, but not included in this report, are the univariate bootstrap Kolmogorov-Smirnov (KS) test (1000 boots) and the multivariate KS test. The univariate KS test has the advantage of comparing treatment and control observations beyond the first two moments (mean and range), while the multivariate KS test examines the probability that both treatment and control observations came from the same distribution. Each test is designed to ensure that apples to apples comparisons are used to estimate the treatment effect.

Section 3: Results and Sensitivity

3.1 Naïve Estimates

Table 2 shows the pre-match means of deforestation and continuous covariates for protected and unprotected observations, as well as their differences. The naïve estimates project deforestation rates to be 6.1% and 2.2% lower in protected areas for Periods 1 and 2 respectively. Though these estimates seem meaningful because they are statistically significant, there are sizeable differences between treatment groups among covariates as well, also statistically significant. The results confirm that the degree of selection bias is nontrivial. The bias is largely consistent with Joppa and Pfaff [11]; location of PAs are overwhelmingly skewed toward higher elevations, steeper slopes, and longer distances from roads. The trends associated with distances from water bodies and streams may seem conflicting; however, large water bodies are used more often by communities than streams. The higher value of land near a reliable fresh water source likely explains why fewer PAs would be located nearby. If a large water body is available, the presence of streams would simply be a hindrance to development offering an explanation for PA's diminished distance to streams. It is worth noting that distance to urban areas was unexpectedly found to be greater for unprotected observations than for protected observations. This anomaly was disregarded because village data was more geographically representative of Madagascar and distance from villages followed the expected trend. Furthermore, population density was greater in unprotected areas, a variable likely to be endogenous to both distances to villages and to urban areas. In general, PAs in Madagascar tend to be positioned in geographic locations characterized by low accessibility. This of course is no accident and has been documented in previous evaluations in countries with marked similarities to Madagascar [11-13].

3.2 OLS Regression

Ordinary Least Squares Regression was used to model the impact protected areas have on deforestation by controlling for the both continuous and factor covariates. This step can help to identify independent terms having the largest impact on deforestation as well as the sign representing their trends. Modeling deforestation in this way also provides an opportunity to observe the sensitivity of the protected area coefficient to model specification. Tables 3A and 3B suggest that the covariate differences illustrated in table 2 can have a profound influence on the estimate of deforestation in PAs. Models 1-6 resulted in a range of treatment effect estimates. In Period 1 the range was -5.1% to -7.0%, and in Period 2 the range was -1.6% to -2.2%. These results suggest that estimates for the second period are both lower and less sensitive to bias. The latter conclusion is further supported by fewer statistically significant covariates, and reduced significance among significant covariates (i.e., model 4).

3.3 Matching

Though there was considerable overlap between the density distributions of protected and unprotected covariate and propensity score values, achieving perfect balance was challenging. Tables 4A and 4B describe the post-match balance statistics for each period. The first two models each period succeeded in reducing covariate differences between treatment groups considerably, but statistically significant differences remained with the following exceptions: population and distance from urban areas in model 1 of period 2, and slope in model 2 of period 2. Propensity score matching was the only method that successfully rendered differences of the covariates between treatment groups statistically insignificant, aside from slope in period 1 and population in period 2.

After matching, the ATT was calculated. Results are summarized in Table 5. Each model from both periods produced deforestation estimates that were statistically significant. For period 1, the matching models including covariates resulted in the lowest treatment effects, -3.83% and -3.88% for models 1 and 2 respectively. These estimates represent more than a 2% deviance from the naïve estimate, -6.06%. While interpretation of those estimates may be hindered by remaining imbalance in their respective matching models, the estimates do suggest that protected areas may be less effective than naïve estimates predict. The 3rd model estimate is only a little over a half of percentage point lower than the naïve estimate.

Estimates in period 2 resulted in more robust estimates across the three models. Differences between models 1 and 3 and models 2 and 3 were only 2 and 3 hundredths of a percentage point respectively. In period 2, the model 3 estimate of -1.52% was in between the other two model estimates. Matching provides convincing evidence that deforestation in protected areas is almost three quarters of a percentage point higher than the naïve estimate.

3.4 Sensitivity

Rosenbaum bounds were calculated for each matching model in Periods 1 and 2 using the R package 'rbounds' [36]. Since matching must make the assumption that observable covariates successfully control for unobserved covariates (or less likely that unobserved covariates only minimally affect the response term), some attention must paid to the consequences of the assumption not holding true. Gamma is an excellent metric for determining how influential an unobserved covariate would have to be in order for the treatment effects described in Table 5, to be statistically insignificant from 0. Gamma values are summarized in Table 5. For period 1, gamma was determined to be 3.4 for each model. For Period 2, gamma was 3.1 for models 1 and 2, and 2.9 for model 3. In this analysis, bias would have to account for deforestation to be between 2.9 and 3.4 times greater in unprotected areas than protected areas for *p* to be greater than α at the .05 level of significance. In Rosenbaum [18], the author reports that existence of a gamma above 2 in an observational study is unlikely.

