Replication Report for Brodie et al. (2023)

Peter Kedron, Lei Song, Wenxin Yang, Amy Frazier

2024-09-05

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 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). Our computational reproduction of the authors' original analysis is presented in *Reproduction Report for Brodie et al. (2023)*. All data and materials needed to fully reproduce the results of this paper are publicly available at GitHub (LINK) with the identifier PROJECT DOI and Figshare (LINK) with the identifier DATA DOI.

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 a binary measure indicating whether a observation site was located inside or outside a PA. Predictor variables measuring PA status, site accessibility, forest structure, understory density, and human development pressure were derived from the World Database of Protected Areas, the NASA Global Ecosystem Dynamics Investigation (GEDI) mission, and the UN Development program Human Development Index.

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-1a: 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**.

> HA-1b: How connected a site is to surrounding protected areas moderates the effect protected area status has on the level of mammal and bird biodiversity observed at that site when adjusting for the confounds of site accessibility, habitat condition, socioeconomic development.

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**..

3 Materials and Procedure

3.1 Computational Reproduction

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 use are available in *Reproduction Report for Brodie et al. (2023)* and can be evaluated in depth by examining the functions called throughout this document. All materials and procedures are publicly available at GitHub **(LINK)** with the identifier **PROJECT DOI**.

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

3.2 Connectivity Measures

We measured site connectivity as the area-weighted flux (AWF) of species observation sites and the surrounding landscape. Area-weighted flux measures flow, weighted by protected area, between all sites as

$$AWF = p_{ij} = e^{-k*d_{ij}}$$

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

To construct our site and PA 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 PA 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 involved excluding PAs that were not yet implemented or had limited conservation value, replacing missing data codes with NA values, converting point-based PAs into circular shapes that correspond to their reported extent, fixing topological errors, and removing overlapping areas (More details can be found in `wdpa_clean` function). Next, we selected only terrestrial PAs within the study area and merged physically connected PAs while separating those labeled as the same PA but not physically connected. The dataset was then clipped to the coastline to retain only terrestrial portions, and the areas for the reorganized PAs were recalculated. Finally, tiny residual areas ($<$0.01 km^2) from the clipping process were removed. For clustering, PAs on the same continent were grouped, under the assumption that mammals cannot cross disconnected landmasses, and only PAs within the same cluster were considered when calculating landscape connectivity. Under the assumption that mammals cannot cross disconnected landmasses, PAs located on the same continent were grouped into clusters. For the purposes of calculating landscape connectivity for mammals, only the PAs within the same cluster (i.e., on the same continent) were considered, ensuring that the analysis reflected realistic movement limitations imposed by geographical barriers.

Once the PAs and their corresponding clusters were prepared, we began calculating the AWF for observation sites. As indicated by the equation, AWF calculation depends on the median dispersal distance. To assess the effect of varying dispersal distances, we computed AWF using a range of median dispersal distances: 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150. For each specified dispersal distance, we identified the PAs within a buffer around each site, with the buffer size set to five times the given 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.

```r
# 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 that analysis did not analyze connectivity, we appended our calculated connectivity measures to that datafile and scaled those variables following the procedures of the original authors. Finally, we identified and removed outlier values from each dataset using **PROCEDURE INFO** and a list provided by Brodie et al.

```r
# Clean data and remove outliers for bird and mammal models
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 STATEMENT OF FIND Efficacy

We first used propensity score matching and a linear mixed effects model to assessed whether the finding that PA preserve vertebrate biodiversity within their boundaries persisted when adjusting for connectivity. Here we report connectivity results using a 100km threshold.

```
# 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 Linear Mixed Effects Model of PA Spillovers

```
# Replicate spillover models for birds w/ connectivity
## Lei, we need to change the outlier procedure in this function
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
## Lei, we need to change the outlier procedure in this function
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
```

```r
}); rpl_spill_mammal <- do.call(c, rpl_spill_mammal)

