Protected areas have a mixed impact on waterbirds, but management helps

Wauchope, H. S.[1,2], Jones, J. P. G.[3], Geldmann, J.[1,4], Simmons, B. I.[1,2], Amano, T.[5,6], Blanco, D.[7], Fuller, R.A.[4], Johnston, A.[8,9], Langendoen, T.[10], Mundkur T.[10], Nagy, S.[10], Sutherland, W.J.[1]

[1]Conservation Science Group, Department of Zoology, University of Cambridge, Cambridge, CB2 3QZ, UK
[2]Centre for Ecology and Conservation, College of Life and Environmental Sciences, University of Exeter, Cornwall Campus, Penryn, TR10 9FE, UK
[3]School of Natural Sciences, College of Engineering and Environmental Science, Bangor University, Bangor, LL57 2UW, UK
[4]Center for Macroecology, Evolution and Climate, GLOBE Institute, University of Copenhagen, Denmark
[5]School of Biological Sciences, University of Queensland, Brisbane, 4072 Queensland, Australia
[6]Centre for Biodiversity and Conservation Science, University of Queensland, Brisbane, 4072 Queensland, Australia
[7]Wetlands International LAC Argentina Office, Capitán General Ramón Freire 1512, Buenos Aires, 1426, Argentina
[8]Cornell Lab of Ornithology, 159 Sapsucker Woods Road, Ithaca, NY 14850, US
[9]Centre for Research into Ecological and Environmental Modelling, University of St Andrews, St Andrews, KY16 9LZ, UK
[10]Wetlands International, Global Office, Horapark 9, 6717 LZ Ede, The Netherlands

Summary Paragraph

International policy is focused on increasing the proportion of the Earth's surface that is protected for nature[1,2]. While studies show that protected areas prevent habitat loss[3–6], there is a surprising lack of evidence for their impact on species' populations: existing studies are local scale or use simple designs that lack appropriate controls[7–13]. We explore how 1506 protected areas have impacted the trajectories of 27,055 waterbird populations across the globe using a robust Before-After-Control-Intervention study design, which compares protected and unprotected populations in the years before and after protection. We show that the simpler study designs typically used to assess protected area effectiveness (before-after and control-intervention) incorrectly estimate impact for 37-50% of populations, such as misclassifying positively impacted populations as negatively impacted, and vice versa. Using our robust study design, we find that protected areas have a decidedly mixed impact on waterbirds, with a strong signal that areas managed for waterbirds or their habitat are more likely to benefit populations, and a weak signal that larger areas are more beneficial than smaller ones. Calls to conserve 30% of the Earth's surface by 2030 are gathering pace[14], but we show that protection alone does not guarantee good biodiversity outcomes. As countries gather to agree the new Global Biodiversity Framework, targets must focus on creating and supporting well-managed protected and conserved areas that measurably benefit populations.

Introduction

Protected areas have been the cornerstone of conservation practice for over a century. Nearly 16% of land and 7% of the ocean are now designated as protected areas[15], and there are prominent calls for the Convention on Biological Diversity to set an area-based target of 30%

coverage from protected areas and other effective area-based conservation measures by 2030[2]. Given the importance to humanity of addressing biodiversity loss[16], it is crucial that the next decade's biodiversity conservation targets are informed by evidence of the most effective conservation strategies and actions[3,17].

Optimizing where protected areas are placed to most efficiently conserve species and their habitat has been a major research theme in conservation science for decades[18]. However, until recently, robust attempts (those making an explicit effort to account for confounding factors) to evaluate the performance of protected areas have been lacking[19,20]. A number of studies have shown that protected areas slow habitat loss, particularly in forests[3–6], however intact habitat does not guarantee the health of populations[21]. Studies attempting to address this problem by quantifying the impact of protected areas on population health and persistence have suffered from a lack of suitable controls[19]. To accurately estimate the impact of a protected area, it is necessary to understand what would have happened in the absence of protection[22] and most do this by using either Before-After or Control-Intervention study designs. Before-After studies compare populations pre- and post-protected area designation[7,13], but cannot ascertain whether the observed difference was caused by the protected area or other factors that changed in the same time period. Control-Intervention studies compare populations between protected and unprotected sites[8–12], but cannot ascertain whether the observed difference was due to the effectiveness of the protected area, or because it was placed where populations were already performing well.

Combining these designs into a Before-After-Control-Impact (BACI) framework – where populations in protected and unprotected sites are compared before and after the date of protected area designation – can overcome these limitations[23], and even approximate causality[24]. The recent emergence of large biodiversity databases in ecology provides an opportunity to test protected-area impact on populations under a BACI framework, but this has not been done.

Using one of the largest global data sets of bird population counts, compiled from citizen science initiatives and NGO- and government-led monitoring programmes in 68 countries, we present the first robust, global-scale assessment of protected area impact on populations. We examined how 1,506 protected areas have impacted the population trajectories of 27,055 waterbird populations, where 'population' is defined as a particular species at a particular site (Fig 1). Waterbirds are an appropriate taxonomic group with which to explore impact, given their broad distribution and ability to respond rapidly to changes in site quality[25]. We asked three questions: 1) How much do the study designs typically used to assess protected area effectiveness cause misleading conclusions, compared to a BACI study design?; 2) What is the impact of protected areas on waterbird populations?; and 3) What factors contribute to protected area impact?

We estimated impact using Before-After, Control-Intervention and BACI study designs. For BACI and Control-Intervention analyses, we matched protected populations to similar unprotected populations using a combination of exact matching and Mahalanobis distance matching (see Methods). We considered the wide range of ways in which populations may respond to protection by counting cases where local immigrations or extinctions had occurred; and using generalized linear models to assess both immediate changes in population numbers and longer-term changes in population trend (an extension of the traditional BACI study design that considers only average change in population size[24]). We used these measures to classify populations into three broad groups: positive, negative, or no impact from protection (see legends of Fig. 3 and Extended Data Figs. 3 & 4 for the full range of population responses and what they were classified as).

To explore the sensitivity of our results to different parameter decisions (such as years of sampling required, the maximum geographical distance between sites, or the strictness of Mahalanobis matching), we ran our entire analysis 21 times: one ‘focal analysis’ using our best guess parameter estimates, plus 20 analyses using estimates sampled from a plausible range for each parameter (‘full parameter analyses’; see methods and Extended Data Table 1).

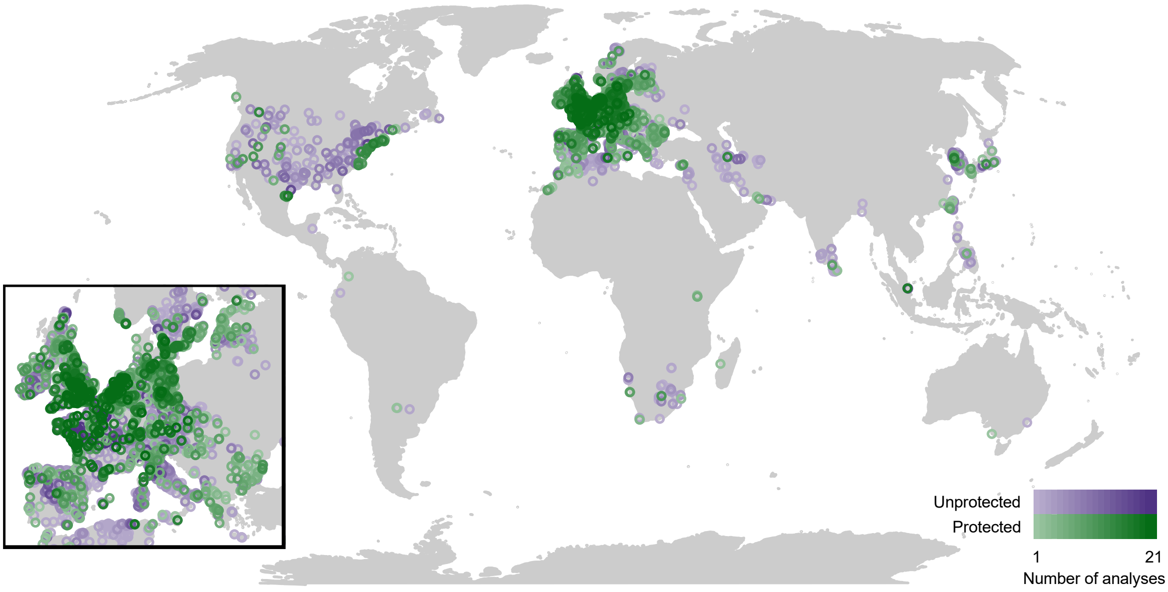


Figure 1. Map of study sites. Locations of protected (green; n=1506) and unprotected (purple; n=3343) sites used across analyses. Darker colours mean a given site was used in a greater number of analyses, to a maximum of 21 (our focal analysis and 20 full parameter analyses; there are 864 protected sites in the focal analysis). See Fig S1 for a map of just the sites used in the focal analysis.

Before-After and Control-Impact estimates

We found that estimates of protected area effectiveness varied markedly based on study design, and that studies using Before-After or Control-Intervention designs can lead to highly misleading conclusions. In our focal analysis, 37% of populations using Before-After, and 50% of populations using Control-Intervention, had different outcomes to those in the BACI analysis (Fig 2). These changes were not simply a result of BACI detecting positive or negative signals where other designs could not: 41% (Before-After) and 57% (Control-Intervention) of populations that were apparently positively impacted were shown to be not impacted, or even negatively impacted under a BACI analysis (Fig 2). Changes to negative impacts were even more striking, with 63% (Before-After) and 92% (Control-Intervention) of apparently negatively impacted populations shown to be not impacted or positively impacted by protection under a BACI analysis (Fig 2). The findings from our full parameter analyses were similar (Extended Data Fig 1). Before-After models were also heavily impacted by regression to the mean (see Supplementary Information 5), an additional reason to consider them unreliable. These results show that relying on Before-After or Control-Intervention studies can distort the picture of a protected area’s impact.

