1 **Title: Using a large-scale biodiversity monitoring dataset to test the effectiveness of**
2 **protected areas in conserving North-American breeding birds**

3

4

5 **Authors:** Victor Cazalis*[1], Soumaya Belghali[1], Ana S.L. Rodrigues[1]

6 * Corresponding author: victor.cazalis@laposte.net

7 [1]: Centre d'Ecologie Fonctionnelle et Evolutive, CEFE UMR 5175, CNRS – Univ. de Montpellier
8 – Univ. Paul-Valéry Montpellier – EPHE – CNRS, Montpellier, France.

9

10

11 **Abstract:**

12 Protected areas currently cover about 15% of the global land area and constitute one of the main
13 tools in biodiversity conservation. Quantifying their effectiveness at protecting species from
14 decline and local extinction involves comparing protected with counterfactual unprotected sites
15 representing "what would have happened to protected sites if they had not been protected".
16 Most studies are based on pairwise comparisons, using as counterfactuals neighbour sites to
17 protected areas, but this choice is often subjective and may be prone to biases. An alternative is
18 to use large-scale biodiversity monitoring datasets, which whereby the effect of protected areas
19 is analysed statistically by controlling for landscape differences between protected and
20 unprotected sites, allowing a more targeted and clearly defined measure of protected areas
21 effect. Here we use the North American Breeding Bird Survey dataset as a case study to
22 investigate protected areas effectiveness in conserving bird assemblages. We analysed the effect
23 of protected areas on species' richness, assemblage-level abundance and abundance of
24 individual species by modelling how these metrics relate to the proportion of each site that is
25 protected, while controlling for site habitat, altitude, productivity and spatial autocorrelation.
26 At the assemblage level, we found no relationship between protection and species richness or
27 overall abundance. At the species level, we found that species that avoid human activities tend
28 to be favoured by protected areas are the one avoiding human activities. Moreover, we found
29 that forest protected areas presented higher abundances of forest species, making the
30 assemblage more typical of this habitat. We did not find that declining species were particularly
31 favoured by protected areas. Our results highlight the complexity of answering the question of
32 protected areas effectiveness, and the necessity to define clearly metrics measured and the
33 controls used.

34

38

39 **Keywords:** biodiversity conservation; biodiversity monitoring; protected areas effectiveness;
40 birds; North-America BBS; conservation effectiveness.

Introduction

The increasing human footprint on natural ecosystems is leading to major declines in species' populations (McRae et al., 2016) and has already resulted in thousands of extinctions (IUCN, 2018), to such an extent that Ceballos et al. (2017) characterised current times as a period of "biodiversity annihilation". Habitat loss and degradation are the most important pressures on biodiversity (Vié et al., 2009; Balmford and Bond, 2005), as a result of anthropogenic activities such as agriculture, urbanisation, industry, transport and recreation (Foley et al., 2005). The most evident response to these threats is to establish areas with restricted, or even no human activities, *i.e.*, to create protected areas (PAs). Modern PAs have their origins in the 19th century and currently represent the most important conservation tool, with about 15% of the global land area already protected to some extent, and coverage planned to reach 17% by 2020 (UNEP-WCMC IUCN, 2016).

Understanding the extent to which PAs are being effective as biodiversity conservation tools is fundamental for guiding future conservation efforts. Accordingly, there is a substantial and large literature on PA effectiveness: as of the 1st October 2018, 260 publications in the Web of Science included in their title "protected AND area* AND effective*". However, within this literature there are disparate approaches to the concept of "effectiveness".

A first set of studies questions whether PAs are effective at representing species or ecosystems, using gap analyses for measuring the overlap between PAs and the distributions of species or ecosystem types (*e.g.*, Rodrigues et al., 2004; Brooks et al., 2004). These studies do not directly quantify the effectiveness of PAs at conserving biodiversity, but the extent to which species or ecosystems are buffered from human impacts under the assumption that PAs are highly effective in doing so. A second set of studies focuses on the means employed locally by PA managers in order to protect biodiversity, for example in terms of staff or money (*e.g.,* Leverington et al., 2010). These analyses do not directly measure PA effectiveness in reducing human impacts, but rather the resources allocated to this purpose. A third type of studies quantifies the effectiveness of PAs at preventing the conversion of natural ecosystems, typically by comparing land use change (e.g., deforestation rates) in protected versus unprotected areas (Nelson and Chomitz, 2009; Andam et al., 2008). These studies quantify PA effects at the habitat or ecosystem level, rather than at the species level. Finally, a set of analyses focuses on measuring the effect of PAs on species themselves, either on the diversity of assemblages or on the abundance of individual species, typically by contrasting protected versus unprotected sites (e.g. Coetzee et al., 2014; Gray et al., 2016; Devictor et al., 2007, discussed below). This fourth approach to PA effectiveness is the focus of the present study.

Assessing the effectiveness of PAs in conserving species can be implemented by comparing population trends (e.g. Gamero et al., 2017; Devictor et al., 2007; Pellissier et al., 2013). Indeed, if PAs are effective, populations in them are expected to be better buffered from threats and thus decline less, or even to increase more, than those outside. Trends however can be misleading, because they are calculated in relation to a reference date (that seldom precedes all anthropogenic impacts) and because they are measured as percentages (which emphasise

84   changes in small numbers). Hence, for example, given a species in three sites: one where it
85   remained stable at 1000 individuals; a second where it initially declined from 1000 to 10 and
86   recently increased to 20; and a third where it first declined from 1000 to 800 and recently
87   recovered to 1000. In terms of recent trends, the second site appears by far the most effective,
88   even though it has the most depleted population, and even though in absolute numbers the
89   population increase in the third site is 20 times more important. In this study, we focus instead
90   on measures of PA effectiveness that assess current state, namely by contrasting population
91   abundances and species diversity (e.g. Coetzee et al., 2014; Kerbiriou et al., 2018; Devictor et
92   al., 2007). Indeed, if PAs have been effective in conserving species, we expect that over time
93   that translates into higher absolute population abundances than in counterfactual areas, as well
94   (if local extinctions have been prevented) in species diversity.

