Journal of Development Economics
Registered Report Stage 1 : Proposal

Conservation and development: Socioeconomic Impact evaluation of Terrestrial Protected Areas in Madagascar based on large national surveys

Iriana Razafimahenina[1 2 4] Florent Bédécarrats[5 6], Ingrid Dallmann[7], Holimalala Randriamanampisoa[1 3 6]

Date of latest draft: 04/12/2024

Abstract

Protected Areas are the most widely used tool for biodiversity conservation. However, their implementation raises concerns about the well-being of local populations, especially when they are very poor and dependent on natural resources, as is the case in Madagascar. This pre-analysis plan outlines the data, methods, and empirical strategies for assessing the impact of terrestrial protected areas on local household well-being and socioeconomic inequalities. Our study relies on geolocated Demographic Health Surveys spanning a 13-years period (2008-2021), complemented by data from the preceding 11 years (1997-2008) to test the parallel trends assumption underlined in the models. We will employ spatio-temporal models, matching techniques, and difference-in-differences methods to rigorously estimate the impacts.

Keywords : Biodiversity Conservation, Well-Being, Demographic and Health Surveys, Spatio-Temporal Models, Geospatial impact evaluation, Madagascar

JEL codes : Q57, I31, C31, Q56, O55

Study pre-registration: On open Science Framework (OSF) with the title 'Conservation and development: socioeconomic impact evaluation of terrestrial Protected Areas in Madagascar based on large national surveys' https://osf.io/bgu5n/

[1] University of Antananarivo, Madagascar
[2] University of Paris Saclay, France
[3] Development Centre for Economic Studies and Research (CERED), Madagascar
[4] French National Research Institute for Sustainable Development (IRD), Madagascar
[5] University of Saint Quentin en Yvelines, France
[6] UMI - Sustainability and Resilience (SOURCE), IRD, France
[7] Agence Française de Développement (AFD), France

Proposed timeline

Phases	Dates
Literature Review, Conceptualization, and writing of the Registered report	May 2024 - January 2025
Retrieve data from selected sources	February 2025
Data cleaning and analysis	February - March 2025
Writing the scientific article	March 2025 - April 2025
Submission to the journal	May 2025

1 Introduction

The reconciliation between conservation and development has been a long-discussed issue within the scientific community (Adams et al., 2004), but its importance has grown considerably over the past decade with the rapid expansion of protected areas (PAs). This issue is particularly relevant for all 195 COP15 signatory states, which have committed to increasing protected areas coverage to 30% of terrestrial land by 2030.

In theory, protected areas can have significant impacts on local livelihoods, both positive and negative. They are recognized as an essential tool for biodiversity conservation (Maxwell et al., 2020), but their creation can deprive nearby communities of access to revenue-generating activities based in natural resources (gathering, hunting, fishing, and harvesting medicinal plants), reduce the amount of land available and restrict economic activities (agriculture, livestock, construction (Kandel et al., 2022). Conversely, they can be accompanied by compensation measures (local development projects, cash transfers), generate economic benefits (jobs in protected areas, tourism), and enhance ecosystem services (increased water resources, erosion control, fire prevention) (Kandel et al., 2022).

Despite these ambivalent potential effects, empirical studies that rigorously assess the impact of protected areas on people's livelihoods are still rare. Of the 1,043 studies reviewed by McKinnon et al. (2016) , only 19 used quantitative methods to evaluate impacts on material living conditions or economic well-being. This meta-analysis shows that the results of studies vary widely depending on the methods used, the context studied, and the location. Kandel et al. (2022) have updated and extended this analysis by focusing on a corpus of 30 quantitative in Asian countries, South American and Africa countries evaluations that specifically address the impact of protected areas on household income. They show that protected areas can have a positive impact on local economies, but that this effect is generally modest and depends on the local context. This variability in impacts highlights the importance of conducting context-specific studies using robust quantitative methods.

Madagascar stands out as a particularly relevant case study for analyzing the relationship between conservation and socioeconomic conditions. The country is the poorest in terms of the first target of the Sustainable Development Goals (SDG 1-1), with the highest proportion of the population living below the international poverty line in the world (Conceição, 2024, pp. 298–299). In 2008, terrestrial protected areas covered 3.6% of Madagascar and 9% of the population lived within 10 km of a protected area. Today, they cover 10.8% and 28% of the population live within 10 km of protected areas[8]. Madagascar is also characterized by a low state capacity (Hanson & Sigman, 2021), which makes it difficult to implement conservation and sustainable development policies and the social measures that should accompany them. These factors, combined with the high dependence of the rural population on natural resources, mean that the impacts of protected areas are potentially different from those observed in less precarious contexts.

However, empirical studies at the national scale are almost non-existent for Madagascar. None of the quantitative impact evaluations identified by McKinnon et al. (2016) covered the country. One of the references consolidated by Kandel et al. (2022) is a multi-country study that includes Madagascar, but it is based on an estimate of an aggregate impact at the commune level and covers only one date. It uses the 1993 census data to match the country's municipalities (Mammides, 2020), without a before-and-after comparison, and in a context where less than 3 % of the territory was covered by protected areas, most of which had been created several decades earlier.

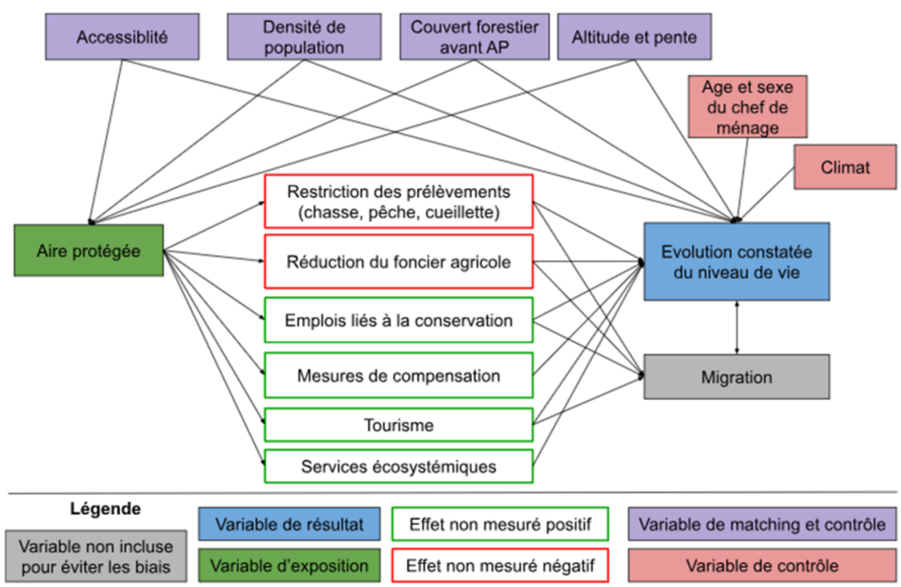
Our contribution to the literature is twofold, both empirical and methodological. Empirically, this study provides an unprecedented national analysis, covering 71 protected areas established between 2008 and 2021, to evaluate the socioeconomic impacts of conservation in contexts of extreme poverty and weak governance. Methodologically, it articulates the state of the art in econometrics, incorporating recent developments to adapt these methods to the study of protected areas. The procedure we propose here could be replicated in other countries, starting with the 39 countries that have at least three geolocated DHS surveys. This approach paves the way for a more systematic evaluation of the impact of protected areas, taking into account the specific context of each country.

2 Research Design

2.1 Hypothesis

Our first hypothesis concerns the overall impact of protected areas in the Malagasy context. In their meta-analysis of 30 studies, Kandel et al. (2022) report a slightly positive average impact, but highlight a large heterogeneity of results across context. Several parameters are likely to influence impact, as represented graphically in Figure 1 in the form of directed acyclic graph (Hünermund & Bareinboim, 2023; Imbens, 2024)

Figure 1: Logic diagram of the theory of change tested in the study



Source:Authors

The factors likely to lead to a decline in well-being seem particularly significant in the Malagasy context, where the population is predominantly rural and living in extreme poverty (the last assessment was in 2012, with 80.7% of the population below the $2.15 a day threshold at 2017 PPP). Six studies conducted in Madagascar between

[8] Calculations by the authors based on the location of the DHS survey clusters. The detailed calculation is provided as supplementary material to the study.

1995 and 2006 estimated the opportunity cost of losing access to protected areas (slash-and-burn agriculture, hunting, gathering, timber, etc.) at between USD 39 and 177 per household per year (Neudert et al., 2017). Golden et al. (2014) estimated that income from hunting accounted for 57 % of household's cash income in areas adjacent to the Makira and Masoala protected areas. Another survey of people living near Makira estimated the value of pharmaceutical use at USD 30-44 per year per household, based on the subsidized price of equivalent treatments in the Malagasy market (Golden et al., 2012).

