# Replication Report for Brodie et al. (2023)

Peter Kedron, Lei Song, Wenxin Yang, Amy Frazier

2024-11-08

## 1 Introduction

This study is a *replication* 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

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) preserve vertebrate biodiversity within their boundaries and in the adjacent unprotected landscape. While Brodie et al. provide statistical evidence of the efficacy of protected area status, they do not assess whether the effect they observe is altered by the connectedness of the protected area network. PA connectivity is widely recognized as a component of effective conservation, a guiding principle of the 30x30 initiative, and a possible confound of the effect of protection on biodiversity. Ecologically, the connectedness of a PA network may affect biodiversity by facilitating gene flow across locations, which may moderate the effect of protection on biodiversity.

To adjust for effect of connectivity on PA efficacy, we computationally reproduced the findings of Brodie et al. using the data shared by the authors, and extended the original analysis by introducing habitat connectivity as an independent predictor of biodiversity and a moderator of the PA effect. We find that . . .

This document presents our replication and extension of Brodie et al. (2023). All data and materials needed to fully reproduce the results of this paper are publicly available in an Open Science Framework project repository with the identifier PROJECT DOI that links to GitHub (LINK) and data and results stored on figshare (LINK) with the identifier DATA DOI. This document is the produce of a computational notebook (RPL_Report.Rmd) available in that project repository. That computational notebook contains the code used to produce the results presented in this report. For brevity and legibility, we have suppressed some longer section of code here. A computational reproduction of the authors' original analysis is also available in the project repository - *Reproduction Report for Brodie et al. (2023)*.

## 2 Study design

We adopt the statistical approach of the original analysis and test the sensitivity of the authors' findings to PA connectivity. Using propensity score matching to control for the confounds of location, site accessibility, and forest structure, Brodie et al. fit a linear mixed-effects model to measure the effect of PA status on bird and mammal biodiversity. The response variable examined in the study were three measures of biodiversity measured for both mammals and birds - Species richness (SR), Functional richness (FR), and Phylogenetic diversity (PD). The response variables were derived from 1,079 sampling locations in the eBird database and 1,365 camera stations deployed across the region.

The primary predictor of interest was binary measures indicating whether a observation site was located inside or outside a PA. The primary predictors of interest for a secondary analysis of spillover and leakage effects

in areas surrounding PAs were the size of the nearest PA and the distance to the nearest PA. This secondary analysis was conducted only for observatio sites located outside PAs. Primary predictors measuring PA status were derived from the World Database of Protected Areas. Additional predictors used to deconfound for site accessibility, forest structure, understory density, and human development pressure were derived from the NASA Global Ecosystem Dynamics Investigation (GEDI) mission, and the UN Development program Human Development Index. Site accessibility was measured using circuit theory-based metrics of proximity to human development1. Forest structure and understory density were measured use three-dimensional metrics derived from GEDI mission.

We conduct two related analyses to evaluate and extend the work of the Brodie et al. (2023).

First, we introduce PA connectivity as a predictor of bird and mammal biodiversity and a moderator of PA efficacy. Brodie et al. find evidence that the legal designation of PAs enhances Southeast Asian bird biodiversity. The authors did not find the same effect for mammals. Our primary hypotheses are:

> HA-1: The protected area status of a site alters the level of mammal and bird biodiversity observed at that site when adjusting for the confounds of site accessibility, habitat condition, socioeconomic development, **and protected area connectivity**.

Second, we examine whether the beneficial spillover effects observed around PAs persist when adjusting for PA connectivity. Brodie et al. find evidence that large PAs are associated with higher biodiversities for mammals and birds in surrounding unprotected areas, but also that the effects for birds are smaller than those for mammals. The authors found that distance to the nearest PA was significantly associated with only mammal biodiversity. Our primary hypotheses are:

> HA-2a: The total area of the protected area located closest to an unprotected observation site alters the level of mammalian of avian biodiversity observed at that site when adjusting for the confounds of site accessibility, habitat condition, and the socioeconomic development, **and protected area connectivity**.

> HA-2b: The distance to the protected area located closest to an unprotected observation site alters on the level of mammalian of avian biodiversity observed at that site when adjusting for the confounds of site accessibility, habitat condition, and the socioeconomic development, **and protected area connectivity**..

While Brodie et al. did not include measures of ecologically-focused landscape connectivity in their analyses, their work does in fact mirror our own interest in connectivity. First, the authors recognize the importance of human-focused connectivity through their site accessibility measure. Second, the analyses of spillovers performed in the study are essentially highly localized forms of connectivity analysis. Functionally, connectivity metrics are largely, if not entirely, determined by distance relationships and area sizes of land covers or land uses in a landscape. The specifications of the mixed effects models testing for spillover or leakage effects alternatively focus attention on the size an adjacent PA or the distance to an adjacent PAs. When the effect of the size of the adjacent PA is the focus, distance to that adjacent PA is controlled for. When the effect of distance to the adjacent PA is the focus, the size of that adjacent PA is controlled for. This approach reduces the ecological analysis of connectivity to the protected landscape adjacent to a site. However, there are several reasons why connections to other protected areas might effect these local effects and the efficacy of PAs - motivating our extended analysis.