Further sensitivity analysis was examined by observing fluctuations in the treatment effect in response to extreme covariate conditions. When determining the best geographic locations for PAs in terms of cost and potential for protection, estimates from this type of post-match analysis are often more practical. Table 6 summarizes the variation in treatment effect as a consequence of being located near or far from primary roads and having extreme or mild slopes. In period 1, PAs located near primary roads experienced more than three times the rate of deforestation than would be expected by that same PA located farther from primary roads. In period 2, mildly sloped observations experienced almost three times the rate of deforestation than would be expected by that same individual in a location characterized by steeper slopes. Table 6 illustrates the sensitivity of the treatment effect to geography; however, it is relatively easy to apply these results to PA management, enforcement, and monitoring, in addition to the aforementioned task of strategically finding an appropriate location for a PA.

Section 4: Discussion and Conclusion

The results of my analysis provide unequivocal evidence that deforestation rates were lower within Madagascar's PAs from 1990 to 2005. However, deforestation rates in PAs did not deviate from unprotected areas by magnitudes reported in some previous studies. I also show that the impact that Malagasy PAs have on deforestation greatly depend on their location. In general bias tends to cause overestimations of the impact PAs may have on reducing deforestation; but at extreme slopes and close to primary roads PAs may be far more effective than even the naïve estimates suggests. It was also clear that deforestation from 2000 to 2005, as compared to 1990 to 2000, occurred at a lower rate. Finally, it appears that characteristics that had statistically significant impacts on deforestation from 1990 to 2000 did not necessarily have an impact from 2000 to 2005.

The importance of this study can be viewed through the lens of an economist, or an ecologist. To an economist, an evaluation of a PA network should carry the same weight as an evaluation of any other government program or private investment. Governments should strive to fund projects with a high likelihood of success as should investors, though there motivations may differ. In the context of this analysis, it should be of great importance that Malagasy PAs are effective, albeit at a rate lower than previously believed. The treatment effects from this analysis can translate directly into a return on investment. To an ecologist, this study should be equally important because funding allocated to global conservation is typically limited. The results from this study could provide global organizations, geared toward ecological preservation with assurance that funding the Malagasy PA network is worthwhile. For a full discussion on evaluating conservation investments see Ferraro and Pattanayak [37].

Unfortunately, it can be difficult to acquire data with the appropriate spatial resolution for the purpose of a particular countrywide analysis covering the time period of interest. It is not uncommon to be forced into choosing one resolution over another. This issue is particularly prevalent in studies focusing on developing countries, because data can be especially scarce. For this analysis, I was unable to find Malagasy road network and population data for 1990. Consequently, period 1 matching will include bias resulting from temporally discordant data. The degree of bias will be a function of the differences between the data at hand, and the data that should have been used. I was also unable to find spatial data describing Malagasy demographics. I believe that per capita income and level of education could be highly correlated with both locations of PAs and deforestation rates. That being said, the resolution of the data used in this analysis is substantially better than that used by many other protected area studies. Village and roads data came directly from the Malagasy government, and forest cover data was 30m resolution.

This study failed to account for the IUCN status of PAs. While this data was available, it was not included in the PA data I used for this analysis. This could cause my analysis to underestimate the effectiveness of the PA network since not all IUCN classifications mandate strict conservation. Some use of forest resources may be sanctioned in a PA for the purpose of subsistence by local villages. Additionally, this report did not address spillover effects by comparing PA observations with unprotected observations within a predefined buffer of the PA network.

Perhaps the greatest weakness of this study is shared by most studies utilizing matching to reduce bias, lack of methodological consensus. Matching is a multistep process. Each decision affecting statistical methodology has implications for causal interpretation of the results. For example, this study relied heavily on the Mahalanobis distance, which relies on normally distributed covariates in order to accurately calculate multivariate distances. If the assumption does not hold, it may be more appropriate to use ranked Mahalanobis distances [18]. Propensity score matching can be problematic as well since there are multiple propensity scores that are equally justifiable. The matching process itself can range from perhaps the most basic, optimal matching, to Genetic matching, which automatically optimizes matching for covariate balance. Even the ostensibly simple task of measuring covariate balance is not consistent. Typically, a good match can be confirmed by reassuring results from a variety of parametric and nonparametric tests including a paired t.test, and univariate and multivariate KS tests. In practice, balance is evaluated using a myriad of parametric and/or nonparametric tests but is often not reported.

Despite the challenges ahead, it should be axiomatic that conservation efforts be evaluated to ensure efficient allocation of capital and preservation of globally unique habitats. There is only to present results with full transparency of methods employed. This will allow disparate approaches to be compared. The purpose of this study is ultimately to inform Malagasy conservation policy, but I hope it will be a helpful contribution current literature as well.

Acknowledgements

I could not have completed this analysis without the guidance of my advisor Alexander Pfaff, and the support of Ranaivo Rasolofoson who provided the data directly from Madagasgar. Finally, if not for my loving wife Sara Champagne, I would have died of starvation and could not have finished this study.

Table 1: Status of Protected Areas and Protected Forests from 1990 to 2005 (calculated spatially using ArcGIS).