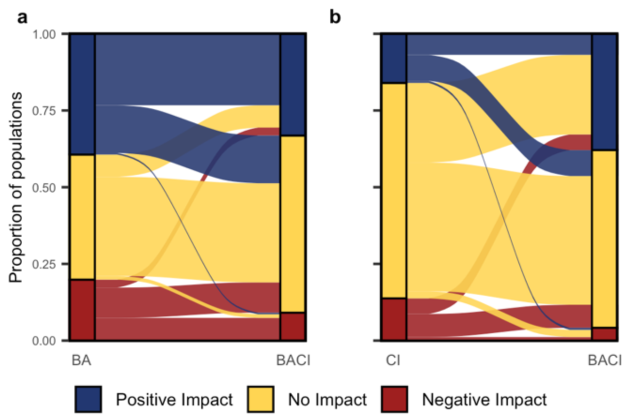


Figure 2. Changes in estimates of protected area impact under different study designs. The change in protected area effectiveness outcome when estimated under a Before-After vs BACI framework (a) or a Control-Intervention vs BACI framework (b). Y axes show proportion of populations in each category under Before-After/Control-Intervention on the left, and BACI on the right. Colour movement shows how our estimate of the impact of protected areas on populations change between study designs. Note that these figures only contain populations where we could obtain both Before-After and BACI (n=6006) or Control-Intervention and BACI (n=3609) estimates of protected area effectiveness. This figure is based on our focal analysis, see Extended Data Fig 1 for changes in outcome across all full parameter analyses.

BACI estimates of protected area impact

We found a mixed impact of protected areas on populations when using a BACI approach. Within nearly all sites, populations showed a range of responses from positive to negative (in the focal analysis the proportion of positively impacted populations within a site ranged from 0 to 1, mean = 0.25 ± 0.21 sd, Fig. 3a). Impacts on populations were similarly variable when grouped by species (in the focal analysis the proportion of positively impacted populations within a species ranged from 0 to 1, mean = 0.36 ± 0.17 sd, Fig. 3b). In our focal analysis, 27% of all populations were positively impacted by protected areas (blues), 21% were negatively impacted (reds), and for 48% we could detect no impact of protection (greys, white, yellows) (our full parameter analyses produced similar results, see Extended Data Fig. 2). Four percent of populations were excluded because of model failure. Of the 48% of cases where we could not detect any difference between protected and unprotected populations, 85% of these (41% of all populations; whites and greys, Fig. 3) were increasing, or had no trend. These cases are difficult to define as a success or failure as, while the protected area did not have a demonstrably positive impact compared to an unprotected area, the protected population appeared to be healthy.

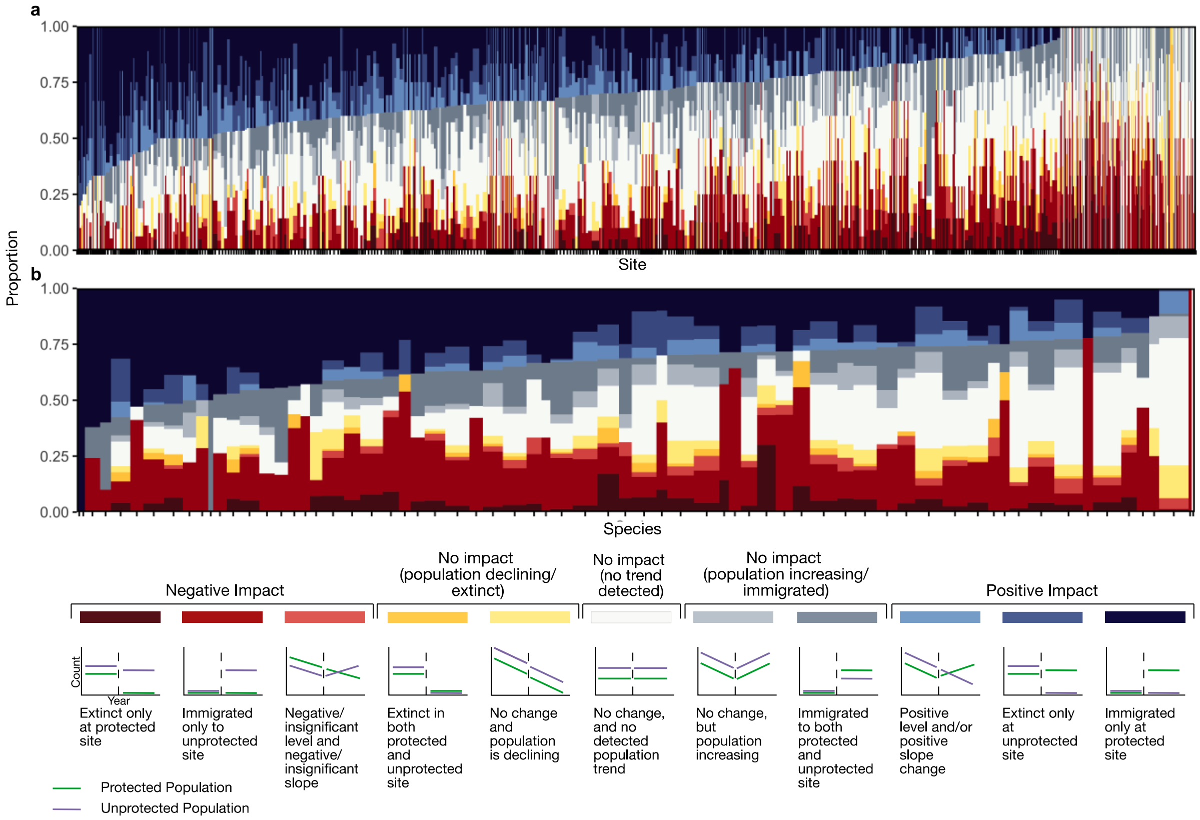


Figure 3. Estimates of protected area impact under a BACI study design. Proportion of populations (n=7313) showing various responses to protection, per site (**panel a**; n=864) and per species (**panel b**; n=67), when calculated in a BACI framework. Each species/site is one vertical bar, with the proportion of their populations in each category shown on the y axis. Bar width is scaled to the number of populations of that species/site in the dataset, log scaled in the case of species, with a wider bar meaning the species/site has more populations. Each colour represents a different way a population can respond to protection, and an example of each response is shown at the bottom. This figure is based on our focal analysis, see Extended Data Fig 3 for the proportion of populations within each broad outcome category across all full parameter analyses.

Regardless, over a quarter of populations showed a negative response (Fig. 3). These are formed from two groups: (i) negatively impacted populations i.e. those that perform worse in protected areas relative to matched controls (21%, reds) and (ii) populations for which there was no positive or negative signal of protection and which were either declining in protected areas at a similar rate to unprotected populations, or where both protected and unprotected populations went locally extinct (7%, yellows). Importantly, half of these negative responses (14% of populations overall), do not occur in sites designated for waterbirds or their habitat (i.e. Ramsar Sites[26] or Special Protected Area – Birds Directive[27] sites) and so we might not necessarily expect a positive impact in these cases and thus should not consider these to be cases where protected areas have not worked.

We consider protected area impact exclusively in the context of how protected areas support the persistence of populations, which ignores the potential benefit of protection on the maintenance of the habitats in which these populations occur. Our dataset was restricted to sites where monitoring occurred: if habitat change meant that waterbirds were no longer found at a site, monitoring would likely cease[28]. Thus, we could not consider such sites as counterfactuals, and so could not account for protected areas having prevented complete habitat conversion. We also do not consider the potential for protected areas to defend against future threats, for instance, protecting a future climate refuge. In sum, it is important to remember that the results presented here about the impact of protected areas on populations are above and beyond these already-known benefits[3–5,29,30].

Our results are also likely to underestimate the positive impact of protection as we were restricted to species for which we were able to obtain adequate matches between protected and unprotected populations, resulting in a bias towards common species (Supplementary Information 10). Common species tend to have more generalist habitat requirements[31] and so may fare better in degraded sites than rarer species. They are also less likely to be the target of specific interventions, which in some cases could actively impede them; for instance, water could be kept at levels appropriate for rare waders, but not for common ducks. To explore whether this affected our results, we assessed whether outcomes varied between regionally threatened and non-threatened species in Europe (Supplementary Information 11; a global analysis was not possible due to data restrictions). We did not find any differences in the impact of protected areas between these groups, possibly because there was only a small set of threatened species in our data, though a recent study[32] similarly found no difference between rare and common species when studying population trends.

Predictors of protected-area impact

We show that the mere designation of a protected area does not necessarily bring benefits to populations. Given this, we used cumulative link mixed models, where the response variable was the impact (positive, no or negative), to investigate which species and protected-area characteristics predict outcomes for populations, based on our BACI framework (see Fig 4). The models had random intercepts for country, site, species, and spatial grid cell. Our explanatory variables included a management variable, which broadly categorized sites as either 'waterbird-managed' (Ramsar or Special Protected Area – Birds Directive sites), or 'mixed-management' (sites either not designated for waterbirds or their habitat, or of unknown management status).

Management for waterbirds was consistently positively correlated with protected area success (Fig. 4). Larger protected areas were also almost always positively correlated with success, though significantly so in only a few analyses (Fig. 4). No other site or species-based characteristic was consistently positively or negatively associated with success (Fig. 4; Extended Data Fig. 5). Depending on the analysis, a large, waterbird-managed area could increase the likelihood of a positive impact on a population anywhere from 1 to 25 percentage points (mean weighted by model confidence = 9 percentage points; see Supplementary Information 13) compared to a small, mixed-management area.

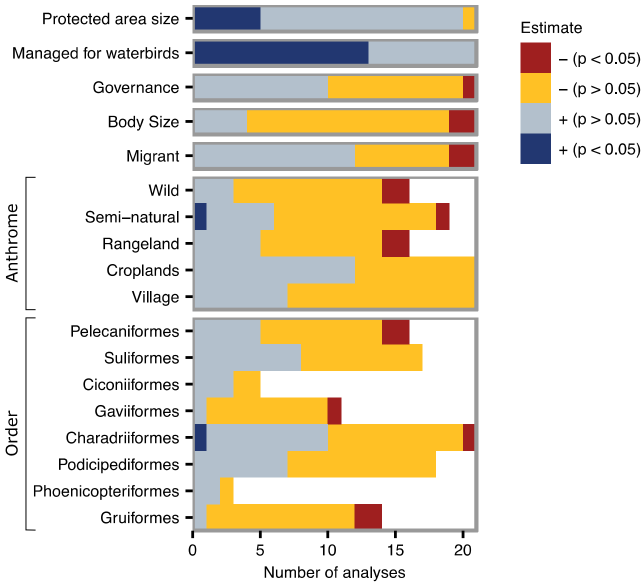


Figure 4. **Predictors of protected area impact.** Number of analyses (20 full parameter analyses plus one analysis with focal parameter estimates), that found significantly positive or negative (blue, red) or insignificantly positive or negative (grey, yellow) relationships between various predictors and protected area impact. Orders are measured relative to Anseriformes, and Anthromes relative to Urban. For odds ratios of each estimate, plus confidence intervals, see Extended Data Fig. 5.

