Does protected area connectivity moderate the efficacy of protection on tropical biodiversity? Evidence from a replication of Brodie et al. 2023

Peter Kedron, Lei Song, Wenxin Yang, Amy Frazier

2024-02-20

Abstract

This study is a *reproduction* of:

> Brodie, J.F., Mohd-Azlan, J., Chen, C. et al. Landscape-scale benefits of protected areas for tropical biodiversity. Nature 620, 807–812 (2023). https://doi.org/10.1038/s41586-023-06410-z

We replicate the analysis of Brodie et al. (2023) and introduce protected area (PA) connectivity as a statistical moderator of the effect PA status has on biodiversity. We find KEY FINDINGS. Using a causal framework that controls for forest structure, site accessibility, and geographic location through matching, Brodie et al. (2023) find evidence that protected areas (PA) do preserve vertebrate biodiversity within their boundaries and in the adjacent unprotected landscape. Brodie et al. provides evidence of the efficacy of protected area status, they do not assess whether the effect they observe is altered by network connectivity.

1 Study metadata

This replication uses the data file provided by Brodie et al. (2023) at https://doi.org/10.6084/m9.figshare.22527298.v1. We independently accessed the World Database on Protected Areas https://www.protectedplanet.net/en/thematic-areas/wdpa?tab=WDPA to construct our site and PA connectivity measures.

- `Key words`: Biodiversity, Conservation, Protected Areas, Connectivity, 30x30
- `Subject`: Ecology and Evolutionary Biology, Natural Resources and Conservation,
- `Date created`: November 8, 2023
- `Date modified`: date of most recent revision
- `Spatial Coverage`: Southeast Asia
- `Spatial Resolution`: Species observations - GPS located point data, GEDI-derived forest structural covariates - 1 km raster, HDI - country-level, Protected Areas - PA Polygons
- `Spatial Reference System`: Specify the geographic or projected coordinate system for the study, e.g. EPSG:4326
- `Temporal Coverage`: 01-2015 to 08-2021
- `Temporal Resolution`: Specify the temporal resolution of your study—i.e. the duration of time for which each observation represents or the revisit period for repeated observations

2 Study design

This study consists of a reproduction of the original work by Brodie et al. (2023) and a replication to analyze how connectivity influences protected area efficacy in preserving tropical biodiversity. We first implement the workflow as described and shared in Brodie et al. (2023) as identically as possible to reproduce the original

study. We then compute and add connectivity measures at the sample points to the original dataset to further examine the effect of connectivity.

Brodie et al. (2023) formulated hypotheses to assess protected area effectiveness and tested them with structural causal modeling. The main focus of their paper was on the overall effectiveness of protected areas (OR-H1) on biodiversity after accounting potential confounders of site accessibiltiy and forest structure.

2.1 Protected area effectiveness on overall biodiversity

> OR-H1: Biodiversity is higher within protected areas than outside (after removing effects of site accessbility and habitat condition).

Debates exist on whether protected areas are effective because of their remote location and good habitat conditions or conservation status itself. Motivated by such, Brodie et al. (2023) estimated protected areas efficacy for conserving tropical mammal and bird diversity after de-confounding the effects of site accessibility and habitat condition (H1). Their outcome variables are various aspects of biodiversity. The treatment variable is whether the observation point is within or outside a reserve. If the treatment variable is significant in a causal model, it supports H1 and adds evidence for the effectiveness of protecting lands.

To do so, they removed confounding effects and built causal models for birds vs mammals, and different biodiversity outcome metrics. Two major confounders identified were site accessibility and habitat quality. Site accessibility was proxied using Human Development Index (HDI), circuit theory-based measures of proximity to human development[1], and an interaction term of the two variables. Habitat quality was proxied using three-dimensional habitat structure metrics derived from the Global Ecosystem Dynamics Investigation (GEDI) mission. Confounding effects were removed using statistical matching based on propensity scores. Biodiversity metrics used as outcome variables include species richness (SR), functional richness (FR), and phylogenetic diversity (PD). PA effects were estimated using separate mixed effects linear regression for different biodiversity metrics and for birds vs mammals.

Brodie et al. (2023) found protection status increased all facets of bird biodiversity but the effects were not significant for mammals.

2.2 Adding connectivity

We ask the research question of whether protected areas moderate the efficacy of protection on tropical biodiversity. We formulate the hypothesis is as follows:

> RPL-H1: Connectivity mdoerates the efficacy of protection on biodiversity.

We build upon the existing structural causal models by Brodie et al. (2023) to test the hypothesis. The original study archetype is quasi-experimental as it uses a matching strategy to perform causal analysis.

3 Materials and procedure

3.1 Computational environment

The reproduction of the original study is conducted in MacBook Pros. Brodie et al. (2023) cleaned up the variables and performed propensity score matching and causal analysis in R. They did not provide scripts on how they derived the biodiversity metrics, circuit theory-based metrics, HDI, and the GEDI metrics. We built upon and annotated R scripts shared by the original study, and added missing information on HDI and GEDI metrics to the scripts. // Is this accurate?

Required packages are as follows:

[1] "circuit theoretical models parameterized with human travel speeds across different terrains and the locations of populations centers and transportation networks" (Brodie et al. 2023)

```r
# library(groundhog)
pkgs <- c("tidyverse", "cowplot", "here", "dagitty", "ggdag", "Hmisc",
          "MatchIt", "modelsummary", "optmatch", "nlme")
# groundhog.library(pkgs, "2024-02-11")
# I don't know why groundhog is not working
lapply(pkgs, require, character.only=TRUE)
```

3.2 Data and variables

In the original study, outcome variables for the causal models are biodiversity metrics (i.e., species richness, functional richness, and phylogenetic diversity) derived from species observations. Mammal observations were assembled by the authors from camera traps in 65 study areas in the study region. Bird observations were gathered from eBird from 2015/01 to 2021/08 following a set of filtering procedures.

The treatment variable for ORIG-H1 was derived from the World Database on Protected Areas (WDPA) by UNEP-WCMC, which is a binary variable on whether the sampling point is within protected areas or not.

Observations were matched based on propoensity scores of their geographic locations (i.e., latitudes and longitudes), forest canopy height, accessibility, and HDI. Predictors in the mixed-effects linear regression models were forest canopy height, site accessibility, HDI, and treatment variables.

We gathered data shared by Brodie et al. (2023). Tabular data of most model inputs were provided in a figshare. Raster files at 1-km resolution for GEDI derived metrics and circuit-based accessibility were shared through a weblink. Variable sources are presented in Table 1.

Table 1. Variables used in Brodie et al. (2023)

Name	Source	Usage
Biodiversity metrics - mammals	Authors	Outcome variable
Biodiversity metrics - birds	eBird	Outcome variable
Protected area boundaries	WDPA	Treatment variables (whether inside PAs)
Ground elevation	NASA GEDI L2B	Predictor - elevation and topography
Circuit-based site accessibility (log transformed)	Authors	Predictor - site accessibility
Human Development Index	Human Development Report 2020	Predictor - site accessibility
Forest structure metrics	NASA GEDI L2A	Predictor - forest structure

3.2.1 eBird

The original study gathered bird observations from eBird.