Several factors that could help improve livelihoods through conservation appear to be fragile in Madagascar, starting with tourism. Naidoo et al. (2019) aggregate data from DHS surveys conducted between 2001 and 2011 in 34 developing countries. Their study is based on matching households near and far from protected areas, but with no pre-post conservation comparison. They highlight positive impacts, but only for a subset of protected areas 'with documented tourism'[9]. According to their study, households living near the protected areas 'with tourism' are 17% wealthier and 16% less likely to be poor than similar households living far from these areas.

However, tourism in Madagascar's protected areas remains low. According to data from Madagascar National Parks (MNP), only 7 protected areas recorded more than 10,000 visitors in 2023 (with a maximum of 30,744 in Isalo), which is low compared to the average of 356,405 visitors per year and per PA recorded in 929 PAs worldwide in the global study by Chung et al. (2018). When new protected areas are created in Madagascar, compensation mechanisms for local populations remain rare, ineffective and insufficient (Bertrand et al., 2012; Rivière, 2017). The most in-depth study on this subject, conducted by Poudyal et al. (2018) with support from the World Bank, focuses on the Ankeniheny Zahamena Corridor (CAZ), created in 2015 to connect several existing protected areas. Five study sites were selected: Two adjacent to the new CAZ protected areas (one eligible for compensation, the other not), two adjacent to long-established protected areas, and one far from the forest boundary. The median cost of the conservation restriction is estimated at USD 2,375 per household per year, representing 27% to 84% of the average annual income. The amounts set aside to compensate beneficiary households were assessed to be insufficient relative to the losses incurred, and 50% of households eligible for compensation received nothing (Poudyal et al., 2016, 2018).

- **Hypothesis 1:** Protected areas in Madagascar, by limiting access to natural resources, have negative impacts on the well-being of nearby households that often exceed the benefits of compensation and ecosystem services, with more adverse impacts than in other countries.

The impact mechanisms represented in Figure 1 are likely to affect households differently depending on their prior characteristics. Compensation measures are generally implemented in the form of projects to promote income-generating activities (agriculture, livestock, handicrafts) in surrounding communities (Poudyal et al., 2018). In the context of such development projects, individuals known as 'development brokers' frequently emerge as intermediaries between local communities and implementing organizations. By mobilizing their social networks and specific skills, these brokers manage to capture a disproportionate share of the benefits of interventions, whether in form of income or access to exclusive opportunities. This dynamic can reinforce pre-existing inequalities within communities, limiting the access of the most vulnerable households to the expected benefits of compensation programs.

[9] The source used to consider that a PA has 'documented tourism' is not reported in Naidoo et al.2019

Although tourism development is often presented as an opportunity for economic growth, it also tends to exacerbate socioeconomic inequalities, particularly in developing countries. Adeniyi et al. (2024) show that in Southern Africa, tourism can initially exacerbate inequalities by concentrating benefits in the most attractive regions, while leaving marginalized communities out of the economic benefits. According to Ghosh and Mitra (2021) the relationship between tourism and inequality follows an inverted Kuznets curve in developing countries, when tourism remains moderate, its growth reduces inequalities, but when tourism becomes massive, further expansion worsens inequalities. Finally, Xuanming et al. (2024) point out that while tourism helps to improve certain socioeconomic indicators, it can also generate inflationary pressures and strain local resources, particularly affecting the most vulnerable households.

- **Hypothesis 2**: Protected areas exacerbate economic inequalities among neighboring communities by creating opportunities that mainly benefit individuals with a higher educational level or a dominant position in the community, allowing them access to rents and jobs related to tourism and associated activities.

IUCN status of protected areas are frequently used to explain differences in effectiveness between them. For example, Naidoo et al. (2019) show that multiple-use protected areas (statuses V and VI) tend to have more beneficial effects than strict areas (statuses I to IV), partly due to greater flexibility in integrating local needs. Beyond status alone, governance plays a central role. Eklund et al. (2017) highlight the importance of transparent and inclusive structures to maximize the positive effects of protected areas on conservation and social justice. Similarly, Eklund et al. (2019) call for management approaches to be adapted to local contexts, with greater involvement of communities in decision-making processes, to better reconcile conservation and development objectives.

This diversity is particularly evident in Madagascar. Although governed by similar formal statuses, protected areas follow different paths depending on the local context and the way in which they are implemented. Froger and Méral (2009) show that the early initiatives of shared governance, gradually introduced with in-depth mediation efforts, achieved encouraging results by strengthening local community support. However, from the 2000s onward, the accelerated deployment of management transfers, driven by quantitative targets, often led to hasty and less contextually adapted implementations, undermining the effectiveness of these mechanisms. These experiences demonstrate that, beyond the protected area status, their establishment period, management approach, and level of community participation significantly influence their socioeconomic impacts.

- **Hypothesis 3**: The impacts of protected areas on well-being and inequalities are heterogeneous, and some protected areas with good levels of local community participation manage to generate greater benefits and distribute them more equitably.

2.2 Basic methodological framework / identification strategy

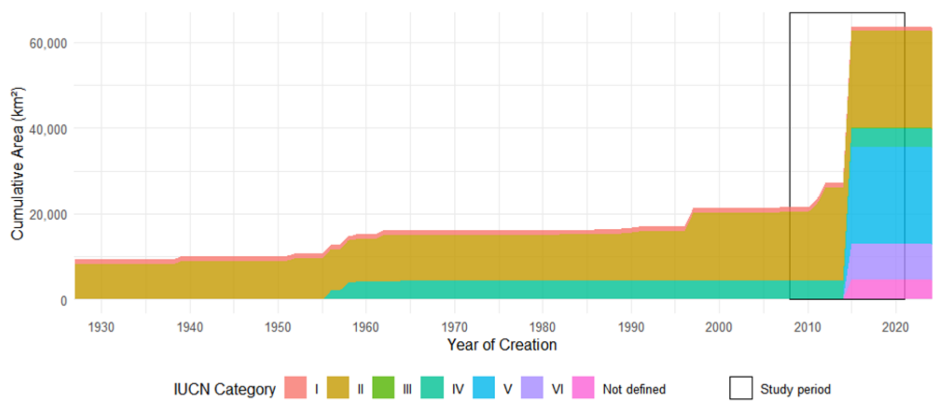
Our evaluative approach is based on a counterfactual measure that estimates the causal effect of the treatment, in this case the presence of protected areas. The counterfactual measure corresponds to a hypothetical scenario describing what would have happened if the intervention under study had not taken place. This approach is based on a comparison between a treatment group (a protected areas) and a control group (unprotected areas with characteristics very similar to those of the protected areas). The study thus fits within the framework of Rubin's causal model Rubin (1974), according to which there are several hypothetical outcomes depending on exposure to the treatment. To ensure comparability between groups, matching techniques are used to assign each unit in the treatment group to a unit with the same observable characteristics in the control group. Using matching techniques reduces endogeneity problems and increases the credibility of the research results (Ma et al., 2020).

We subsequently use the difference-in-difference method to estimate the causal effect attributable to the creation of protected areas. This method allows us to compare the observed changes in the treatment group and control group, while controlling for pre-existing differences between these two groups. By comparing the differences in local households' livelihood before and after the creation of the protected areas, we isolate the specific effect of the creation of these protected areas. Matching and difference-in-difference methods are often used together to reduce selection bias. Several studies have used this combination of methods to evaluate the impact of conservation on land use and livelihoods (Ma et al., 2020; Schleicher et al., 2020a).

2.3 Intervention

This study evaluates the impact of terrestrial protected areas creation on rural household well-being between 2008 and 2021. These time frames were chosen on the basis of the availability of geolocalised data on household living conditions and coincide with a period of strong expansion of protected areas in the country, as shown in Figure 2

Figure 2: Evolution of protected areas in Madagascar and study period



Source: Calculations by authors based on data of the Service de la Gouvernance des Aires Protégées (SGAP) of the Ministère de l'environnement et Développement Durable (MEDD)

Protected areas in Madagascar were first created in 1927 under the French colonial administration and, until the early 2000s, were mainly characterized by strict conservation (IUCN statuses I, II, III and IV). At the fifth IUCN Parks Summit in Durban in 2003, the Malagasy government committed to a tripling of protected areas. This declaration led to a wave of protected area creations, with 28 provisional creation decrees published between April 2006 and December 2007, and a global decree bringing the number of new protected areas to 97 in 2008. These decrees did not designate a manager, leaving the management to the Ministry of the Environment. This seems to have been the general practice, and managers were in place in the majority of the newly protected areas

Table 1: Number and Area of Protected Areas created before and after 2008

[H]

Enactment period by decree	**Number of Protected Areas**			**Area (km²)**		
	Terrestrial	Non-terrestrial	Total	Terrestrial	Non-terrestrial	Total
Starting in 2008	71	9	80	47,482	6,815	54,298
Before 2008	43	0	43	21,801	0	21,801
Total	114	9	123	69,283	6,815	76,099

Source: Calculations by authors based on data from the Service de la Gouvernance des Aires Protégées (SGAP) of the Ministère de l'environnement et Développement Durable (MEDD)

in the following years. However, it was not until 2015 that a final decree ratified these creations. Some of these protected areas were the subject to early management transfer decrees, between 2011 and 2015. Uncertainty remains about the exact date of these early transfers, but since our study period spans a broader interval (2008-2021), its helps compensates for any inaccuracy in determining the exact start date of treatment, i.e actual conservation. In this study, the treatment is defined as an area that is declared as protected by a decree and that is operationally managed.