These values are likely to underestimate the positive impact of management. Our classification of sites into waterbird-managed sites and mixed-management sites is a simple metric of diverse on-the-ground practices (a more nuanced classification is not possible at the global scale) and inevitably, some mixed-management sites are likely to be managed for waterbirds, and management quality will vary within waterbird-managed sites[33,34]. Both these factors would reduce the observed difference between the two management classifications, meaning the true difference is likely higher. That waterbird-managed sites perform better emphasizes the need for effective management to avoid negative outcomes, and suggests that policy needs to focus on setting and adhering to ambitious management targets.

The weak positive association between protected area size and impact adds a new element to the ‘Single Large or Several Small’ protected area debate that considers which is better for conserving biodiversity. Studies have agreed that several smaller protected areas typically provide higher species richness than a few large areas[35], but that larger areas are critical for persistence of larger species[36]. Our results demonstrate the importance of larger protected areas for supporting populations of waterbirds through time. This is concerning given many protected areas across the world are small and many are currently being downsized[37].

While our analysis includes data from 68 countries across 6 continents, the data are biased towards Europe, North America and East Asia; a common problem in large-scale ecological studies[38]. There are a number of initiatives in less-studied areas of the world to increase the supply and quality of ecological data[39–43]; supporting and incorporating efforts such as these will be vital to informing truly global evaluations of conservation effectiveness.

Our results show a mixed impact of protected areas, supporting concerns raised over protected area efficacy in recent years[44,45]. We had expected that, given their ability to move between sites[25], waterbirds would show a more immediate and positive signal of protection than other non-mobile taxa, such as reptiles, where positive signals might not be apparent until multiple generation cycles of improved breeding rate had occurred. The lack of signal could be due to poor or limited management of many protected areas, or it could be due to forces that cannot be controlled within the borders of a protected area. Waterbirds rely on water, and threats such as pollution, upstream dam installation and sea level rise cannot be managed by a protected area, and can have devastating consequences[46–48]. Terrestrial taxa will be less impacted by such threats and therefore may experience more positive responses to protection[49], although beyond border threats are not limited to those affecting water: climate change, air pollution and disease have the potential to impact all species[49]. Finding solutions to conserving species in the face of these more ubiquitous threats is a key conservation challenge.

Conclusions

The parties to the UN Convention on Biological Diversity will soon decide on the post-2020 Global Biodiversity Framework, which will set nature conservation policy for the decade ahead. It is likely to include a commitment to protect and conserve 30% of Earth protected by 2030 (and there are growing calls for this to reach 50% by 2050[14]). Researchers have warned that such calls must consider the social and political context in which conservation operates, or risk undermining conservation support[50]. Our results raise additional concerns about the '30 by 30' approach by showing protection alone does not guarantee optimal biodiversity outcomes. Halting biodiversity loss requires improvements to the performance of existing protected areas, and action to address ubiquitous threats beyond area borders. Ever-increasing area-based targets must be accompanied by equally ambitious targets that ensure protected area effectiveness.

Acknowledgements

We thank the coordinators, thousands of volunteer counters, and funders of the International Waterbird Census. This data collection effort is funded by the Ministry of the Environment of Japan, Environment Canada, AEWA Secretariat, EU LIFE+ NGO Operational Grant, MAVA Foundation, Swiss Federal Office for Environment and Nature, French Ministry of Environment and Sustainable Development, UK Department of Food and Rural Affairs, Norwegian Nature Directorate, Dutch Ministry of Economics, Agriculture and Innovation, DOB Ecology and Wetlands International members. CBC Data is provided by National Audubon Society and through the generous efforts of Bird Studies Canada and countless volunteers across the western hemisphere. HSW was funded by a Cambridge-Australia Poynton Scholarship, Cambridge Department of Zoology J.S. Gardiner Studentship and Cambridge Philosophical Society Grant. HSW and BIS are funded by the Royal Commission for the Exhibition 1851. WJS was funded by Arcadia, The David and Claudia Harding Foundation and MAVA. JPGJ was supported by a visiting fellowship to Fitzwilliam College Cambridge. This work was performed using resources provided by the Cambridge Service for Data Driven Discovery (CSD3) operated by the University of Cambridge Research Computing Service (www.csd3.cam.ac.uk), provided by Dell EMC and Intel using Tier-2 funding from the Engineering and Physical Sciences Research Council (capital grant EP/P020259/1), and DiRAC funding from the Science and Technology Facilities Council (www.dirac.ac.uk). Finally, the authors would like to acknowledge the use of the University of Exeter High-Performance Computing (HPC) facility in carrying out this work.

Author Contributions

H.S.W., J.P.G.J., J.G., B.I.S., T.A., R.A.F., A.J. and W.J.S. conceived the study. D.B., T.L., T.M. and S.N. provided waterbird count data which H.S.W. and T.A. collated. H.S.W. performed analysis, produced figures and wrote the text with advice from all authors, especially J.P.G.J., J.G., B.I.S., T.A., A.J. and W.J.S. All authors contributed to the review of the manuscript before submission for publication and approved the final version.

Competing Interests

The authors declare no competing interests.

Code availability

The code used to produce all analysis and figures are archived on Zenodo, doi:10.5281/zenodo.5794483. Code are also available on GitHub https://github.com/hannahwauchope/PAImpact , this is the recommended mode of access as it will contain any updates or clarifications.

Data Availability

The waterbird count data used in this study are collated and managed by Wetlands International and the National Audubon Society, and are available on request (http://iwc.wetlands.org/index.php/ ; http://netapp.audubon.org/cbcobservation/ respectively). All the data that pertain to explanatory variables are freely available, as specified in Extended Data Tables 1 and 2.

Methods

We published a pre-analysis plan for this paper laying out our planned analysis before we looked in detail at the data[51]. Pre-analysis plans are useful to reduce the risk of cherry picking or HARKing (hypothesizing after results are know) which has led to a replication crisis in science[52]. As much as possible, we have followed the methods we set out, however we discovered a number of factors we had not considered (for instance, the potential for immigrations and extinctions and the fact that both trend and immediate change must be considered, see[24]). The conceptual basis of our revised methodology is described in detail in Wauchope et al[24], and Supplementary Information 7 describes the choices we've made that deviate from the pre-analysis plan and why.

Overview

A brief summary of our workflow is as follows: we took yearly counts of 749 waterbird species at 45,745 sites across the world from the International Water Census and Christmas Bird Count. Of these, we wanted to find populations, here defined as a certain species at a certain site, that occurred in a protected area *and* where yearly counts had begun before the protected area was designated. For our Before-After (hereafter BA) analysis, we then assessed how each population at each of those sites changed from before to after the protected area was designated. For Control-Intervention (hereafter CI) and BACI analysis, we matched each of these protected populations to unprotected populations surveyed over the same period, that were similar based on a number of site and species characteristics. For CI, we compared populations in the years after the protected area was designated between unprotected and protected population pairs. For BACI, we compared change in protected populations from before to after protected area designation, and then compared this to the before-after change in matched unprotected populations over the same period.

Whether BA, CI or BACI, we then classified the impact to the population as positive, negative or no impact from protection. Next, we looked to see whether our conclusions about impact varied when we analysed a population in a BA, CI or BACI framework. We found BA and CI analyses to be unreliable, so discarded them at this point. Finally, we looked to see whether there were correlates that predicted protected area impact have, by running cumulative link models on BACI data. These correlated outcome (Positive, Negative or No impact) to a range of site and species level predictors such as protected area size, species body size, land use type and whether the site was managed for waterbirds. Finally, we ran sensitivity tests varying a range of parameters that were used to make analytical decisions to test the robustness of conclusions.

All analysis was completed using R v4.0.3[53] and QGIS v3.10[54], data figures and base maps were produced using the R package ggplot2[55], impact legends were produced using Inkscape[56].

Time Series Preparation

We took site-specific annual counts from two long term surveys: the International Waterbird Census (IWC), coordinated by Wetlands International, and the Christmas Bird Count (CBC), run by the National Audubon Society. We used Wetland International's definition of "waterbird", and took any species from the corresponding families (list of families in Supplementary Information 2). Our initial dataset consisted of 749 species at 45,475 sites, spanning 1940 to 2018. We then restricted our data to only sites surveyed in December to February. We imputed zeroes, by taking any site where a species has been observed, and recording any years where the species was not mentioned as '0' years.

As CBC data is not standardized for effort, we required that these species showed a log-linear relationship with effort (i.e. the rate of new individuals detected in a search slows with increased effort). For each species, we ran a simple negative binomial generalized linear model in R, using the glm.nb function from package MASS [57], using all available CBC data for that species:

$$\log\left(E(Count_i)\right) = \beta \log(h_i) \tag{1}$$

Where *Count* is all counts of a species and h_i is the number of survey hours for each count. We retained CBC data for all species where there was a significant positive relationship between count and effort.

Protected (and Unprotected) Area Data

We first created a dataset of counts at protected sites. We took our protected area data from the World Database on Protected Areas ('Protected Planet')[58], downloading the full dataset of all protected areas globally, and overlaying our sites to determine which fell in protected areas. Some coastal site coordinates fell just outside the land cover layer that protected areas are aligned to, so we snapped all sites to the base terrestrial layer used by Protected Planet[59], but by no more than 10km. We removed any sites where the designation status was proposed, and any UNESCO biosphere reserves as these are often not afforded formal protection[60]. We next removed any sites where there was no information about designation date. In some cases, there were multiple protected area data entries for a site, in these cases we took the earliest designation year given. Finally, we reduced the count dataset to only the 10 years before and after the designation date of whichever protected area the survey site fell within, requiring that at least 7 years before and after were surveyed (we tested the number of years restricted from 5-15 years, and number of years measured from 4-13; Extended Data Table 1a and b, respectively).