- `Title`: eBird.
- `Abstract`: A community science platform for reporting bird sightings.
- `Spatial Coverage`: Tropical region (overlapping countries of Brunei, Cambodie, China, Indonesia, Laos, Malaysia, Singapore, Thailand, and Vietnam).
- `Spatial Resolution`: Vector.
- `Spatial Reference System`: Not specified.
- `Temporal Coverage`: 2015/01 - 2021/08.
- `Temporal Resolution`: Not applicable.
- `Lineage`: Brodie et al. (2023) queried and subset data directly from eBird website or its R package or API.
- `Distribution`: eBird webpage and other download methods.
- `Constraints`: Non-commercial use.

- `Data Quality`: Although a direct data quality layer is not associated, Brodie et al. (2023) followed recommendations from existing studies to filter out data points.

Variables constructed were as follows:

Table 2 Variables created from bird observations via eBird.

Label	Alias	Definition	Type	Accuracy	Domain	Missing Data Value(s)	Missing Data Frequency
SR.mean	Species richness	Number of species	Float	Unknown	Equal or greater than 0	Not applicable	Unknown
maxFRic	Functional richness	Diversity of species functional traits	Float	Unknown	Equal or greater than 0	Not applicable	Unknown
asymptPD	Phylogenetic diversity	Cumulative evolutionary time of the species assemblage	Float	Unknown	Equal or greater than 0	Not applicable	Unknown

3.2.2 The World Database on Protected Areas

The original study used protected area boundaries to derive the three treatment variables (Table 3). Brodie et al. (2023) did not specify how they cleaned up protected area boundary. We dissolved PAs and then converted them from multi-parts to single-parts to avoid double counting. For mammals, we excluded marine protected areas.

// need to confirm with Lei.

- `Title`: The World Database on Protected Areas (WDPA).
- `Abstract`: A global database on protected areas (PAs) and other effective conservation measurers (OECM).
- `Spatial Coverage`: Tropical region (overlapping countries of Brunei, Cambodie, China, Indonesia, Laos, Malaysia, Singapore, Thailand, and Vietnam).
- `Spatial Resolution`: Vector.
- `Spatial Reference System`: WGS 84.
- `Temporal Coverage`: Latest.
- `Temporal Resolution`: Updated monthly.
- `Lineage`: Brodie et al. (2023) queried and subset data directly from eBird website or its R package or API.
- `Distribution`: WDPA webpage.
- `Constraints`: Non-commercial use.
- `Data Quality`: Unknown.

Table 3 Variables created from protected area boundaries.

Label	Alias	Definition	Type	Accuracy	Domain	Missing Data Value(s)	Missing Data Frequency
PA	Within or outside PAs	Whether the point is inside a PA or not	Binary	Not applicable	1 for inside and 0 for outside	Not applicable	Unknown
PA_size_km2	Functional richness	Diversity of species functional traits	Float	Unknown	Equal or greater than 0	Not applicable	Unknown
dist_to_PA	Phylogenetic diversity	Cumulative evolutionary time of the species assemblage	Float	Unknown	Equal or greater than 0	Not applicable	Unknown

3.2.3 GEDI L2 metrics

The Global Ecosystem Dynamics Investigation (GEDI) is a spaceborne light detection and ranging (LiDAR) mission monitoring forest structure on earth. The original study derived both ground elevation and forest structure metrics from the Level 2 dataset of GEDI. Level 2 GEDI data are at footprint level so Brodie et al. (2023) used krigging interpolation to create wall-to-wall layers (1-km resolution).

L2A includes elevation data. The original study computed slope and topographic position index (TPI) to represent topographic traits at each site. The original study used five L2B metrics, which were canopy height (relative height at 95%), plant area volume density (PAVD) between 0 and 5 m (represents understory density), cumulative plant area index from ground to canopy top, foliage height diversity of plant area index, and proportional cover. They found the five forest structure metrics to be highly correlated and only kept canopy height and understory density in models.

- `Title`: The Global Ecosystem Dynamics Investigation Level 2 Elevation and Height Metrics.
- `Abstract`: Global footprint level observations from GEDI on ground elevation and forest structure.
- `Spatial Coverage`: Tropical region (overlapping countries of Brunei, Cambodie, China, Indonesia, Laos, Malaysia, Singapore, Thailand, and Vietnam).
- `Spatial Resolution`: Footprints are of 25-m resolution and extrapolated into 1-km resolution.
- `Spatial Reference System`: WGS 84.
- `Temporal Coverage`: 2019/04/17 to 2022/04/12
- `Temporal Resolution`: Not applicable.
- `Lineage`: Brodie et al. (2023) queried and subset data directly from eBird website or its R package or API.
- `Distribution`: Original GEDI L2 metrics can be derived from NASA website and Brodie et al. (2023) shared krigged results on a webpage.
- `Constraints`: Non-commercial use.
- `Data Quality`: Original GEDI L2 metrics have quality and degrade flags and Brodie et al. (2023) kept only data points of satisfying quality.

Table 4 Variables derived from GEDI L2 metrics.

Label	Alias	Definition	Type	Accuracy	Domain	Missing Data Value(s)	Missing Data Frequency
elev	Elevation	Ground elevation at the site (krigged)	Integer	Unknown	Equal to or greater than 0 (terrestrial observations)	Unknown	Unknown
slope	Slope	Slope of topography	Float	Unknown	0 to 90	Unknown	Unknown
TPI	Topographic Position Index	Difference between the elevation of a focal raster cell with those of its neighbors (not mentioned in paper)	Float	Unknwon	Not bounded	Unknown	Unknown
rh_95_a0.pred	Relative height at 95%	Roughly the top canopy height (krigged)	Float	Unknown	Equal to or greater than 0	Unknown	Unknown
pavd_0_5.pred	Plant area volume density from 0 to 5 m	A proxy of understory forest density (krigged)	Float	Unknown	Equal to or greater than 0	Unknown	Unknown
pai_a0.pred	Plant area index	Cumulative PAI from ground to canopy (krigged)	Float	Unknown	Equal to or greater than 0	Unknown	Unknown
fhd_pai_1m_a0.pred	Foliage height diversity	Shannon's diversity of PAI across heights (krigged)	Float	Unknown	Equal to or greater than 0	Unknown	Unknown
cover_a0.pred	Proportional coverage	Openness or closeness of canopy (krigged)	Float	Unknown	0 to 1	Unknown	Unknown

3.2.4 Human Development Index (HDI)

Human Development Index for each country was included in the causal models as a simple metric of socioeconomic level.