Table 1 presents the distribution of protected areas (in number and area) according to their period of designation by decree, taking 2008 as the reference year. In the treatment period (2008 to 2021), 71 protected areas were created covering 47,282 km².

In the context of our study, the population considered as treated encompasses households living in rural areas within 10 km of a protected area created between 2008 and 2021, according to the GPS coordinates provided in the Demographic Health Surveys (DHS) data. These GPS coordinates correspond to the centroids of the enumeration areas surveyed. To protect respondent confidentiality, these coordinates are first randomly shifted using the following procedure: An offset angle between 0 and 360 is randomly drawn, then an offset distance is randomly drawn, between 0 and 2 km in urban areas and between 0 and 5 km in rural areas. For 1% of rural clusters, the distance drawn is between 0 and 10 km (Skiles et al., 2013). The 10 km threshold used in our analysis corresponds to the distance most commonly used in similar impact evaluations (see references in Appendix A). Households in the control group are those living in a rural area more than 10 km away from a protected area created between 2008 and 2021, and they exhibit very similar characteristics or share significant traits with households in the treatment group. To check the robustness of the analysis, we also test other distances (5 km and 15 km).

We decided to exclude rural populations living within 10 km of protected areas created before 2008, as they are considered treated before the study period. Urban areas were also excluded from the study because the mechanisms by which protected areas could affect the living conditions of urban populations are likely very different from those rural populations, which are mainly farmers. Due to the marked differences in living standards between rural and urban areas, including urban households would also risk unnecessarily increasing the variance of our estimates. Furthermore, according to the 2018 census, urban population accounts for 24% of the Malagasy population (INSTAT, 2020a), and this proportion is likely even lower among population living near protected areas.

In some cases, the 10 km zone around a newly created protected area may overlap with the 10 km zone around an existing older protected areas. In such a case, the treatment date taken into account is the date of the first creation in chronological order. If the older protected area was created before 2008, the overlapping zone will be excluded from the analysis.

2.4 Sample and statistical power

2.4.1 Sample

The unit of analysis in our study is the household. This is indeed a level at which a significant portion of individual resources are pooled, and it is at this level that data on living standards – the outcome variable of our study – are available in national surveys (Deaton, 1997).

The data on household living conditions used for this study comes from surveys conducted by the '*Institut National de la Statistique de Madagascar'* (INSTAT) as part of the Demographic Health Surveys (DHS) programs. The DHS surveys rely on two-stage stratified sampling. The population of interest was divided into 23 study areas corresponding to the 22 regions of Madagascar, the capital Antananarivo (considered separately) and the Analamanga region without the capital (to isolate the effect of the capital on the regional results). Except for the capital, two strata were created in each study area: The urban stratum and the rural stratum. A total of 45 sampling strata were created. At the first stage, enumeration areas (referred to as 'clusters' hereafter) were randomly selected within each area with a probability proportional to the population of the cluster according to the most recent census. A complete enumeration of households was conducted in each selected cluster, with 657 clusters in 2021, 600 in 2008 and 270 in 1997. At the second stage, a sample of households was randomly drawn from within these clusters to be invited to participate in the survey. This method helps to reduce survey costs while ensuring the representativeness of the results.

Our study will be based on samples from the DHS surveys, which form repeated cross-sections over 13 years 'treatment' period of (2008-2021). To validate the comparability hypothesis between our treatment and control groups during the treatment period, we will analyze their trends over an 11 years 'pre-treatment' period (1997-2008).

2.4.2 Statistical power

The calculation of statistical power aims to estimate the probability that the study detects an effect when it exists. The expected power threshold α is commonly set at 0.8, which means that the study has an 80 % chance of rejecting the null hypothesis when an effect is present. Furthermore, the significance threshold α is set at 0.05, indicating a 5 % risk of incorrectly rejecting the null hypothesis when it is true.

The minimal detectable effect represents the smallest difference that the study can detect between the treatment and control groups, taking into account the parameters chosen for power and the specification threshold. To measure this effect, we use the standardized effect size, also known as Cohen's d, defined as follows:

$$d = \frac{\bar{X}_1 - \bar{X}_2}{s}$$

Table 2: Statistical power of rural household percentile wealth index study

[h]

Metric	**Value**
Intra-Cluster Correlation	0.4783
Adjusted Standard Deviation	28.8680
Cohen's d	0.2582
Minimal Detectable State in Outcome Variable unit	7.4548

Source: Authors' calculations from DHS 2021 data

where $\bar{X}_1$ and $\bar{X}_2$ are the average of the outcome variable for the treatment and control groups, and s is the combined standard deviation of the two groups. The outcome variable used for the power calculation is the rural household percentile of the wealth index (see Section 4.1.1, calculated based on the wealth index variable present in the DHS survey.

We use the average and standard deviation values from the treatment group for both groups as the matching process has not yet been performed. We assume that the control group, once matched, will be similar to the treatment group in terms of variability and sample size.

Prior the power calculation, we segmented the wealth index into percentiles rather than quintiles to better capture the continuity of the variable and maximize the sensitivity of the econometric analysis. The approach, already used by some researchers (Staveteig & Mallick, 2014), preserves greater granularity in the variable while improving the precision of the estimates. By increasing the number of classes, we will reduce the risk of bias associated with artificial discontinuities and improve the ability of our models to capture the marginal effects of conservation policies on living standards.

With these characteristics, we calculate that the minimal detectable effect is 7.5, indicating that the study is able to detect a minimum change of 7.5 percentiles in the wealth index between the groups (see Table 2)

3 Data

3.1 Data collection and processing

Considering the long-term, large-scale, complex, and politically sensitive nature of the intervention to be evaluated, along with the availability of relevant data at the national level, experimentation appears neither feasible not appropriate. As a result, our study is observational in nature. It uses secondary data on the socioeconomic conditions of households, their geographical environment and their location in relation to protected areas.

3.1.1 Data on household socioeconomic conditions

The DHS covers a wide range of topics, including demographic characteristics, living conditions, health, education, sanitation, and hygiene. DHS surveys were performed in Madagascar in 1997, 2008 and 2021. Additionally, the DHS program accompanied INSTAT surveys in 2011, 2013 and 2016, known as the Malaria Indicator Surveys (MIS). These surveys, which focus specifically on malaria related issues, do not include certain health or demographic questions that are included in the DHS surveys. However, all the variables relating to household living conditions that we will use in our study from the DHS surveys are also present in the MIS surveys.

The MIS data can therefore be mobilized for this study. However, the periods covered by MIS only allow for comparisons beyond a simple ‘before’ and ‘after’ framework, which complicates the analysis and necessitates methods with limited methodological consensus. Since 2020, the state of the art in difference-in-difference methods has been challenged by the recognition that classical two-way fixed effects methods can yield spurious results when the effects of an intervention are heterogeneous and assessed over multiple periods. Since then, a dozen alternative approaches have been developed, but econometricians continue to debate their reliability (Roth et al., 2023). To avoid undermining the credibility of our results, we will initially limit ourselves to two study periods (DHS 2008 and 2021). Only if this approach lacks sufficient statistical power will we consider incorporating additional data.

The data on the household geophysical environment mobilized are DHS Data and maybe MIS if statistical power is sufficient for the analysis.

3.1.2 Data on the household geophysical environment

Household geophysical environment data will be obtained using the R package mapme.biodiversity (Görgen & Bhandari, 2022). This package automates the retrieval and processing of large raster-format data to produce a series of indicators applied to user-defined polygons for specified periods. It has the advantage of providing a list of data defined by the most recognized sources in the literature, automatically retrieving them, and applying state of the art methods to calculate spatio-temporal indicators from these sources. It can handle large datasets using efficient routines and parallel computing. The data mobilized by this channel are Global Forest Change datasets for forest cover rate, Nasa SRTM data for slope and elevation, Worldpop data for population density, and accessibility data in 2000.