We next created a dataset of counts at unprotected sites for CI and BACI analysis. For Christmas Bird Count data, surveying is conducted in a circle with a radius of 12.07km. If there is a protected site in this circle, we cannot be sure that the counts are not being biased by protection. Therefore, we only counted sites as unprotected if no protected area occurred in the entire circle. For IWC data, we included sites that were at least 1km from a protected area, to avoid any confounding of results from spill-over effects[61] (we sensitivity tested this threshold from 500m to 5km; Extended Data Table 1c). We consider sites to be unprotected until the point in time when a protected area was designated at that site. For instance, a site, A, could be designated as a protected area in the year 2000, but this would mean that counts before this point, say, from the 1980s, would be of waterbirds at a site not experiencing any benefit of protection. We could therefore match a protected site from the 1980s to Site A's counts in the 1980s, and treat A's counts as unprotected at this time.

BA, CI and BACI Datasets

In all cases, we defined the "after" period as being the years after, but not including, the designation date of the PA. We also defined cases of 'all zeros' to account for local immigrations and extinctions. Waterbirds are highly mobile and can quickly immigrate to, or emigrate from a site. In these cases we cannot assess a change in trend between, for instance, a before period where there are individuals absent and an after period when they have immigrated to the site (for a detailed explanation of why immigrations and extinctions pose a problem for trend analysis, see[24]). Theoretically, we should only consider cases with only zero counts in a before or after period as 'all zero' local immigrations or extinctions, but because waterbirds are able to appear as vagrants at a site, we chose to classify cases where at least 70% of years were zero counts as all zeroes. We tested this threshold from 60 – 80% (Extended Data Table 1d). It's important to note that any sites where the species had never occurred would not be included in the dataset, so even in cases of all zeroes the species is known to be able to occur at the site.

To create the BA dataset, we took all protected populations where there were cases of counts (as opposed to all zeroes) in either the before period, after period or both. We subset the BA dataset to only protected populations that also occur in the BACI dataset.

To create the CI dataset, we took all protected populations with counts (as opposed to all zeroes) in the after period, and matched these to unprotected populations also with counts over the same time period (see matching below). We subset the CI dataset to only protected populations that also occur in the BACI dataset.

To create the BACI dataset, we matched protected and unprotected populations, requiring that at least one period (either protected before, protected after, unprotected before, or unprotected after) had counts (as opposed to all zeroes).

Matching

<u>Data preparation</u>

We developed a statistical matching method to achieve matching of BACI and CI analyses. The covariates we used for matching, how we prepared them and justification for their use are given in Extended Data Table 2, broadly they encompass variables related to climate, land use and human impact. We removed highly correlated variables by first calculating the variance inflation factor (using the VIF function from the usdm package in R[62]) of all covariates, and iteratively removing variables with a VIF greater than four until none were over four[63]. We next removed variables with a Pearson's Correlation Coefficient of over 0.7.

For BACI, we matched only on covariates in the years prior to designation (as protected and unprotected sites might be expected to differ in the years after protected area designation, especially on covariates related to human impact). For CI, we matched on covariates only in the years after designation, as we choose to be blind to the 'before' period in this analysis.

We then proceeded with matching, separately for each species. The following describes the procedure for one species.

<u>Mahalanobis Distances</u>

We used Mahalanobis distance matching to evaluate how similar protected and unprotected sites were. Though Mahalanobis distance has been criticized in the past for performing poorly when matching on many covariates[64,65], recent criticisms of the most commonly used matching method, Propensity Score Matching[66], meant we were interested to test other options and found Mahalanobis distance matching to perform markedly better in comparisons (Supplementary Information 9).

Mahalanobis Distance (md) computes the distance between points in multivariate space. The Mahalanobis distance between two sets of points is calculated as follows:

$$md_{(x,y)} = \sqrt{(\boldsymbol{x}-\boldsymbol{y})^T\boldsymbol{S}^{-1}(\boldsymbol{x}-\boldsymbol{y})} \tag{2}$$

Where **x** and **y** are vectors containing values for each covariate (in our case, therefore, the list of covariate values for sites x & y) and **S** is the covariance matrix of the covariates.

This formula requires each site to have one value for each covariate, so we took means of the values for the years pre- (BACI) or post- (CI) designation.

For each species, we created a large matrix with protected sites in columns and unprotected sites in rows, with Mahalanobis distance values populating the rows. Because we wanted to match exactly on the years only prior to protected area designation, we first created separate matrices (using function mahal from R package DOS[67]), each containing only protected areas designated in a certain year (See Extended Data Fig. 6a, b for an example). Mahalanobis distance requires at least two protected sites to work (to be able to calculate the covariance matrix), and so we could not build Mahalanobis distance matrices for years where only one protected area in our dataset was designated. This resulted in a minimal loss of sites.

These Mahalanobis distance matrices were then combined into the larger distance matrix containing all the sites across all designation years that the species occurred in (Extended Data Fig. 6c).

<u>Exact Matching</u>

We required that sites were exactly matched on a number of criteria, where sites failed they were excluded from the Mahalanobis distance matrix (Extended Data Fig. 6d). For each protected site, we removed unprotected sites not of the same anthrome category, continent, and migratory status. We also removed any sites greater than 500km from the protected area (we tested this value from 100km to 2500km; Extended Data Table 1e).

For BACI analysis, we needed to satisfy the parallel trends assumption[24,68], which specifies that the trends of control and intervention populations in the 'before' period must be parallel. To test this, we modelled the difference in trends between each protected and potential unprotected matched site. We used a negative binomial glm (glm.nb, R package MASS[57]),:

$$\log\left(E\left(Count_{i,j}\right)\right) = \alpha + \beta_1 Y_i + \beta_2 CI_j + \beta_3 Y_i CI_j + \text{offset}(\log(h_i)) + \epsilon \qquad (3)$$

Where the count of the population in year *i* at site *j* is predicted by the Year (*Y*), a binary term that is 1 for the protected site and zero for the unprotected site (*CI*) and the interaction between the two. Log of effort is included as an offset for CBC data (effort is held at 1 for IWC data). We also checked for temporal autocorrelation and adjusted the model if it was present (see "Temporal Autocorrelation" below). If the interaction coefficient (β_3) was significant ($p<0.05$), then there was a significant difference between the trend of the two populations, and the unprotected population was discarded.

If no unprotected sites met the exact match criteria, the protected site did not have a match and was excluded (e.g. Extended Data Fig. 6d, Site E).

Picking Matches

Next, we ran an optimized greedy nearest-neighbour algorithm to select, from the Mahalanobis distance matrix (with any sites not satisfying exact match criteria excluded), the unprotected site with the smallest Mahalanobis distance. We ran this without replacement, meaning each protected site could be matched to only one unprotected site, to ensure no pseudoreplication. A greedy algorithm works through the dataset, picking the best match for each successive protected site and removing the matched unprotected site from the potential matching pool as it goes. However, greedy algorithms have a tendency to get stuck in local optima[69], so to account for this, we ran the greedy algorithm 1000 times, each time randomizing the order of protected sites that the greedy algorithm would work through. We found the global distance for each iteration and used the set with the smallest global distance (Extended Data Fig. 6e, e.g. with randomisations in the figure a smaller global distance would be detected).

Evaluating Match Quality

Once we had our matched sets for each species, we needed to ensure that the matches were of a high enough quality to be used. This was done by assessing the covariate balance between matched and unmatched sites for each species using the 'standardised difference in means' (SDiM), which is calculated using the following formula[70]:

$$d_{cov} = \frac{\overline{\boldsymbol{T}}_{cov} - \overline{\boldsymbol{C}}_{cov}}{\sqrt{\frac{var(\boldsymbol{T}_{cov}) - var(\boldsymbol{C}_{cov})}{2}}} \qquad (4)$$

Where $\mathbf{T}_{cov}$ is the values of covariate *cov* for protected sites (mean from the years before and equal to designation), $\mathbf{C}_{cov}$ is the same for unprotected sites, *var* is the variance of each of these and d_{cov} is the standardized mean difference between protected and unprotected sites. We assessed the SDiMs to see whether they were below 0.25 for all covariates[65,71] (we sensitivity tested this threshold from 0.1 to 0.25; Extended Data Table 1f). If they were not, the matched pair with the greatest distance was removed and the SDiM checked again. Once all covariates had a SDiM of <0.25 (or the relevant sensitivity value), the remaining matched pairs were considered the 'final' matched dataset for that species (Extended Data Fig. 6f). If less than 80% of the sites that a species occurred in were remaining, we discarded the species, to ensure that the matched set was not biased to a certain subset of all sites for that species (we sensitivity tested this value from 50-90%; Extended Data Table 1g).

Assessing Protected Area Impact

Following the framework set out by[24], we defined a number of ways that a population could respond to protection. Broadly, populations can respond to a protected area by immigrating to the area, going locally extinct from the area, showing a change in trend, or by showing an immediate change, i.e. an immediate increase or decrease in the number of individuals (See legends of Fig 3, Extended Data Figs 3 and 4).

For comparing BA, a population could show an immediate change or change in trend, or the population could immigrate to the site or go locally extinct at the site (Extended Data Fig. 3). For comparing BACI, the BA changes were compared between protected and unprotected sites. For example, a population could be stable in the period before protection, and declining in the period after – this would be a negative BA trend change (Extended Data Fig. 3). But if a matched unprotected population was also stable in the before period, but declining at a *faster* rate in the after period, then the BACI trend change would be positive (Fig. 3), as the protected area had slowed the decline of the protected species, even if it hadn't halted it. If the unprotected population was declining at a similar rate to the protected population in the after period, this would be a case of no impact under a BACI framework (Fig. 3). For comparing CI, only the difference in trend between protected and unprotected populations was considered (Extended Data Fig. 4).

All BA, CI or BACI time periods with all zeroes were categorised as immigrations or extinctions, for instance, in BACI analysis if protected population had no counts in the before period, but did in the after period, while the matched unprotected site had no counts in the before and after period, this would be classified as a local immigration (and a positive impact of the protected area).

For time periods with all counts we ran the following models. In all cases *Y* represents the year, centred around the year of protected area designation so that year of designation equals zero. *BA* is a binary term that is 0 in the years before protected area designation, and 1 in the years after; note that this isn't included in the CI model as only 'after' years are used. *CI* is a binary term that is 0 for the unprotected population and 1 for the protected population; note that the *CI* term isn't included in the BA model as this model does not include unprotected populations. Finally, each model includes an offset term for effort (h), to account for variable effort in CBC data. For IWC data, effort is always set to 1 and so does not contribute to the model. All models were negative binomial glms, run using R package MASS[57].