# 3  Materials and Procedure

## 3.1  Replication Workflow

This replication attempt builds on our computational reproduction of Brodie et al. (2023). We use the same computational environments, data, and data preparation procedure adopted in that work. Complete details

of the materials and methods used are available in *Reproduction Report for Brodie et al. (2023)* and can be evaluated in depth by examining the functions called throughout this document. Overall our workflow (Fig. 1) uses seven functions to prepare, process, and statistically analyze our data. Additional functions used to create figures are called in our computational notebook and available in the project repository.
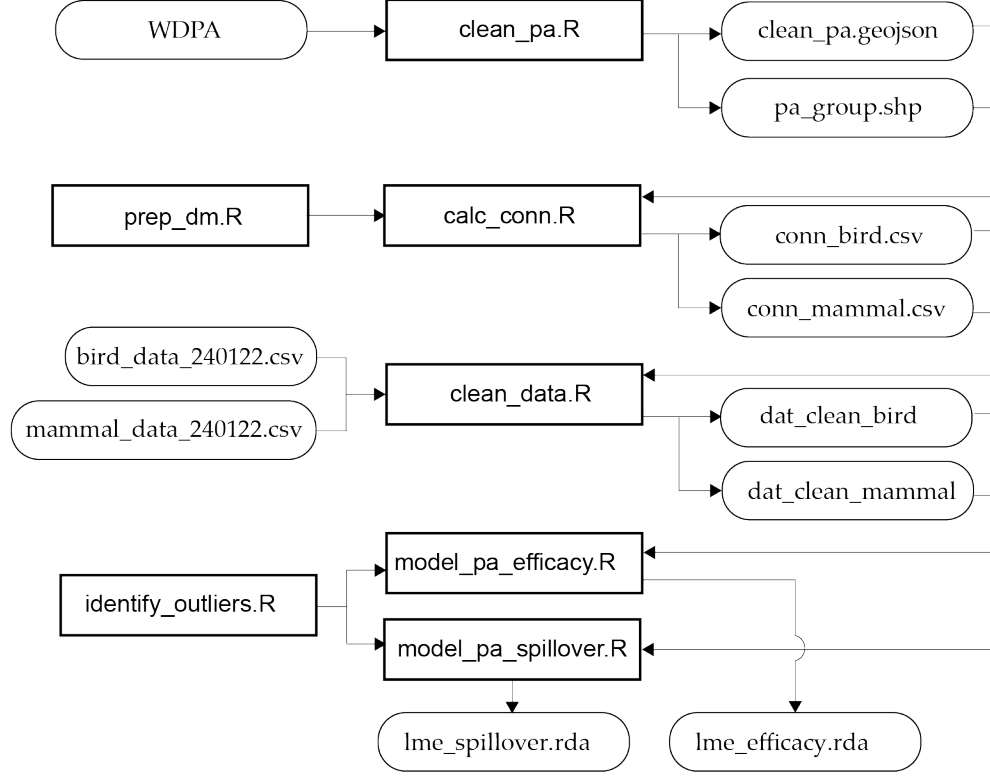


Figure 1: Replication Workflow

## 3.2 Connectivity Measures

We measure site connectivity as the area-weighted flux (AWF) of species observation sites and the surrounding landscape. In this application, $AWF$ measures flow, weighted by PA area, between all observation sites as,

$$AWF = \sum_{j}^{n} \sum_{j,i \neq j}^{n} p_{ij} \cdot a_i \cdot a_j$$

$$p_{ij} = e^{-\theta \cdot d_{ij}}$$

$$\theta = \frac{-\ln(0.5)}{d_{med}}$$

where $p_{ij}$ is the probability of dispersal between two sites $i$ and $j$, $a_i$ and $a_j$ are the areas of PAs $i$ and $j$, $d_{ij}$ is the observed distance between $i$ and $j$, and $d_{med}$ is a median dispersal distance at which $p_{ij} = 0.5$.

To construct our site connectivity measures, we independently gathered data on PA geographies from the the World Database on Protected Areas (https://www.protectedplanet.net/en/thematic-areas/wdpa?tab=WDPA). The WDPA dataset was pre-processed using several steps to ensure accuracy and relevance for

the study. First, we used the `wdpa_fetch` and `wdpa_clean` functions from the R package `wdpar` to query and clean the data. This process excluded PAs that were not yet implemented or had limited conservation value, replaced missing data codes with NA values, converted point-based PAs into circular shapes that correspond to their reported extent, fixed topological errors, and removed overlapping areas. Details about this procedures are available in the `wdpa_clean` function.

Next, we restricted our analysis by subsetting to those terrestrial PAs located within the study area. We then merged administratively distinct, but physically connected PAs and segmented PAs that were part of the same administrative unit, but not physically connected. The dataset was then clipped to the coastline to retain only the terrestrial portions of the PAs The areas for the reorganized PAs were recalculated. This created PAs that better matched real world conditions and provided more representative measures of PA size for spillover and connectivity analyses.

Finally, small residual areas ($<0.01$ km²) from the clipping process were removed. We then grouped PAs on the same contiguous landmass, under the assumption that mammals cannot cross locations disconnected by large water bodies. Only PAs located on the same landmass were considered when calculating landscape connectivity for mammals to allow the analysis to reflect realistic movement limitations imposed by geographical barriers.