3.1.3 Data on protected areas

Most studies of protected areas rely on the World Database on Protected Areas (WDPA), widely considered as the most comprehensive registry of marine and terrestrial protected areas (Bingham et al., 2019). However, in Madagascar, after triangulated data with other sources – including records from the official Malagasy body overseeing protected areas, official decrees, and documentation from protected area managers – we found that WDPA data are inaccurate for about one-third of protected areas. The inaccuracies include classifications of IUCN status, incorrect creation dates, and delineations that include buffer zones extending beyond the boundaries of protected areas. Such lack of reliability could undermine the credibility of the results if we use WDPA data. Thus, a study of the degradation of the effectiveness of protected areas during COVID-19 pandemic(Eklund et al., 2022), which relied on WDPA data, was heavily criticized by Andrianambinina (2023) for using flawed sources, leading to errors that compromised its conclusions.

The authoritative Malagasy body that administers data on the boundaries of Madagascar’s protected areas is a department of the Ministry of the Environment and Sustainable Development called the *Service de la Gouvernance des Aires Protégées* (SGAP). These are the data we want to prioritize for our study, and we obtained the permission for use.

4 Analysis

4.1 Variables used in the analysis

The data on the socioeconomic conditions of households and their geophysical environment described above are used to construct the variables in the impact evaluation model.

4.1.1 Outcome variables

In this analysis, two outcome variables are considered: household living standards (primary variable) and inequalities in living standards at the level of the surveyed localities (secondary variable).

- **Main outcome variable: Household living standards**

The first outcome variable, household living standard, is estimated from the wealth index, calculated specifically for rural areas (variable coded hv270a in the DHS data). The wealth index is defined in the DHS data catalogue as : 'A *composite measure of a household's cumulative living standard. The wealth index is calculated using easy-to-collect data on a household's ownership of selected assets, such as televisions and bicycles; materials used for housing construction; and types of water access and sanitation facilities. Generated with a statistical procedure known as principal components analysis, the wealth index places individual households on a continuous scale of relative wealth. DHS separates all interviewed households into five wealth quintiles to compare the influence of wealth on various population, health and nutrition indicators. As a response to criticism that a single wealth index is too urban in its construction and not able to distinguish the poorest of the poor from other poor households, this variable provides an urban- and rural-specific wealth index'* (Program/ICF, 2018). *As described above, we will translate this wealth index into an integer between 1 and 100, corresponding to the household's wealth percentile relative to the distribution of the whole sample.*

According to the DHS variable catalog by survey generation, the variable hv270a is available for the 2021 DHS survey, but not for previous years. However, the procedure for calculating the wealth index is well-documented (Rustein & Kiersten, 2004), and the DHS program provides the SPSS codes for overall calculations encompassing both rural and urban areas (variable hv270), as well as the weights associated with each asset category for the variable hv270[10]. These elements will enable us to reproduce and validate, in R, the calculation of the variable hv270 (wealth index) for 2008 and 1997, and then extrapolate the method to drive the variable hv270a (rural wealth index) for these years.

- **Secondary outcome variable: inequality of household living standards**

In addition to the evaluation impact of protected areas on household living standards, we will seek to understand their influence on socioeconomic inequalities within the affected populations. To do this, we propose to use a standardized Z-Score of the wealth index, allowing for the comparison of the relative distribution of wealth around the mean within the study population, at the level of each survey cluster.

The Z-Score Z_i for each household i is calculated from the wealth index using the following formula:

[10] Available at https://www.dhsprogram.com/topics/wealth-index/Wealth-Index-Construction.cfm/, consulted on 30/08/2024

$$Z_i = \frac{W_i - \mu_W}{\sigma_W}$$

where W_i is the wealth index for household i , μ_W represents the average wealth index for all households surveyed in each rural cluster, and σ_W is the standard deviation of households in the cluster.

4.1.2 Matching variables

Matching covariates are used to select units not exposed to conservation (control groups) that are comparable to the exposed units (treatment group). Appropriate covariates for the matching process are variables likely to influence both the probability of treatment (whether a protected area has been created near the household) and the outcome (household living standard and inequalities among households). The literature shows that protected areas tend to be created in less dense, less accessible, higher and steeper regions (Joppa & Pfaff, 2010a). These variables may also affect living standards : areas that are more dense, flat, low-lying and accessible (in terms of travel time and geography) tend to be wealthier(Gallup et al., 1999). We will use the following list of matching variables :

- Forest cover rates in 2000 correspond to the percentage coverage by vegetation with a height of 5 meters or more (Hansen et al., 2013). This variable is provided by the Global Forest Change dataset, which indicates for each pixel of one arc-second (approximately 30 meters at the equator) an estimate of forest cover in 2000 (Hansen et al., 2013). In Madagascar, as in other countries, protected areas have been preferentially created by targeting forest zones(Carvalho et al., 2020; Wilson et al., 2006). Globally, Naidoo et al. (2019) indicates that forest cover in an area is negatively correlated with the living standards of its inhabitants.

- Slope and elevation are calculated using a digital terrain model from NASA-SRTM (Shuttle Radar Topography data), which provides an elevation estimate for arc-second pixels (NASA JPL, 2020). Slope is measured as a percentage for each plot, while elevation is measured in meters. These topological variables influence the location of protected areas (Joppa & Pfaff, 2010b), as well as the agricultural potential of an area and therefore the living standards of rural populations (Canavire-Bacarreza & Hanauer, 2013).

- Population density in 2000 corresponds to the estimated number of inhabitants per km^2 based on Worldpop data. Worldpop data provide estimates of population density for the year 2000 at a spatial resolution of about 1 km. They use modeling techniques and combines census data with various geospatial datasets (WorldPop & CIESIN, 2018).

- Accessibility in 2000 corresponds to the estimated travel time for households to the nearest cities, measured in minutes. These accessibility data to cities are compiled by the Joint Research Center (JRC), with 2000 as the reference year (Uchida & Nelson, 2011). Accessibility to cities determines the ability to benefit from the services, products and opportunities they offer and is therefore a key factor in determining living standards in rural areas (INSTAT, 2020b; Weiss et al., 2018).

Each of these variables will be calculated for a circle of 10 km radius around the GPS coordinates of the survey household cluster.

4.1.3 Control variables

Matching variables listed above will be integrated as control variables in the difference-in-difference regression stage to increase the robustness of the model by minimizing the bias that would arise from residual imbalances. Other variables, such as the age and sex of the head of the household or rainfall, do not introduce bias as they do not influence treatment assignment (the location of protected areas), but they increase the variance of the results. They are added as control variables to reduce the unexplained variability of the model and thus more accurately estimates the treatment effect.

The variables on age and sex of the head of household are taken from the DHS surveys. These characteristics determine the economic opportunities and resource management capacity of households (Lo Bue et al., 2022). Sex of the head of household is recoded as v151 and v152 *in the DHS survey.*

We also integrate the Standardized Precipitation Evapotranspiration Index (SPEI) (Vicente-Serrano et al., 2010) for the year preceding the survey. This index is calculated based on a long-term reference (1981-2010) to quantify excess or deficit of rainfall. We use the SPEI for the year prior to the survey. The SPEI is calculated from monthly precipitation and minimum and maximum temperature data from worldclim, using an improved version of the Hargreaves method defined by Droogers and Allen (2002), as implemented in the SPEI R package (Vicente-Serrano et al., 2010). This variable will be calculated for a circle with a radius of 10 km around the GPS coordinates of the cluster.

We chose not to include migration among the controls, as this variable is likely to be influenced by the direct consequences of protected areas (reduction in land use, jobs related to conservation, etc.), but also by the evolution of living standards in areas surrounding the protected areas (see Figure 1). This position in the logical diagram presented in Figure 1 makes migration a 'collider', which would introduce an artificial bias if it were taken as a control variable (Hünermund & Bareinboim, 2023; Imbens, 2020).

4.1.4 Weights

In the DHS data, each observation is associated with a sampling weight, reflecting the inverse of its probability of selection in the sample. We will take these weights into account to maintain the representativeness of our estimates relative to the national population. In addition, when we use the genetic matching algorithm (see below), observations assigned an additional weight aimed at improving comparability between the treatment and control groups. To obtain a final weighting that accounts for both representativeness (sampling weight) and comparability (matching weight), we multiply these two types of weights. Thus, each observation is assigned a final weight that incorporates both dimensions.

4.2 Statistical methods and models

The statistical methods used in this analysis combine matching and difference-in-difference (DID) to estimate the causal effect of protected areas on household livelihoods.

4.2.1 Matching

Matching maximizes comparability between treatment and control groups in terms of observable characteristics (Ho et al., 2007). This method increases the similarity between households living in or within 10 km of protected areas (treatment group) and households not affected by protected areas (control group) (Desbureaux, 2021; Schleicher et al., 2020b).

We will perform the matching using the genetic matching method proposed by the GenMatch() function in the R package MatchIt. This variant of nearest neighbor matching combines all variables into a single measure of distance known as Mahalanobis distance. This distance measures the difference between units in the matched groups to quantify the similarity between the two groups of observations, while accounting for correlations between the covariates and their covariances (Diamond & Sekhon, 2013, p. 935).