# Concatenate them together
mods <- list("bird" = do.call(c, list(rpl_efficacy_bird, rpl_spill_bird)),
             "mammal" = do.call(c, list(rpl_efficacy_mammal, rpl_spill_mammal)))
```

Table 1: Results from mixed-effects linear regression for species richness (SR), functional richness (FR), and phylogenetic diversity (PD).

	Bird			Mammal		
Variable	SR	FR	PD	SR	FR	PD
All sites						
R^2	0.341	0.294	0.308	0.355	0.303	0.454
(Intercept)	139.750 (19.699; 0.000)	218.642 (13.758; 0.000)	2.936 (0.206; 0.000)	10.535 (1.493; 0.000)	13.317 (2.002; 0.000)	2.182 (0.247; 0.000)
Forest canopy height	28.893 (3.106; 0.000)	36.996 (3.022; 0.000)	0.168 (0.035; 0.000)	0.133 (0.229; 0.561)	1.035 (0.362; 0.004)	0.030 (0.028; 0.293)
Site accessibility	5.500 (2.648; 0.038)	9.688 (2.554; 0.000)	-0.054 (0.030; 0.072)	-0.344 (0.370; 0.353)	-0.465 (0.587; 0.429)	0.001 (0.046; 0.980)
HDI	-7.330 (15.128; 0.643)	-14.201 (10.521; 0.219)	0.173 (0.158; 0.310)	0.016 (0.954; 0.988)	0.939 (1.228; 0.487)	0.099 (0.156; 0.561)
PA	7.380 (6.674; 0.269)	10.789 (6.514; 0.098)	**0.254 (0.075; 0.001)**	-0.614 (0.962; 0.523)	-1.157 (1.532; 0.450)	-0.019 (0.119; 0.873)