95

96   Three recent studies investigated the effects of PAs on the state of species abundance and/or
97   diversity, through meta-analyses of studies that made pairwise comparisons between protected
98   and unprotected sites (Geldmann et al., 2013; Coetzee et al., 2014; Gray et al., 2016). The
99   underlying studies used in these meta-analyses did not necessarily aim to measure PA
100   effectiveness, more often they investigated the effects of anthropogenic pressure, using PAs as
101   benchmarks (e.g. Sinclair et al., 2002; Bihn et al., 2008; Wunderle et al., 2006, all used in
102   Coetzee meta-analysis). The meta-analyses considered that unprotected sites acts as
103   counterfactuals to the protected sites (i.e., by assuming that the latter would be in a similar
104   condition to the former had it not been protected), measuring the effect of protection as the
105   observed difference between the two. Pairwise comparisons often compare neighbouring sites,
106   which presents the advantage of ensuring that both have broadly similar environmental
107   characteristics (e.g. same climate), but do not necessarily take in account the fact that PAs tend
108   to be biased in their location towards higher altitudes and lower productivity areas (Joppa and
109   Pfaff, 2009). To account for this, Gray et al. (2016) controlled for the differences in altitude,
110   slope and agricultural suitability. Controlling for these factors means that their results are less
111   influenced by PAs' location biases and, therefore, that they reflect more strongly the effects of
112   protection itself. Another potential bias resulting from pairwise comparisons of neighbouring
113   sites arises from the leakage effect, whereby the human activities that would have taken place
114   inside a PA are displaced to areas around it, artificially inflating the perceived effectiveness of
115   PAs (Ewers and Rodrigues, 2008). This effect is difficult to control for, but should be reduced
116   if the counterfactual sites are not immediately adjacent to the PAs.

117

118   An important decision when choosing a suitable spatial counterfactual to a PA, one that strongly
119   affects the definition and thus the measure of PA effectiveness, is whether to control for habitat
120   type or not. Indeed, not considering it could lead to comparing sites that are not expected to
121   have similar biodiversity regardless of protection (e.g. protected grassland vs unprotected
122   forest), while not considering it overlooks the effect of PAs through preventing habitat changes
123   (e.g. deforestation or urbanization). For instance, given a hypothetical PA covering a natural
124   grassland, possible counterfactuals include an unprotected natural grassland (same habitat, but
125   unprotected), as well as a diversity of unprotected sites with different habitats, for example an
126   extensive pasture (same vegetation structure, but with relatively low-level anthropogenic use),
127   an herbaceous cropland (same vegetation structure but highly transformed), or an urbanised

128 area (a wholly different ecosystem). This choice is certain to have a major impact on the
129 differences observed, and thus on the measure of PA effectiveness, but it is not necessarily
130 obvious what the best counterfactual should be. In theory, it is the site that best represents "what
131 would have happened to the PA in the absence of protection"; in practice, this is not necessarily
132 easily determined. All three meta-analyses include comparisons where habitat has not been
133 controlled for, meaning that the counterfactual's habitat may be different or similar to the
134 protected site's habitat. Additionally, a subset of Gray et al. (2016)'s analyses focuses on
135 comparisons between protected and unprotected sites with matched habitats. In this, the
136 measure of PA effectiveness concerns protection from habitat degradation rather than
137 protection from habitat conversion.

139 Another key consideration in analysing PA effectiveness is the biodiversity metrics of interest.
140 The three meta-analyses applied a diversity of metrics, some at the level of species'
141 assemblages, some focused on individual species. Gray et al. (2016) used only assemblage-
142 level metrics and found higher species richness and overall abundance inside PAs than outside,
143 but no difference in rarefaction-based richness (i.e. number of species for a given number of
144 individuals) nor in the proportion of endemic species. When matching sites with similar
145 habitats, species richness was only higher in young and small PAs than in unprotected sites (no
146 difference between other protected and unprotected sites), suggesting that the effect of PAs on
147 habitat degradation was light. Conversely, Geldmann et al. (2013) considered only species-level
148 metrics (presence, abundance, nest survival) and found contrasted but mainly positive effects
149 of PAs. Finally, Coetzee et al. (2014) considered both levels; at the assemblage level, they found
150 higher species richness and overall abundance in protected than in unprotected sites; at the
151 species level, they found that individual species abundances were typically higher inside PAs.

153 In this study, we use a different approach for quantifying PA effectiveness, one which is not
154 based on pairwise comparisons, but instead takes advantage of a large dataset compiling bird
155 counts across a near-continental area: the North American Breeding Bird Survey (Pardieck et
156 al., 2017). This approach has already been used in other geographical areas, with other datasets
157 (e.g. Devictor et al., 2007 with French birds; Kerbiriou et al., 2018 with French bats; Duckworth
158 and Altwegg, 2018 with South-African birds) with heterogeneous results but mainly showing
159 positive effects of PAs. In this approach, instead of pairing sites, the effect of PAs is quantified
160 through statistical models in which covariates control for differences between protected and
161 unprotected sites. This removes the subjectivity in the choice of counterfactuals, by making it
162 clear which variables are controlled for, and the measure of effectiveness being investigated. In
163 our study, we control for altitude and productivity in order to reduce the effect of PA location
164 biases. We estimate PA effectiveness on two levels of biodiversity: on species' assemblages,
165 through indices of richness and summed abundance; and on individual species, by estimating
166 the effect of PAs on species' abundance for the most common species. At the assemblage level,
167 we expect to find higher species diversity inside PAs. Indeed, as human activities are causing
168 species population declines and local extinctions (Ceballos et al., 2017), and as PAs are
169 expected to buffer against these activities, this should predictably lead to overall higher species
170 richness and higher total abundance inside PAs, as found by Coetzee et al. (2014) and Gray et
171 al. (2016). At the species level, we expect individual species' abundances to be higher in PAs.

172　　However, given differences in species' habitat requirements, this result cannot be expected to
173　　hold universally (*i.e.,* species are not all expected to be more abundant in all PAs). For example,
174　　we expect protected forests to have a positive effect on forest species, but not on grassland
175　　species. To take this into account, we control in our analyses for broad vegetation structure
176　　(forest, shrub, herbaceous), by investigating separately the effects of PAs dominated by a
177　　particular vegetation structure on species with different habitat requirements. Additionally, we
178　　expect species with overall declining populations (thus more affected by anthropogenic
179　　activities), and species that avoid human presence (more sensitive to human disturbance) to
180　　present higher abundances inside PAs.

Methods

As stated in the introduction, in this study we will use the term PAs effectiveness as the difference in diversity or abundance between protected and unprotected sites, acknowledging that it includes both effectiveness to select the most interesting sites for conservation when implementing PAs and the effectiveness in create more positive or less negative biodiversity trends inside PAs.

Bird data

We used data from the North-American Breeding Bird Survey (BBS), a long-term volunteer-based monitoring scheme in Canada, the USA, and Mexico (Pardieck et al., 2017). Here we studied only Canada and the USA, as few Mexican routes are monitored. This program is based on the annual monitoring of 25-mile routes during the breeding season. Each route is split into 50 stops; at each stop, the observer counts every bird heard or seen during three minutes, before moving to the next stop.