Rosenbaum (1983) also suggests using a distance measure in combination with allowed maximum distances between matched units (called 'calipers') to avoid creating pairs with excessively large differences. In practice, the caliper is commonly set at an interval of 0.25 standard deviations of the Mahalanobis distance, which corresponds to the most common recommendation in the literature.

Matching models rely on the assumption that the distributions of the covariates are similar between the treatment and control groups, making the evaluation of the covariate balance crucial. The validity of the estimates depends directly on this balance, highlighting the importance of conducting tests to measure it.

e will conduct a two-step balance test : before matching to determine the extent to which the raw data are initially unbalanced and to check whether balance has been achieved between the covariates after matching. This balance is checked using the Standardized Mean Difference (SMD), which measures the difference between the means of the covariates in the treatment and control groups in order to compare the relative balance of variables measured in different units, divided by a standardization factor to put the difference on a common scale for all covariates. The closer the SMD is to 0, the better the balance between the groups for the covariate in question. Although there is no consensus in the literature, an SMD greater than 0.1 can be considered indicative of a significant imbalance(Austin, 2009). If the SMD exceeds 0.1, we will increase the caliper interval to achieve a SMD ≤ 0.1.

After matching, a visual test of the quality of the matching can be performed using the quantile-quantile plot (Q-Q plot) for each matching covariate. The graphical method provides an overview of the distribution of the covariates. The plot facilitates the identification of imbalances at different levels of the distribution. When the Q-Q plot points are close to the diagonal, the quantiles of the two distributions are similar, indicating good balance between the groups.

In the matching technique, the observations that did not match are excluded. Descriptive statistics will be performed on this excluded population to fully understand the analysis universe.

The result of genetic matching is a selection of matched observations and associated weights. These weights will be combined with the survey weights.

4.2.2 Difference-in-difference

The difference-in-difference principle is to compare the wealth index of control and treatment households before and after the establishment of protected areas. Our treatment began in 2011, so we use 2008 DHS data for the pre-treatment and 2021 DHS data for the post-treatment. This method relies on the parallel trends assumption, meaning that, in the absence of treatment, treatment and control groups would have experienced similar changes over time. To verify this assumption, we include two pre-treatment, points (DHS 1997 and DHS 2008 data) and assess trends between these years.

The simplified equation for the difference-in-difference (Daw & Hatfield, 2018) is as follows :

$$Y_{it} = \alpha_0 + \beta_1 \text{Treatment}_i + \beta_2 \text{Post}_t + \delta D_{it} + \varepsilon_{it}$$

With the parameters defined as:

Y_{it}: Wealth index for household i in year t (1997, 2008 or 2021)

α_0: Intercept, representing the baseline wealth index

β_1: Coefficient associated with the variable $Treatment_i$

$Treatment_i$: Binary variable that takes the value of 1 for households i affected by a protected area (treatment households) and 0 for households i not affected by protected areas (control households)

β_2: Coefficient associated with the variable $Post_t$

$Post_t$: Binary variable that takes the value of 1 for the post-treatment year 2021 and 0 for the pre-treatment year 2008

δ: Coefficient of interest representing the effect of protected area on household living standards in 2008 and 2021

D_{it}: Interaction term between treatment and period that is equal to 1 for treatment households i after the treatment year t and 0 for the other households.

ε_{it}: Error term

The difference-in-difference method relies on the assumption of parallel trends. To validate this assumption, we will use as a reference, among the rural households surveyed in 1997, those living in an area located in or within of a protected area created between 2008 and 2021 (placebo treatment group) and those matched to them using the method described above (placebo control group). We will graphically represent the evolution of the treatment and control groups in 1997, 2008 and 2021 for visual confirmation. To confirm the validity of the parallel trends assumption, we will conduct a placebo test 1997-2008 outcomes, as well as matching and control variables, while defining the treatment variable for post-2008 protected areas. If the estimated effect between 1997 and 2008 is null or statistically insignificant, this supports the parallel trends assumption.

Additionally, if we incorporate data from MIS 1997, MIS 2011, and MIS 2013, we could apply an estimator suitable for staggered adoption and multiple study periods accounting for potential heterogeneous treatment effects (Borusyak et al., 2024). This will only be performed if power is insufficient with two periods of data. .

Robust standard errors will be clustered at the enumeration area level, addressing intra-cluster correlations to ensure accurate inference. Observations with missing values for regression variables will be excluded. If exclusions exceed 2% of the sample, we will analyze their characteristics of excluded observations and discuss potential corrective measures to address bias associated with exclusion.

4.3 Robustness and sensitivity tests

4.3.1 Alternative distance thresholds

We set a threshold of 10 km to define the treatment group, as is common practice in the literature (Section 7) . To verify the robustness of our conclusions, we also conduct additional analyses using distances of 5 km and 15 km. At 5km, smaller sample sizes may lead to wider confidence. At 5km, smaller sample sizes may lead to wider confidence intervals, limiting statistical power. However, we will assess whether the direction and magnitude of the estimated impacts remain consistent across these thresholds. This comparison will help determine how proximity to protected areas affects treatment effect.

4.3.2 Multiple outcome and multiple hypothesis testing

This study evaluates three hypotheses: H1 (overall impact on living standards), H2 (impact on inequalities) and H3 (heterogeneous impact based on governance). H1 is our primary hypothesis. To reduce the risk of Type I errors in secondary hypothesis H2 and H3, we will apply Benjamini-Hochberg's (1995) false discovery rate. This correction ensures that conclusions drawn from H2 and H3 are robust to multiple testing concerns.

4.3.3 Pseudo panel approach

To account for unobserved factors or household-level shocks potentially correlated with the outcome variable, we will use a pseudo-panel approach to estimate fixed effects models at the level of household cohorts (Deaton, 1985). T This method estimates fixed effects models by grouping individuals into exogenous, time-invariant cohorts (e.g., by household head's year of birth, sex, and region). These grouping variables ensure within-cohort variation does not bias results. Following Jung et al. (2019), we will construct these cohorts as follows: '*In order to get consistent estimates from the pseudo-panel approach, grouping variables need to be exogenous, time-invariant, and available for all household in the data (Verbeek,2008). We use household head's sex, birth year, and the region as grouping variables, with birth year divided by quartile, variables commonly used in the literature (e.g., Bernard, et al. 2011 ; Pless and Fell, 2017). Owing to the variation in cohort size between years, which leads to heteroscedasticity, we weigh the observations using the inverse of the square root of cohort size (Dargay, 2007 ; Pless and Fell, 2017)'.*

The model equation is as follows:

$$\overline{Y}_{ct} = \theta_c + D_{ct} + \sum_{k=2015}^{2021} \beta \overline{Dist}_{ct} \mathbb{1}(t = k) + \delta \overline{X}_{ct} + \epsilon_i$$

Y_{ct}: Average value of the household i wealth index within the cohort c at the period t

θ_c: Fixed effect controlling for the time-invariant cohort-specific characteristics

D_{ct} :Time-fixed effects at cohort level

$Dist_{ct}$: Average distance between the location of households and the boundaries of protected areas (where$Dist_{ct} \triangleleft 10km$ within the cohort).

X_{ct} : Average of the control variables (rainfall, drought)

ϵ_i : Error terms

4.3.4 Rosenbaum sensitivity analysis

To evaluate the robustness of the results against potential biases arising from unobserved confounding variables, we will perform a sensitivity analysis using Rosenbaum's method (Rosenbaum, 2002, pp. 105–170). This analysis tests the extent to which potential unobserved biases could alter the results (Keele, 2010, p. 7). A sensitivity parameter Γ measures the degree of deviation from random treatment assignment. We will test a range of Γ values to assess how robust our conclusions are to unobserved confounding. If results remain significant for plausible values of Γ, we can conclude that unobserved bias is unlikely to explain our findings.

4.3.5 Heterogeneous effects

To investigate heterogeneous effects of protected areas, we adopt the following approaches. First, we introduce differentiating variables (sex and age of the head of household, rainfall) to test whether the impact of protected areas varies according to these characteristics (e.g. during wet periods vs dry periods). We also analyze the potential influence of governance by distinguishing between so-called 'strict' protected areas (IUCN categories I-IV) and 'multi-use' areas (IUCN categories V-VI). Furthermore, the use of a pseudo-panel (grouping individuals into exogenous cohorts, for example by age or region) will allow us to examine whether specific cohorts exhibit a more pronounced impact while better controlling for unobserved factors. Next, we study intra-community inequalities using Z index (z-score) of the wealth index; this allows us to see if the effect of protected areas is more concentrated on certain households rather than uniformly affecting the entire local population. Finally, we test different distances (5 km, 10 km, and 15 km) to evaluate whether proximity to protected areas determines the magnitude of the impact. All of these analyses contribute to a better understanding of the factors and populations for which conservation generates differentiated effects

5 Interpreting results

This study evaluates how the creation of protected areas affects the living standards and socioeconomic inequalities of rural households in Madagascar. The results aim to provide actionable guidelines for conservation policies that balance biodiversity protection with the needs and rights of local communities.