Connectivity	-5.738 (3.324; 0.085)	**-8.226 (3.235; 0.011)**	**0.215 (0.038; 0.000)**	2.524 (1.565; 0.107)	4.216 (2.508; 0.093)	0.208 (0.207; 0.315)
PA|Connectivity	-0.786 (6.781; 0.908)	1.020 (6.891; 0.882)	0.109 (0.076; 0.154)	-1.301 (1.436; 0.365)	-1.351 (2.270; 0.552)	-0.079 (0.176; 0.653)
Outside protected areas - 'PA size' effect						
R^2	0.224	0.334	0.336	0.420	0.347	0.585
(Intercept)	137.886 (13.341; 0.000)	215.240 (30.996; 0.000)	2.963 (0.317; 0.000)	8.431 (2.097; 0.000)	8.095 (2.685; 0.003)	1.752 (0.326; 0.000)
Forest canopy height	21.427 (2.886; 0.000)	22.076 (3.057; 0.000)	-0.021 (0.040; 0.600)	0.464 (0.420; 0.270)	0.686 (0.787; 0.384)	0.010 (0.050; 0.846)
Site accessibility	-7.627 (3.597; 0.034)	-9.698 (4.249; 0.023)	-0.210 (0.051; 0.000)	-0.860 (0.443; 0.053)	-2.834 (0.732; 0.000)	-0.092 (0.051; 0.075)
HDI	12.163 (10.810; 0.304)	-3.258 (25.004; 0.901)	0.236 (0.255; 0.392)	0.504 (1.225; 0.702)	3.639 (1.488; 0.071)	0.034 (0.196; 0.871)
Distance to PA	-9.368 (3.483; 0.007)	-15.675 (4.455; 0.000)	-0.207 (0.049; 0.000)	-0.520 (0.458; 0.256)	-2.000 (0.786; 0.011)	-0.013 (0.054; 0.807)
PA size (binary)	6.439 (7.432; 0.387)	8.993 (9.779; 0.358)	**0.459 (0.108; 0.000)**	2.362 (1.800; 0.190)	**6.268 (2.792; 0.025)**	0.421 (0.218; 0.054)