Given the length of BBS routes, they often intersect multiple land use types (e.g. forested, urban, agriculture, with different bird assemblages), and they are rarely wholly contained within protected sites (most of the routes that cross PAs do so only in small fractions of their length). As a result, whole BBS routes are not particularly suited sampling units for investigating how PAs affect bird species. We chose instead to focus on small sections of BBS routes – sequences of five stops, covering about 2.5 miles – in order to obtain field sampling units that are less heterogeneous in land type and for which there is a stronger correspondence between the presence or PAs and the bird assemblages detected. For each route, we only used the first sequence of five stops, because the only precisely georeferenced point we had access to was the starting stop of each route. Indeed, even if in principle additional stops are spaced about 0.5 miles from each other, in practice this distance can vary, making the location of additional stops in each route progressively more imprecise. Henceforth, and for simplicity, we use the term "routes" to refer to these initial sections of five stops rather than to entire routes.

We excluded bird taxa that are not well detected by this diurnal road-based monitoring scheme (aquatic and nocturnal birds), those that correspond to non-indigenous species, and hybrids. Overall, we analysed 400 species.

We focused on routes sampled at least 5 years between 2007 and 2016, obtaining a set of 3,427 routes analysed. For routes sampled more than five years, we analysed only five (randomly selected) years of data, thus ensuring a consistent sampling effort across all routes. For each species, the abundances were summed across the five points and the five years, giving a single value per species per route. We winsorized the abundances of each species (*i.e.,* values above 95% quantiles were reduced to the 95% quantile value) to limit the impact of extreme values.

Landscape data

For each route, we analysed the properties of the landscape within a 500 m buffer around the route's 2.5-mile track (total area ca. 6 km^2), which we considered as a suitable description of the environment affecting the composition of birds detected by the BBS and which corresponds

224   broadly to the bird detection radius of the BBS. Small et al. (2012) showed that the immediate
225   landscape composition (buffer of 0.4 km) of BBS routes was similar to large-scale landscape
226   composition (buffer of 10 km), so this choice is not expected to strongly affect the results.
227
228   Protected area is defined by the IUCN as "a clearly defined geographical space, recognised,
229   dedicated and managed, through legal or other effective means, to achieve the long term
230   conservation of nature with associated ecosystem services and cultural values". They are
231   categorised by the IUCN within seven categories based on their protection level from Ia "strictly
232   protected areas set aside to protect biodiversity […], where human visitation, use and impacts
233   are strictly controlled and limited to ensure protection of the conservation values." to VI which
234   "conserve ecosystems and habitats together with associated cultural values and traditional
235   natural resource management systems. They are generally large, with most of the area in a
236   natural condition, where a proportion is under sustainable natural resource management and
237   where low-level non-industrial use of natural resources compatible with nature conservation is
238   seen as one of the main aims of the area" (UNEP-WCMC and IUCN, 2018). We used the PAs'
239   shapefile, including both locations and IUCN categories of PAs, which was provided by the
240   World Database on Protected Areas (UNEP-WCMC and IUCN, 2018). We calculated the
241   proportion of area inside each route's buffer that falls within a PA (all IUCN-categories
242   combined, and dissolved to avoid double-counting of areas under multiple PA designations).
243   We have also run analyses considering stricter PAs only (categories I-IV), as the effectiveness
244   can vary with protection level.
245
246   For each route, we obtained values according to four environmental variables: net primary
247   productivity, altitude, human footprint, and type of vegetation structure. The first three are
248   continuous variables, available as raster files, and we obtained a mean values across all pixels
249   that overlap the respective buffer: net primary productivity as the mean during spring months
250   (Mars to June) between 2004 and 2015 according to the monthly Net Primary Productivity
251   Terra/Modis (NASA, 2017; resolution 0.1 degree, about 62 km^2 at 45°N); altitude using
252   GLOBE Digital Elevation Model (National Geophysical Data Center, 1999; resolution 0.008
253   degree, about 0.40 km^2 at 45°N); human footprint from the 2009 Global terrestrial Human
254   Footprint (Venter et al., 2016; resolution 0.01*0.008, about 0.50 km^2 at 45°N). We defined the
255   vegetation structure as a categorical variable with three types: forest, shrub and herbaceous. We
256   started by reclassifying the land cover classes in the Global Land Cover 2000 layer (Bartholomé
257   and Belward, 2005; resolution 0.009 degree, about 0.50 km^2 at 45°N) into the three vegetation
258   structure types: forest from land cover classes 1-9 (N=1,749 routes); shrub, 11-12 (N=409);
259   herbaceous, 13-16 (which includes croplands; N=1,140). We then obtained the main vegetation
260   structure type for each route as the dominant in the buffer. Routes which were dominated by
261   other land use classes (burned trees, 10; mosaic, 17-18; bare areas, 19; water areas 20-21;
262   artificial, 22) were not analysed because too scarce. Routes used in analyses are mapped in
263   Appendix S1.
264

265   ## Statistical analyses
266   We estimated the effect of PAs on each of two assemblage indices (species richness and summed
267   abundance) and on the abundance of individual species using General Additive Models
268   (GAMs). Models all had identical structures, with the response variable modelled as function
269   of the proportion of PAs inside the buffer, interacting with vegetation structure type. We added
270   smoothed terms controlling for productivity and altitude, as PAs are globally biased towards
271   high altitude and low productivity areas (Joppa and Pfaff, 2009), as well as longitude and
272   latitude in order to correct for spatial autocorrelation:
273   Response ~ PA * vegetation + s(productivity, altitude, longitude, latitude)

274   ### Assemblage level
275   For each route, and across all 400 bird species analysed, we calculated two assemblage indices,
276   in each case using the cumulative number of species or individuals seen across the 5 stops, over
277   5 years: species richness ($\mu = 28.6 \pm 9.5$ species); summed bird abundance across all species ($\mu =$
278   249 ± 88 individuals). We then used a GAM to model each of these two assemblage variables
279   against the above-mentioned covariates, assuming a Gaussian distribution for richness and a
280   negative binomial distribution for abundance.