Once the PAs and their corresponding clusters were prepared, we began calculating the $AWF$ for observation sites. Calculation of $AWF$ depends on the specification of a median dispersal distance. To assess the effect of varying dispersal distances, we computed $AWF$ using a range of median dispersal distances between 10km and 150km. For each specified dispersal distance, we identified the PAs around each site and within a buffer set to five times the dispersal distance. For mammals, only PAs in the same continent as the observation site were included. The Euclidean distances between the observation site and the selected PAs were then used to calculate the AWF for that site.

This procedure is implemented as:

```
source(here("procedure/code/kick_off.R"))
# Load required packages and scripts
kick_off("procedure/code")

# Calculate site connectivity for birds and mammals
calc_conn(taxon = "bird", src_dir = src_dir, dst_dir = dst_dir)
calc_conn(taxon = "mammal", src_dir = src_dir, dst_dir = dst_dir)
```

## 3.3 Preparation of the Analytical Data Files

Preparation of our analytical data file followed the procedures outlined in Brodie et al. (2023) and implemented in our computational reproduction. Because Brodie et al. did not directly analyze landscape connectivity, we appended our calculated connectivity measures to the authors' original datafile and scaled those variables following the procedures used for other predictors.

```
# Clean data for statistical analysis
conn_metrics <- "awf_ptg"
src_dir <- "data/raw/public"
conn_dir <- "data/derived/public"
dst_dir <- "data/derived/public"
dat_clean_bird <- clean_data("bird", conn_metrics, src_dir, conn_dir,
    dst_dir)
dat_clean_mammal <- clean_data("mammal", conn_metrics, src_dir,
    conn_dir, dst_dir)
```

# 4  Statistical Results

## 4.1  Implementation of the Linear Mixed Effects Model of PA Efficacy

We first use propensity score matching and a linear mixed effects model to assess whether the finding that PA preserve vertebrate biodiversity within their boundaries persisted when adjusting for connectivity (**HA-1**). Here we report connectivity results using a 100km threshold. Sensitivity analyses for alternative threshold are included in section 6.

```r
# Make a common catalog
var_catalog <- data.frame(response_variable = c("asymptPD", "maxFRic",
    "SR.mean"), name = c("PD", "FR", "SR"))

# Reproduce models for birds
dat_clean_bird <- subset(dat_clean_bird, med_dist == 100)

rpl_efficacy_bird <- lapply(1:nrow(var_catalog), function(i) {
    model_pa_efficacy(dat_clean_bird, "connec", "bird", var_catalog[[i,
        "response_variable"]], "auto")
})
names(rpl_efficacy_bird) <- sprintf("mod_bird_eff_%s", var_catalog$name)

# Reproduce models for mammals
dat_clean_mammal <- subset(dat_clean_mammal, med_dist == 50)

rpl_efficacy_mammal <- lapply(1:nrow(var_catalog), function(i) {
    model_pa_efficacy(dat_clean_mammal, "connec", "mammal", var_catalog[[i,
        "response_variable"]], "auto")
})
names(rpl_efficacy_mammal) <- sprintf("mod_mammal_eff_%s", var_catalog$name)
```

## 4.2  Implementation of the Linear Mixed Effects Models of PA Spillovers

We use propensity score matching and a linear mixed effects model to assess whether The the spillover effects of the total area of the nearest PA (**HA-2a**) or distance to the nearest PA (**HA-2b**) persist when accounting for PA connectivity to the larger PA network. It is worth noting that although the text and data file of the original article identify PA size and PA distance as continuous measures, they are in fact converted to binary indicators in the authors' code file. Accordingly, we match this procedure and use the same threshold values applied in the binary conversion by Brodie et al. For these specifications, we again report connectivity results using a 100km threshold. The authors did however include non-binary size and distance alternatively in their models - the PA size effect model included a continuous measure of distance as a control, the distance to PA model included a continuous measure of PA size as a control. To control for multicollinearity between connectivity and continuous measures of PA size and distance to the nearest PA, we removed these measures from our models.

```r
# Replicate spillover models for birds w/ connectivity
rpl_spill_bird <- lapply(1:nrow(var_catalog), function(i) {
    mods <- lapply(c("BigPA", "CloseToPA"), function(bnr_var) {
        model_pa_spillover(dat_clean_bird, "connec", "bird",
            bnr_var, var_catalog[[i, "response_variable"]], "auto")
    })
    names(mods) <- sprintf("mod_bird_%s_%s", c("size", "dist"),
```

```
        var_catalog[[i, "name"]])
    mods
})
rpl_spill_bird <- do.call(c, rpl_spill_bird)

# Replicate spillover models for mammals w/ connectivity
rpl_spill_mammal <- lapply(1:nrow(var_catalog), function(i) {
    mods <- lapply(c("BigPA", "CloseToPA"), function(bnr_var) {
        model_pa_spillover(dat_clean_mammal, "connec", "mammal",
            bnr_var, var_catalog[[i, "response_variable"]], "auto")
    })
    names(mods) <- sprintf("mod_mammal_%s_%s", c("size", "dist"),
        var_catalog[[i, "name"]])
    mods
})
rpl_spill_mammal <- do.call(c, rpl_spill_mammal)

# Concatenate the model results
mods <- list(bird = do.call(c, list(rpl_efficacy_bird, rpl_spill_bird)),
    mammal = do.call(c, list(rpl_efficacy_mammal, rpl_spill_mammal)))
```

## 4.3 Statistical Results of the Replication Attempts

Overall, the statistical results of our replication (Table 1) were generally similar to those produced in the original analyses of Brodie et al. (Table 2). Focusing on the primary predictors of interest for the all sites PA efficacy model, our analysis preserved the direction of the originally estimated PA effects for both birds and mammals.