Connectivity	**14.136 (4.319; 0.001)**	1.357 (5.782; 0.814)	**0.229 (0.061; 0.000)**	0.426 (2.671; 0.873)	3.870 (3.880; 0.319)	-0.364 (0.331; 0.271)
PA size|Connectivity	0.217 (6.884; 0.975)	14.093 (9.520; 0.139)	-0.037 (0.101; 0.716)	-0.699 (3.815; 0.855)	-5.110 (5.807; 0.379)	0.219 (0.463; 0.636)
Outside protected areas - 'Distance to PA' effect						
R^2	0.403	0.425	0.411	0.347	0.395	0.348
(Intercept)	108.899 (15.852; 0.000)	191.708 (20.064; 0.000)	2.933 (0.198; 0.000)	10.066 (2.229; 0.000)	10.527 (4.017; 0.009)	1.830 (0.322; 0.000)
Forest canopy height	23.308 (2.338; 0.000)	32.584 (3.033; 0.000)	0.180 (0.031; 0.000)	0.438 (0.342; 0.201)	0.518 (0.497; 0.297)	-0.071 (0.041; 0.088)
Site accessibility	3.070 (2.370; 0.196)	-18.477 (3.194; 0.000)	-0.264 (0.032; 0.000)	-0.405 (0.535; 0.449)	1.316 (0.881; 0.136)	0.071 (0.072; 0.324)
HDI	-0.152 (12.629; 0.991)	-17.271 (16.678; 0.340)	0.218 (0.158; 0.217)	0.038 (1.201; 0.976)	1.477 (2.269; 0.551)	0.093 (0.184; 0.640)
PA size	16.453 (4.350; 0.000)	34.652 (7.179; 0.000)	0.376 (0.058; 0.000)	0.328 (0.821; 0.689)	-0.248 (1.331; 0.852)	0.075 (0.106; 0.475)