- `Title`: Human Development Index
- `Abstract`: An index on the level of human development by country.
- `Spatial Coverage`: Tropical region (overlapping countries of Brunei, Cambodie, China, Indonesia, Laos, Malaysia, Singapore, Thailand, and Vietnam).
- `Spatial Resolution`: Not applicable.
- `Spatial Reference System`: Not applicable.
- `Temporal Coverage`: 2020.
- `Temporal Resolution`: Not applicable.
- `Lineage`: Direct query thourgh the official website.
- `Distribution`: Acquired directly through Human Development Report 2020.
- `Constraints`: Non-commercial use.
- `Data Quality`: Unknown.
- `Variables`: HDI
  - `Label`: Not included in the files shared by the original study, we added HDI values and labeled as 'HDI'.
  - `Alias`: Human Development Index
  - `Definition`: A metric on the level of human development.
  - `Type`: Float
  - `Accuracy`: Unknown
  - `Domain`: 0 to 1
  - `Missing Data Value(s)`: Not applicable
  - `Missing Data Frequency`: Not applicable

3.3 Prior observations

At the beginning of this analysis, we had observed the dataset provided by Brodie et al with their publication. We noticed the following caveats in the script and raw data shared by the original study. 1) HDI was missing. 2) The process of computing biodiversity metrics and GEDI metrics was unclear. 3) The procedure for preprocessing and cleaning PA boundaries was unclear.

We did not manipulate the data before beginning our reproduction attempt.

3.4 Bias and threats to validity

Given the research design as described in the original paper and primary data shared, we find that potential spatial autocorrelation of sample points were not addressed. In addition, uncertainty issues were not discussed thoroughly in the paper, which includes 1) the representativeness of biodiversity measures, 2) the validity of forest structure measures created through krigging, and 3) accessibility represented by country-level Human Development Index.

// should we add a map/moran's I result here? // other concerns to add?

3.5 Data transformations

We did the following data transformations:

1) Cleaned up PAs

3.5.1 Clean protected areas (takes too long to run so I skipped many lines –> should discuss)

As Brodie et al. did not specify how they pre-processed protected area boundaries, we did the following steps to remove issues such as overlapping and double counting areas.

```
# Load libraries, easy to switch to use groundhog
# something wrong with the path in rmd --> reset path
# requires PhantomJS: webdriver::install_phantomjs()
pkgs <- c("sf", "dplyr", "terra", "optparse", "wdpar", "here")
sapply(pkgs, require, character.only = TRUE)
```

```
##       sf    dplyr    terra optparse    wdpar     here
##     TRUE     TRUE     TRUE     TRUE     TRUE     TRUE
```

```
sf_use_s2(FALSE) # deal with buffering odd

# Command line inputs
option_list <- list(
    make_option(c("-s", "--src_dir"),
                action = "store", default = "data/raw/public", type = 'character',
                help = "The source directory for reading data [default %default]."),
    make_option(c("-d", "--dst_dir"),
                action = "store", default = 'data/derived/public', type = 'character',
                help = paste0("The path to save the csv [default %default].")))
opt <- parse_args(OptionParser(option_list = option_list))

# Directories and paths
src_dir <- opt$src_dir
dst_dir <- opt$dst_dir
```

First, we assumed the geographic coordinate system adopted by the original paper from the field names of their variables (i.e., utm) and used UTM Zone 46 (EPSG:32646). We used the spatial extent of a GEDI metric layer provided to derive the study area, and added a 100 km buffer to avoid potential edge effects. // is it 100 km?

```
# Read samples and raster template
## According to the pairs of lat/lon and east/north, they used
## UTM Zone 46 (EPSG:32646) for projection, so here we will use the same one.
fnames <- list.files(file.path(src_dir, "training"), full.names = TRUE)
print(file.path())
```

```
## character(0)
```

```
pts <- do.call(rbind, lapply(fnames, function(fname){
    read.csv(fname) %>%
        select(all_of(names(.)[
            stringr::str_detect(names(.), "station|country|lat|long")])) %>%
        st_as_sf(coords = c(3, 2), crs = 4326)
}))

template <- rast(
    file.path(src_dir, "GEDIv002_20190417to20220413_cover_krig.tiff")) %>%
    extend(c(100, 100)) # add a buffer
# is the buffer 100 km?
values(template) <- 1:ncell(template)

########################## Query Protected areas (PAs) ##########################
```

```
## Within this part, most of the issues related to PAs are solved
## 1. Project the PAs or relevant layers to use precise distance.
## 2. Trim the PAs to terrestrial only.
## 3. Separate or union polygons and re-index them.

# Query and clean PAs
#raw_pas <- c("KHM", "CHN", "IDN", "LAO", "MYS",
#             "SGP", "THA", "VNM", "BRN") %>%
#    lapply(wdpa_fetch, wait = TRUE,
#           download_dir = rappdirs::user_data_dir("wdpar")) %>%
#    bind_rows()
#raw_pas <- wdpa_clean(raw_pas, crs = "EPSG:32646",
#                      geometry_precision = 100000)
```

Second, we removed marine protected areas and cleaned protected area boundaries by countries' administrative boundaries. We computed sizes of each protected areas after cleaning up PAs (e.g., issues such as overlap).

```
#raw_pas <- raw_pas %>% filter(MARINE != "marine")
## Note: now the No is 1638

# Trim the polygons to terrestrial only
## Clip all not marine polygons to administrative border.
## Use the same data source in examples of R package wdpar.

#raw_adm_bry <- lapply(
#    c("KHM", "CHN", "IDN", "LAO", "MYS",
#      "SGP", "THA", "VNM", "BRN"), function(iso){
#          file_path <- tempfile(fileext = "rds")
#          lk <- "https://biogeo.ucdavis.edu/data/gadm3.6/Rsf"
#          download.file(sprintf("%s/gadm36_%s_0_sf.rds", lk, iso), file_path)
#          readRDS(file_path)}) %>% bind_rows()

## Process the boundary a bit
#adm_bry <- raw_adm_bry %>%
#    st_set_precision(100000) %>%
#    sf::st_make_valid() %>%
#    st_set_precision(100000) %>%
#    st_combine() %>%
#    st_union() %>%
#    st_set_precision(100000) %>%
#    sf::st_make_valid() %>%
#    st_transform(st_crs(raw_pas)) %>%
#    sf::st_make_valid()

## Clip PAs to administrative boundary

#clean_pas <- raw_pas %>% st_intersection(adm_bry)
## No of PAs drop from 1638 to 1618

# Crop PAs to extent (study region's bounding box)
## WARNING: remember to re-run this step again after union/separate the polygons

#bbox <- st_as_sfc(st_bbox(template)) %>% st_transform(st_crs(clean_pas))
#clean_pas <- clean_pas %>%
```

```
#    slice(unique(unlist(suppressMessages(st_intersects(bbox, .)))))
## Note: No of PAs drop further to 1260

# Union/separate polygons.
# ^ Wenxin: not 100% sure what this mean?

#clean_pas <- st_cast(st_union(clean_pas), "POLYGON") %>%
#    st_as_sf() %>% rename(geometry = x)
## No of PAs change from 1260 to 4270 (lose administrative meaning)

# Crop to extent again

#clean_pas <- clean_pas %>%
#    slice(unique(unlist(suppressMessages(st_intersects(bbox, .))))) %>%
#    mutate(index = 1:nrow(.))
## No of PAs change from 4270 to 4259

# Re-calculate the area

#clean_pas <- clean_pas %>%
#    mutate(REP_AREA = st_area(.) %>% units::set_units("km2"))

# Clean the tiny ones with "partial" marine type due to a high chance to be the
# noisy remaining of the clip

#false_pas <- st_join(clean_pas, raw_pas %>% select(MARINE)) %>%
#    filter(MARINE != "terrestrial" & REP_AREA < 0.01 %>% units::set_units("km2"))

#clean_pas <- clean_pas %>% filter(!index %in% unique(false_pas$index))
```