281   ### Species level
282   We excluded the rarest species from this analysis, keeping only the 149 species observed on
283   more than 100 routes, in order to have enough statistical power. For each species, we only
284   analysed routes within the species' distribution within our study area. We obtained an
285   approximation of this distribution by delimiting the 90 % spatial kernel of the routes where the
286   species was observed, using the *'adehabitat'* R package (Calenge, 2006). We treated all routes
287   inside the kernel where the species was not observed as having zero abundance.
288
289   We modelled each species' abundance using a GAM as mentioned above, with a Poisson
290   distribution. We then calculated for each species a "PA effect" (PAE), measured as the difference
291   in predicted abundance between a fully protected and an unprotected route with all control
292   variables fixed to their median values. We calculated PAE separately for each of the three types
293   of vegetation structure, to obtain for each species a value of PAE_{For} for routes dominated by
294   forest, PAE_{Shrub} for shrub routes, and PAE_{Herb} for herbaceous routes.
295
296   For each type of vegetation structure, we studied PAE values in order to understand the factors
297   explaining which species are favoured or not by PAs. To do so, we used a linear model (LM)
298   and two phylogenetic linear models (phyBM and phyL, see below) with species-level
299   covariates. We considered three covariates: species' habitat preference, population trend, and
300   human-affinity. We extracted from Del Hoyo et al. (2013) species' main habitat (11 categories;
301   see Fig.2). We used species' population trends in North America between 1966 and 2015,
302   calculated for each species by Sauer et al. (2017) from the BBS data (negative number for
303   declining species, positive for increasing species). We winsorized these values, folding down
304   the 2.5% extreme values on each side, bringing estimates to a Gaussian distribution. Finally,
305   we estimated for each species a human-affinity index, as the median human footprint of the
306   routes where the species was observed, weighted by species' abundance on the route.
307

308   The two phylogenetic models used are the Brownian motion model (phyBM) and the Lambda

309   model (phyL), both implemented in the '*phylolm*' R package (Tung Ho and Ané, 2014). To

310   obtain the bird phylogeny, we selected randomly 100 phylogenetic trees over 10,000 from Jetz

311   et al. (2012) and calculated a maximum clade credibility tree using *Tree Annotator* from *Mr*

312   *Bayes* (Drummond et al., 2012) with no burnin, and node heights calculated with the median.

Results

Assemblage-level analyses

At the assemblage level, species richness and summed abundance differed very significantly between vegetation structure types (respectively $P<2.10^{-16}$, $P=1.10^{-6}$), underlying the importance of accounting for habitat differences when studying PAs effect.

However, neither species richness, nor summed abundance were significantly affected by the proportion of PAs in the buffer (respectively $P=0.13$, $P=0.13$), or its interaction with vegetation structure type (respectively $P=0.20$, $P=0.29$). This lack of significance between assemblage indices and the proportion of PAs in the buffer was also true when not controlling for vegetation structure (respectively $P=0.68$, $P=0.058$).

Species-level analyses

PAE_{For} – the predicted difference in a given species' abundance between protected *versus* unprotected forest routes – differed significantly depending on the species' main habitat, under both LM and phyL models but not under phyBM model. Hence, the first two models indicate that within forest routes, species with any type of forest as main habitat (mixed, deciduous, forest, conifer; Table 1 and Fig.2) are predicted to have significantly higher abundances when routes are protected. We found no significant PA effect within forest routes for species favouring other habitat types.

In all models, species' population trends between 1966 and 2015 did not significantly explain PAE_{For} (Table 1). In contrast, species' human-affinity was significantly negatively correlated with PAE_{For} (*i.e.,* species with lower affinity to humans had higher effects of PAs in forested routes; Table 1, Fig.3). This effect was also significant when only forest species were considered (green dots in Fig.3; see Supporting Information in Appendix S2 for additional test).

The effect of PAs within shrub routes (PAE_{Shrub}) and within herbaceous routes (PAE_{Herb}) was not affected by species' main habitat under any of the models (Supporting information, Appendix S3). PAE_{Shrub} decreased significantly with species' trend under all models (*i.e.,* declining species had higher effects of PAs in shrub routes), whereas it decreased with human-affinity only under model phyBM. PAE_{Herb} was not significantly correlated with any of the three covariates.

These results, however, need to be interpreted taking into account that shrub or herbaceous protected routes were rare in our dataset: on average, each species' kernel included only 10 shrub and 7 herbaceous routes protected by 50% or more, contrasted with 60 protected forest routes (Fig.1; see Appendix S4 in Supporting Information). The lack of significance in models with PAE_{Shrub} and PAE_{Herb} might thus be due to the limited number of protected routes in the sample, whereas the significant correlations between PAE_{Shrub} and both species' trends and human-affinity might not be robust.

355 Results, both at the assemblage and at the species levels, were similar but less significant when
356 we considered only PAs of stricter management, as defined by IUCN categories I-IV (Dudley,
357 2008; see Supporting Information, Appendix S6). For shrub and herbaceous routes, the number
358 of protected routes was even smaller than when all PAs were considered, leading to aberrant
359 results.

360

361 Table 1: Model summaries regarding the estimated effect of PAs on species within forest routes (PAE$_{For}$): linear
362 model (LM), phylogenetic linear model with Brownian motion model (phyBM), phylogenetic linear model with
363 Lambda model (phyL). The top part gives estimates and P-values for all covariates, the bottom part gives estimates
364 and P-values for all species' habitat preferences, with trend and human-affinity fixed to zero. N corresponds to the
365 number of species in each case.

366 * P-values for the habitat variable as a whole could not be obtained, as Anova tables are not implemented in the *'phylolm'* package.

	LM		PhyBM		PhyL	
Model variables						
	Estimate	*P*	*Estimate*	*P*	*Estimate*	*P*
Habitat	-	**2.10^{-8}**	-	*NA**	-	*NA**
Trend	0.027	0.602	0.039	0.544	0.027	0.59
Human-affinity	-0.079	**0.005**	-0.065	**0.022**	-0.076	**0.008**
Main habitats						
Mixed forest (N=16)	0.98	**1.10^{-4}**	0.80	0.48	0.96	**2.10^{-4}**
Forest (N=7)	0.77	**0.0167**	0.69	0.55	0.75	**0.02**
Deciduous forest (N=18)	0.86	**0.0022**	0.59	0.60	0.83	**0.003**
Conifer forest (N=22)	0.83	**4.10^{-5}**	0.38	0.73	0.79	**2.10^{-4}**
Semi open (N=27)	0.29	0.23	0.18	0.88	0.27	0.27
Riparian (N=18)	0.28	0.27	0.11	0.92	0.26	0.32
Generalist (N=2)	0.49	0.37	0.03	0.98	0.46	0.40
Shrub (N=19)	0.07	0.78	-0.22	0.84	0.046	0.85
Arid (N=5)	-0.27	0.44	-0.22	0.85	-0.29	0.41
Open (N=14)	-0.41	0.13	-0.36	0.75	-0.43	0.11
Urban (N=1)	-0.87	0.27	-0.95	0.69	-0.91	0.25