Distance to PA (binary)	**18.124 (6.092; 0.003)**	**17.983 (7.639; 0.019)**	-0.007 (0.082; 0.933)	**3.240 (1.506; 0.032)**	0.878 (2.312; 0.704)	0.128 (0.193; 0.506)
Connectivity	-1.594 (4.852; 0.743)	-2.719 (7.985; 0.734)	**0.237 (0.065; 0.000)**	2.264 (3.080; 0.463)	5.816 (5.042; 0.249)	0.025 (0.392; 0.950)
Distance to PA|Connectivity	12.376 (9.584; 0.197)	-13.242 (10.699; 0.216)	0.013 (0.129; 0.916)	2.968 (2.724; 0.276)	-0.618 (4.227; 0.884)	0.027 (0.344; 0.937)

Table 2: Comparison of PA effacicy

	Bird			Mammal		
Change	SR	RF	FD	SR	RF	FD
Direction	$\equiv$	$\equiv$	$\equiv$	$\equiv$	$\equiv$	$\equiv$
Magnitude	$\downarrow$	$\downarrow$	$\downarrow$	$\uparrow$	$\uparrow$	$\downarrow$
Significance	s $\rightarrow$ ns	s $\rightarrow$ ns	$\equiv$ s	$\equiv$ ns	$\equiv$ ns	$\equiv$ ns

Note:
s is for statistically significant, and ns is for not statistically significant. $\equiv$ indicates the same, $\downarrow$ is for decrease, and $\uparrow$ is for increase.

Table 3: Comparison of PA spillover

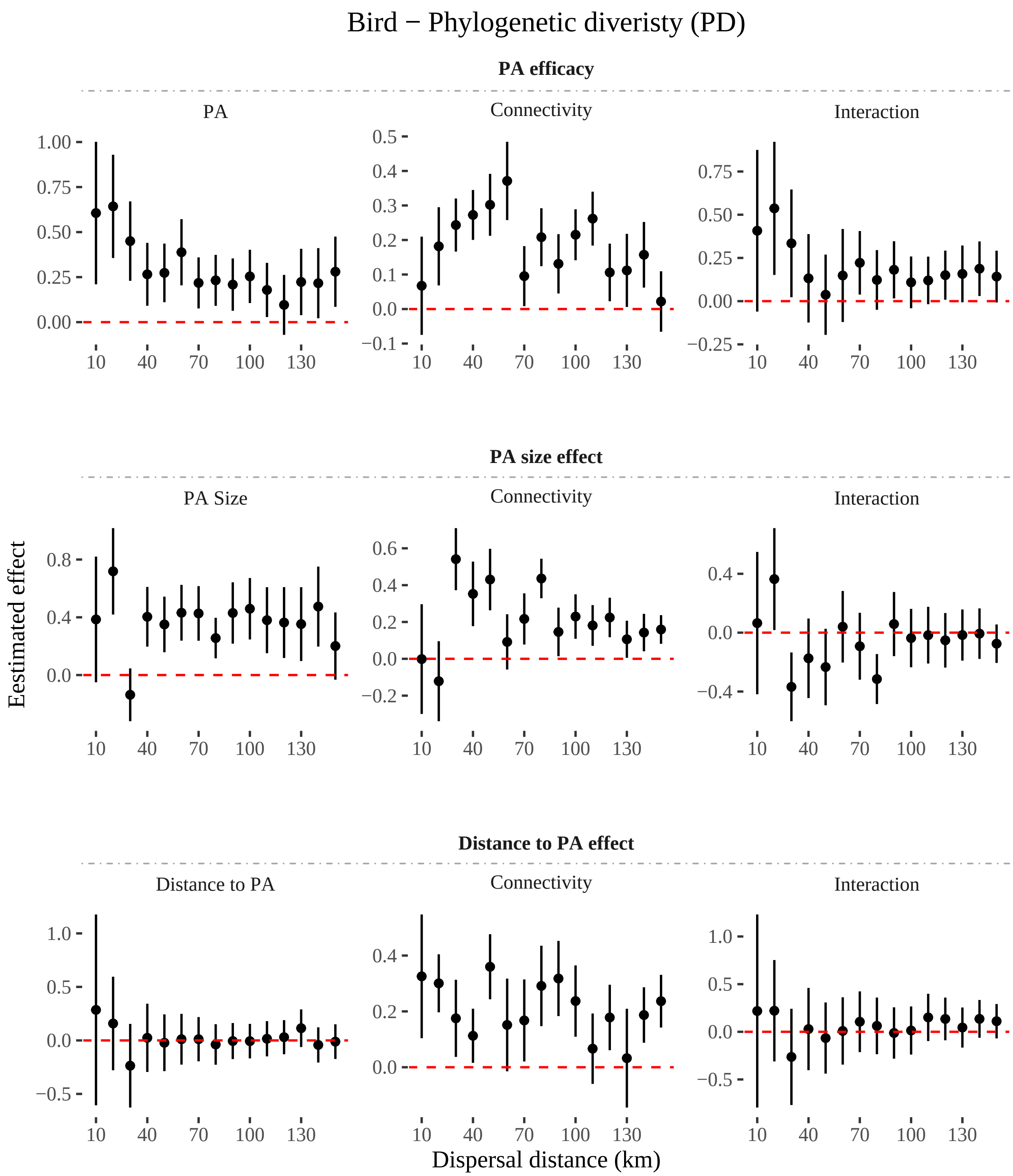
	Bird			Mammal		
Change	SR	RF	FD	SR	RF	FD
Outside protected areas - 'PA size' effect						
Direction	$\equiv$	$\equiv$	$\equiv$	$\equiv$	$\equiv$	$\equiv$