Year	Total Forested Area (ha)	Total Protected Area (ha)	Protected Forest Area (ha)	Percent of Forest Protected	Annual Percent of Deforestation in PAs
1990	1,009	1,352,960	83	8.2	
2000	930	1,820,551	115	12.3	$-.85$
2005	906	2,757,713	191	21.1	$-.54$

Table 2: Sample Means for 1990-2000, 2000-2005, and observations located in PAs added from 1991-2000.

	1990-2000				**2000-2005**		
Variables	Protected	Unprotected	Difference	PAs Added between 1991 and 2000	Protected	Unprotected	Difference
Deforestation	0.0215	0.082	−0.0606***	0.0028	0.0062	0.028	−0.0218***
Primary Road Distance	14624	14249	374.65**	21431	16865	14367	2498.5***
Secondary Road Distance	4353.8	3080.9	1272.9***	4615.5	4472.3	3054.1	1418.2***
Elevation	715.13	499.32	215.81***	428.28	642.79	503.4	139.39***
Slope	4.3417	2.5065	1.8351***	3.4038	4.0854	2.4889	1.5964***
Village Distance	4080.6	3612.5	468.05***	3659.9	3970.1	3603.9	366.21***
Urban Area Distance	83052	85817	−2765.5***	59297	77855	87084	−9229.6***
Population	15.296	16.274	−0.9778***	12.276	14.224	16.007	−1.7835***
Streams Distance	5596.6	5894.7	−298.14***	4758.8	5333	5885.2	−552.18***
Waterbody Distance	38987	36186	2800.8***	48020	42152	35916	6235.5***
Observations	8110	89444		3559	12269	87360	

† significant at $p < .10$; $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$

Table 3A: Ordinary Least Square Regression Models for Deforestation from 1990-2000.

Coefficients	Model (1)	Model (2)	Model (3)	Model (4)	Model (5)	Model (6)
Protected Areas	-0.0606^{***}	-0.0534^{***}	-0.0509^{***}	-0.0593^{***}	-0.0699^{***}	-0.0509^{***}
	(0.00309)	(0.00311)	(0.00312)	(0.00317)	(0.0032)	(0.00296)
Primary Road Distance		$-2.13e{-}06^{***}$	$-1.84e{-}06^{***}$	$-1.57e{-}06^{***}$	$-1.48e{-}06^{***}$	$-1.16e{-}06^{***}$
		$(7.53e-08)$	$(8.1e-08)$	$(8.64e-08)$	$(8.64e-08)$	$(8.04e-08)$
Secondary Road Distance			$-2.11e{-}06^{***}$	$-1.52e{-}06^{***}$	$-1.64e{-}06^{***}$	$1.76e{-}06^{***}$
			$(3.07e-07)$	$(3.13e-07)$	$(3.13e-07)$	$(2.89e-07)$
Elevation			$-5.99e-06^{*}$	$-2.21e{-}05^{***}$	$-1.55e{-}05^{***}$	$6.2e-06^{*}$
			$(2.43e-06)$	$(2.72e-06)$	$(2.76e-06)$	$(2.97e-06)$
Slope		-0.00353^{***}	-0.00287^{***}	-0.00246^{***}	-0.00227^{***}	-0.000983^{**}
		(0.000288)	(0.000322)	(0.000336)	(0.00034)	(0.000319)
Village Distance			$2.52e{-}06^{***}$	$2.62e{-}06^{***}$	$2.22e{-}06^{***}$	$1.98e{-}06^{***}$
			$(3.8e-07)$	$(3.81e-07)$	$(3.81e-07)$	$(3.53e-07)$
Urban Area Distance			$-1.43e-08$	$-1.04e-08$	$-2.4e-07^{***}$	$-2.63e{-}07^{***}$
			$(1.63e-08)$	$(1.76e-08)$	$(2.02e-08)$	$(1.96e-08)$
Population			0.000567^{***}	0.000478^{***}	0.000417^{***}	0.00028^{***}
			$(3.8e-05)$	$(3.83e-05)$	$(3.87e-05)$	$(3.58e-05)$
Streams Distance		$-3.75e-07^{**}$	$-7.57e{-}07^{***}$	$-8.94e{-}07^{***}$	$-1.48e{-}06^{***}$	$-9.62e-08$
		$(1.24e-07)$	$(1.31e-07)$	$(1.38e-07)$	$(1.41e-07)$	$(1.43e-07)$
Waterbody Distance			$-2.43e{-}07^{***}$	$-5.02e{-}07^{***}$	$-5.17e{-}07^{***}$	$-4.88e{-}07^{***}$
			$(3.94e-08)$	$(4.26e-08)$	$(4.42e-08)$	$(4.12e-08)$
X-Coordinate					$-9.25e{-}08^{***}$	$-1.62e-08$
					$(9.46e-09)$	$(1.06e-08)$
Y-Coordinate					$-1.22e{-}07^{***}$	$-1.75e{-}07^{***}$
					$(6.02e-09)$	$(6.34e-09)$
Administrative Boundary (Dummy)				Yes	Yes	Yes
Forest Classification (Dummy)						Yes
WWF Eco Class (Dummy)						Yes
Intercept (Unprotected Areas)	0.082^{***}	0.123^{***}	0.121^{***}	0.249^{***}	1.3^{***}	1.56^{***}
	(0.00089)	(0.00176)	(0.00291)	(0.00985)	(0.0498)	(0.246)
N	97554	97554	97554	97554	97554	97554
R^2	0.004	0.015	0.019	0.024	0.029	0.188
adj. R^2	0.004	0.015	0.018	0.024	0.029	0.188
Resid. sd	0.266	0.265	0.264	0.263	0.263	0.240

Standard errors in parentheses

† significant at $p < .10$; $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$

Table 3B: Ordinary Least Square Regression Models for Deforestation from 2000-2005.