Hypothesis 1 assumes that, on average, protected areas reduce the living standards of riparian households. If our matching and difference-in-differences estimates show a decline in the wealth index for these households, this would indicate that restricted access to natural resources, coupled with insufficient compensation or benefit-sharing mechanisms, has adversely affected their living standards. This conclusion would then invite public authorities and conservation stakeholders to put in place more inclusive mechanisms, such as better sharing of tourism revenues, targeted cash transfers or maintaining regulated access to certain resources, to prevent increased vulnerability.

If the z-score results show that conservation exacerbates inequalities (H2), this would suggest that 'local elites' disproportionately benefit from financial opportunities linked to tourism or protected area management. In such cases, corrective measures would be necessary to prevent benefits from being concentrated among a small group. Examples include enforcing local employment quotas or ensuring transparent participatory governance.

If the analysis does not show a decline in livelihoods, this would indicate that some households have successfully diversified their incomes or benefited from ecosystem services such as water access, erosion control, soil fertility, or pollination. Such an observation would strengthen the thesis that, in certain favorable institutional or market contexts, investment in biodiversity can coexist with dynamic local development. This observation would then encourage the expansion of conservation while replicating the approaches that are most effective in stimulating or maintaining household livelihoods. If, in addition, the analysis demonstrates the absence of negative effects on inequalities, or even an effect that promotes their alleviation, this would strengthen the social and political acceptability of a rapid expansion of protected areas.

Finally, hypothesis 3 concerns the heterogeneity of effects based on the type of governance of protected areas. If the study finds that certain management approaches—such as those involving local communities– lead to more favorable and equitable outcomes, these results would support scaling up participatory models that have been promoted over the past fifteen years. Conversely, if participatory governance does not appear to play a decisive role, this would suggest prioritizing other explanatory factors, such as donor support or local context, before drawing broader conclusions about the social effectiveness of conservation.

These results will contribute to the debate on the compatibility of conservation and poverty reduction by identifying how institutional configurations or governance models influence the vulnerability of rural households. They will provide new insights for analyzing biodiversity protection policies in low-income countries.

6 Bibliography

Adams, William. M., Aveling, R., Brockington, D., Dickson, B., Elliott, J., Hutton, J., Roe, D., Vira, B., & Wolmer, W. (2004). Biodiversity Conservation and the Eradication of Poverty. *Science*, *306*(5699), 1146–1149. https://doi.org/10.1126/science.1097920

Adeniyi, O., Adekunle, W., Afolabi, J., & Kumeka, T. T. (2024). Tourism-income inequality Nexus in Africa: Evidence from SADC countries. *Current Issues in Tourism*, *27*(18), 2899–2917. https://doi.org/10.1080/13683500.2023.2227377

Andrianambinina, O. D., Rafanoharana, S. C., Rasamuel, H. A. T., Waeber, P. O., Ganzhorn, J. U., & Wilmé, L. (2023). Decrease of deforestation in Protected Areas of Madagascar during the Covid-19 years. *Madagascar Conservation & Development*, *18*(1), 15–21. https://www.ajol.info/index.php/mcd/article/view/262765

Austin, P. C. (2009). Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. *Statistics in Medicine*, *28*(25), 3083–3107. https://doi.org/10.1002/sim.3697

Bertrand, A., Serpantié, G., Randrianarivelo, G., Montagne, P., Toillier, A., Karpe, P., Andriambolanoro, D., & Derycke, M. (2012). Contre un retour aux barrières : Quelle place pour la gestion communautaire dans les nouvelles aires protégées malgaches ? *Les Cahiers d'Outre-Mer*, *257*(1), 85–123. https://doi.org/10.4000/com.6493

Bingham, H. C., Juffe Bignoli, D., Lewis, E., MacSharry, B., Burgess, N. D., Visconti, P., Deguignet, M., Misrachi, M., Walpole, M., & Stewart, J. L. (2019). Sixty years of tracking conservation progress using the World Database on Protected Areas. *Nature Ecology & Evolution*, *3*(5), 737–743. https://idp.nature.com/authorize/casa?redirect_uri=https://www.nature.com/articles/s41559-019-0869-3&casa_token=Zmluq7JmhxUAAAAA:Id6cAJrCDpmtydpiga8LlmkB5wb9a1O_bwCGwCeA1gz314-F6VT5ekCOeVMJo_Zrpbh_uuoXf0fwD3Txjg

Borusyak, K., Jaravel, X., & Spiess, J. (2024). Revisiting Event-Study Designs: Robust and Efficient Estimation. *The Review of Economic Studies*, rdae007. https://doi.org/10.1093/restud/rdae007

Canavire-Bacarreza, G., & Hanauer, M. M. (2013). Estimating the impacts of Bolivia's protected areas on poverty. *World Development*, *41*, 265–285. https://www.sciencedirect.com/science/article/pii/S0305750X12001702

Carvalho, F., Brown, K. A., Gordon, A. D., Yesuf, G. U., Raherilalao, M. J., Raselimanana, A. P., Soarimalala, V., & Goodman, S. M. (2020). Methods for prioritizing protected areas using individual and aggregate rankings. *Environmental Conservation*, *47*(2), 113–122. https://doi.org/10.1017/S0376892920000090

Chung, M. G., Dietz, T., & Liu, J. (2018). Global relationships between biodiversity and nature-based tourism in protected areas. *Ecosystem Services*, *34*, 11–23. https://doi.org/10.1016/j.ecoser.2018.09.004

Conceição, P. (Ed.). (2024). *Breaking the gridlock: Reimagining cooperation in a Polarized world*. UNDP.

Daw, J. R., & Hatfield, L. A. (2018). Matching and Regression to the Mean in Difference-in-Differences Analysis. *Health Services Research*, *53*(6), 4138–4156. https://doi.org/10.1111/1475-6773.12993

Deaton, A. (1985). Panel data from time series of cross-sections. *Journal of Econometrics*, *30*(1-2), 109–126. https://doi.org/10.1016/0304-4076(85)90134-4

Deaton, A. (1997). *The Analysis of Household Surveys: A Microeconometric Approach to Development Policy*. World Bank Publications.

Desbureaux, S. (2021). Subjective modeling choices and the robustness of impact evaluations in conservation science. *Conservation Biology*, *35*(5), 16151626. https://doi.org/10.1111/cobi.13728

Diamond, A., & Sekhon, J. S. (2013). Genetic Matching for Estimating Causal Effects: A General Multivariate Matching Method for Achieving Balance in Observational Studies. *Review of Economics and Statistics*, *95*(3), 932–945. https://doi.org/10.1162/REST_a_00318

Droogers, P., & Allen, R. G. (2002). Estimating Reference Evapotranspiration Under Inaccurate Data Conditions. *Irrigation and Drainage Systems*, *16*(1), 33–45. https://doi.org/10.1023/A:1015508322413

Eklund, J., & Cabeza, M. (2017). Quality of governance and effectiveness of protected areas: Crucial concepts for conservation planning. *Annals of the New York Academy of Sciences*, *1399*(1), 27–41. https://doi.org/10.1111/nyas.13284

Eklund, J., Coad, L., Geldmann, J., & Cabeza, M. (2019). What constitutes a useful measure of protected area effectiveness? A case study of management inputs and protected area impacts in Madagascar. *Conservat Sci and Prac*, *1*(10), e107. https://doi.org/10.1111/csp2.107

Eklund, J., Jones, J. P., Räsänen, M., Geldmann, J., Jokinen, A.-P., Pellegrini, A., Rakotobe, D., Rakotonarivo, O. S., Toivonen, T., & Balmford, A. (2022). Elevated fires during COVID-19 lockdown and the vulnerability of protected areas. *Nature Sustainability*, *5*(7), 603–609.

Froger, G., & Méral, P. (2009). Les temps de la politique environnementale à Madagascar : Entre continuité et bifurcations. In *Diversité des politiques de développement durable : Temporalités et durabilités en conflit à Madagascar, au Mali et au Mexique* (pp. 45–67). Karthala ; GEMDEV. https://www.documentation.ird.fr/hor/fdi:010048701

Gallup, J. L., Sachs, J. D., & Mellinger, A. D. (1999). Geography and Economic Development. *International Regional Science Review*, *22*(2), 179–232. https://doi.org/10.1177/016001799761012334

Ghosh, S., & Mitra, S. K. (2021). Tourism and inequality: A relook on the Kuznets curve. *Tourism Management*, *83*, 104255. https://doi.org/10.1016/j.tourman.2020.104255

Golden, C. D., Bonds, M. H., Brashares, J. S., Rodolph Rasolofoniaina, B. J., & Kremen, C. (2014). Economic valuation of subsistence harvest of wildlife in Madagascar. *Conservation Biology*, *28*(1), 234–243. https://doi.org/https://doi.org/10.1111/cobi.12174

Golden, C. D., Rasolofoniaina, B. R., Anjaranirina, E. G., Nicolas, L., Ravaoliny, L., & Kremen, C. (2012). *Rainforest pharmacopeia in Madagascar provides high value for current local and prospective global uses*.