For the all sites model testing **H1**, the PA effect estimates for mammals remained statistically indistinguishable from zero.
However, we observed shifts in the estimated magnitude of the PA effect for the bird models and a change in the statistical significance of the PA effect for bird models estimated with FR. Specifically, the estimated magnitude of the PA effect for birds dropped to 58 percent, 68 percent, 89 percent of the originally estimated value respectively for SR, FR, and PD. FR also became statistically indistinguishable from zero at the p=0.05 level adopted by the original authors. This observed shift in bird model coefficients is consistent with connectivity acting as a confound of the PA effect.

Connectivity was not a consistent predictor of bird or mammal biodiversity across the three response variables. While estimated coefficients for mammals we positive, only FR was statistically significantly different from zero. Estimated coefficients for bird models were negative and significant for SD, negative and nonsignificant for FR, and positive and significant for PD. This estimate inconsistency may be the product of several factors. First, by design biodiversity measures include a large number of species likely to have different habitat preferences and dispersal distances. Without richer data, the use of a single connectivity cannot match that variation possible leading to the misestimation of effect sizes. Second, the specifications used in the original analysis are simplifications of the biodiversity data generating process. These models likely omit important factors that may well be correlated with connectivity and many of the other measures included in the analysis.

The result of our 'PA-size' models testing **H2-a** were consistent with those of the original authors. Controlling for connectivity, our estimates for the PA size effect for birds where of the same direction, general magnitude, and significance as those of the original authors.
We also found connectivity to have a significant positive relationship with the SR and PD measures of bird biodiversity. Interestingly, connectivity had the opposite sign of the continuous distance to PA measures included by the original authors. The results of our mammal models were similarly consistent. While

Table 1: Results from mixed-effects linear regression for species richness (SR), functional richness (FR), and phylogenetic diversity (PD). Following Brodie et al., bolding indicates predictors with estimated coefficients with p<0.05

| | Bird | | | Mammal | | |
|---|---|---|---|---|---|---|
| Variable | SR | FR | PD | SR | FR | PD |
| **All sites** | | | | | | |
| $R^2$ | 0.220 | 0.241 | 0.185 | 0.365 | 0.314 | 0.430 |
| (Intercept) | 146.605 | 225.835 | 2.992 | 10.157 | 12.172 | 2.198 |
| | (16.188; 0.000) | (12.942; 0.000) | (0.184; 0.000) | (1.380; 0.000) | (1.890; 0.000) | (0.232; 0.000) |
| Forest canopy height | 19.284 | 32.634 | 0.118 | 0.001 | 1.059 | 0.053 |
| | (3.074; 0.000) | (3.035; 0.000) | (0.036; 0.001) | (0.209; 0.995) | (0.365; 0.004) | (0.028; 0.060) |
| Site accessibility | 7.653 | 0.905 | -0.129 | -0.275 | -0.140 | 0.004 |
| | (2.648; 0.004) | (2.566; 0.724) | (0.031; 0.000) | (0.332; 0.407) | (0.547; 0.799) | (0.043; 0.929) |
| HDI | -9.355 | -17.972 | 0.158 | -0.046 | 1.434 | 0.162 |
| | (12.531; 0.480) | (10.160; 0.120) | (0.145; 0.312) | (0.894; 0.962) | (1.244; 0.313) | (0.150; 0.340) |
| Connectivity | **-8.292** | -4.573 | **0.201** | 2.299 | **4.141** | 0.266 |
| | **(3.646; 0.023)** | (3.665; 0.212) | **(0.055; 0.000)** | (1.279; 0.072) | **(2.043; 0.043)** | (0.160; 0.096) |
| PA | **14.806** | 10.988 | **0.339** | 0.383 | 0.631 | -0.003 |
| | **(6.605; 0.025)** | (6.643; 0.098) | **(0.083; 0.000)** | (0.473; 0.419) | (0.787; 0.423) | (0.061; 0.964) |
| **Outside protected areas - 'PA size' effect** | | | | | | |
| $R^2$ | 0.247 | 0.376 | 0.197 | 0.588 | 0.356 | 0.608 |
| (Intercept) | 139.738 | 234.163 | 3.191 | 6.622 | 12.317 | 1.981 |
| | (18.113; 0.000) | (29.797; 0.000) | (0.225; 0.000) | (2.537; 0.009) | (3.176; 0.000) | (0.331; 0.000) |
| Forest canopy height | 17.817 | 16.100 | -0.073 | 0.385 | 0.199 | 0.029 |
| | (3.118; 0.000) | (3.468; 0.000) | (0.047; 0.118) | (0.284; 0.175) | (0.580; 0.731) | (0.039; 0.466) |
| Site accessibility | -4.363 | -13.908 | -0.231 | -1.157 | -1.349 | 0.010 |
| | (3.717; 0.241) | (4.059; 0.001) | (0.053; 0.000) | (0.503; 0.022) | (0.737; 0.067) | (0.053; 0.852) |
| HDI | 9.644 | 9.794 | 0.323 | 0.202 | 1.521 | 0.061 |
| | (14.641; 0.535) | (24.043; 0.698) | (0.182; 0.127) | (1.550; 0.903) | (1.848; 0.457) | (0.200; 0.776) |
| PA size (binary) | 2.469 | **18.144** | **0.392** | **4.704** | -0.150 | -0.031 |
| | (6.367; 0.698) | **(7.698; 0.019)** | **(0.098; 0.000)** | **(1.015; 0.000)** | (1.720; 0.931) | (0.121; 0.801) |
| Connectivity | **16.402** | 2.905 | **0.272** | -1.100 | 1.812 | -0.207 |
| | **(4.190; 0.000)** | (5.010; 0.562) | **(0.062; 0.000)** | (2.444; 0.653) | (3.533; 0.608) | (0.280; 0.460) |
| **Outside protected areas - 'Distance to PA' effect** | | | | | | |
| $R^2$ | 0.297 | 0.247 | 0.238 | 0.293 | 0.260 | 0.272 |
| (Intercept) | 112.919 | 187.726 | 2.898 | 12.256 | 15.199 | 2.330 |
| | (15.767; 0.000) | (14.419; 0.000) | (0.212; 0.000) | (1.680; 0.000) | (1.933; 0.000) | (0.200; 0.000) |
| Forest canopy height | 22.068 | 40.489 | 0.144 | 0.484 | 2.670 | 0.188 |
| | (3.008; 0.000) | (3.164; 0.000) | (0.041; 0.000) | (0.487; 0.320) | (0.600; 0.000) | (0.044; 0.000) |
| Site accessibility | 11.218 | -4.559 | -0.183 | -2.351 | -1.795 | -0.143 |
| | (3.070; 0.000) | (3.197; 0.154) | (0.041; 0.000) | (0.581; 0.000) | (0.736; 0.015) | (0.058; 0.015) |
| HDI | -3.223 | -11.637 | 0.225 | 2.626 | 3.467 | 0.252 |
| | (12.531; 0.806) | (11.619; 0.355) | (0.170; 0.235) | (1.168; 0.088) | (1.419; 0.071) | (0.148; 0.163) |
| Distance to PA (binary) | 13.859 | **22.754** | 0.134 | 1.695 | **3.568** | **0.247** |
| | (7.670; 0.071) | **(7.112; 0.001)** | (0.088; 0.130) | (1.068; 0.113) | **(1.324; 0.007)** | **(0.102; 0.015)** |
| Connectivity | 1.933 | 1.879 | 0.124 | 4.479 | **9.016** | 0.473 |
| | (4.915; 0.694) | (6.829; 0.783) | (0.088; 0.163) | (2.666; 0.093) | **(3.200; 0.005)** | (0.250; 0.059) |