367
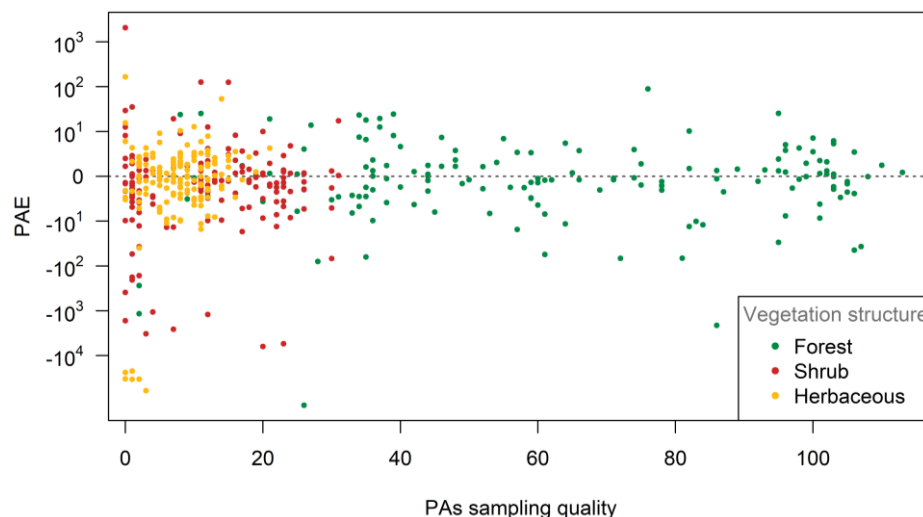
368

369 Figure 1: Estimated Protected Areas effect per species (PAE) (represented on a log scale in both negative and
370 positive values), against PAs sampling quality, per vegetation structure type of the routes. PAs sampling quality
371 was quantified as the number of routes within the species' kernel with at least 50% of the buffer area covered by
372 PAs. Each point in the plot corresponds to a species, and each species can be represented up to three times, one
373 for each vegetation structure type.
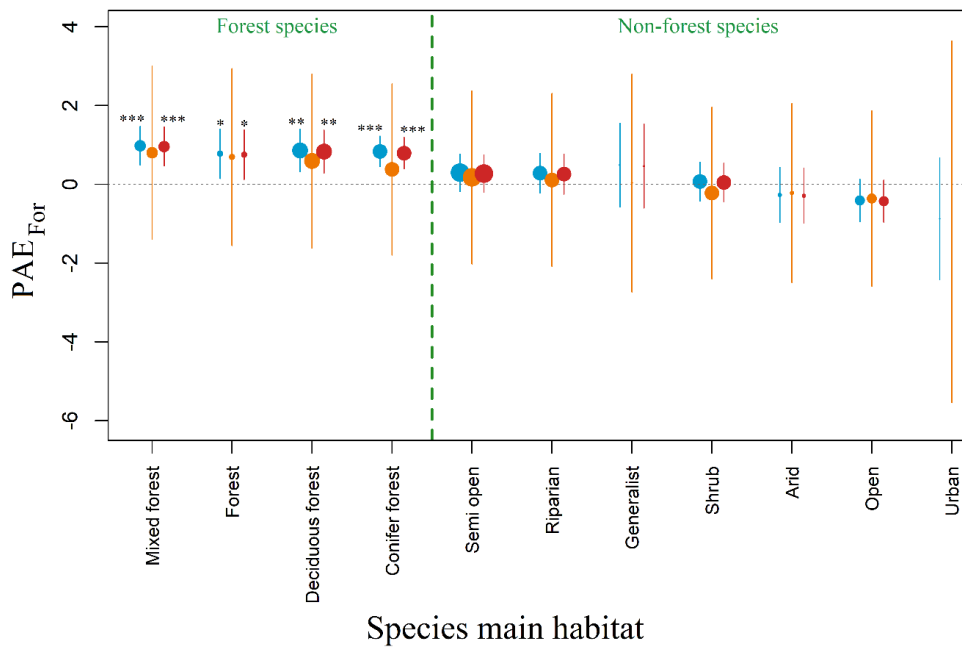
Figure 2: Estimated effect of PAs on species within BBS forest routes (PAE_{For}) per species' main habitat (Del Hoyo et al., 2013), estimated with three models: linear model (LM, blue), phylogenetic linear model with Brownian motion model (phyBM, orange), and phylogenetic linear model with Lambda model (phyL, red). Estimates were all calculated with species' population trend and human-affinity fixed to zero. Error bars represent 95% CI; dots sizes are proportional to the number of species in each habitat group. Stars indicate significant effects for the particular model, for the particular species' main habitat (P: $0.05 < * < 0.01 < ** < 0.001 < ***$). Habitat types are ordered from the highest to the lowest PAE_{For} values under phyBM model.
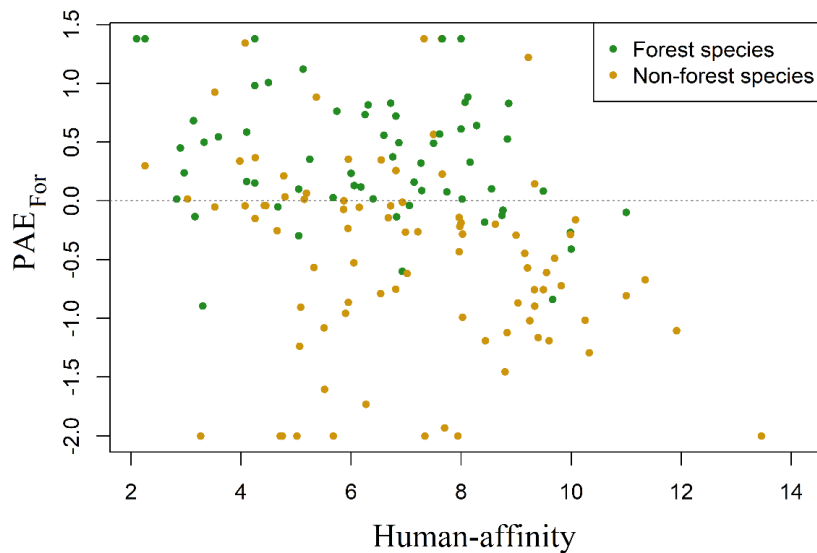
Figure 3: Species' human-affinity (higher for species found preferably in areas of higher human footprint) against the estimated effect of PAs on species within BBS forest routes (PAE_{For}, high for species whose abundance in forest routes is higher in protected rather that in unprotected areas). Forest species (green) are species whose main habitat is "forest", "conifer forest", "mixed forest" or "deciduous forest"; non-forest species (brown) are all other species.