<u>BA</u>

$$\log(E(Count_i)) \sim \beta_0 + \beta_1 BA_i + \beta_2 Y_i + \beta_3 BA_i Y_i + \text{offset}(\log(h_i)) + \epsilon \qquad (5)$$

β_1 gives the immediate change and β_3 gives the trend change[24].

CI

$$\log(E(Count_i)) \sim \beta_0 + \beta_1 CI_i + \beta_2 Y_i + \beta_3 CI_i Y_i + \text{offset}(\log(h_i)) + \epsilon \tag{6}$$

β_3 gives the difference in trend between protected and unprotected sites.

BACI

$$\log(E(Count_{i,j})) \sim \beta_0 + \beta_1 BA_i + \beta_2 CI_j + \beta_3 T_i + \beta_4 BA_i CI_j + \beta_5 BA_i Y_i + \beta_6 CI_j Y_i + \beta_7 BA_i CI_j Y_i + \text{offset}(\log(h_i)) + \epsilon \tag{7}$$

B_4 gives the immediate change and β_7 gives the trend change[24]. We excluded any cases where β_6 was significant as this indicates a significant difference between protected and unprotected trends in the before period, meaning the parallel trends assumption is not satisfied. Though we checked for this in matching, running a full model containing 'after' data as well (cf only before data, as in matching) meant that very occasionally this term became significant, presumably because of an increase in power.

In a small proportion of populations, models failed to converge. In these cases, we removed the population from analysis.

Temporal Autocorrelation

Time series data are vulnerable to the effects of temporal autocorrelation, where counts in one year are impacted by counts in the years before, and as a result are not independent, as models assume. Being mobile, we expect less temporal autocorrelation in waterbird data than for sessile species (waterbird population numbers can change markedly at a site year to year), but nevertheless we checked for, and accounted for, temporal autocorrelation in our data. For each population model (whether BA, CI or BACI; and also for the models used to check for parallel trends in the matching stage), we checked for temporal autocorrelation using three implementations of the Durbin-Watson test in R: durbinWatsontest from package car[72], testTemporalAutocorrelation from package DHARMa[73], and dwtest from package lmtest[74]. Though each of these implementations performs the same test, variations in methodology meant we found some population models had significant temporal autocorrelation under one, but not another. To be conservative, we decided that if a population had significant autocorrelation under any of the three tests, we considered there to be temporal autocorrelation. If this was the case, we re-ran the population model as a negative binomial generalised linear mixed model (using glmer.nb from package lme4[75]) including a random intercept for Year for BA analyses, and Site:Year for CI and BACI analyses, to account for the autocorrelation.

Classifying Outcomes

We then classified outcomes. We aimed to be generous for assigned positive outcomes, and so for BA and BACI, a significantly ($p<0.05$) positive immediate or trend change (even if the other was significantly negative) meant that the protected area was classed as having had a positive impact on the population. If both immediate and trend were insignificant, then the protected area had had no impact. And if either was negative and the other insignificant, or if both were significantly negative, the protected area was classed as having had a negative impact. We conducted a supplementary analysis to see whether relaxing this p-value would result in detecting more positive impacts, see Supplementary Information 12.

For CI, a significantly positive difference between protected and unprotected trends was classed as a positive impact, significantly negative was a negative impact, and an insignificant difference no impact.

Drivers of change

To explore the predictors of protected area effectiveness, we considered body mass, species migratory status, taxonomic order, the broad anthrome category (i.e. land use type) of the protected area, protected area size, and country governance[11]. See Extended Data Table 3 for details of how each covariate was obtained.

To test how these covariates might correlate to protected area effectiveness, we ran cumulative link mixed effects models that allow for ordinal predictors and random factors, with the response term being a three-level factor: negative impact, no impact, or positive impact. To account for spatial autocorrelation, we included a random intercept for "grid cell", with sites each assigned to a gridcell of size 2* Max distance between protected and unprotected sites (Table 1e). In this way errors are grouped by sites that are closer together. In some of the 21 analyses, typically those with smaller sample sizes, including both country and grid cell as random factors meant the model could not converge; in these cases we retained only country as a random factor. We used the clmm function from R package 'Ordinal'[76]. The model specification was as follows:

$$\begin{aligned} Impact_{i,j,k} \sim\ & \beta_1 MigatoryStatus_i + \beta_2 \log{(BodyMass_)}_i + \beta_3 Order_i + \beta_4 Anthrome_j \\ & + \beta_5 RamsarSPA_j + \beta_6 \log(PA\ Area)_j + \beta_7 MeanGovernance_k + (1|i) \\ & + (1|k) + (1|k{:}j) + (1|m) + (1|m{:}j)\ \epsilon \end{aligned}$$

Where *i, j*, *k* and *m* are species, site, country and gridcell, respectively. In some sensitivity tests some covariates did not have sufficient populations to be able to test them, in these cases certain levels of the covariate were removed (e.g. if there were not enough populations of a particular taxonomic order) or in some cases the entire covariate was removed. Not all protected areas have area data reported, and so we had to run models only on the subset of data where area data was available. To ensure this reduced set was not misrepresenting the full dataset, we also ran models without the protected area Area covariate and on the full dataset; results were broadly similar (Supplementary Information 8), and in the case of BACI, waterbird-managed sites were more strongly positively associated with outcomes.

We estimated the effect size of management and protected area size using the function ggpredict from R package ggeffects[77], which returns odds ratios from the cumulative link mixed models. We estimated effect size for water-bird managed vs mixed-managed sites, and for 5 quintiles of log(protected area size): 0.05, 0.25, 0.5, 0.75 and 0.95. For the effect size reported in the manuscript, we compared the chance of a positive impact on a population in a mixed-management site in the 0.05th size quintile to the chance of a positive impact on a population in a waterbird-managed site in the 0.95th quintile.

Finally, some covariates violated the proportional odds ratio assumption upon which cumulative link models rest. To check for the impact of this we ran individual binomial generalized linear mixed-effects models (using function glmer from R package lme4[75]) to conduct pairwise comparisons of outcome levels. These models supported the general conclusions made in this paper (see Supplementary Information 13 for further details).

Full Parameter Analyses

The focal analysis inevitably is based on somewhat arbitrary modelling choices. We therefore ran our models an additional 20 times with a range of parameter values for decisions such as: the number years of counts required before and after protection, the threshold at which we classify All Zeroes, the maximum distance between protected and unprotected sites for an acceptable match and how similar we required matched sites to be (Extended Data Table 1). Testing all parameter combinations was computationally impractical so we used a latin hypercube sampling method[78]. This is a way to adequately sample a high dimensional parameter space when random sampling is prohibitively inefficient; it creates multiple combinations of covariates that together evenly sample the entire n dimensional sample space. We randomly created 20 parameter combinations (using function randomLHS from the R package 'lhs'[79]), which are displayed in Extended Data Table 1. We call these analyses our 'full parameter' analyses.

References

1. High Ambition Coalition for Nature and People. 50 Countries Announce Bold Commitment to Protect at Least 30% of the World's Land and Ocean by 2030. *Campaign for Nature* (2021).
2. Waldron A, Adams V, Allan J, Arnell A, Asner G, Atkinson S, Baccini A, Baillie J, et al. *Protecting 30% of the planet for nature: costs, benefits and economic implications*. (2020).
3. Geldmann, J., Manica, A., Burgess, N. D., Coad, L. & Balmford, A. A global-level assessment of the effectiveness of protected areas at resisting anthropogenic pressures. *Proceedings of the National Academy of Sciences* **116**, 23209 LP – 23215 (2019).
4. Nelson, A. & Chomitz, K. M. *Protected area effectiveness in reducing tropical deforestation*. (2009).
5. Scharlemann, J. P. W. *et al.* Securing tropical forest carbon: the contribution of protected areas to REDD. *Oryx* **44**, 352–357 (2010).
6. Feng, Y. *et al.* Assessing the effectiveness of global protected areas based on the difference in differences model. *Ecological Indicators* **130**, 108078 (2021).
7. Laurance, W. F. *et al.* The fate of Amazonian forest fragments: A 32-year investigation. *Biological Conservation* **144**, 56–67 (2011).
8. Laurance, W. F. *et al.* Averting biodiversity collapse in tropical forest protected areas. *Nature* **489**, 290 (2012).
9. Terraube, J., Van doninck, J., Helle, P. & Cabeza, M. Assessing the effectiveness of a national protected area network for carnivore conservation. *Nature Communications* **11**, 2957 (2020).
10. Barnes, M. D. *et al.* Wildlife population trends in protected areas predicted by national socio-economic metrics and body size. *Nature Communications* **7**, 12747 (2016).
11. Amano, T. *et al.* Successful conservation of global waterbird populations depends on effective governance. *Nature* **553**, 199 (2018).
12. Kleijn, D., Cherkaoui, I., Goedhart, P. W., van der Hout, J. & Lammertsma, D. Waterbirds increase more rapidly in Ramsar-designated wetlands than in unprotected wetlands. *Journal of Applied Ecology* **51**, 289–298 (2014).