the direction of our coefficient estimates changes for mammal SR and FR, both also remained statistically indistinguishable from zero. The statistical relationship between connectivity and mammal biodiversity was also not distinguishable from zero.

The results of our 'Distance to PA' models testing **H2-b** were the least consistent with those of the original authors. While we estimated coefficients with consistent directions, we observed shifts in statistical significance for the SR and FR bird models and for the FR and PD mammal models. Location close to a PA became a significant positive predictor of mammal FR and PD in our models. Connectivity itself demonstrated a consistently positive relationship across all models and for birds and mammals, but was only statistically significant for the mammal FR model.

Table 2: Comparison of PA efficacy and spillover

| Change | Bird | | | Mammal | | |
|---|---|---|---|---|---|---|
| | SR | FR | PD | SR | FR | PD |
| **All sites** | | | | | | |
| Direction | ≡ | ≡ | ≡ | - → + | - → + | ≡ |
| Magnitude | ↓ | ↓ | ↓ | ↑ | ↑ | ↓ |
| Significance | ≡ s | s → ns | ≡ s | ≡ ns | ≡ ns | ≡ ns |
| **Outside protected areas - 'PA size' effect** | | | | | | |
| Direction | ≡ | ≡ | ≡ | ≡ | + → - | + → - |
| Magnitude | ↓ | ↓ | ↑ | ↑ | ↓ | ↓ |
| Significance | ≡ ns | ≡ s | ≡ s | ≡ s | s → ns | s → ns |
| **Outside protected areas - 'Distance to PA' effect** | | | | | | |
| Direction | ≡ | ≡ | - → + | ≡ | ≡ | ≡ |
| Magnitude | ↑ | ↑ | ↑ | ↑ | ↑ | ↑ |
| Significance | s → ns | ns → s | ≡ ns | ≡ ns | ns → s | ns → s |

*Note:*
s is for statistically significant, and ns is for not statistically significant.
≡ indicates the same, ↓ is for decrease, and ↑ is for increase.

# 5  Discussion

- PA size statistics. What is the difference in the mean PA sizes between brodie and our combined PAs? Std? Could this link with the PA size effects?