Magnitude	$\downarrow$	$\downarrow$	$\uparrow$	$\uparrow$	$\downarrow$	$\downarrow$
Significance	$\equiv$ ns	s $\rightarrow$ ns	$\equiv$ s	s $\rightarrow$ ns	$\equiv$ s	s $\rightarrow$ ns
Outside protected areas - 'Distance to PA' effect						
Direction	$\equiv$	$\equiv$	$\equiv$	$\equiv$	$\equiv$	$\equiv$
Magnitude	$\uparrow$	$\uparrow$	$\downarrow$	$\uparrow$	$\downarrow$	$\downarrow$
Significance	$\equiv$ s	ns $\rightarrow$ s	$\equiv$ ns	ns $\rightarrow$ s	$\equiv$ ns	$\equiv$ ns

Note:
s is for statistically significant, and ns is for not statistically significant. $\equiv$ indicates the same, $\downarrow$ is for decrease, and $\uparrow$ is for increase.

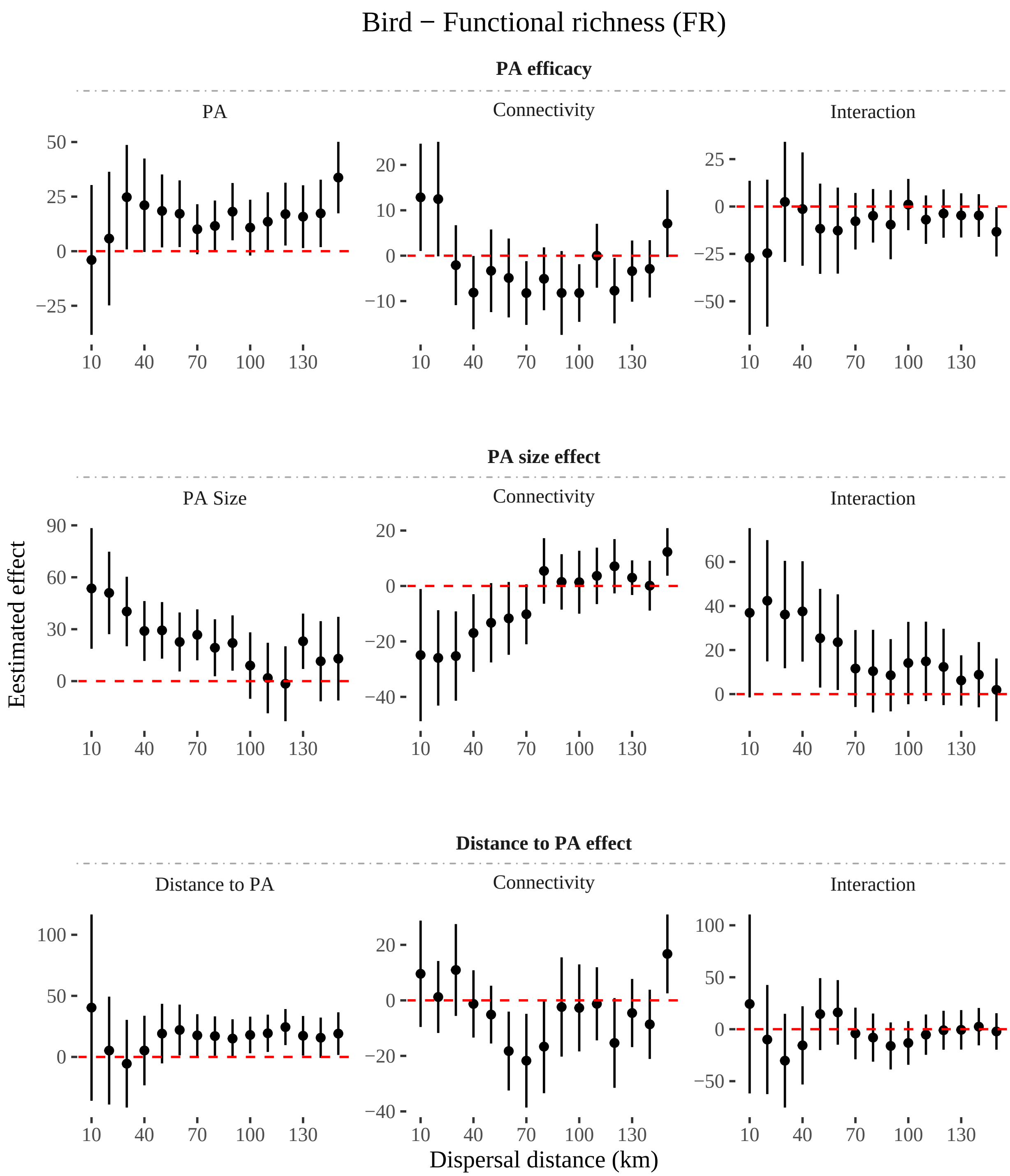
5 Sensitivity Analyses

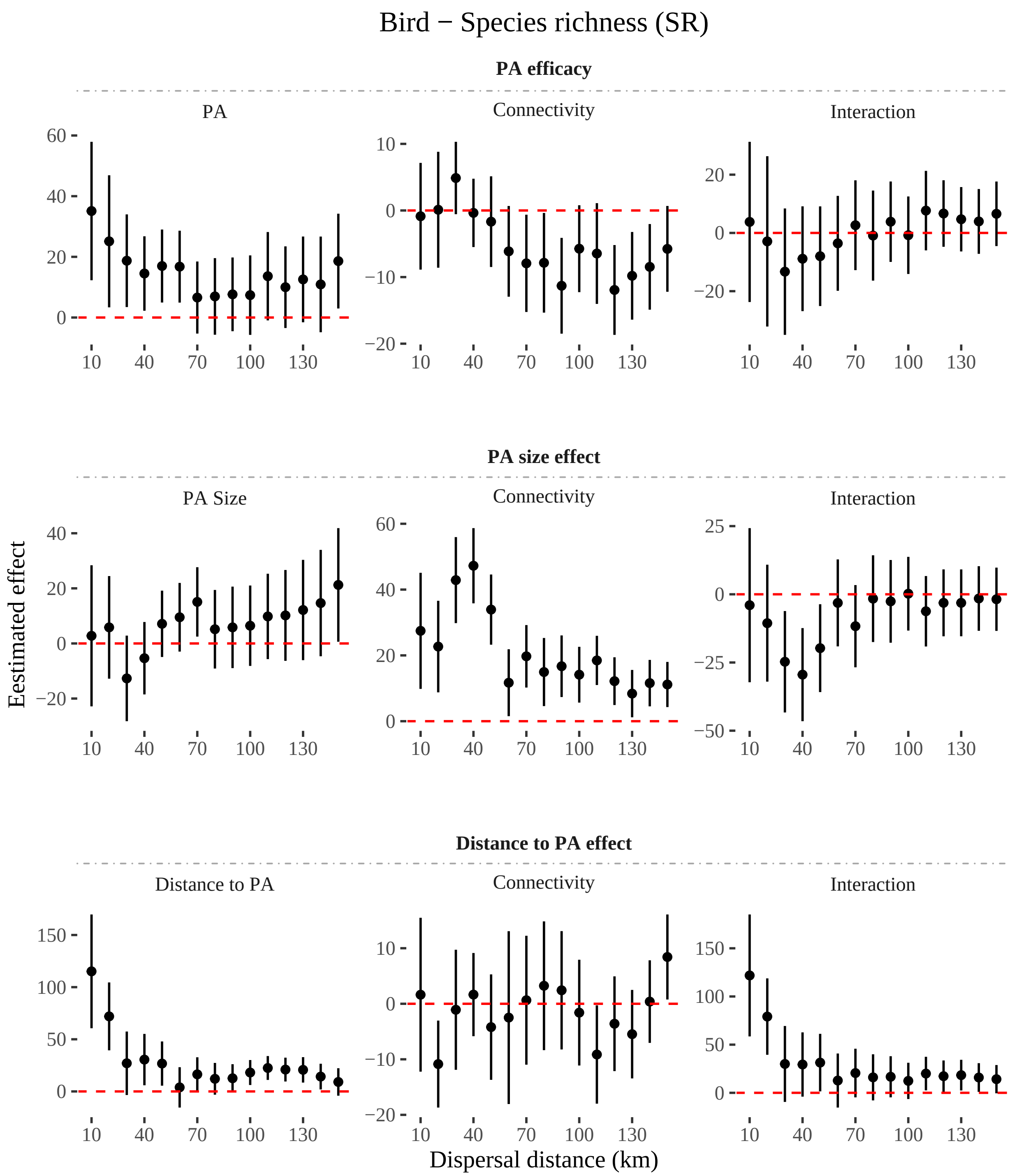
Compile models runs all the functions above and produces x…

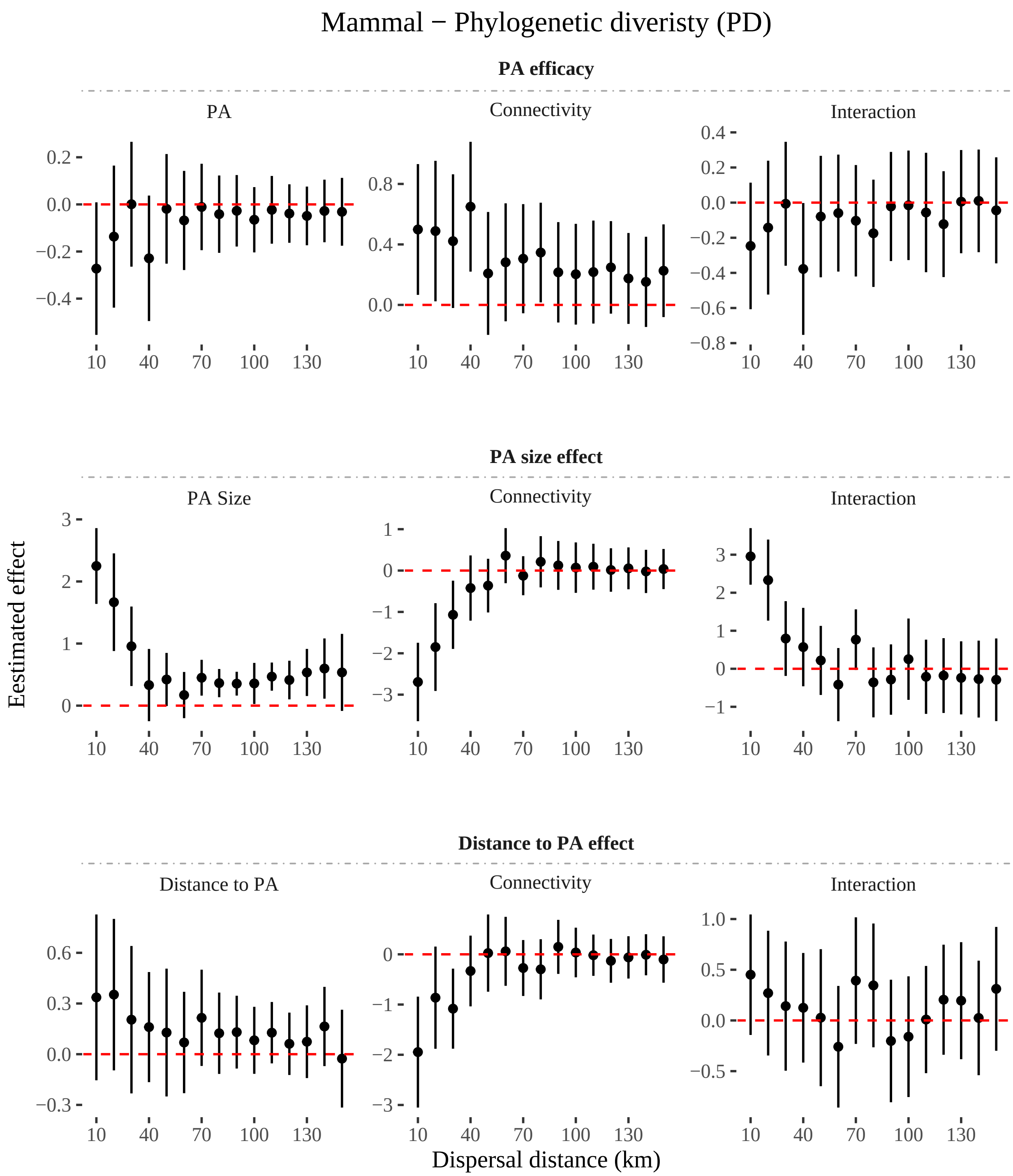
```
for (dist in seq(10, 150, 10)){
    compile_models(taxon = "bird", med_dist = dist,
                   src_dir = "data/derived/public",
                   dst_dir = "results")
    compile_models(taxon = "mammal", med_dist = dist,
                   src_dir = "data/derived/public",
                   dst_dir = "results")
}