Görgen, D. A., & Bhandari, O. P. (2022). *Efficient Monitoring of Global Biodiversity Portfolios*. https://doi.org/10.32614/CRAN.package.mapme.biodiversity

Hansen, M. C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S. A., Tyukavina, A., Thau, D., Stehman, S. V., Goetz, S. J., Loveland, T. R., & others. (2013). High-resolution global maps of 21st-century forest cover change. *Science*, *342*(6160), 850853. http://www.sciencemag.org/content/342/6160/850.short

Hanson, J. K., & Sigman, R. (2021). Leviathan's Latent Dimensions: Measuring State Capacity for Comparative Political Research. *The Journal of Politics*, *83*(4), 1495–1510. https://doi.org/10.1086/715066

Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2007). Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference. *Political Analysis*, *15*(3), 199236. https://doi.org/10.1093/pan/mpl013

Hünermund, P., & Bareinboim, E. (2023). Causal inference and data fusion in econometrics. *The Econometrics Journal*, utad008.

Imbens, G. W. (2020). Potential outcome and directed acyclic graph approaches to causality: Relevance for empirical practice in economics. *Journal of Economic Literature*, *58*(4), 11291179.

Imbens, G. W. (2024). Causal Inference in the Social Sciences. *Annual Review of Statistics and Its Application*, *11*(Volume 11, 2024), 123–152. https://doi.org/10.1146/annurev-statistics-033121-114601

INSTAT. (2020a). *Résultats globaux du troisième recensement général de la population et de l'habitation de 2018 à Madagascar (RGPH-3)*. Institut national de la statistique.

INSTAT. (2020b). *Résultats globaux du troisième recensement général de la population et de l'habitation de 2018 à madagascar (RGPH-3)*. Institut national de la statistique.

Joppa, L., & Pfaff, A. (2010a). Reassessing the forest impacts of protection: The challenge of nonrandom location and a corrective method. *Annals of the New York Academy of Sciences*, *1185*(1), 135–149. https://doi.org/10.1111/j.1749-6632.2009.05162.x

Joppa, L., & Pfaff, A. (2010b). Reassessing the forest impacts of protection: The challenge of nonrandom location and a corrective method. *Annals of the New York Academy of Sciences*, *1185*(1), 135149. https://doi.org/10.1111/j.1749-6632.2009.05162.x

Jung, S., Liao, C., Agrawal, A., & Brown, D. G. (2019). Evidence on Wealth-Improving Effects of Forest Concessions in Liberia. *Journal of the Association of Environmental and Resource Economists*, *6*(5), 961–1000. https://doi.org/10.1086/704614

Kandel, P., Pandit, R., White, B., & Polyakov, M. (2022). Do protected areas increase household income? Evidence from a Meta-Analysis. *World Development*, *159*, 106024. https://doi.org/10.1016/j.worlddev.2022.106024

Keele, L. (2010). *An overview of rbounds: An R package for Rosenbaum bounds sensitivity analysis with matched data.*

Lo Bue, M. C., Le, T. T. N., Santos Silva, M., & Sen, K. (2022). Gender and vulnerable employment in the developing world: Evidence from global microdata. *World Development*, *159*, 106010. https://doi.org/10.1016/j.worlddev.2022.106010

Ma, B., Zhang, Y., Hou, Y., & Wen, Y. (2020). Do Protected Areas Matter? A Systematic Review of the Social and Ecological Impacts of the Establishment of Protected Areas. *IJERPH*, *17*(19), 7259. https://doi.org/10.3390/ijerph17197259

Mammides, C. (2020). Evidence from eleven countries in four continents suggests that protected areas are not associated with higher poverty rates. *Biological Conservation*, *241*, 108353. https://www.sciencedirect.com/science/article/pii/S0006320719312777?casa_token=1saHx-9SppkAAAAA:sw9KzbZvzqu2WLub5u-K06mA2kgTygSvTi5AEsjBz0rUm8h3h9SKsdId52pG5VEr4SobaFTfguA

Maxwell, S. L., Cazalis, V., Dudley, N., Hoffmann, M., Rodrigues, A. S., Stolton, S., Visconti, P., Woodley, S., Kingston, N., & Lewis, E. (2020). Area-based conservation in the twenty-first century. *Nature*, *586*(7828), 217–227.

McKinnon, M. C., Cheng, S. H., Dupre, S., Edmond, J., Garside, R., Glew, L., Holland, M. B., Levine, E., Masuda, Y. J., Miller, D. C., Oliveira, I., Revenaz, J., Roe, D., Shamer, S., Wilkie, D., Wongbusarakum, S., & Woodhouse, E. (2016). What are the effects of nature conservation on human well-being? A systematic map of empirical evidence from developing countries. *Environ Evid*, *5*(1), 8. https://doi.org/10.1186/s13750-016-0058-7

Naidoo, R., Gerkey, D., Hole, D., Pfaff, A., Ellis, A. M., Golden, C. D., Herrera, D., Johnson, K., Mulligan, M., Ricketts, T. H., & Fisher, B. (2019). Evaluating the impacts of protected areas on human well-being across the developing world. *Science Advances*, *5*(4), eaav3006. https://doi.org/10.1126/sciadv.aav3006

NASA JPL. (2020). *NASADEM Merged DEM Global 1 arc second V001*. NASA EOSDIS Land Processes Distributed Active Archive Center. https://doi.org/10.5067/MEASURES/NASADEM/NASADEM_HGT.001

Neudert, R., Ganzhorn, J. U., & Waetzold, F. (2017). Global benefits and local costs–The dilemma of tropical forest conservation: A review of the situation in Madagascar. *Environmental Conservation*, *44*(1), 82–96.

Poudyal, M., Rakotonarivo, O. S., Razafimanahaka, J. H., Hockley, N., & Jones, J. P. G. (2018). Household economy, forest dependency & opportunity costs of conservation in eastern rainforests of Madagascar. *Sci Data*, *5*(1), 180225. https://doi.org/10.1038/sdata.2018.225

Poudyal, M., Ramamonjisoa, B. S., Hockley, N., Rakotonarivo, O. S., Gibbons, J. M., Mandimbiniaina, R., Rasoamanana, A., & Jones, J. P. (2016). Can REDD+ social safeguards reach the 'right'people? Lessons from Madagascar. *Global Environmental Change*, *37*, 31–42. https://www.sciencedirect.com/science/article/pii/S095937801630005X

Program/ICF, T. D. (2018). *Standard recode manual for DHS-7*. USAID.

Rivière, M. (2017). *Les (dé)connexions du développement : ethno-géographie systémique de l'aide au développement et à la conservation forestière à Amindrabe, Madagascar* [PhD thesis]. https://theses.hal.science/tel-01720849

Rosenbaum, P. (1983). *The central role of the propensity score in observational studies for causal effects*.

Rosenbaum, P. (2002). *Observational Studies*.

Roth, J., Sant'Anna, P. H. C., Bilinski, A., & Poe, J. (2023). What's trending in difference-in-differences? A synthesis of the recent econometrics literature. *Journal of Econometrics*, *235*(2), 2218–2244. https://doi.org/10.1016/j.jeconom.2023.03.008

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, *66*(5), 688–701. https://doi.org/10.1037/h0037350

Rustein, S., & Kiersten, J. (2004). *The DHS Wealth Index*. ICF International.

Schleicher, J., Eklund, J., D. Barnes, M., Geldmann, J., Oldekop, J. A., & Jones, J. P. (2020b). Statistical matching for conservation science. *Conservation Biology*, *34*(3), 538–549.

Schleicher, J., Eklund, J., D. Barnes, M., Geldmann, J., Oldekop, J. A., & Jones, J. P. (2020a). Statistical matching for conservation science. *Conservation Biology*, *34*(3), 538–549.

Skiles, M. P., Burgert, C. R., Curtis, S. L., & Spencer, J. (2013). Geographically linking population and facility surveys: Methodological considerations. *Population Health Metrics*, *11*(1), 14. http://www.biomedcentral.com/content/pdf/1478-7954-11-14.pdf

Staveteig, S., & Mallick, L. (2014). *Intertemporal comparisons of poverty and wealth with DHS data: A harmonized asset index approach*. ICF International.