Discussion

389

390   We compared the effect of PA coverage on bird species diversity, using assemblage indices
391   (species richness, summed abundance) and individual species' abundances.

392

393   At the assemblage level, we did not find significant differences in species richness or summed
394   abundance between protected and unprotected sites, irrespective of whether vegetation structure
395   was taken into account or not. In one sense, this is not surprising, particularly when it comes to
396   species richness: according to the intermediate disturbance hypothesis, an area with low human-
397   induced disturbance can have higher species richness than a pristine area (Roxburgh et al.,
398   2004). Accordingly, Hiley et al., (2016) found lower alpha avian diversity in Mexican PAs than
399   in unprotected areas. However, our results contrast with previous studies investigating this
400   question such as Coetzee et al. (2014) or Gray et al. (2016), which found a positive effect of
401   PAs on species richness and summed abundance, including in North America (Coetzee et al.,
402   2014). These two studies being meta-analyses, it is possible that a publication bias against
403   studies showing negative or null effects of PAs (discussed by Coetzee et al., 2014) artificially
404   increased the difference they measured. This is even more so the case given that the underlying
405   studies of the meta-analyses were often designed to measure the effect of anthropogenic
406   pressures, using PAs as benchmarks, rather than measuring the effectiveness of PAs (e.g.
407   Sinclair et al., 2002; Bihn et al., 2008; Wunderle et al., 2006, all used in Coetzee meta-analysis),
408   and may thus have focused on particularly intact protected sites and/or in highly degraded non-
409   protected sites. Conversely, our study may not be representative of studies at a global scale, for
410   example if North American birds are less sensitive to human activities than other taxa in North
411   America and/or birds in other regions, or if there is less contrast in human impacts in protected
412   *versus* unprotected areas in North America than elsewhere. In addition, the lack of difference
413   between protected and unprotected sites in terms of richness and abundance could also
414   potentially be explained by a difference in species' detectability (Boulinier et al., 1998) if PAs
415   protect mainly species that are difficult to detect. This detection problem should not affect our
416   result at the species level.

417

418   Even if overall species richness and abundance are similar, PAs may nonetheless have an effect
419   on avian assemblages if different species respond differently to protection. We found that in
420   routes whose vegetation is dominated by forests, PAs seem to have an overall positive effect
421   on species' abundance, but only for those species with forest as their main habitat. Forest PAs
422   thus seem to maintain a more forest-typical bird assemblage than comparable unprotected
423   forests. This effect was significant with the linear model, and with one (phyL), but not the other
424   (phyBM) of the two phylogenetic linear models. This suggests that much of the effect attributed
425   to habitat preferences under the linear model can actually be considered as phylogenetic
426   difference, which is not surprising as bird habitat preferences and phylogeny are correlated.
427   Phylogenetic models could theoretically allow us to measure the effectiveness of PAs in
428   protecting species across phylogeny, and to check if some taxa were not effectively protected
429   (*e.g.,* they could highlight that a given family is not protected by PAs). However, to draw such
430   conclusions, we would need to know how species are affected by PAs in each vegetation
431   structure types, which is not the case here. Therefore, phylogenetic models give little

432   information here, only highlighting that the difference in PAE_{For} between species habitat
433   preferences is correlated with phylogeny. Moreover, all models indicate that species with low
434   human-affinity (*i.e.,* species that avoid human-impacted areas) are more favoured by forest PAs.
435   This highlights that forest PAs protect species that are the most sensitive to human pressure,
436   while species with high human-affinity, often benefiting from human presence, showed lower
437   abundances inside PAs. Contrary to our expectation, and to previous results for common French
438   birds (Devictor et al., 2007), we found no correlation between species' population trends over
439   the past 50 years and PAE_{For}. This may reflect the fact that our model included only relatively
440   common species (*i.e.,* observed on at least 100 routes in the studied years). It is thus possible
441   that the most endangered species are favoured by PAs, but that we could not measure it.

443   Our models suggested that PAs in shrub areas have a beneficial impact on declining species and
444   those with low human-affinity, whereas we did not find significant results for herbaceous areas.
445   Given the scarcity of protected routes within both of these vegetation structure types, we do not
446   consider these results robust or informative of the effectiveness of PAs, but they nonetheless
447   emphasise the biases of BBS routes against shrub areas and herbaceous PAs (Appendix S4).

449   Given that PAs located in forests are not expected to favour the same species as PAs located in
450   grasslands or shrub lands, we controlled for vegetation structure in our analyses of PA effects.
451   However, this control masked the effect PAs may have had in preventing changes in vegetation
452   structure (and associated changed in bird assemblages). For instance, given the vegetation
453   structure categorisation we applied, the counterfactual for a protected forest was an unprotected
454   forest, which does not take into account the possibility that the PA may have prevented the
455   forest from being cleared. In other words, our approach does not measure the effect PAs can
456   have on species diversity by preventing habitat destruction (that modifies vegetation structure
457   type), only the effects PAs can have in preventing habitat degradation (not modifying the
458   vegetation structure type), for example from natural forest to exploited forest, or from natural
459   grassland to croplands.

461   Pairwise comparisons of protected *versus* unprotected sites, and thus the meta analyses from
462   Geldmann et al. (2013), Coetzee et al. (2014) and Gray et al. (2016), can take into account the
463   combined effects of habitat destruction and habitat degradation on species diversity, given that
464   the counterfactual chosen may well have a different habitat structure than the protected site
465   (e.g., a protected forest compared with an unprotected cropland). Nonetheless, defining the
466   effectiveness measured in these meta-analyses is not straightforward, as it depends heavily on
467   the choice of counterfactuals in underlying studies, which are defined directly by authors
468   depending on their objectives. For instance, as discussed before, numerous studies used in the
469   meta-analyses compare a highly degraded site with a protected site used as benchmark, in order
470   to estimate the impact of anthropogenic degradation, which can lead to an overestimate of PA
471   effectiveness. Other studies aimed to estimate PA effectiveness directly (e.g. Wasiolka and
472   Blaum, 2011; Lee et al., 2007), but their choice of counterfactual was subjectively based on
473   what authors considered likely to have happened to the protected site had it not been protected
474   (Coetzee et al., 2014). Finally, some other studies used in meta-analyses were not particularly
475   interested in differences between protected and unprotected sites, protection was only used as