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



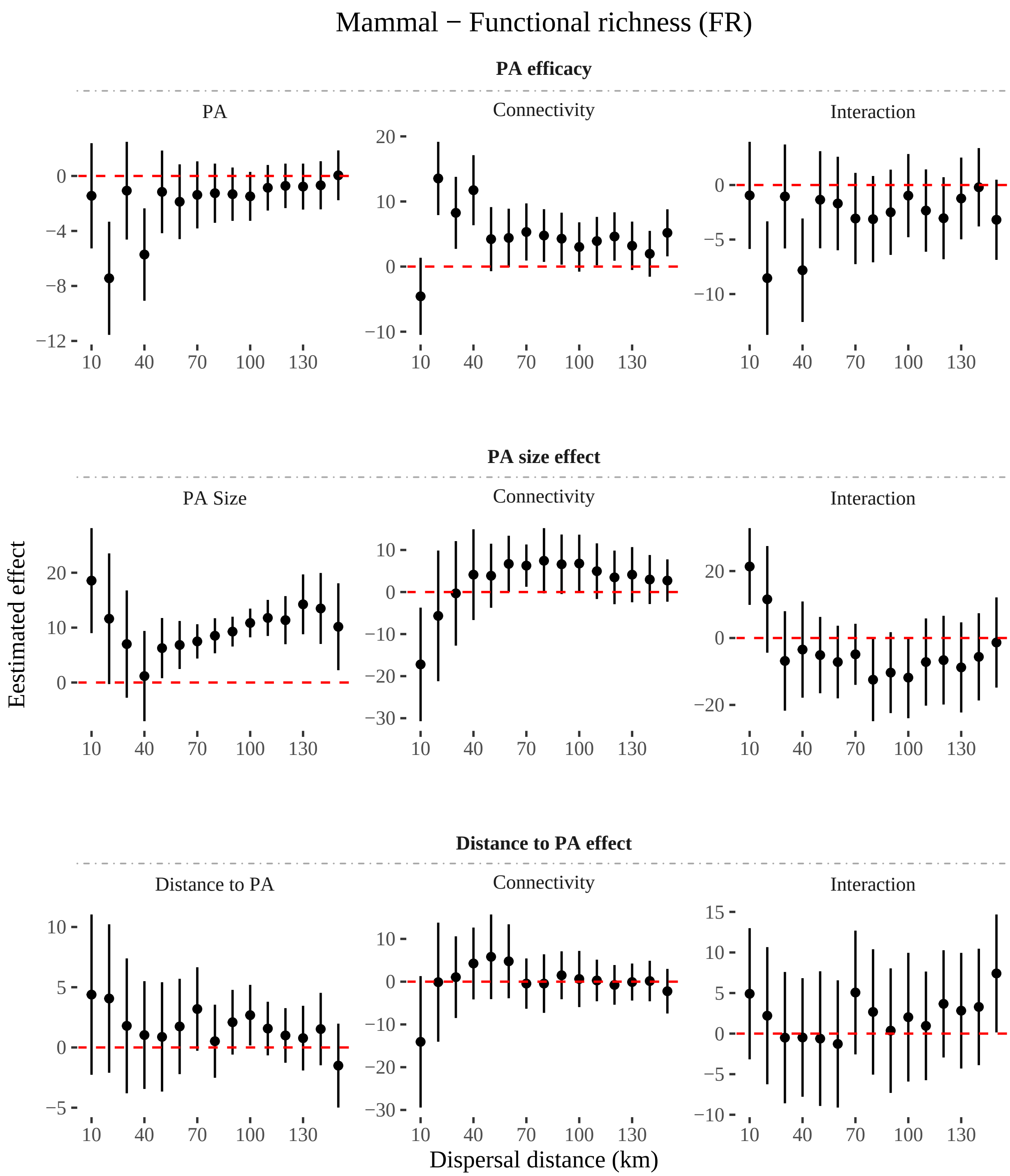
Bird − Phylogenetic diveristy (PD)
PA efficacy
PA
Connectivity
Interaction
PA size effect
PA Size
Connectivity
Interaction
Distance to PA effect
Distance to PA
Connectivity
Interaction
Eestimated effect
Dispersal distance (km)


Bird − Functional richness (FR)
PA efficacy
PA
Connectivity
Interaction
PA size effect
PA Size
Connectivity
Interaction
Distance to PA effect
Distance to PA
Connectivity
Interaction
Eestimated effect
Dispersal distance (km)
10
40
70
100
130