Uchida, H., & Nelson, A. (2011). Agglomeration Index: Towards a New Measure of Urban Concentration. *Urbanization and Development: Multidisciplinary Perspectives*. https://doi.org/10.1093/acprof:oso/9780199590148.003.0003

Vicente-Serrano, S. M., Beguería, S., & López-Moreno, J. I. (2010). A Multiscalar Drought Index Sensitive to Global Warming: The Standardized Precipitation Evapotranspiration Index. *Journal of Climate*, *23*(7), 1696–1718. https://doi.org/10.1175/2009JCLI2909.1

Weiss, D. J., Nelson, A., Gibson, H. S., Temperley, W., Peedell, S., Lieber, A., Hancher, M., Poyart, E., Belchior, S., & Fullman, N. (2018). A global map of travel time to cities to assess inequalities in accessibility in 2015. *Nature*, *553*(7688), 333–336.

Wilson, K. A., McBride, M. F., Bode, M., & Possingham, H. P. (2006). Prioritizing global conservation efforts. *Nature*, *440*(7082), 337340. https://www.nature.com/articles/nature04366

WorldPop, & CIESIN. (2018). *Global 1km Population*. Global High Resolution Population Denominators Project. https://doi.org/10.5258/SOTON/WP00647

Xuanming, P., Dossou, T. A. M., Dossou, K. P., & Alinsato, A. S. (2024). The impact of tourism development on social welfare in Africa: Quantile regression analysis. *Current Issues in Tourism*, *27*(7), 1159–1172. https://doi.org/10.1080/13683500.2023.2214351

7 Appendix

Table 3: Comparable study parameters of the impact of spatialized policies on household livelihoods

References	Title	Study coverage	Treatment variable	Outcome variables	Matching method	Matching variables	Control variables	Buffer zone	Estimation methods	Comments
Wittermyer et al., 2008 (Science)	Accelerated Human Population Growth at Protected Area Edges	306 PAs in 45 countries in Africa and Latin America	Protected areas	average annual rates of population growth	Does not use matching	Does not use matching	Does not use matching	calculated average annual rates of population growth within a 10-km buffer of 306 rural (10) International Union for Conservation of Nature (IUCN) category I and II PAs and nature World Heritage Sites	Wilcoxon test	

Andrade et al, 2012 (Ecology and Society)	Protected Areas and Local Communities: an Inevitable Partnership toward Successful Conservation Strategies?	South and Central America, 25 in Africa and 24 in Asia	Protected areas	Level of compliance of local community with park's policies	Does not use matching	Does not use matching	Does not use matching	Applies a buffer zone of 10 km around each PA, making it possible to calculate the population density for each PA within a radius of 10 km from its border	Regression model	The analysis conduct a systematic review and meta-analysis to identify which factors may influence the level of community compliance with PA regulations.The variable is: Age, Area, Existence of buffer zones, IUCN category, Population density, Country adult literacy rates, GDP PPP per capita, PA budget per country, level of local community participation in PA management

Miranda et al., 2014 (Inter-American Bank)	Effects of Protected Areas on Forest Cover Change and Local Communities	Peru (2000-2005)	Protected areas	Deforestation disturbance, Capita income,capita expenditure,poverty rate,extreme poverty rate	Quasi-experimental matching techniques + Mahalanobis distance matching	Elevation, Slope, Aspect, Average precipitation, Average maximum temperature, Average mean temperature, Distance to nearest, Population center $> 10000 hab$, Distance to nearest population center, Proportion land suitable for forestry, Water source in house, Electric lighting literacy, Primary school, Education, Employment in agriculture or forestry, Population density	Elevation, Slope, Aspect, Average precipitation, Average maximum temperature, Average mean temperature, Distance to nearest, Population center $> 10000 hab$, Distance to nearest population center, Proportion land suitable for forestry, Water source in house, Electric lighting literacy, Primary school, Education, Employment in agriculture or forestry, Population density	Two distance defined are households located less than 5 km from the boundary of the protected zone and households located less than 10 km away	Not mentionned	

Oldekop et al., 2016 (conservation biology)	A global assessment of the social and conservation outcomes of protected areas	165 PAs	Protected areas	PA proprety (protection management, governance, geographical region, size, biome) and Social impacts (displacement, monetary, livelihood, cultural, compensation, conflict, empowerment, unequal distribution of impact)	Does not use matching	Does not use matching	Does not use matching	10 km of PA boundaries	Regression model	The analysis conduct a meta-analysis to assess how Pas affect the well-being of local people, the factors associated with these impacts, the relationship between Pas' conservation and socioeconomic outcomes.

Chung et al., 2018 (Elsevier)	Global relationships between biodiversity and nature-based tourism in protected areas	929 PAs in 50 countries	Protected areas	average annual visitor numbers for each PA	Does not use matching	Does not use matching	Does not use matching	Buffer zone variables with 10-km distance increments across PAs boundaries	Regression model	We designated 10-km distance intervals from the PA boundary (0-km) to 50-km buffer zones.Then, we extracted numerical values from the PA boundary and each of five rings (0–10, 10–20, 20–30, 30–40 and 40–50 km) using the spatial dataset

Naidoo et al. 2019 (Science Advances)	Evaluating the impacts of protected areas on human well-being across the developing world	84 developing countries	Households near protected areas	Children height-for-age, stunting Wealth index, Probability of being poor (<100)	Genetic matching	Tree cover, Distance to nearest road, Year of DHS survey, Elevation, Annual precipitation, Population density, Education, level male-headed household, Anthropogenic pressure index, Urban household, Child breastfed time to nearest, water source	Tree cover, Distance to nearest road, Year of DHS survey, Elevation, Annual precipitation, Population density, Education, level male-headed household, Anthropogenic pressure index, Urban household Child breastfed time to nearest, water source	Households living within 10 km of a PA close enough to be affected by PA presence	Bayesian hierarchical regression	No comparison before and after treatment

Jung et al.,2019 (Journal of the Association of Environmental and Resource Economists)	Evidence on Wealth-Improving Effects of Forest Concessions in Liberia	Liberia (2007-2013)	Households within 5 km of concession boundaries	Wealth score quartile (DHS)	Mahalanobis distance matching + caliper method after matching	under five, Household head sex, Household head age, livestock, Bank, Road, Forest, Town, Other concessions	Religion, Education	The distance thresholds used are rather conservative in identifying affected households, using all observations within 10 km of the concession boundaries	Difference-in-difference(event study specification)	Mention quartiles but wealth index are normally available in quintiles

Nowakowski et al., 2023 (nature sustainability)	Co-benefits of marine protected areas for nature and people	Meso American region (2005-2018)	Proximity to marine protected areas	Fish abundance (biomass, diet), Wealth index, Stunting of children (height for age)	Mahalanobis distance matching + caliper method after matching	Fish abundance, Coral cover, Reef, Habitat, Depth human Footprint Chlorophyll-a, climate change, MPA fishing restrictions, MPA enforcement, MPA age, MPA are, Distance to coat, Natural land cover, travel time to cities, Population density, elevation, Education level, Country (HN),Distance to coast, MPA fishing restrictions	Fish abundance, Coral cover, Reef, Habitat, Depth human Footprint Chlorophyll-a climate change, MPA fishing restrictions, MPA enforcement, MPA age, MPA area, Distance to coat, Natural land cover, travel time to cities, Population density, elevation, Education level, Country (HN), Distance to coast, MPA fishing restrictions	Use of statistical matching of DHS survey clusters close (≤10 km) to MPAs with those far (>10 km) from MPAs	Bayesian GLMMs	

Source: Authors

8 Administrative information

8.1 Funding

The study is performed in the framework of the BETSAKA project. The BETSAKA project is cofunded by the Development Impact Lab of the German KfW Development Bank; the *Agence Française de Développement* (AFD) and the French Research Institute for Sustainable Development (IRD).

8.2 Institutional Review Board (Ethical approval)

The project includes the analysis of personal data collected through surveys. In order to manage these data effectively and ethically, a data management plan will be implemented. This plan will comply with the good practices defined by authoritative bodies such as the French Ministry of Research in its National Plan for Open Science and the European Commission in the guidelines of the European Research Council (ERC). These good practices aim to ensure that the collection, storage, processing and sharing of data are conducted in a manner that respects privacy, ensures confidentiality and maintains data integrity. The main aspects of these guidelines include obtaining informed consent from participants, the anonymity of personal data when necessary, and establishing clear protocols for data access and sharing. All participants in the research project will therefore be informed and trained to respect these rules and best practices.

The implementation of the data management plan is based on the DMP-Opidor system developed by INIST-CNRS, recommended by the '*Agence Nationale de la Recherche et l'Institut de Recherche et de Développement' (IRD).* DMP-Opidor is a tool designed to help researchers create, manage and share their data management plan. It provides a structured framework that guides researchers through the various stages of data management, ensuring compliance with legal and ethical requirements.

8.3 Declaration of interest

Iriana Razafimahenina, Florent Bédécarrats, Ingrid Dallmann and Holimalala Randriamanampisoa have declared that they have no conflicts of interest.

9 Acknowledgements

We greatly appreciate the contribution of the researchers and conservation officers in Madagascar through the discussions and exchanges we were able to have.