# 6  Sensitivity Analyses

Compile models runs all the functions above and produces x...

```r
for (dist in seq(10, 150, 10)) {
    try({
        message(paste0(dist, " - bird"))
        compile_models(taxon = "bird", outliers = "auto", med_dist = dist,
            src_dir = "data/derived/public", dst_dir = "results")
        message(paste0(dist, " - mammal"))
        compile_models(taxon = "mammal", outliers = "auto", med_dist = dist,
            src_dir = "data/derived/public", dst_dir = "results")
    })
}

# Generate plots
plot_dist_effect(src_dir = "results", dst_dir = "results/figures")
```

# Bird – Phylogenetic diveristy (PD)

## PA efficacy



## PA size effect



## Distance to PA effect



Dispersal distance (km)
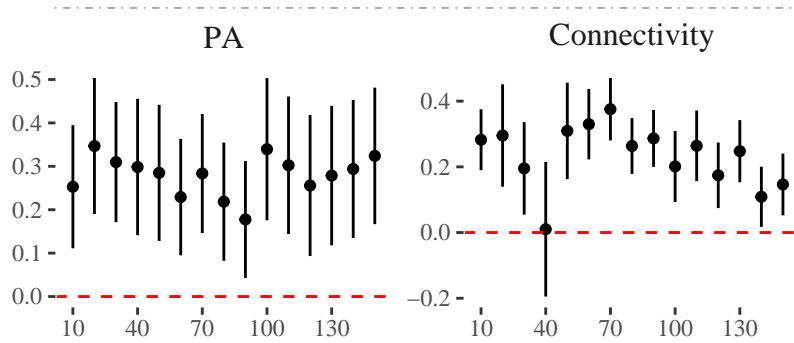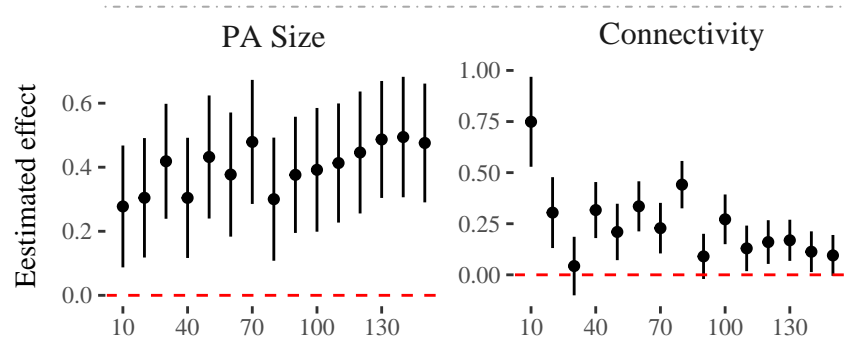
Eestimated effect
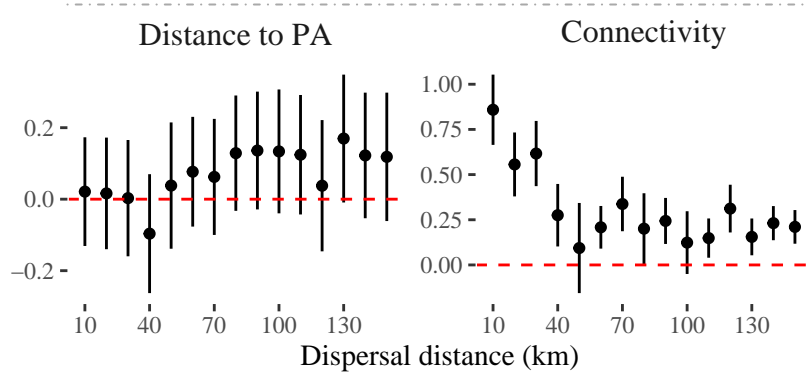
# Bird – Functional richness (FR)