476 a covariate explaining potentially some noise around the signal the authors were interested in
477 (e.g. Naidoo, 2004; McCarthy et al., 2010). Because of the diversity of approaches used in these
478 meta-analyses, it is difficult to define precisely what has been measured as PA effectiveness.
479 Although our approach does not allow us to measure the full effects of PAs, the difference we
480 measured between protected and unprotected sites is defined statistically depending on the
481 covariates included, which allows to understand clearly what is being included in measured
482 effects of PA. A main advantage of large biodiversity monitoring datasets (such as breeding
483 bird-monitoring schemes) in relation to pairwise comparisons is thus the possibility of applying
484 a well-defined and repeatable control.
485
486 More broadly, our results highlight that clearly measuring PA effectiveness in conserving
487 species diversity is impossible without defining precisely what is expected from them. In this
488 study, we measured PAs effectiveness as the difference in abundance or richness between
489 protected and unprotected sites. This definition assumes that PAs are expected to protect
490 globally species diversity, and therefore gathers our ability to protect richest areas and to reduce
491 human impacts on biodiversity in these areas. If PAs are expected to present higher diversity in
492 terms of assemblage metrics (species richness or summed abundance), then we found no
493 evidence in our analyses that PAs are effective. If PAs are expected to protect all species'
494 populations, then we did not find they were effective either, as for about half of the 149 species
495 studied here we found a negative effect of PAs in forest. However, our results show that North-
496 American forest PAs present higher abundances in forest species when compared with
497 unprotected forest sites (especially for species with low affinity to human activities). That this
498 result holds even though we found no significant difference in total abundance suggests that
499 bird assemblages in protected forests are more forest-typical than those in unprotected forests.
500 Our results thus indicate that forest PAs in North-American are contributing to prevent forest
501 habitat degradation, and associated losses in the abundance of forest specialist species. BBS
502 routes do not currently cover sufficiently well other habitats besides forest to allow us to
503 investigate whether the same result applies to PAs with a different vegetation structure, but
504 datasets with a bigger proportion of sampling points inside PAs, across all habitats, would help
505 investigating this question.
506
507 Overall, our results emphasize the complexity of resolving a question that seemed so
508 straightforward, and whose answer seemed so intuitive. In practice, understanding whether PAs
509 are effective or not, and quantifying such effects, involves defining clearly what effect is being
510 tested, on which facet of species diversity, and how to obtain appropriate counterfactuals.

Data accessibility statement:

All data used in the study (birds and landscape covariates) are public and accessible to anybody.
All sources are given with references.

References:

Andam, K.S., Ferraro, P.J., Pfaff, A., Sanchez-Azofeifa, G.A., and Robalino, J.A. (2008). Measuring the effectiveness of protected area networks in reducing deforestation. Proc. Natl. Acad. Sci. *105*, 16089–16094.

Balmford, A., and Bond, W. (2005). Trends in the state of nature and their implications for human well-being: Trends in the state of nature. Ecol. Lett. *8*, 1218–1234.

Bartholomé, E., and Belward, A.S. (2005). GLC2000: a new approach to global land cover mapping from Earth observation data. Int. J. Remote Sens. *26*, 1959–1977.

Bihn, J.H., Verhaagh, M., Brändle, M., and Brandl, R. (2008). Do secondary forests act as refuges for old growth forest animals? Recovery of ant diversity in the Atlantic forest of Brazil. Biol. Conserv. *141*, 733–743.

Boulinier, T., Nichols, J.D., Sauer, J.R., Hines, J.E., and Pollock, K.H. (1998). Estimating species richness: the importance of heterogeneity in species detectability. Ecology *79*, 1018–1028.

Brooks, T.M., Bakarr, M.I., Boucher, T., Da Fonseca, G.A.B., Hilton-Taylor, C., Hoekstra, J.M., Moritz, T., Olivieri, S., Parrish, J., Pressey, R.L., et al. (2004). Coverage Provided by the Global Protected-Area System: Is It Enough? BioScience *54*, 1081.

Calenge, C. (2006). The package "adehabitat" for the R software: A tool for the analysis of space and habitat use by animals. Ecol. Model. *197*, 516–519.

Ceballos, G., Ehrlich, P.R., and Dirzo, R. (2017). Biological annihilation via the ongoing sixth mass extinction signaled by vertebrate population losses and declines. Proc. Natl. Acad. Sci. 201704949.

Coetzee, B.W.T., Gaston, K.J., and Chown, S.L. (2014). Local Scale Comparisons of Biodiversity as a Test for Global Protected Area Ecological Performance: A Meta-Analysis. PLOS ONE *9*, e105824.

Del Hoyo, J., Elliott, A., and Sargatal, J. (2013). Handbook of the birds of the World (Barcelona, Spain).

Devictor, V., Godet, L., Julliard, R., Couvet, D., and Jiguet, F. (2007). Can common species benefit from protected areas? Biol. Conserv. *139*, 29–36.

Drummond, A.J., Suchard, M.A., Xie, D., and Rambaut, A. (2012). Bayesian Phylogenetics with BEAUti and the BEAST 1.7. Mol. Biol. Evol. *29*, 1969–1973.

548   Duckworth, G.D., and Altwegg, R. (2018). Effectiveness of protected areas for bird
549   conservation depends on guild. Divers. Distrib.

550   Dudley, N. (2008). Guidelines for applying protected area management categories (Iucn).

551   Ewers, R.M., and Rodrigues, A.S.L. (2008). Estimates of reserve effectiveness are
552   confounded by leakage. Trends Ecol. Evol. *23*, 113–116.

553   Foley, J.A., DeFries, R., Asner, G.P., Barford, C., Bonan, G., Carpenter, S.R., Chapin, F.S.,
554   Coe, M.T., Daily, G.C., Gibbs, H.K., et al. (2005). Global Consequences of Land Use.
555   Science *309*, 570–574.

556   Gamero, A., Brotons, L., Brunner, A., Foppen, R., Fornasari, L., Gregory, R.D., Herrando, S.,
557   Hořák, D., Jiguet, F., Kmecl, P., et al. (2017). Tracking Progress Toward EU Biodiversity
558   Strategy Targets: EU Policy Effects in Preserving its Common Farmland Birds. Conserv. Lett.
559   *10*, 395–402.

560   Geldmann, J., Barnes, M., Coad, L., Craigie, I.D., Hockings, M., and Burgess, N.D. (2013).
561   Effectiveness of terrestrial protected areas in reducing habitat loss and population declines.
562   Biol. Conserv. *161*, 230–238.

563   Gray, C.L., Hill, S.L.L., Newbold, T., Hudson, L.N., Börger, L., Contu, S., Hoskins, A.J.,
564   Ferrier, S., Purvis, A., and Scharlemann, J.P.W. (2016). Local biodiversity is higher inside
565   than outside terrestrial protected areas worldwide. Nat. Commun. *7*, 12306.