13. Reyes-Arriagada, R. *et al.* Population Trends of a Mixed-Species Colony of Humboldt and Magellanic Penguins in Southern Chile after Establishing a Protected Area. *Avian Conservation and Ecology* **8**, 13 (2013).
14. Bukart, K. Motion 101 passes at IUCN, calls for protecting 50% of Earth's lands and seas. *One Earth* https://www.oneearth.org/motion-101-passes-at-iucn-calls-for-protecting-50-of-earths-lands-and-seas/ (2021).
15. UNEP-WCMC and IUCN. *Protected Planet Report 2020*. (2021).
16. IPBES. *Global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services.* (IPBES secretariat, 2019).
17. Barnes, M. D., Glew, L., Wyborn, C. & Craigie, I. D. Prevent perverse outcomes from global protected area policy. *Nature Ecology & Evolution* **2**, 759–762 (2018).
18. Pressey, R. L., Cabeza, M., Watts, M. E., Cowling, R. M. & Wilson, K. A. Conservation planning in a changing world. *Trends in Ecology & Evolution* **22**, 583–592 (2007).
19. Geldmann, J. *et al.* Effectiveness of terrestrial protected areas in reducing habitat loss and population declines. *Biological Conservation* **161**, 230–238 (2013).
20. Rodrigues, A. S. L. & Cazalis, V. The multifaceted challenge of evaluating protected area effectiveness. *Nature Communications* **11**, 5147 (2020).
21. Redford, K. H. The Empty Forest. *BioScience* **42**, 412–422 (1992).
22. Ferraro, P. J. Counterfactual thinking and impact evaluation in environmental policy. *New Directions for Evaluation* **2009**, 75–84 (2009).
23. Adams, V. M., Barnes, M. & Pressey, R. L. Shortfalls in Conservation Evidence: Moving from Ecological Effects of Interventions to Policy Evaluation. *One Earth* **1**, 62–75 (2019).
24. Wauchope, H. S. *et al.* Evaluating Impact Using Time-Series Data. *Trends in Ecology & Evolution* **36**, 196–205 (2021).
25. Kingsford, R. T., Roshier, D. A. & Porter, J. L. Australian waterbirds time and space travellers in dynamic desert landscapes. *Marine and Freshwater Research* **61**, 875–884 (2010).
26. The Ramsar Convention Secretariat. Managing Ramsar Sites. *ramsar.org* https://www.ramsar.org/sites-countries/managing-ramsar-sites (2014).
27. European Commission. The Birds Directive. *EU Nature Law* https://ec.europa.eu/environment/nature/legislation/birdsdirective/index_en.htm.
28. Zhang, W., Sheldon, B. C., Grenyer, R. & Gaston, K. J. Habitat change and biased sampling influence estimation of diversity trends. *Current Biology* **31**, 3656-3662.e3 (2021).
29. Bruner, A. G., Gullison, R. E., Rice, R. E. & da Fonseca, G. A. B. Effectiveness of parks in protecting tropical biodiversity. *Science* **291**, 125–128 (2001).
30. Carranza, T., Balmford, A., Kapos, V. & Manica, A. Protected Area Effectiveness in Reducing Conversion in a Rapidly Vanishing Ecosystem: The Brazilian Cerrado. *Conservation Letters* **7**, 216–223 (2014).
31. Rabinowitz, D. Seven forms of rarity. in *The Biological Aspects of Rare Plant Conservation* (ed. Synge, H.) 205–217 (John Wiley & Sons, Ltd, 1981).
32. Daskalova, G. N., Myers-Smith, I. H. & Godlee, J. L. Rare and common vertebrates span a wide spectrum of population trends. *Nature Communications* **11**, 4394 (2020).
33. Hettiarachchi, M., Morrison, T. H. & McAlpine, C. Forty-three years of Ramsar and urban wetlands. *Global Environmental Change* **32**, 57–66 (2015).

34. Munishi, P., Chuwa, J., Kilungu, H., Moe, S. & Temu, R. Management effectiveness and conservation initiatives in the Kilombero Valley Flood Plains Ramsar Site, Tanzania. *Tanzania Journal of Forestry and Nature Conservation* **81**, 1–10 (2012).
35. Fahrig, L. Why do several small patches hold more species than few large patches? *Global Ecology and Biogeography* **29**, 615–628 (2020).
36. Newmark, W. D. Extinction of Mammal Populations in Western North American National Parks. *Conservation Biology* **9**, 512–526 (1995).
37. Mascia, M. B. & Pailler, S. Protected area downgrading, downsizing, and degazettement (PADDD) and its conservation implications. *Conservation Letters* **4**, 9–20 (2011).
38. Di Marco, M. *et al.* Changing trends and persisting biases in three decades of conservation science. *Global Ecology and Conservation* **10**, 32–42 (2017).
39. African Conservation Foundation. African Conservation Foundation. https://africanconservation.org (2020).
40. Aves Argentina. Aves Argentina. https://www.avesargentinas.org.ar.
41. Birds Caribbean. Birds Caribbean. https://www.birdscaribbean.org.
42. The Amazon Conservation Team. The Amazon Conservation Team. https://www.amazonteam.org/brazil/ (2021).
43. Wetlands International. Asian Waterbird Census. https://south-asia.wetlands.org/our-approach/healthy-wetland-nature/asian-waterbird-census/.
44. Gill, D. A. *et al.* Capacity shortfalls hinder the performance of marine protected areas globally. *Nature* **543**, 665 (2017).
45. Geldmann, J. *et al.* A global analysis of management capacity and ecological outcomes in terrestrial protected areas. *Conservation Letters* **11**, e12434 (2018).
46. Kingsford, R. T., Bino, G. & Porter, J. L. Continental impacts of water development on waterbirds, contrasting two Australian river basins: Global implications for sustainable water use. *Global Change Biology* **23**, 4958–4969 (2017).
47. Jia, Q., Wang, X., Zhang, Y., Cao, L. & Fox, A. D. Drivers of waterbird communities and their declines on Yangtze River floodplain lakes. *Biological Conservation* **218**, 240–246 (2018).
48. Lehikoinen, A., Rintala, J., Lammi, E. & Pöysä, H. Habitat-specific population trajectories in boreal waterbirds: Alarming trends and bioindicators for wetlands. *Animal Conservation* **19**, 88–95 (2016).
49. Boyd, C. *et al.* Spatial scale and the conservation of threatened species. *Conservation Letters* **1**, 37–43 (2008).
50. Schleicher, J. *et al.* Protecting half of the planet could directly affect over one billion people. *Nature Sustainability* **2**, 1094–1096 (2019).
51. Wauchope, H. *et al.* Quantifying the impact of protected areas on near-global waterbird population trends, a pre-analysis plan. *PeerJ Preprints* **7, e27741v**, (2019).
52. Nosek, B. A., Ebersole, C. R., DeHaven, A. C. & Mellor, D. T. The preregistration revolution. *Proceedings of the National Academy of Sciences* **115**, 2600–2606 (2018).
53. R Core Team. R: A Language and Environment for Statistical Computing. (2020).
54. QGIS.org. QGIS Geographic Information System. (2021).
55. Hadley Wickham. *ggplot2: Elegant Graphics for Data Analysis*. (Springer-Verlag, 2016).
56. Inkscape Project. Inkscape. (2020).
57. Venables, W. N. & Ripley, B. D. *Modern Applied Statistics with S*. (Springer, 2002).
58. UNEP-WCMC & IUCN. The World Database on Protected Areas (WDPA)/The Global Database on Protected Areas Management Effectiveness (GD-PAME). www.protectedplanet.net (2019).

59. NOAA. Global Self-consistent, Hierarchical, High-resolution Geography Database (GSHHG). (2017).
60. Coetzer, K. L., Witkowski, E. T. F. & Erasmus, B. F. N. Reviewing Biosphere Reserves globally: effective conservation action or bureaucratic label? *Biological Reviews* **89**, 82–104 (2014).
61. Ament, J. M. & Cumming, G. S. Scale dependency in effectiveness, isolation, and social-ecological spillover of protected areas. *Conservation Biology* **30**, 846–855 (2016).
62. Naimi, B., Hamm, N. A. S., Groen, T. A., Skidmore, A. K. & Toxopeus, A. G. Where is positional uncertainty a problem for species distribution modelling? *Ecography* **37**, 191–203 (2014).
63. Salmerón Gómez, R., García Pérez, J., López Martín, M. D. M. & García, C. G. Collinearity diagnostic applied in ridge estimation through the variance inflation factor. *Journal of Applied Statistics* **43**, 1831–1849 (2016).
64. Gu, X. S. & Rosenbaum, P. R. Comparison of Multivariate Matching Methods: Structures, Distances, and Algorithms. *Journal of Computational and Graphical Statistics* **2**, 405–420 (1993).
65. Stuart, E. A. Matching methods for causal inference: A review and a look forward. *Statistical science : a review journal of the Institute of Mathematical Statistics* **25**, 1–21 (2010).
66. King, G. & Nielsen, R. Why Propensity Scores should not be used for matching. *Political Analysis* **27**, 435–454 (2019).
67. Rosenbaum, P. R. DOS: Design of Observational Studies. (2018).
68. Linden, A. A matching framework to improve causal inference in interrupted time-series analysis. *Journal of Evaluation in Clinical Practice* **24**, 408–415 (2018).
69. Simmons, B. I., Hoeppke, C. & Sutherland, W. J. Beware greedy algorithms. *Journal of Animal Ecology* **88**, 804–807 (2019).
70. Austin, P. C. Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. *Statistics in medicine* **28**, 3083–3107 (2009).
71. Rubin, D. B. Using Propensity Scores to Help Design Observational Studies: Application to the Tobacco Litigation. *Health Services and Outcomes Research Methodology* **2**, 169–188 (2001).
72. Fox, J. & Weisberg, S. *An R Companion to Applied Regression*. (Sage, 2019).
73. Hartig, F. DHARMa: Residual Diagnostics for Hierarchical (Multi-Level / Mixed) Regression Models. (2021).
74. Zeileis, A. & Hothorn, T. Diagnostic Checking in Regression Relationships. *R News* **2**, 7–10 (2002).
75. Bates, D., Maechler, M., Bolker, B. & Walker, S. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* **67**, 1:48 (2015).
76. Christensen, R. ordinal—Regression Models for Ordinal Data. (2019).
77. Lüdecke, D. ggeffects: Tidy Data Frames of Marginal Effects from Regression Models. *The Journal of Open Source Software* **3**, 772 (2018).
78. McKay, M. D., Beckman, R. J. & Conover, W. J. Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code. *Technometrics* **21**, 239–245 (1979).
79. Carnell, R. lhs: Latin Hypercube Samples. (2020).
80. Harris, I., Jones, P. D., Osborn, T. J. & Lister, D. H. Updated high-resolution grids of monthly climatic observations - the CRU TS3.10 Dataset. *International Journal of Climatology* **34**, 623–642 (2014).