## PA efficacy



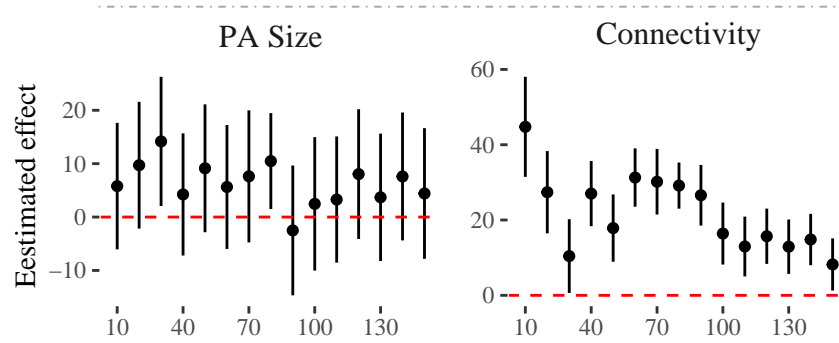## PA size effect



## Distance to PA effect



Dispersal distance (km)
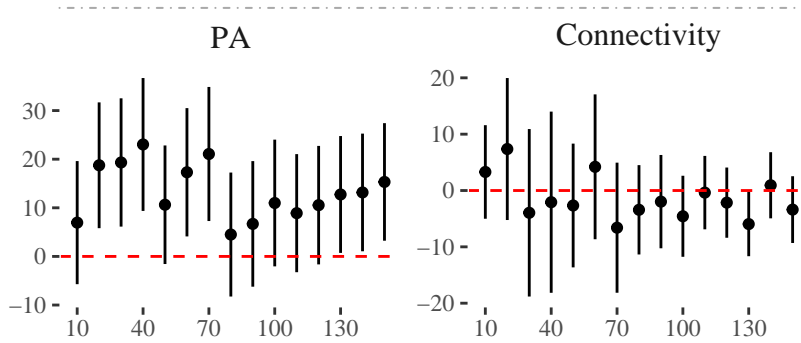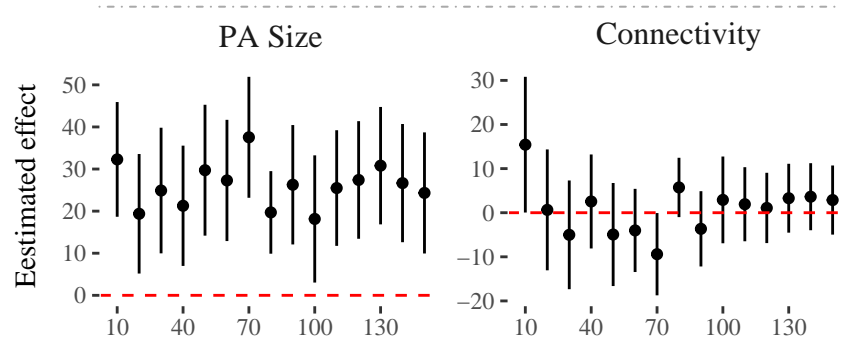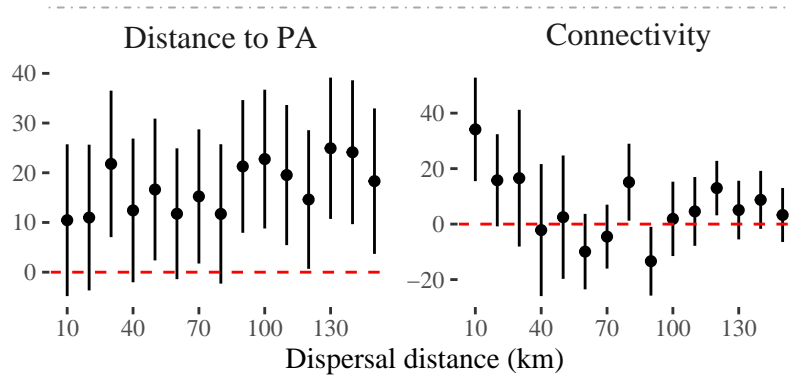
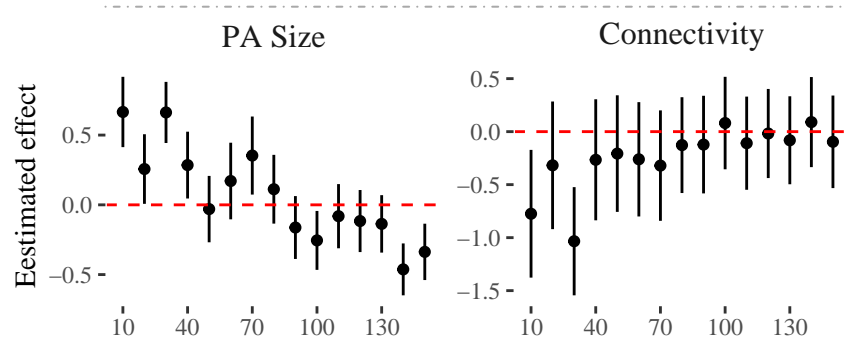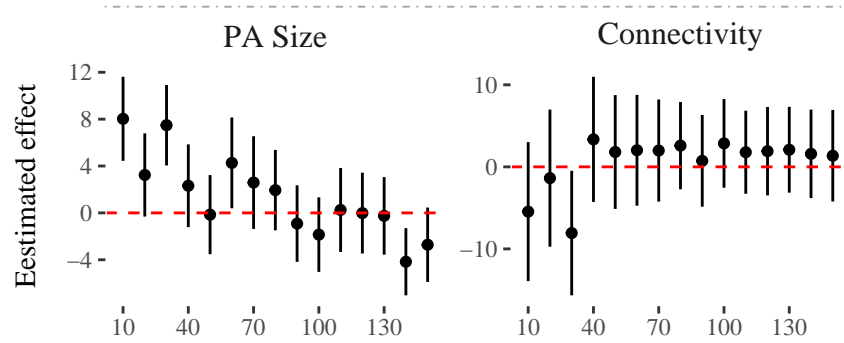# Bird – Species richness (SR)

## PA efficacy



## PA size effect



## Distance to PA effect



Dispersal distance (km)
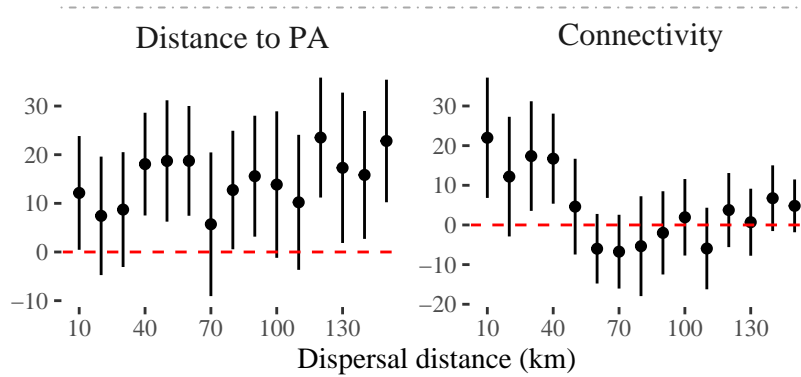
# Mammal – Phylogenetic diveristy (PD)
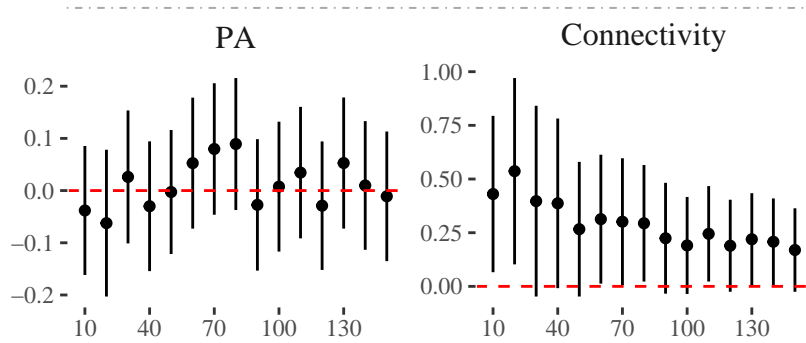
## PA efficacy