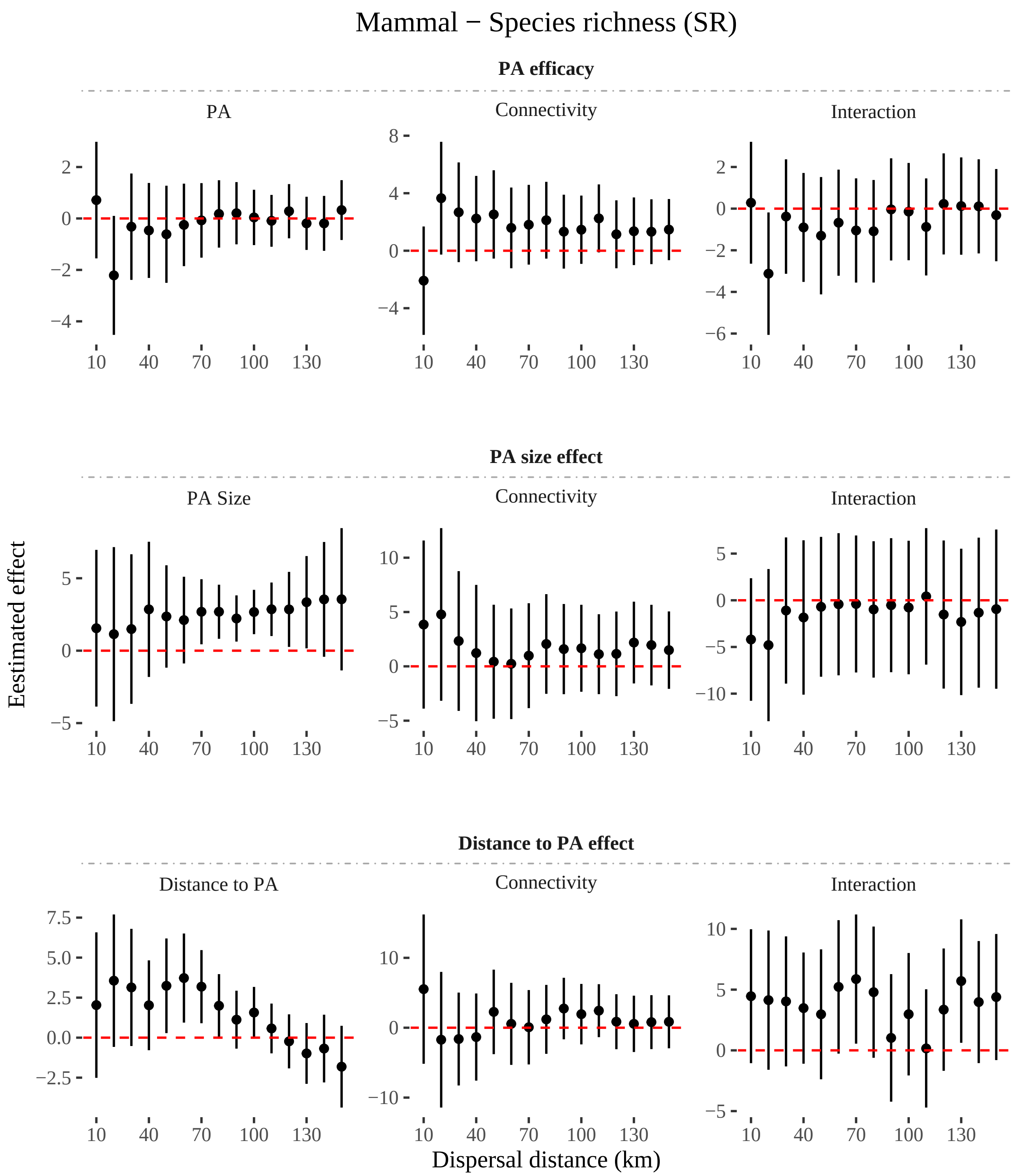

Bird − Species richness (SR)
PA efficacy
PA
Connectivity
Interaction
PA size effect
PA Size
Connectivity
Interaction
Distance to PA effect
Distance to PA
Connectivity
Interaction
Eestimated effect
Dispersal distance (km)



Mammal − Phylogenetic diveristy (PD)
PA efficacy
PA
Connectivity
Interaction
PA size effect
PA Size
Connectivity
Interaction
Distance to PA effect
Distance to PA
Connectivity
Interaction
Eestimated effect
Dispersal distance (km)



Mammal − Functional richness (FR)
PA efficacy
PA
Connectivity
Interaction
PA size effect
PA Size
Connectivity
Interaction
Distance to PA effect
Distance to PA
Connectivity
Interaction
Eestimated effect
Dispersal distance (km)



6 Discussion