81. Lu, C. & Tian, H. Global nitrogen and phosphorus fertilizer use for agriculture production in the past half century: Shifted hot spots and nutrient imbalance. *Earth System Science Data* **9**, 181–192 (2017).
82. Joppa, L. N. & Pfaff, A. High and Far: Biases in the Location of Protected Areas. *PLOS ONE* **4**, e8273 (2009).
83. Hurtt, G. C. *et al.* Harmonization of land-use scenarios for the period 1500-2100: 600 years of global gridded annual land-use transitions, wood harvest, and resulting secondary lands. *Climatic Change* **109**, 117–161 (2011).
84. Lloyd, C. T., Sorichetta, A. & Tatem, A. J. Data Descriptor: High resolution global gridded data for use in population studies Background & Summary. 1–17 (2017) doi:10.1038/sdata.2017.1.
85. Kaufmann, D. & Kraay, A. Worldwide Governance Indicators (WGI). *The World Bank Group* www.govindicators.org (2019).
86. Pekel, J.-F., Cottam, A., Gorelick, N. & Belward, A. S. High-resolution mapping of global surface water and its long-term changes. *Nature* **540**, 418–422 (2016).
87. Sandvik, B. World Borders Dataset. *Thematic Mapping* http://thematicmapping.org/downloads/world_borders.php (2009).
88. BirdLife International. Species Distribution Data Download. http://www.birdlife.org/datazone/info/spcdownload.
89. Wilman, H. *et al.* EltonTraits 1.0: Species-level foraging attributes of the world's birds and mammals. *Ecology* **95**, 2027 (2014).
90. WWF International. *Management Effectiveness Tracking Tool*. (2007).

Extended Data Table 1. **Parameter estimates and sample sizes across analyses.** Shows focal parameter estimates, plus 20 estimates from full parameter samples. Parameters are a) the maximum number of years of data the sample can have, to either side of protected area (PA) designation; b) the minimum number of years that must be sampled, to either side of protected area designation; c) the closest distance an unprotected site can be to a protected area before it is excluded from analysis; d) the proportion of counts that must be zeroes for the time period to be classified as "All Zeroes"; e) the maximum distance between paired protected and unprotected sites; f) the standardised difference in means threshold for BACI and CI matching; g) the proportion of populations that must be matched successfully to retain a species, for BACI and CI matching. h), i), j) show the number of protected sites/species/populations in that analysis run (note that BA and CI will generally be a subset of these). See Supplementary Information 4 for a further taxonomic break down of species in the focal analysis.

Analysis	a) Total years to either side of PA designation	b) Min number of measured years to either side of PA designation	c) Min distance to PA for unprotected sites	d) Proportion of counts that are zero for period to be classified as "All Zeroes"	e) Max distance between protected and unprotected sites (matching, BACI/CI)	f) Standardised difference in means threshold (matching, BACI/CI)	g) Proportion of species' populations that must be matched to retain species (matching, BACI/CI)	h) *N* Protected Sites	i) *N* Species	j) *N* Popula
Focal	10	7	1.00	0.70	500	0.25	0.70	864	67	7313
1	10	10	0.50	0.68	272	0.12	0.53	209	23	951
2	6	5	1.78	0.71	2091	0.19	0.72	933	77	12475
3	14	13	0.95	0.72	2500	0.25	0.50	282	63	4325
4	7	4	2.95	0.69	587	0.23	0.71	1328	68	6050
5	9	6	2.61	0.70	1986	0.15	0.87	953	17	2709
6	11	10	4.30	0.60	1542	0.21	0.68	395	34	1937
7	12	10	4.73	0.78	785	0.19	0.55	469	55	3784
8	5	4	1.36	0.76	100	0.24	0.63	492	51	1402
9	8	6	2.22	0.79	1390	0.20	0.84	952	66	11198
10	13	11	3.49	0.61	1121	0.14	0.53	309	11	677
11	15	10	3.37	0.62	454	0.17	0.90	493	32	1781
12	12	7	3.83	0.77	2199	0.11	0.81	543	4	686
13	9	8	1.95	0.64	1874	0.22	0.66	592	74	7465
14	5	4	1.56	0.74	1154	0.18	0.85	1115	74	13901
15	11	11	4.08	0.68	648	0.14	0.77	122	7	235
16	6	5	0.88	0.65	1676	0.16	0.64	930	74	11922
17	8	5	4.43	0.62	2402	0.23	0.60	1242	77	8738
18	10	9	3.13	0.80	1401	0.15	0.75	404	16	1184
19	15	10	5.00	0.66	192	0.10	0.57	392	33	1894
20	13	10	2.40	0.75	1010	0.13	0.80	334	5	519

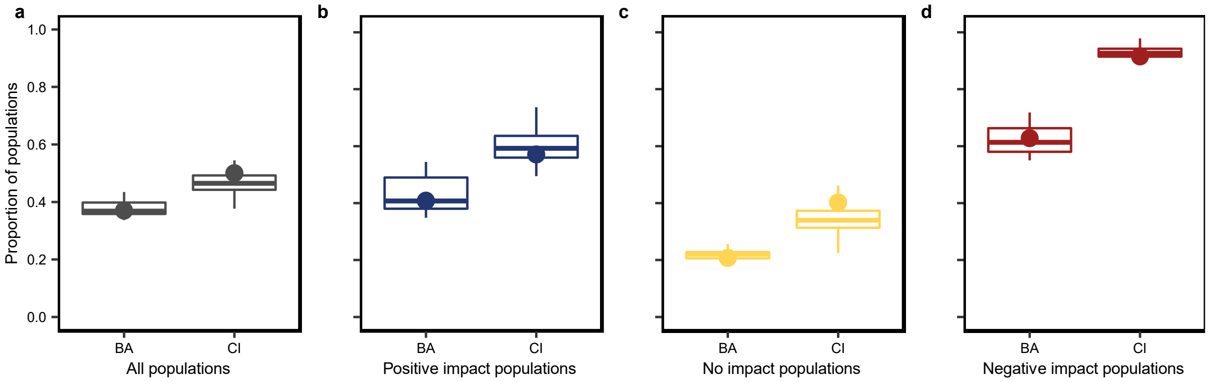
Extended Data Table 2. Covariates used to perform site matching. First, the three categorical variables (anthrome, region and migratory status) were used for exact matching. Next, all continuous variables were assessed for collinearity and highly collinear variables were removed. The remaining continuous variables were used to calculate mahalanobis distance.

Category and reason for inclusion	Variable	Data source	Resolution	Data transformation
Climate. This is a key variable that can determine suitability of a site for a species (meaning it is good to balance on) and also likelihood of being designated a PA.	Total annual precipitation (mm)	CRU TS4.01 [80]	0.5º, monthly (1961-2016)	Yearly sum of Jan-Dec
	Total precipitation December – February (mm)			Sum of Dec previous year and Jan & Feb current year
	Mean annual temperature (ºC)			Mean, min, max of months Jan-Dec
	Minimum annual temperature (ºC)			
	Maximum annual temperature (ºC)			
	Mean temperature December – February (ºC)			Mean, min, max of Dec previous year and Jan & Feb current year
	Minimum temperature December – February (ºC)			
	Maximum temperature December – February (ºC)			
Fertiliser input. Eutrophication can affect waterbird populations[48], can be a metric of distance to farming land and therefore human impact as well as a measure of the potential value of land for uses other than protection.	Nitrogen (g N/m^2 cropland/yr)	Lu & Tian[81]	0.5º, yearly (1961–2013)	NA
	Phosphorous (g P/m^2 cropland/yr)			
Land use. This is a direct measure of nearness to human impact, important for impacts to bird populations but also for likelihood of protected area designation – protected area s are less likely to be designated in areas suitable for agriculture and farming[82].	Anthrome (categorical)	HYDE 3.2.001 [83]	5', centennial (10,000BC-1600AD) decadal (1700-2000), yearly (2001-2016)	Pre-2000 data taken from nearest decade
	Grazing land (km^2/gridcell)			Temporal linear interpolation to obtain yearly data between decades of 1960-2000
	Irrigated land (not rice; km^2/gridcell)			
	Irrigated land (rice; km^2/gridcell)			
	Pasture land (km^2/gridcell)			
	Rangeland (km^2/gridcell)			
	Rainfed crop land (no rice; km^2/gridcell)			
	Rainfed crop land (rice; km^2/gridcell)			
Human presence.	Human population density (inhabitants/km^2 pergridcell)			
	Total built up area (km^2 per gridcell)			

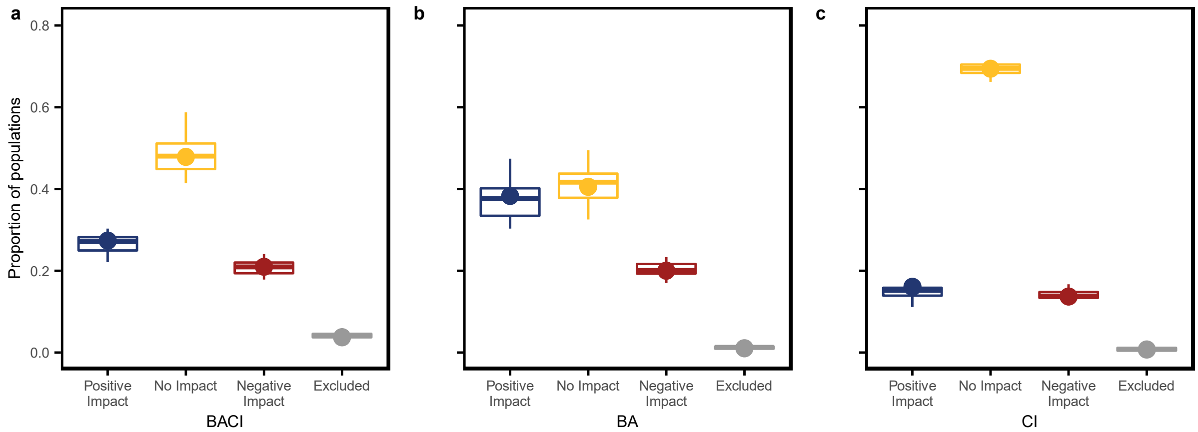
Protected areas are more likely to be designated in areas far from humans[82], and human presence can also affect waterbird numbers either directly through hunting or through habitat degradation.	Rural human population count (inhabitants/gridcell)			
	Urban human population count (inhabitants/gridcell)			
	Travel time to nearest city	WorldPop [84]	1km, yearly	Spatial bilinear interpolation to 5’ grid cells
Governance. Governance in a country is a significant predictor of protected area effectiveness[11], meaning it is important we compare protected area s with similar governance.	Mean of the six World Governance Index metrics (Control of Corruption, Government Effectiveness, Political Stability and Absence of Violence/Terrorism, Rule of Law, Regulatory Quality, Voice and Accountability)	World Bank [85]	By country, 1996, 1998, 2000, and yearly 2002-2016	Mean taken across all years because data is only available from 1996. Therefore, just one value per site for all years.
Water. Water presence is an important covariate for waterbirds, which rely on it for survival.	Surface water (presence/absence)	Pekel et al[86]	30m, 1985-2005	Converted to 5’ gridcells by taking sum of ‘presence’ $30m^2$ cells in each
Elevation. Protected areas are biased towards where they can least prevent land conversion[82] which often results in them being in high elevation regions. Higher elevation sites are also likely to have less pressure and thus have lower biodiversity losses regardless of whether they are protected areas or not.	Elevation	WorldPop [84]	1km, NA	Spatial bilinear interpolation to 5’ grid cells
Global Region. Because we are aiming to compare trends inside and outside protected areas, we wanted populations to at least be in similar regions to reduce unknown variance in comparisons.	Continent (categorical)	TM World Borders [87]	NA	NA
Migratory Status In some cases species have some migratory and some resident populations. To ensure we were not comparing between populations of different migratory types we exact matched on migratory status.	Migratory Status	Birdlife.org[88]	Species range polygons delineated by different migratory types	We classified each population (site species combination) based on the polygon the site fell within

Extended Data Table 3. Covariates used to assess what factors affect protected area impact.