Finally, we clustered contiguous PAs whose boundaries touch each other for computing connectivity of mammal species and saved cleaned PA outputs.

```
########################## Cluster Protected areas (PAs) ##########################
## 4. Cluster the PAs that are connected.
## The assumption here is that it is difficult (if not possible) for mammals
## to pass the ocean or huge waterbodies.
## Warning: this is only for mammal species
# Wenxin: Was this (or the entire PA cleaning) done only for connectivity analysis? Should I still incl

#adm_reorg <- raw_adm_bry %>%
#    st_transform(st_crs(clean_pas)) %>%
#    st_buffer(10) %>% st_union() %>% st_buffer(-10) %>% st_cast("POLYGON") %>%
#    st_intersection(bbox) %>% st_as_sf() %>% mutate(group = 1:nrow(.)) %>%
#    mutate(area = st_area(.) %>% units::set_units("km2"))

#clean_pas <- st_join(clean_pas, adm_reorg) # No is 2873

## Clean further
#false_pas <- clean_pas %>%
#    filter(area > 5e+05 %>% units::set_units("km2") &
#               REP_AREA < 1 %>% units::set_units("km2"))
#clean_pas <- clean_pas %>% filter(!index %in% unique(false_pas$index)) # 2198
```

```
## Re-index
#clean_pas <- clean_pas %>% mutate(index = 1:nrow(.)) %>%
#    select(-area)

#adm_reorg <- adm_reorg %>% filter(group %in% clean_pas$group) %>%
#    rename(geometry = x) %>% select(group)

# Save out
# st_write(clean_pas, file.path(dst_dir, "clean_pas.geojson"))
# st_write(adm_reorg, file.path(dst_dir, "pa_groups.shp"))

# skipping and reading existing ones in case needed
clean_pas <- st_read(file.path(dst_dir, 'clean_pas.geojson'))
```

```
## Reading layer `clean_pas' from data source
##   `/Users/wenxinyang/Desktop/GitHub/RPl-Brodie-2023/data/derived/public/clean_pas.geojson'
##   using driver `GeoJSON'
## Simple feature collection with 2198 features and 3 fields
## Geometry type: POLYGON
## Dimension:     XY
## Bounding box:  xmin: 745225 ymin: -840825 xmax: 3601414 ymax: 2894781
## Projected CRS: WGS 84 / UTM zone 46N
```

```
adm_reorg <- st_read(file.path(dst_dir, 'pa_groups.shp'))
```

```
## Reading layer `pa_groups' from data source
##   `/Users/wenxinyang/Desktop/GitHub/RPl-Brodie-2023/data/derived/public/pa_groups.shp'
##   using driver `ESRI Shapefile'
## Simple feature collection with 737 features and 1 field
## Geometry type: POLYGON
## Dimension:     XY
## Bounding box:  xmin: 743023 ymin: -841329 xmax: 3602001 ymax: 2944096
## Projected CRS: WGS 84 / UTM zone 46N
```

3.5.2 Conceptualization and causal diagram

Before testing the hypotheses, we re-constructed the causal diagram for ORIG-H1.

```
# Load libraries
# library(groundhog)
pkgs <- c("dagitty", "ggdag")
# groundhog.library(pkgs, "2023-12-05")
lapply(pkgs, require, character.only=TRUE)
```

```
## [[1]]
## [1] TRUE
##
## [[2]]
## [1] TRUE
```