Coefficients	Model (1)	Model (2)	Model (3)	Model (4)	Model (5)	Model (6)
Protected Areas	-0.0218^{***}	-0.0183^{***}	-0.0184^{***}	-0.0161^{***}	-0.0185^{***}	-0.016^{***}
	(0.00151)	(0.00153)	(0.00154)	(0.00158)	(0.00159)	(0.00161)
Primary Road Distance		$-1.96e{-}07^{***}$	$-1.09e-07^{*}$	$1.08e-07^{*}$	$1.35e-07^{**}$	$6.74e-08$
		$(4.32e-08)$	$(4.65e-08)$	$(4.97e-08)$	$(4.98e-08)$	$(5.06e-08)$
Secondary Road Distance			$-6.83e{-}07^{***}$	$-7.56e{-}07^{***}$	$-7.69e{-}07^{***}$	$-5.64e-07^{**}$
			$(1.81e-07)$	$(1.85e-07)$	$(1.85e-07)$	$(1.86e-07)$
Elevation			$-1.86e{-}05^{***}$	$-1.41e{-}05^{***}$	$-1.24e{-}05^{***}$	$-1.27e{-}05^{***}$
			$(1.43e-06)$	$(1.59e-06)$	$(1.62e-06)$	$(1.91e-06)$
Slope		-0.00157^{***}	$-0.000326^{\dagger}$	-0.00028	-0.000277	-0.000439^{*}
		(0.000168)	(0.000189)	(0.000197)	(0.000199)	(0.000204)
Village Distance			$2.05e{-}06^{***}$	$1.93e{-}06^{***}$	$1.82e{-}06^{***}$	$1.95e{-}06^{***}$
			$(2.25e-07)$	$(2.26e-07)$	$(2.26e-07)$	$(2.29e-07)$
Urban Area Distance			$-9.72e-09$	$4.05e-09$	$-6.74e{-}08^{***}$	$-1.11e{-}07^{***}$
			$(9.5e-09)$	$(1.03e-08)$	$(1.2e-08)$	$(1.26e-08)$
Population			$-2.24e-06$	$-1.72e-06$	$-3.65e-05$	$-9.17e{-}05^{***}$
			$(2.65e-05)$	$(2.68e-05)$	$(2.74e-05)$	$(2.78e-05)$
Streams Distance		$9.21e{-}07^{***}$	$6.02e{-}07^{***}$	$1.99e-07^{*}$	$2.56e-08$	$-1.38e-07$
		$(7.24e-08)$	$(7.65e-08)$	$(8.04e-08)$	$(8.28e-08)$	$(9.15e-08)$
Waterbody Distance			$6.74e-08^{**}$	$2.1e-08$	$9.29e-09$	$1.6e-08$
			$(2.29e-08)$	$(2.46e-08)$	$(2.56e-08)$	$(2.61e-08)$
X-Coordinate					$-2.27e{-}08^{***}$	$-4.95e{-}08^{***}$
					$(5.57e-09)$	$(6.75e-09)$
Y-Coordinate					$-3.92e{-}08^{***}$	$-3.21e{-}08^{***}$
					$(3.53e-09)$	$(4.08e-09)$
Administrative Boundary (Dummy)				Yes	Yes	Yes
Forest Classification (Dummy)						Yes
WWF Eco Class (Dummy)						Yes
Intercept (Unprotected Areas)	0.028^{***}	0.0293^{***}	0.0294^{***}	0.0845^{***}	0.418^{***}	0.465^{***}
	(0.000531)	(0.00103)	(0.00174)	(0.00617)	(0.0293)	(0.0658)
N	99629	99629	99629	99629	99629	99629
R^2	0.002	0.006	0.008	0.012	0.013	0.020
adj. R^2	0.002	0.006	0.008	0.012	0.013	0.020
Resid. sd	0.157	0.157	0.156	0.156	0.156	0.156

Standard errors in parentheses

† significant at $p < .10$; $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$

Table 4A: Matching Balances for 1990-2000.