## PA size effect



## Distance to PA effect
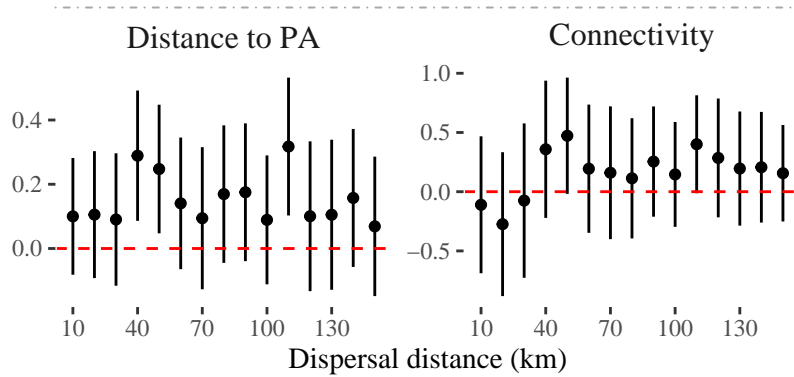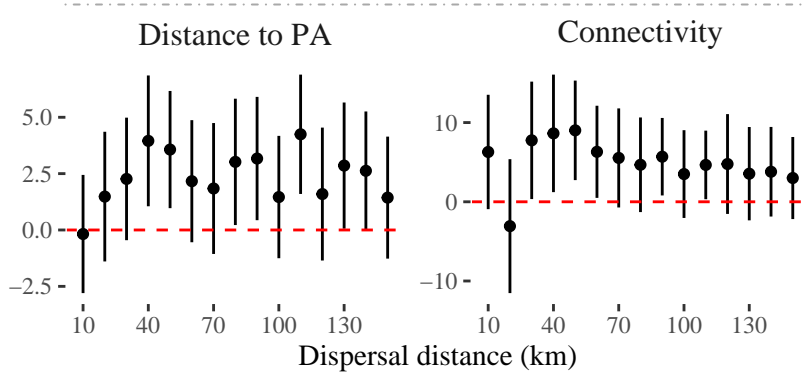
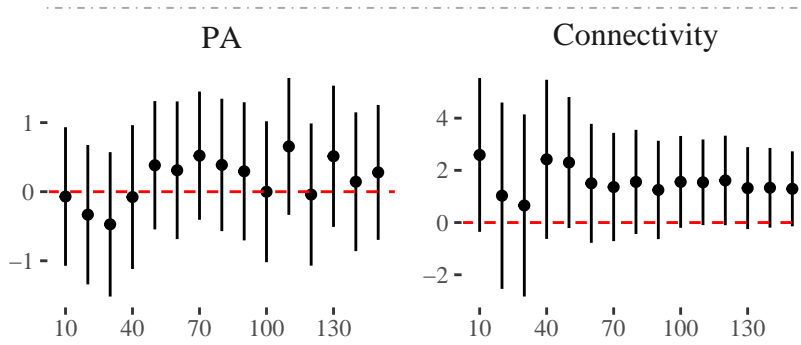# Mammal – Functional richness (FR)

**PA efficacy**

PA

Connectivity

**PA size effect**

PA Size

Connectivity

Eestimated effect

**Distance to PA effect**

Distance to PA

Connectivity

Dispersal distance (km)

# Mammal – Species richness (SR)

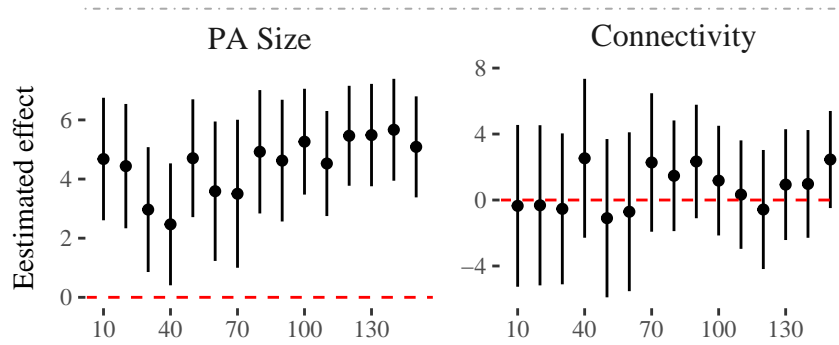## PA efficacy



## PA size effect



## Distance to PA effect