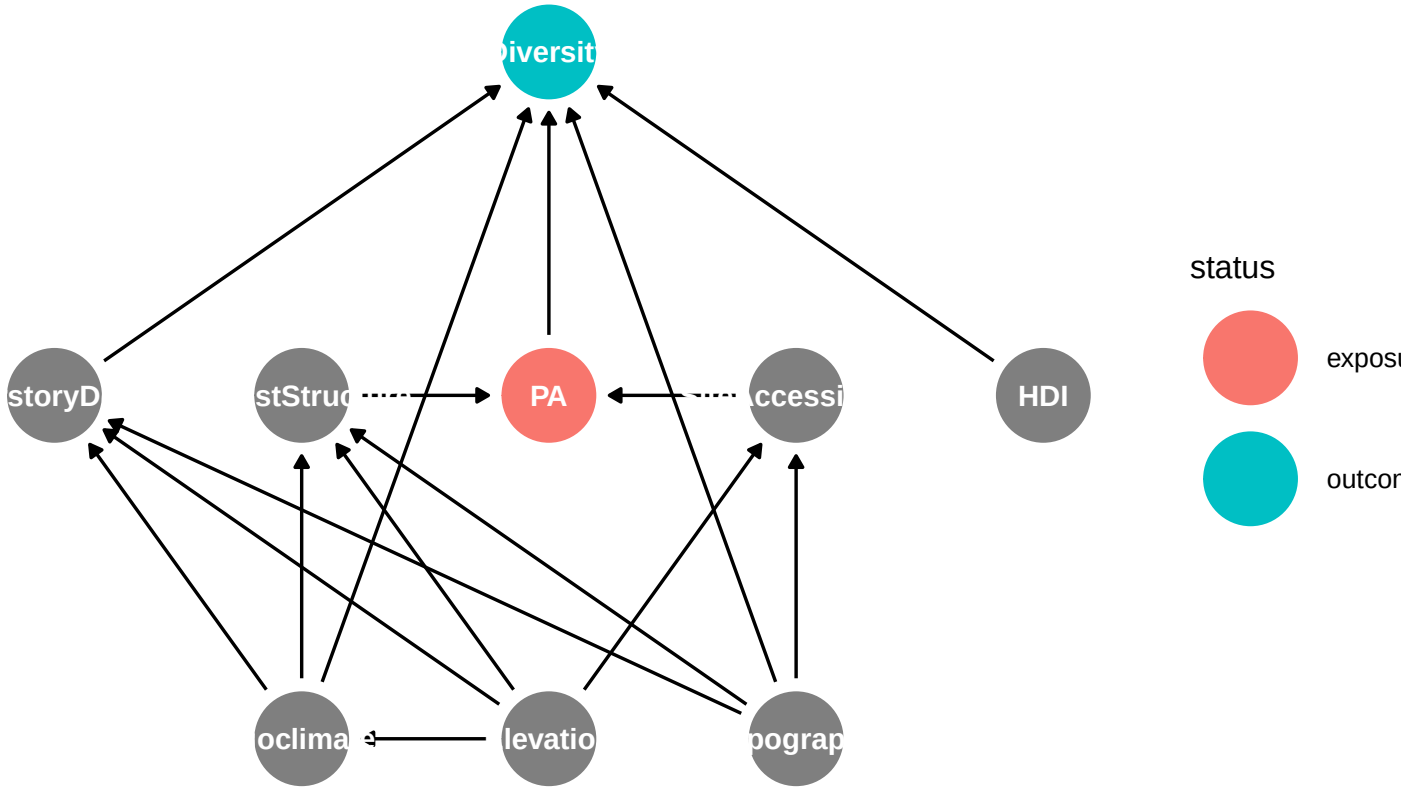
```
# ------------------------------------------------------------------------------
# Structural Causal Modelling
# ------------------------------------------------------------------------------

# Create the directed acyclic graph (DAG) of potential causal pathways with the
# addition of connectivity as a moderator of the PA effect. Assign PA as
```

```
# the exposure and diversity as the outcome.
# Connectivity -> Diversity
dagBrodie <- dagitty("dag {
  PA -> Diversity
  ForestStructure -> PA
  SiteAccessibility -> PA
  Bioclimate -> ForestStructure -> PA -> Diversity
  Bioclimate -> UnderstoryDensity -> Diversity
  Bioclimate -> Diversity
  Elevation -> Bioclimate
  Elevation -> ForestStructure
  Elevation -> UnderstoryDensity
  Elevation -> SiteAccessibility
  Topography -> UnderstoryDensity
  Topography -> ForestStructure
  Topography -> SiteAccessibility
  Topography -> Diversity
  HDI -> Diversity
  PA [exposure]
  Diversity [outcome]
               }"
)

# Organize the data into a visual hierarchy
coordinates( dagBrodie ) <-  list(x = c(Diversity = 3,
                                        UnderstoryDensity = 1,
                                        ForestStructure = 2,
                                        PA = 3,
                                        SiteAccessibility = 4,
                                        HDI = 5,
                                        Bioclimate = 2,
                                        Elevation = 3,
                                        Topography = 4
                                        #, Connectivity = 6
                                        ),
                                  y = c(Diversity = 3,
                                        UnderstoryDensity = 2,
                                        ForestStructure = 2,
                                        PA = 2,
                                        SiteAccessibility = 2,
                                        HDI = 2,
                                        Bioclimate = 1,
                                        Elevation = 1,
                                        Topography = 1
                                    #    , Connectivity = 2
                                        ))

# Plot the DAG to confirm the visual structure and exposure
ggdag_status(dagBrodie) + theme_dag()
```




```
# Identify the set of adjustment variables needed to identify
# the effect of PA on biodiversity and Test whether connectivity fulfills the
# adjustment criterion
adjustmentSets(dagBrodie)
```

```
## { Bioclimate, Topography, UnderstoryDensity }
## { Bioclimate, Elevation, Topography }
## { Elevation, ForestStructure, Topography }
## { ForestStructure, SiteAccessibility }
```

```
# isAdjustmentSet(dagBrodie, c("Connectivity"))

# Plot the alternative adjustment sets
# ggdag_adjustment_set(dagBrodie,
#                      node_size = 20,
#                      text_col = "black"
# ) + ggplot2::theme(legend.position = "bottom")
```

3.5.3 Data preparation for bird models

We performed a series of data preparation procedures based on scripts nad raw data shared by Brodie et al. (2023) for birds and mammals separately. Steps included 1) adding country-specific HDI values which were missing from the raw data, 2) standardizing variables, and 3) following the original script to discard different outliers for different models.

```
# rm(list = ls())

# Load the data provided by Brodie et al. (2023)
dat_brodie_bird <- data.frame(read.csv(
                          here("data/raw/public/training/bird_data_230326.csv"),
```

```
                          header = T))

# Simplify the variable names of site identifier and geographic coordinates
names(dat_brodie_bird)[names(dat_brodie_bird) == "site"] <- "station"
names(dat_brodie_bird)[names(dat_brodie_bird) == "lat_wgs84"] <- "lat"
names(dat_brodie_bird)[names(dat_brodie_bird) == "long_wgs84"] <- "long"

# Search for HDI in the column names
grep("hdi", names(dat_brodie_bird), value = TRUE)
```

```
## character(0)
```

```
grep("HDI", names(dat_brodie_bird), value = TRUE)
```

```
## character(0)
```

```
# Assign stations the HDI value of its country
# Reference: Human Development Report 2020: The Next Frontier-Human Development
# and the Anthropocene (United Nations Development Programme, 2020).
# https://hdr.undp.org/data-center/human-development-index#/indicies/HDI
# https://hdr.undp.org/sites/default/files/2021-22_HDR/HDR21-22_Statistical_Annex_HDI_Table.xlsx

dat_HDI <- data.frame(
    country = unique(dat_brodie_bird$country),
    HDI = c(0.593, 0.768, 0.705, 0.607, 0.803, 0.939, 0.800,
            0.703, 0.829))
dat_brodie_bird <- left_join(dat_brodie_bird, dat_HDI, by = "country")

# Create dataframe containing the subset of variable used in the analysis
dat_bird <- dat_brodie_bird %>% select(station, country, PA, utm_east, utm_north,
                            Hansen_recentloss,access_log10, HDI, dist_to_PA,
                            PA_size_km2, rh_95_a0.pred, pavd_0_5.pred,
                            pai_a0.pred, fhd_pai_1m_a0.pred, cover_a0.pred,
                            agbd_a0.pred,asymptPD, maxFRic, SR.mean)

# Add the connectivity variables for each station calculated with
# calc_conn_metrics.R
dat_conn_metrics_bird <- data.frame(read.csv(
                                here("data/derived/public/conn_flux_bird_10_150.csv"),
                                header = T))
dat_bird <- left_join(dat_bird, dat_conn_metrics_bird, by = "station")

# Scale subset of continuous variables in dat
# Peter - Need to add the connectivity measures to the scaling list when they
# are introduced
dat_scale_bird <- data.frame(scale(subset(dat_bird, select = c("utm_east", "utm_north",
                                                    "HDI", "access_log10",
                                                    "PA_size_km2",
                                                    "dist_to_PA",
                                                    "rh_95_a0.pred",
                                                    "pavd_0_5.pred",
                                                    "pai_a0.pred",
                                                    "fhd_pai_1m_a0.pred",
                                                    "cover_a0.pred",
                                                    "agbd_a0.pred",
```

```
                                                    "awf_ptg"),
                                    ),
                                center = TRUE, scale = TRUE))

# Append scaled variables to data with .z suffixes
dat_bird[paste0(names(dat_scale_bird), '.z')] <- dat_scale_bird

# Rename scaled, predicted relative canopy height at 95% (rh_95_a0.pred.z)
# as forest_structure to match DAG
names(dat_bird)[names(dat_bird) == "rh_95_a0.pred.z"] <- "forest_structure"

# Rename scaled, predicted plant area volume density between 0m and 5m
# (pavd_0_5.pred.z) as understory_density to match DAG
names(dat_bird)[names(dat_bird) == "pavd_0_5.pred.z"] <- "understory_density"

# Exclude stations that underwent recent forest loss as defined by
# Hansen et al. (2013)
dat_clean_bird <- subset(dat_bird, Hansen_recentloss == 0)
```