Coefficients	Covariate Match	Covariate & P.S. Match	P.S. Match
	Model (1)	Model (2)	Model (3)
Primary Road Distance			
Initial Difference	374.65^{***}	374.65^{***}	374.65^{***}
Current Difference	182.59^{***}	196.16^{***}	$-271.48^{\dagger}$
Percent Change	51.264	47.643	172.46
Secondary Road Distance			
Initial Difference	1272.9^{***}	1272.9^{***}	1272.9^{***}
Current Difference	230.78^{***}	199.05^{***}	$-79.851^{\dagger}$
Percent Change	81.87	84.363	106.27
Elevation			
Initial Difference	215.81^{***}	215.81^{***}	215.81^{***}
Current Difference	31.569^{***}	29.745^{***}	$-2.9849^{\dagger}$
Percent Change	85.371	86.217	101.38
Slope			
Initial Difference	1.8351^{***}	1.8351^{***}	1.8351^{***}
Current Difference	0.1709^{***}	0.1283^{***}	-0.1247^{**}
Percent Change	90.687	93.009	106.79
Village Distance			
Initial Difference	468.05^{***}	468.05^{***}	468.05^{***}
Current Difference	132.76^{***}	120.1^{***}	$4.2844^{\dagger}$
Percent Change	71.636	74.34	99.085
Urban Area Distance			
Initial Difference	-2765.5^{***}	-2765.5^{***}	-2765.5^{***}
Current Difference	978.49^{***}	1181.9^{***}	$562.99^{\dagger}$
Percent Change	135.38	142.74	120.36
Population			
Initial Difference	-0.9778^{***}	-0.9778^{***}	-0.9778^{***}
Current Difference	0.425^{***}	0.479^{***}	$-0.2703^{\dagger}$
Percent Change	143.47	148.98	72.357
Streams Distance			
Initial Difference	-298.14^{***}	-298.14^{***}	-298.14^{***}
Current Difference	120.02^{***}	66.902^{***}	$23.579^{\dagger}$
Percent Change	140.26	122.44	107.91
Waterbody Distance			
Initial Difference	2800.8^{***}	2800.8^{***}	2800.8^{***}
Current Difference	306.46^{***}	428.56^{***}	$90.153^{\dagger}$
Percent Change	89.058	84.699	96.781
Observations	8110	8107	8083
Number of Matches	1	1	2
P.S. Caliper	-	.50	.10
Distance Tolerance	1e-05	1e-05	1e-07

† significant at $p < .10$; $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$

Table 4B: Matching Balances for 2000-2005.

Coefficients	Covariate Match	Covariate & P.S. Match	P.S. Match
	Model (1)	Model (2)	Model (3)
Primary Road Distance			
Initial Difference	2498.5^{***}	2498.5^{***}	2498.5^{***}
Current Difference	157.33^{***}	173.57^{***}	$-35.575^{\dagger}$
Percent Change	93.703	93.053	101.42
Secondary Road Distance			
Initial Difference	1418.2^{***}	1418.2^{***}	1418.2^{***}
Current Difference	289.37^{***}	190.65^{***}	85.144^{*}
Percent Change	79.597	86.557	93.996
Elevation			
Initial Difference	139.39^{***}	139.39^{***}	139.39^{***}
Current Difference	20.179^{***}	23.931^{***}	$-4.795^{\dagger}$
Percent Change	85.523	82.831	103.44
Slope			
Initial Difference	1.5964^{***}	1.5964^{***}	1.5964^{***}
Current Difference	0.1395^{***}	$0.0085^{\dagger}$	$0.0369^{\dagger}$
Percent Change	91.26	99.466	97.688
Village Distance			
Initial Difference	366.21^{***}	366.21^{***}	366.21^{***}
Current Difference	153.8^{***}	122.91^{***}	$18.157^{\dagger}$
Percent Change	58.003	66.437	95.042
Urban Area Distance			
Initial Difference	-9229.6^{***}	-9229.6^{***}	-9229.6^{***}
Current Difference	-213.01^{*}	668.85^{***}	$803.34^{\dagger}$
Percent Change	97.692	107.25	108.7
Population			
Initial Difference	-1.7835^{***}	-1.7835^{***}	-1.7835^{***}
Current Difference	$0.0622^{\dagger}$	0.1675^{***}	-0.3434^{**}
Percent Change	103.49	109.39	80.746
Streams Distance			
Initial Difference	-552.18^{***}	-552.18^{***}	-552.18^{***}
Current Difference	117.24^{***}	49.763^{***}	$0.451^{\dagger}$
Percent Change	121.23	109.01	100.08
Waterbody Distance			
Initial Difference	6235.5^{***}	6235.5^{***}	6235.5^{***}
Current Difference	304.78^{***}	291.07^{***}	$-125.35^{\dagger}$
Percent Change	95.112	95.332	102.01
Observations	12269	12252	12250
Number of Matches	1	1	1
P.S. Caliper	-	.10	.01
Distance Tolerance	1e-05	1e-05	1e-07

† significant at $p < .10$; $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$

Table 5: Multiple Estimates of Protection's Impact on Deforestation and Sensitivity to Bias. All estimates are presented as a percent.

Estimation Method	**1990-2000**				**2000-2005**			
	Deforestation (total)	Deforestation (yearly)	Γ	Sammple Size	Deforestation (total)	Deforestation (yearly)	Γ	Sammple Size
Naive Estimate	-6.06 ***	-0.606		Full Sample	-2.18 ***	-0.437		Full Sample
OLS Regression	-5.93 ***	-0.593		Full Sample	-1.61 ***	-0.321		Full Sample
Covariate Match	-3.83 ***	-0.383	3.4	8110	-1.5 ***	-0.299	3.1	12269
P.S. and Covariate Match	-3.88 ***	-0.388	3.4	8107	-1.55 ***	-0.31	3.1	12252
P.S. Match	-5.47 ***	-0.547	3.4	8083	-1.52 ***	-0.303	2.9	12250

† significant at $p < .10$; $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$

Table 6: Estimation of treatment effects for extreme cases of Primary Road Distance and Slope for Periods 1 and 2.