566   Hiley, J.R., Bradbury, R.B., and Thomas, C.D. (2016). Impacts of habitat change and
567   protected areas on alpha and beta diversity of Mexican birds. Divers. Distrib. *22*, 1245–1254.

568   IUCN (2018). IUCN Red List of Threatened Species. Versio 2018.1. http://iucnredlist.org.

569   Jetz, W., Thomas, G.H., Joy, J.B., Hartmann, K., and Mooers, A.O. (2012). The global
570   diversity of birds in space and time. Nature *491*, 444–448.

571   Joppa, L.N., and Pfaff, A. (2009). High and Far: Biases in the Location of Protected Areas.
572   PLoS One *4*.

573   Kerbiriou, C., Azam, C., Touroult, J., Marmet, J., Julien, J.-F., and Pellissier, V. (2018).
574   Common bats are more abundant within Natura 2000 areas. Biol. Conserv. *217*, 66–74.

575   Lee, T.M., Sodhi, N.S., and Prawiradilaga, D.M. (2007). THE IMPORTANCE OF
576   PROTECTED AREAS FOR THE FOREST AND ENDEMIC AVIFAUNA OF SULAWESI
577   (INDONESIA). Ecol. Appl. *17*, 1727–1741.

578   Leverington, F., Costa, K.L., Pavese, H., Lisle, A., and Hockings, M. (2010). A Global
579   Analysis of Protected Area Management Effectiveness. Environ. Manage. *46*, 685–698.

580   McCarthy, J.L., McCarthy, K.P., Fuller, T.K., and McCarthy, T.M. (2010). Assessing
581   Variation in Wildlife Biodiversity in the Tien Shan Mountains of Kyrgyzstan Using Ancillary
582   Camera-trap Photos. Mt. Res. Dev. *30*, 295–301.

583   McRae, L., Freeman, R., and Marconi, V. (2016). "The Living Planet index" in: Living Planet
584   Report 2016: Risk and resilience in a new era (ed. Oerlemans N). (Gland, Switzerland).

585 Naidoo, R. (2004). Species richness and community composition of songbirds in a tropical
586 forest-agricultural landscape. Anim. Conserv. *7*, 93–105.

587 NASA (2017). Net Primary Productivity (Terra/Modis) NASA Earth Observations (monthly
588 data 2004-2015).

589 National Geophysical Data Center (1999). Global Land One-kilometer Base Elevation
590 (GLOBE), version 1.

591 Nelson, A., and Chomitz, K.M. (2009). Protected Area Effectiveness in Reducing Tropical
592 Deforestation (Washington DC, USA: Independent Evaluation Group, The World Bank).

593 Pardieck, K.L., Ziolkowski, D.J., Lutmerding, M., Campbell, K., and Hudson, M.-A.R.
594 (2017). North American Breeding Bird Survey Dataset 1966 - 2016, version 2016.0.

595 Pellissier, V., Touroult, J., Julliard, R., Siblet, J.P., and Jiguet, F. (2013). Assessing the Natura
596 2000 network with a common breeding birds survey: Assessing Natura 2000 with a common
597 breeding birds survey. Anim. Conserv. *16*, 566–574.

598 Rodrigues, A.S.L., Andelman, S.J., Bakarr, M.I., Boitani, L., Brooks, T.M., Cowling, R.M.,
599 Fishpool, L.D.C., da Fonseca, G.A.B., Gaston, K.J., Hoffmann, M., et al. (2004).
600 Effectiveness of the global protected area network in representing species diversity. Nature
601 *428*, 640–643.

602 Roxburgh, S.H., Shea, K., and Wilson, J.B. (2004). The intermediate disturbance hypothesis:
603 patch dynamics and mechanisms of species coexistence. Ecology *85*, 359–371.

604 Sauer, J.R., Niven, D.K., Hines, J.E., Ziolkowski, D.J.J., Pardieck, K.L., Fallon, J.E., and
605 Link, W.A. (2017). The North American Breeding Bird Survey, Results and Analysis 1966 -
606 2015. Version 2.07.2017 (Laurel, MD: USGS Patuxent Wildlife Research Center).

607 Sinclair, A.R.E., Mduma, S.A.R., and Arcese, P. (2002). Protected areas as biodiversity
608 benchmarks for human impact: agriculture and the Serengeti avifauna. Proc. R. Soc. Lond.
609 Ser. B-Biol. Sci. *269*, 2401–2405.

610 Small, M.F., Veech, J.A., and Jensen, J.L.R. (2012). Local landscape composition and
611 configuration around North American Breeding Bird Survey routes: *Ecological Archives*
612 E093-215. Ecology *93*, 2298–2298.

613 Tung Ho, L. si, and Ané, C. (2014). A Linear-Time Algorithm for Gaussian and Non-
614 Gaussian Trait Evolution Models. Syst. Biol. *63*, 397–408.

615 UNEP-WCMC, and IUCN (2018). Protected Planet: [WDPA-shapefile-polygons; The World
616 Database on Protected Areas (WDPA)/The Global Database on Protected Areas Management
617 Effectiveness (GD-PAME)] [On-line], [downloaded 02/10/2018], Cambridge, UK: UNEP-
618 WCMC and IUCN. Available at: www.protectedplanet.net.

619 UNEP-WCMC IUCN (2016). Protected Planet Report (Cambridge UK and Gland,
620 Switzerland: UNEP-WCMC and IUCN).

621   Venter, O., Sanderson, E.W., Magrach, A., Allan, J.R., Beher, J., Jones, K.R., Possingham,
622   H.P., Laurance, W.F., Wood, P., Fekete, B.M., et al. (2016). Global terrestrial Human
623   Footprint maps for 1993 and 2009. Sci. Data *3*, 160067.

624   Vié, J.-C., Hilton-Taylor, C., and Stuart, S.N. (2009). Wildlife in a changing world: an
625   analysis of the 2008 IUCN red list of threatened species (Gland, Switzerland : Barcelona,
626   Spain: IUCN ; Lynx Edicions).

627   Wasiolka, B., and Blaum, N. (2011). Comparing biodiversity between protected savanna and
628   adjacent non-protected farmland in the southern Kalahari. J. Arid Environ. *75*, 836–841.

629   Wunderle, J.M., Henriques, L.M.P., and Willig, M.R. (2006). Short-Term Responses of Birds
630   to Forest Gaps and Understory: An Assessment of Reduced-Impact Logging in a Lowland
631   Amazon Forest [1]: Reduced-Impact Logging Effects on Amazon Birds. Biotropica *38*, 235–
632   255.

633