3.5.4 Data preparation for mammal models

We performed the same steps for mammals.

```
# Load the data provided by Brodie et al. (2023)
dat_brodie_mam <- data.frame(read.csv(
    here("data/raw/public/training/mammal_data_230326.csv"), header = T))

# Simplify the variable names of site identifier and geographic coordinates
names(dat_brodie_mam)[names(dat_brodie_mam) == "site"] <- "station"
names(dat_brodie_mam)[names(dat_brodie_mam) == "lat_wgs84"] <- "lat"
names(dat_brodie_mam)[names(dat_brodie_mam) == "long_wgs84"] <- "long"

# Search for HDI in the column names
grep("hdi", names(dat_brodie_mam), value = TRUE)
```

```
## character(0)
```

```
grep("HDI", names(dat_brodie_mam), value = TRUE)
```

```
## character(0)
```

```
# Assign stations the HDI value of its country
# Reference:
# Human Development Report 2020: The Next Frontier-Human Development and the
# Anthropocene (United Nations Development Programme, 2020).
# website: https://hdr.undp.org/data-center/human-development-index#/indicies/HDI
# https://hdr.undp.org/sites/default/files/2021-22_HDR/HDR21-22_Statistical_Annex_HDI_Table.xlsx
dat_HDI <- data.frame(
    country = unique(dat_brodie_mam$country),
    HDI = c(0.803, 0.768, 0.800, 0.705, 0.939,  0.703))
dat_brodie_mam <- left_join(dat_brodie_mam, dat_HDI, by = "country")

# Create dataframe containing the subset of variable used in the analysis
dat_mam <- dat_brodie_mam %>% select(station, study_area, country, PA, utm_east, utm_north,
                             Hansen_recentloss,access_log10, HDI, dist_to_PA,
```

```
                            PA_size_km2, rh_95_a0.pred, pavd_0_5.pred,
                            pai_a0.pred, fhd_pai_1m_a0.pred, cover_a0.pred,
                            agbd_a0.pred,asymptPD, maxFRic, SR.mean)

# Add the connectivity variables for each station calculated with
# calc_conn_metrics.R
dat_conn_metrics <- data.frame(read.csv(
    here("data/derived/public/conn_flux.csv"),
    header = T))
dat_mam <- left_join(dat_mam, dat_conn_metrics, by = "station")

# Scale subset of continuous variables in dat
# Peter - Need to add the connectivity measures to the scaling list when they
# are introduced
dat_scale_mam <- data.frame(scale(subset(dat_mam, select = c("utm_east", "utm_north",
                                                    "HDI", "access_log10",
                                                    "PA_size_km2",
                                                    "dist_to_PA",
                                                    "rh_95_a0.pred",
                                                    "pavd_0_5.pred",
                                                    "pai_a0.pred",
                                                    "fhd_pai_1m_a0.pred",
                                                    "cover_a0.pred",
                                                    "agbd_a0.pred",
                                                    "awf_rst_ptp2"),
),
center = TRUE, scale = TRUE))

# Append scaled variables to data with .z suffixes
dat_mam[paste0(names(dat_scale_mam), '.z')] <- dat_scale_mam

# Rename scaled, predicted relative canopy height at 95% (rh_95_a0.pred.z)
# as forest_structure to match DAG
names(dat_mam)[names(dat_mam) == "rh_95_a0.pred.z"] <- "forest_structure"

# Rename scaled, predicted plant area volume density between 0m and 5m
# (pavd_0_5.pred.z) as understory_density to match DAG
names(dat_mam)[names(dat_mam) == "pavd_0_5.pred.z"] <- "understory_density"

# Exclude stations that underwent recent forest loss as defined by
# Hansen et al. (2013)
dat_clean_mam <- subset(dat_mam, Hansen_recentloss == 0)
```

3.6 Analysis

After data preparation, we built separate linear mixed effects models for birds and mammals using species richness, functional richness, and phylogenetic diversity values. We showed significance of variables using p-values and set thresholds of 0.001 (***), 0.01 (**), and 0.05 (*).

// do we add our follow-up analyses here?