Time Period	Estimation Method	Primary Road Distance (High)	Primary Road Distance (Low)	Slope (High)	Slope (Low)
1990-2000	Mean Comparison	-0.041^{***}	-0.075^{***}	-0.041^{***}	-0.067^{***}
	OLS Regression	-0.029^{***}	-0.078^{***}	-0.051^{***}	-0.066^{***}
	P.S. Match	-0.025^{***}	-0.078^{***}	-0.047^{***}	-0.066^{***}
	Observations	39683	57871	31466	66088
2000-2005	Mean Comparison	-0.019^{***}	-0.024^{***}	-0.0085^{***}	-0.028^{***}
	OLS Regression	-0.02^{***}	-0.015^{***}	-0.0087^{***}	-0.024^{***}
	P.S. Match	-0.014^{***}	-0.016^{***}	-0.0079^{***}	-0.022^{***}
	Observations	41147	58482	32319	67310

† significant at $p < .10$; $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$

Figure 1: Administrative Provinces of Madagascar

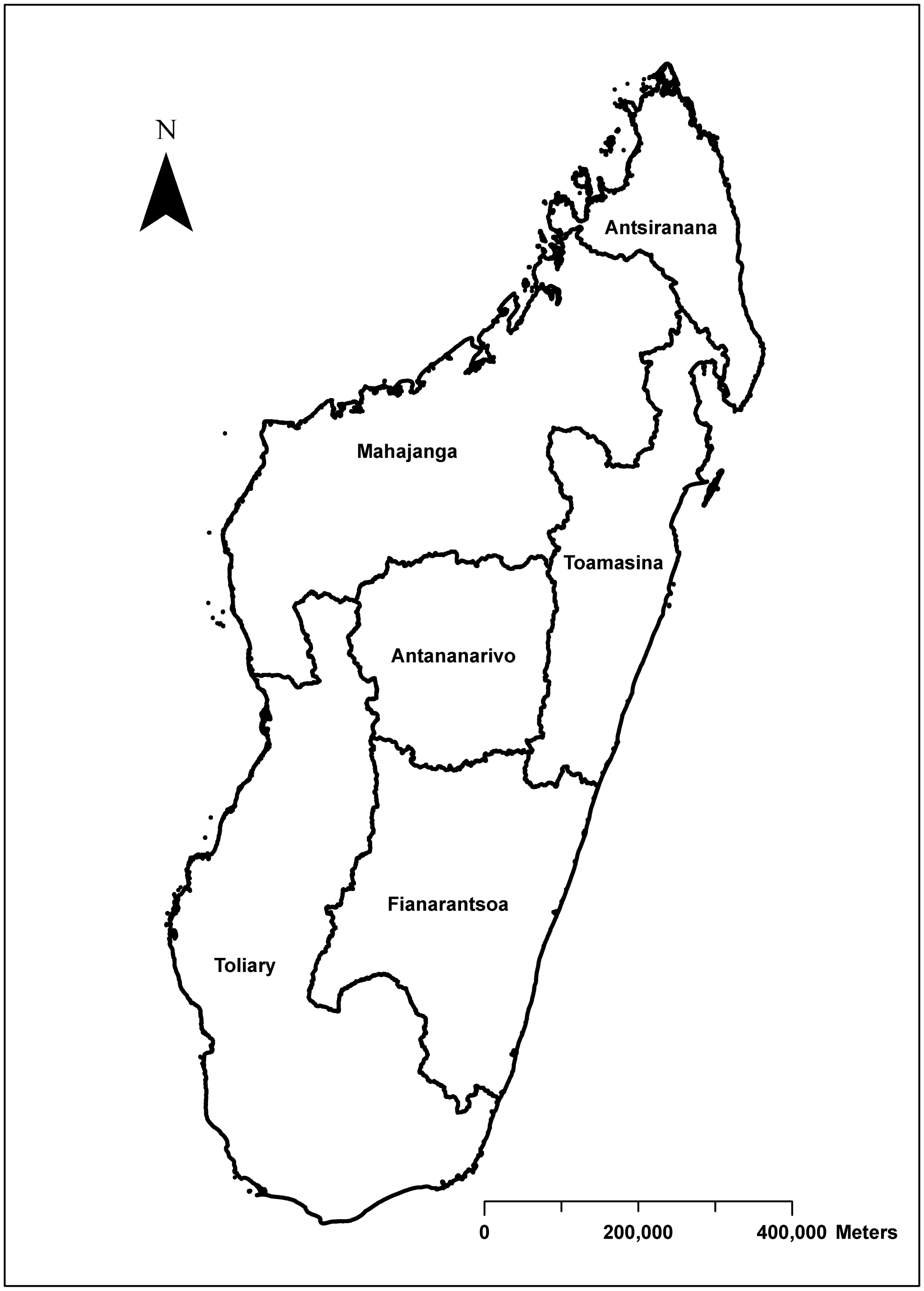
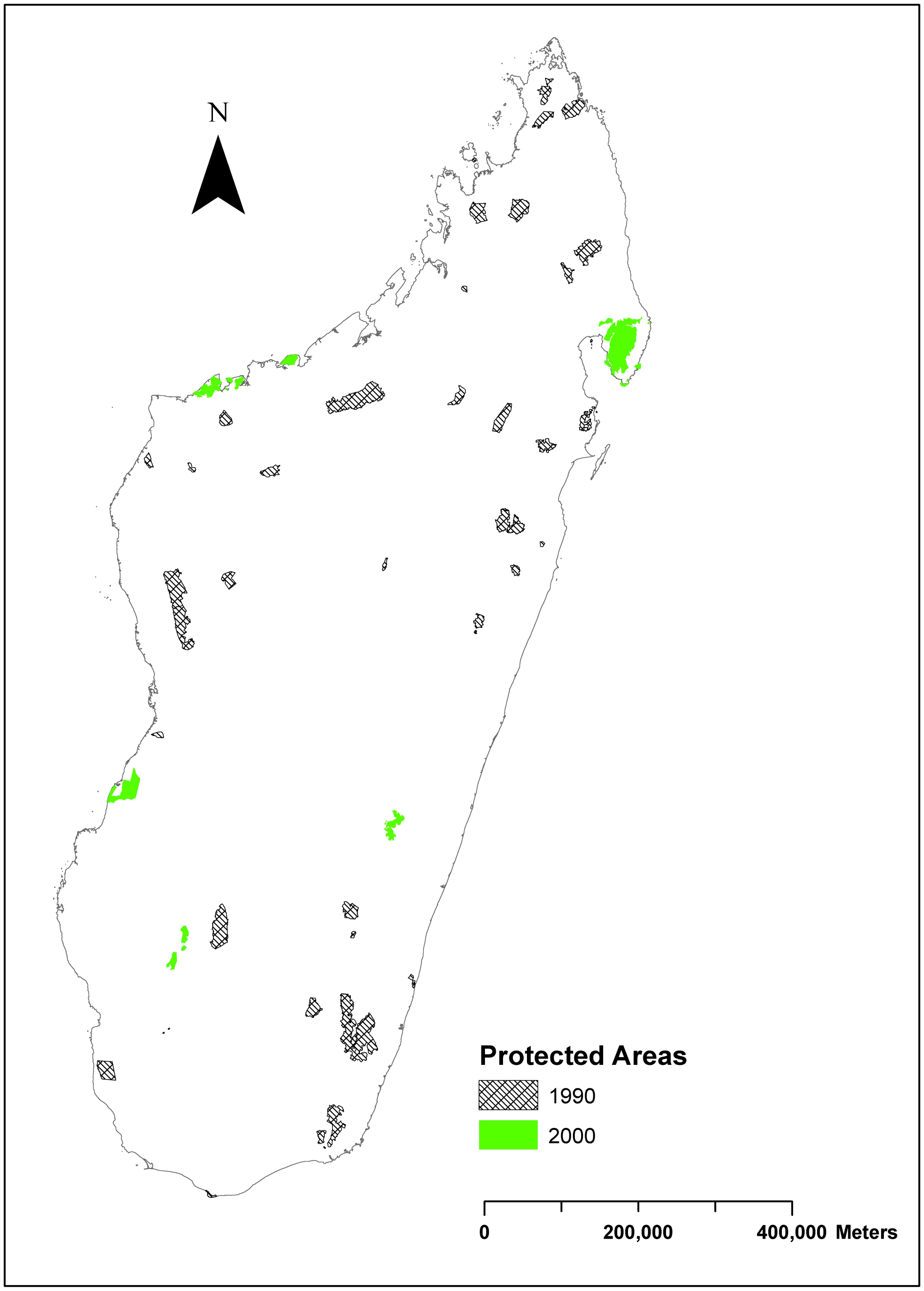


Figure 2: Protected Area Network



References

[1] Scales, I.R., Lost in translation: conflicting views of deforestation, land use and identity in western Madagascar, *The Geographical Journal*, **178** (2012) 67-79.

[2] Harper, G.J., et al., Fifty years of deforestation and forest fragmentation in Madagascar, *Environmental Conservation*, **34** (2007) 325-333.

[3] Kremen, C., et al., Aligning conservation priorities across taxa in Madagascar with high-resolution planning tools, *Science*, **320** (2008) 222-226.

[4] Mittermeier, R.A., et al., Conserving the world's biological diversity Conservation International,WorldWildlife Fund, International Union for the Conservation of Nature, World Resources Institute, World Bank, Washington, (1990).

[5] Raik, D., Forest Management in Madagascar: An Historical Overview, *Madagascar Conservation & Development*, **2** (2007) 5-10.

[6] Ganzhorn, J.U., et al., The state of lemur conservation in Madagascar, *Primate Conservation*, **17** (1997) 70-86.

[7] Thomas F. Allnutt, et al., Madagascar Digital Conservation Atlas Report, (2009).

[8] Toillier, A., et al., Livelihood Strategies and Land Use Changes in Response to Conservation: Pitfalls of Community-Based Forest Management in Madagascar, *Journal of Sustainable Forestry*, **30** (2011) 20-56.

[9] Consiglio, T., et al., Deforestation and plant diversity of Madagascar's littoral forests, *Conserv Biol*, **20** (2006) 1799-1803.

[10] Miller, J.S. & Porter Morgan, H.A., Assessing the effectiveness of Madagascar's changing protected areas system: a case study of threatened Boraginales, *Oryx*, **45** (2011) 201-209.

[11] Joppa, L.N. & Pfaff, A., High and far: biases in the location of protected areas, *PLoS ONE*, **4** (2009) e8273.

[12] Andam, K.S., et al., Measuring the effectiveness of protected area networks in reducing deforestation, *Proceedings of the National Academy of Sciences of the United States of America*, **105** (2008) 16089-16094.

[13] Joppa, L. & Pfaff, A., Reassessing the forest impacts of protection: the challenge of nonrandom location and a corrective method, *Ann N Y Acad Sci*, **1185** (2010) 135-149.

[14] Joppa, L.N. & Pfaff, A., Global protected area impacts, *Proc Biol Sci*, **278** (2011) 1633-1638.

[15] Vogt, N.D., et al., Understanding the stability of forest reserve boundaries in the West Mengo region of Uganda, *Ecol. Soc.* , **11** (2006) 1–22.

[16] Oliveira, P., et al., Land-use allocation protects the Peruvian Amazon, *Science*, **317** (2007) 1233–1236.

[17] Mas, J.F., Assessing protected area effectiveness using surrounding (buffer) areas environmentally similar to the target area, *Environ. Monit. Assess*, **105** (2005) 69–80.

[18] Rosenbaum, P.R., Design of observational studies, Springer, New York, 2010.

[19] Abadie, A. & Imbens, G.W., Large Sample Properties of Matching Estimators for Average Treatment Effects, *Econometrica*, **74** (2006) 235-267.

[20] Rosenbaum, P.R. & Rubin, D.B., The Central Role of the Propensity Score in Observational Studies for Causal Effects, *Biometrika*, **70** (1983) 41-55.

[21] Sekhon, J.S., Opiates for the Matches: Matching Methods for Causal Inference, *Annual Review of Political Science*, **12** (2009) 487-508.

[22] Cochran, W. & Rubin, D., Controlling bias in observational studies: a review, *Sankhya, Ser. A* **35** (1973) 417–446.

[23] Rubin, D., Using multivariate sampling and regression adjustment to control bias in observational studies, *J. Am. Stat. Assoc* **74** (1979) 318–328.

[24] Rubin, D., Bias reduction using Mahalanobis-metric matching, *Biometrics* **36** (1980) 293–298.

[25] Joshua D. Angrist & Pischke, J.-S., Mostly Harmless Econometrics: An Empiricist's Companion, Princeton University Press, Princeton, New Jersey, 2009.

[26] Abadie, A. & Imbens, G., Technical report: Simple and bias-corrected matching estimators, in, University of California, Berkeley Department of Economics, 2002.

[27] Rubin, D.B., The use of matched sampling and regression adjustments to remove bias in observational studies, *Biometrics* **29** (1973) 185–203.

[28] Sekhon, J.S., Multivariate and Propensity Score Matching Software with Automated Balance Optimization: The Matching Package for R, *Journal of Statistical Software*, **42** (2011) 1-52.

[29] Abadie, A. & Imbens, G.W., Bias-Corrected Matching Estimators for Average Treatment Effects, *Journal of Business and Economic Statistics*, **29** (2011) 1-11.

[30] Sekhon, J.S. & Grieve, R., A Nonparametric Matching Method for Covariate Adjustment with Application to Economic Evaluation, *Experiments in Political Science 2008 Conference Paper*, (2009).

[31] Jasjeet, S.S., Multivariate and Propensity Score Matching Software with Automated Balance Optimization: The Matching package for R, *Journal of Statistical Software*, **42** (2011).

[32] Olson, D.M., et al., Terrestrial Ecoregions of the World: A New Map of Life on Earth, *BioScience* (2001) 933-938. .

[33] Madagascar Vegetation Map, in: Justin Moat, P. Smith (Eds.) MADAGASCAR VEGETATION ATLAS, Royal Botanic Gardens, Kew, 2007.

[34] R Development Core Team, R: A Language and Environment for Statistical Computing, in, R Foundation for Statistical Computing, Vienna, Austria, 2011.

[35] Rosenbaum, P.R. & Rubin, D.B., Constructing a Control Group Using Multivariate Matched Sampling Methods That Incorporate the Propensity Score, *The American Statistician*, **39** (1985) 33-38.

[36] Keele, L.J., rbounds: Perform Rosenbaum bounds sensitivity tests for matched and unmatched data, (2011).

[37] Ferraro, P.J. & Pattanayak, S.K., Money for nothing? A call for empirical evaluation of biodiversity conservation investments, *PLoS biology*, **4** (2006) e105.