3.6.1 Bird models

Then we run the PD, FR, and SR models for birds.

```r
# ----------------- Analysis of phylogenetic diversity (PD) --------------------
dat_PD_efficacy_bird <- subset(dat_clean_bird, med_dist == 150)

# Remove high-leverage outliers identified by Brodie et al.
### Brodie et al. identified outlier using the hatvalue function. We should run
### the analysis with their outlier set removed, but also run the hatvalues
### analysis to identify the outliers for our particular specification thereby
### mirroring their procedure.

PD_efficacy_outliers_bird <- c("L2422371", "L3776738", "L2521761", "L6127181",
                               "L3865754")
dat_PD_efficacy_bird <- dat_PD_efficacy_bird[! dat_PD_efficacy_bird$station %in%
                                       PD_efficacy_outliers_bird, ]

# Select variables for analysis and restrict to rows with complete values
# Peter - We need to add the eventual connectivity measures to this selection
# so they are in place for our extended analysis

dat_PD_efficacy_bird <- dat_PD_efficacy_bird %>% select(asymptPD, PA, country, utm_east,
                          utm_north, utm_east.z, utm_north.z, forest_structure,
                          access_log10.z, HDI.z, awf_ptg.z)
dat_PD_efficacy_bird <- dat_PD_efficacy_bird[complete.cases(dat_PD_efficacy_bird), ]

# Perform propensity score matching following the DAG developed in the
# structural causal modeling  and retrieve the matched dataset
match_PD_bird <- matchit(PA ~ utm_east.z + utm_north.z + forest_structure +
                         access_log10.z + HDI.z,
                     data = dat_PD_efficacy_bird, method = "full",
                     distance = "glm", link = "probit", replace = F)
dat_matched_PD_bird <- match.data(match_PD_bird)

# Run original Brodie linear mixed effects model with exponential spatial
# correlation structure for the residuals
mod_PD_efficacy_bird <- lme(asymptPD ~ forest_structure + access_log10.z
                        + HDI.z + PA, random = list(~1 | country),
                        data = dat_matched_PD_bird, weights = ~I(1/weights),
                        correlation = corExp(form = ~utm_east + utm_north,
                                             nugget = TRUE))
# summary(mod_PD_efficacy_bird)


# ----------------- Analysis of Functional Richness (FR) -----------------------
dat_FR_efficacy_bird <- subset(dat_clean_bird, med_dist == 100)

# Remove high-leverage outliers identified by Brodie et al.
### Brodie et al. identified outlier using the hatvalue function. We should run
### the analysis with their outlier set removed, but also run the hatvalues
### analysis to identify the outliers for our particular specification thereby
### mirroring their procedure.

FR_efficacy_outliers_bird <- c("L921125", "L2422371", "L4331944", "L13465594")
dat_FR_efficacy_bird <- dat_FR_efficacy_bird[! dat_FR_efficacy_bird$station %in%
                                       FR_efficacy_outliers_bird, ]
```

```r
# Select variables for analysis and restrict to rows with complete values
# Peter - We need to add the eventual connectivity measures to this selection
# so they are in place for our extended analysis

# No variable named maxFR in the csv, only maxFRic, not fully sure if they are the same
dat_FR_efficacy_bird <- dat_FR_efficacy_bird %>%
    select(maxFRic, PA, country, utm_east,
           utm_north, utm_east.z, utm_north.z, forest_structure,
           access_log10.z, HDI.z, awf_ptg.z)
dat_FR_efficacy_bird <- dat_FR_efficacy_bird[complete.cases(dat_FR_efficacy_bird), ]

# Perform propensity score matching following the DAG developed in the
# structural causal modeling  and retrieve the matched dataset
match_FR_bird <- matchit(PA ~ utm_east.z + utm_north.z + forest_structure +
                         access_log10.z + HDI.z,
                     data = dat_FR_efficacy_bird, method = "full",
                     distance = "glm", link = "probit", replace = F)
dat_matched_FR_bird <- match.data(match_FR_bird)

# Run original Brodie linear mixed effects model with exponential spatial
# correlation structure for the residuals
mod_FR_efficacy_bird <- lme(maxFRic ~ forest_structure + access_log10.z
                       + HDI.z + PA, random = list(~1 | country),
                       data = dat_matched_FR_bird, weights = ~I(1/weights),
                       correlation = corExp(form = ~utm_east + utm_north,
                                            nugget = TRUE))
# summary(mod_FR_efficacy)

# ------------------- Analysis of species richness (SR) ------------------------

dat_SR_efficacy_bird <- subset(dat_clean_bird, med_dist == 100) # Wenxin: why is the med_dist different

# Remove high-leverage outliers identified by Brodie et al.
### Brodie et al. identified outlier using the hatvalue function. We should run
### the analysis with their outlier set removed, but also run the hatvalues
### analysis to identify the outliers for our particular specification thereby
### mirroring their procedure.

SR_efficacy_outliers_bird <- c("L4789498", "L921125", "L1122096",
                           "L7010824", "L3865754", "L3776738")
dat_SR_efficacy_bird <- dat_SR_efficacy_bird[! dat_SR_efficacy_bird$station %in%
                                       SR_efficacy_outliers_bird, ]

# Select variables for analysis and restrict to rows with complete values
# Peter - We need to add the eventual connectivity measures to this selection
# so they are in place for our extended analysis

dat_SR_efficacy_bird <- dat_SR_efficacy_bird %>%
    select(SR.mean, PA, country, utm_east,
           utm_north, utm_east.z, utm_north.z, forest_structure,
           access_log10.z, HDI.z, awf_ptg.z)
dat_SR_efficacy_bird <- dat_SR_efficacy_bird[complete.cases(dat_SR_efficacy_bird), ]
```

```
# Perform propensity score matching following the DAG developed in the
# structural causal modeling  and retrieve the matched dataset
match_SR_bird <- matchit(PA ~ utm_east.z + utm_north.z + forest_structure +
                         access_log10.z + HDI.z,
                     data = dat_SR_efficacy_bird, method = "full",
                     distance = "glm", link = "probit", replace = F)
dat_matched_SR_bird <- match.data(match_SR_bird)

# Run original Brodie linear mixed effects model with exponential spatial
# correlation structure for the residuals
mod_SR_efficacy_bird <- lme(SR.mean ~ forest_structure + access_log10.z
                        + HDI.z + PA, random = list(~1 | country),
                        data = dat_matched_SR_bird, weights = ~I(1/weights),
                        correlation = corExp(form = ~utm_east + utm_north,
                                             nugget = TRUE))
# summary(mod_SR_efficacy_bird)

# Summarize the model outputs in a table
# msummary(list('Bird PD'=mod_PD_efficacy_bird, 'Bird FR'=mod_FR_efficacy_bird, 'Bird SR'=mod_SR_effica
```

3.6.2 Mammal models

```
# ------------------ Analysis of phylogenetic diversity (PD) ---------------------
dat_PD_efficacy_mam <- subset(dat_clean_mam, med_dist == 100)

# Select variables for analysis and restrict to rows with complete values
# Peter - We need to add the eventual connectivity measures to this selection
# so they are in place for our extended analysis

dat_PD_efficacy_mam <- dat_PD_efficacy_mam %>%
    select(asymptPD, PA, study_area, country, utm_east,
           utm_north, utm_east.z, utm_north.z, forest_structure,
           access_log10.z, HDI.z, awf_rst_ptp2.z)
dat_PD_efficacy_mam <- dat_PD_efficacy_mam[complete.cases(dat_PD_efficacy_mam), ]

# Perform propensity score matching following the DAG developed in the
# structural causal modeling  and retrieve the matched dataset
match_PD_mam <- matchit(PA ~ utm_east.z + utm_north.z + forest_structure +
                         access_log10.z + HDI.z,
                     data = dat_PD_efficacy_mam, method = "full",
                     distance = "glm", link = "probit", replace = F)
dat_matched_PD_mam <- match.data(match_PD_mam)

# Run original Brodie linear mixed effects model with exponential spatial
# correlation structure for the residuals
mod_PD_efficacy_mam <- lme(asymptPD ~ forest_structure + access_log10.z
                        + HDI.z + PA, random = list(~1 | country, ~1 | study_area),
                        data = dat_matched_PD_mam, weights = ~I(1/weights),
                        correlation = corExp(form = ~utm_east + utm_north,
                                             nugget = TRUE))
# summary(mod_PD_efficacy_mam)
```

```r
# ----------------- Analysis of Functional Richness (FR) ----------------------

dat_FR_efficacy_mam <- subset(dat_clean_mam, med_dist == 100)

# Select variables for analysis and restrict to rows with complete values
# Peter - We need to add the eventual connectivity measures to this selection
# so they are in place for our extended analysis

# No variable named maxFR in the csv, only maxFRic, not fully sure if they are the same
dat_FR_efficacy_mam <- dat_FR_efficacy_mam %>%
    select(maxFRic, PA, study_area, country, utm_east,
           utm_north, utm_east.z, utm_north.z, forest_structure,
           access_log10.z, HDI.z, awf_rst_ptp2.z)
dat_FR_efficacy_mam <- dat_FR_efficacy_mam[complete.cases(dat_FR_efficacy_mam), ]

# Perform propensity score matching following the DAG developed in the
# structural causal modeling  and retrieve the matched dataset
match_FR_mam <- matchit(PA ~ utm_east.z + utm_north.z + forest_structure +
                        access_log10.z + HDI.z,
                    data = dat_FR_efficacy_mam, method = "full",
                    distance = "glm", link = "probit", replace = F)
dat_matched_FR_mam <- match.data(match_FR_mam)

# Run original Brodie linear mixed effects model with exponential spatial
# correlation structure for the residuals
mod_FR_efficacy_mam <- lme(maxFRic ~ forest_structure + access_log10.z
                       + HDI.z + PA, random = list(~1 | country, ~1 | study_area),
                       data = dat_matched_FR_mam, weights = ~I(1/weights),
                       correlation = corExp(form = ~utm_east + utm_north,
                                            nugget = TRUE))
# summary(mod_FR_efficacy_mam)

# ------------------- Analysis of species richness (SR) ------------------------

dat_SR_efficacy_mam <- subset(dat_clean_mam, med_dist == 100)

# Select variables for analysis and restrict to rows with complete values
# Peter - We need to add the eventual connectivity measures to this selection
# so they are in place for our extended analysis

dat_SR_efficacy_mam <- dat_SR_efficacy_mam %>%
    select(SR.mean, PA, study_area, country, utm_east,
           utm_north, utm_east.z, utm_north.z, forest_structure,
           access_log10.z, HDI.z, awf_rst_ptp2.z)
dat_SR_efficacy_mam <- dat_SR_efficacy_mam[complete.cases(dat_SR_efficacy_mam), ]

# Perform propensity score matching following the DAG developed in the
# structural causal modeling  and retrieve the matched dataset
match_SR_mam <- matchit(PA ~ utm_east.z + utm_north.z + forest_structure +
                        access_log10.z + HDI.z,
                    data = dat_SR_efficacy_mam, method = "full",
                    distance = "glm", link = "probit", replace = F)
dat_matched_SR_mam <- match.data(match_SR_mam)
```

	Bird SR	Bird FR	Bird PD	Mammal SR	Mammal FR	Mammal PD
(Intercept)	129.164***	218.684***	2.905***	9.583***	11.547***	2.195***
	(10.092)	(13.685)	(0.230)	(1.248)	(2.037)	(0.259)
forest_structure	19.652***	33.546***	0.080*	0.333	1.410***	0.070*
	(2.129)	(3.044)	(0.031)	(0.210)	(0.389)	(0.029)
access_log10.z	10.537***	5.225*	−0.123***	−0.393	−0.174	−0.033
	(1.631)	(2.342)	(0.024)	(0.328)	(0.554)	(0.043)
HDI.z	−4.076	−15.732	−0.026	−0.445	−0.264	0.050
	(7.681)	(10.058)	(0.176)	(0.833)	(1.334)	(0.162)
PA	31.602***	25.523***	0.371***	0.322	−0.601	−0.084
	(4.889)	(6.616)	(0.075)	(0.494)	(0.845)	(0.065)
SD (Intercept country)	23.571	28.456	0.550	2.198	3.754	0.562
SD (Observations)	61.000	83.423	0.990	5.878	10.017	0.764
SD (Intercept study_area)				3.917	5.714	0.453
Num.Obs.	1099	1101	1100	1293	1297	1297
R2 Marg.	0.159	0.190	0.054			
R2 Cond.			0.277			
AIC	13 466.0	13 690.0	3854.8	8890.4	10 233.5	3616.1
BIC	13 511.0	13 735.1	3899.8	8942.0	10 285.2	3667.8
ICC			0.2			
RMSE	64.66	65.83	0.90	6.41	10.49	0.76

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

```
# Run original Brodie linear mixed effects model with exponential spatial
# correlation structure for the residuals
mod_SR_efficacy_mam <- lme(SR.mean ~ forest_structure + access_log10.z
                       + HDI.z + PA, random = list(~1 | country, ~1 | study_area),
                       data = dat_matched_SR_mam, weights = ~I(1/weights),
                       correlation = corExp(form = ~utm_east + utm_north,
                                            nugget = TRUE))
```

4 Results

```
#summary(mod_SR_efficacy_mam)
msummary(list('Bird SR'=mod_SR_efficacy_bird, 'Bird FR'=mod_FR_efficacy_bird, 'Bird PD'=mod_PD_efficacy
```

```
# Note: if rendering this part of the code returns errors for tex rendering
# Please consider going through steps on this debugging page:
# https://yihui.org/tinytex/r/#debugging

# Also: https://github.com/travis-ci/travis-ci/issues/10166
# sudo tlmgr install
# could be really helpful
```

We compared reproduction results with Extended Table 1 in Brodie et al. (2023) and found general similar results. We identified significant PA effects on bird species diversity and non significant effects on mammal species diversity as reported in the paper. PA effect sizes for SR (ORIG:27.04, RPR:31.6), FR (ORIG:24.02, RPR:25.52), PD (ORIG:0.38, RPR:0.37) from the reproduction are similar to those reported in Brodie et al. (2023). Significance of other variables and their coefficient values were also similar to the original study.

5 Discussion

The goal of the report is to reproduce analysis and results from Brodie et al. (2023) on the effect of protected areas on preserving tropical bird and mammal biodiversity after removing confounding effects of site accessibility and forest structure. The original paper provided scripts and data which allowed for reproductions. The scripts were in general reproducible. The data file contained most information but did not include raw observation data, a meta data file, or include HDI.

We found supporting evidence for ORIG-H1 and produced similar results as the original study but were unable to reproduce exact results as Brodie et al. (2023). Reasons for the differences could be HDI values, data version issues, or computation environment.

Through reconstructing the causal diagram and reproduction, we found that connectivity from sampling points to protected areas was not accounted for in the original study.

6 Integrity Statement

The authors of this preregistration state that they completed this preregistration to the best of their knowledge and that no other preregistration exists pertaining to the same hypotheses and research.

7 Acknowledgements

- `Funding Name`: name of funding for the project
- `Funding Title`: title of project grant
- `Award info URI`: web address for award information
- `Award number`: award number

This report is based upon the template for Reproducible and Replicable Research in Human-Environment and Geographical Sciences, DOI:%5B10.17605/OSF.IO/W29MQ](https://doi.org/10.17605/OSF.IO/W29MQ)

8 References