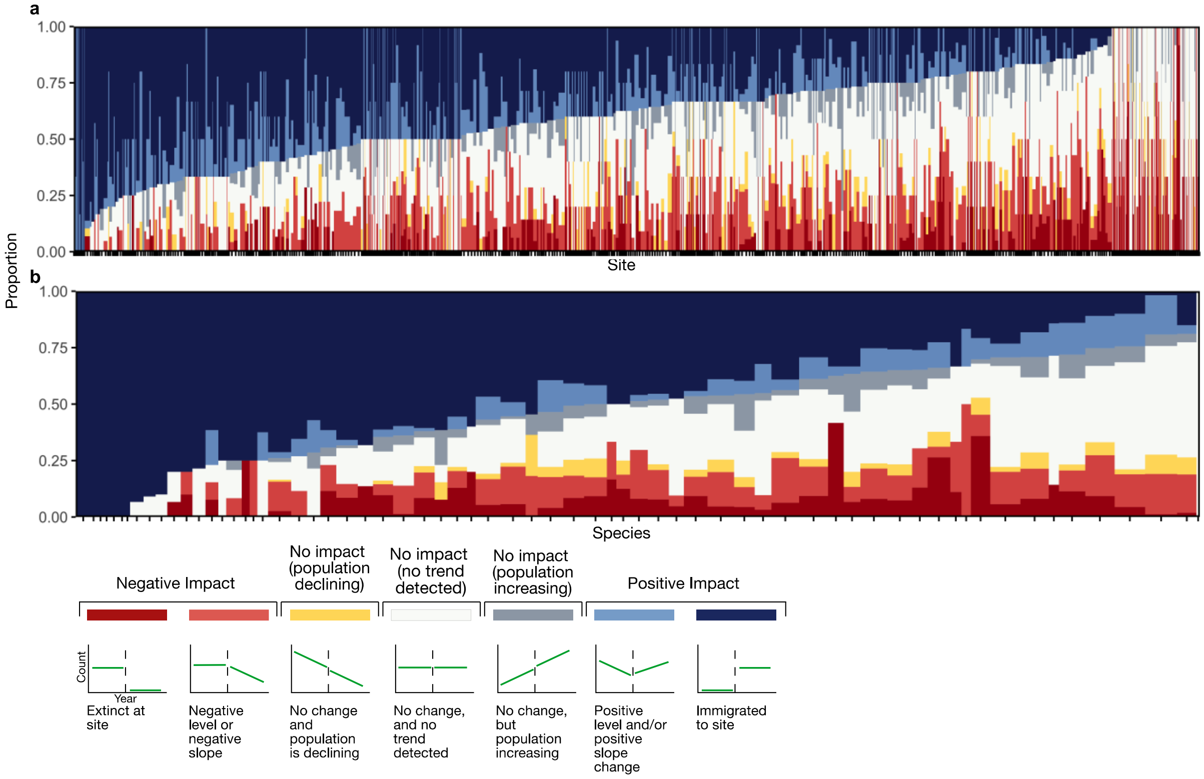
Category	Variable and Reason for Inclusion	Category/Levels	Source
Species	Body Mass We expected larger species to respond better to protected areas[10], due to the fact that larger bodied species are more vulnerable to hunting.	Continuous	Wilman et al[89]
	Taxonomic group Different taxonomic groups may respond differently to protection, so we looked for differences between orders.	Categorical: Order	Birdlife.org[88]
Species (nested within Site)	Migration Status. Because migrants are affected by other stressors than just those in their wintering site, we expect migrants will show less responsiveness to protected areas (and it is beyond the scope of this study to consider migratory networks). Some species are migrants in parts of their range and non-migrant in others, so categorised each population at each site separately.	Categorical: Non-migrant, Migrant	Birdlife.org[88]
Site (nested in Country)	Anthrome. We expected that sites in more remote regions (i.e. semi-natural, wild) will show less responsiveness to protection, as these sites are less likely to have been being exploited in the absence of protection.	Categorical: Urban, Village, Croplands, Rangeland, Semi-natural, Wild	HYDE[83] (see Extended Data Table 2)
	Protected area size. We expected larger protected areas to perform better, because of reduced edge effects. In some cases, sites occurred in multiple protected areas that were of different sizes and had been designated at different times. In these cases, we used the size of the largest size protected area, if that area was designated earliest. If not, we took the mean of all areas.	Continuous	World Database on Protected Areas[58]
	Protected area Management. The best way to assess management would be with the Management Effectiveness Tracking Tool (METT[90]) but unfortunately this is biased away from Europe and the USA, unlike our dataset, and only a few of the protected areas in our dataset are included in the METT. Instead, we chose to compare sites we know to be managed for birds to other sites, acknowledging that some of the 'other' sites may also be managed for waterbirds, but not having the power to ascertain management status of all cases. We created a category comprised of Ramsar and Special Protected Area (Birds Directive) sites, which encompasses 55-57% of populations. A full list of waterbird-managed and other sites is given in Supplementary Information 6.	Categorical	World Database on Protected Area[58]
Country	Governance. We expected sites in better governed areas to respond better to protection[11].	Continuous	World Bank[85] (see Extended Data Table 2)



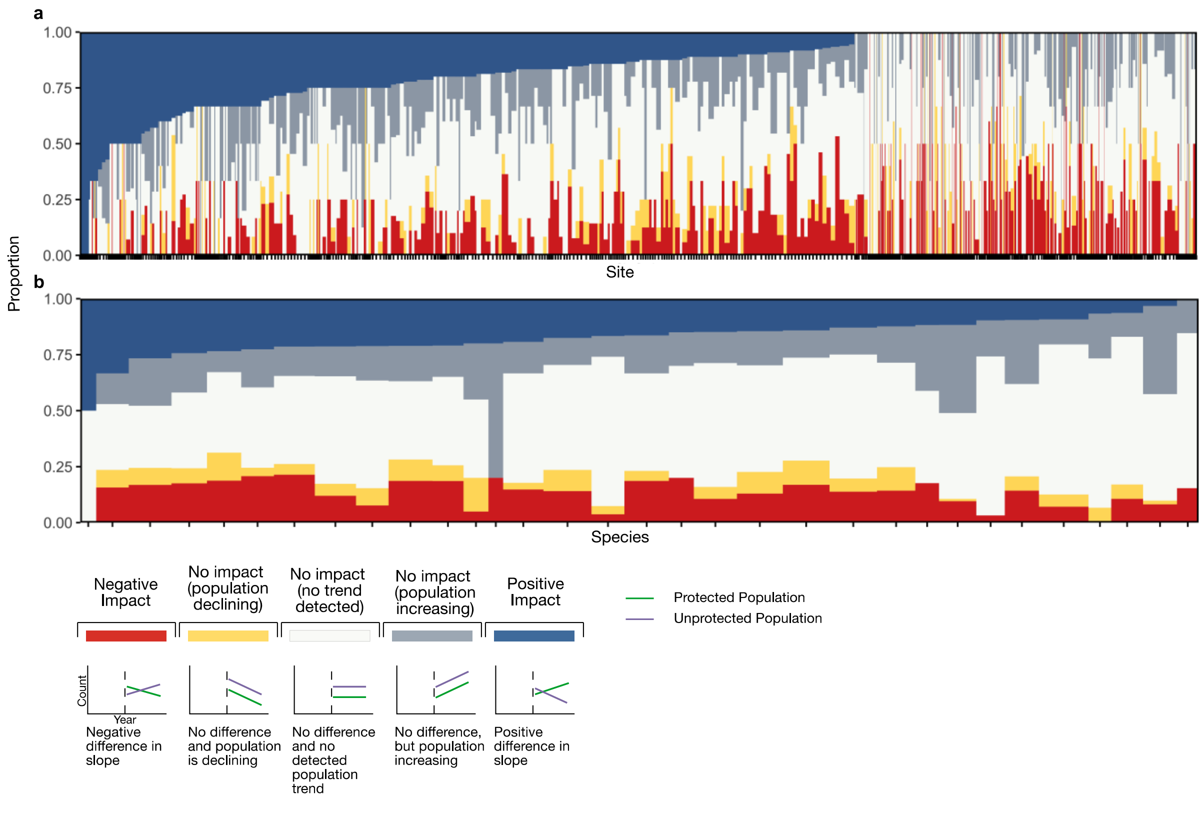
Extended Data Figure 1. Changes in estimates of protected area impact under different study designs, for all analyses. Proportion of Before-After (BA) or Control-Intervention (CI) populations that changed outcome when analysed under a BACI framework, by each analysis (n=21; 20 full parameter, plus one focal analysis). Shown for all populations (a), then the proportion of positive (b), no (c) or negative impact populations (d) that changed in outcome. Each point is an analysis, with boxplots showing distribution (box bounded by 25th and 75th percentiles, centre shows 50th percentile, whiskers extend to 1.5*IQR above 75th percentile, for maxima, or below 25th percentile, for minima). Large points show focal analysis estimates.



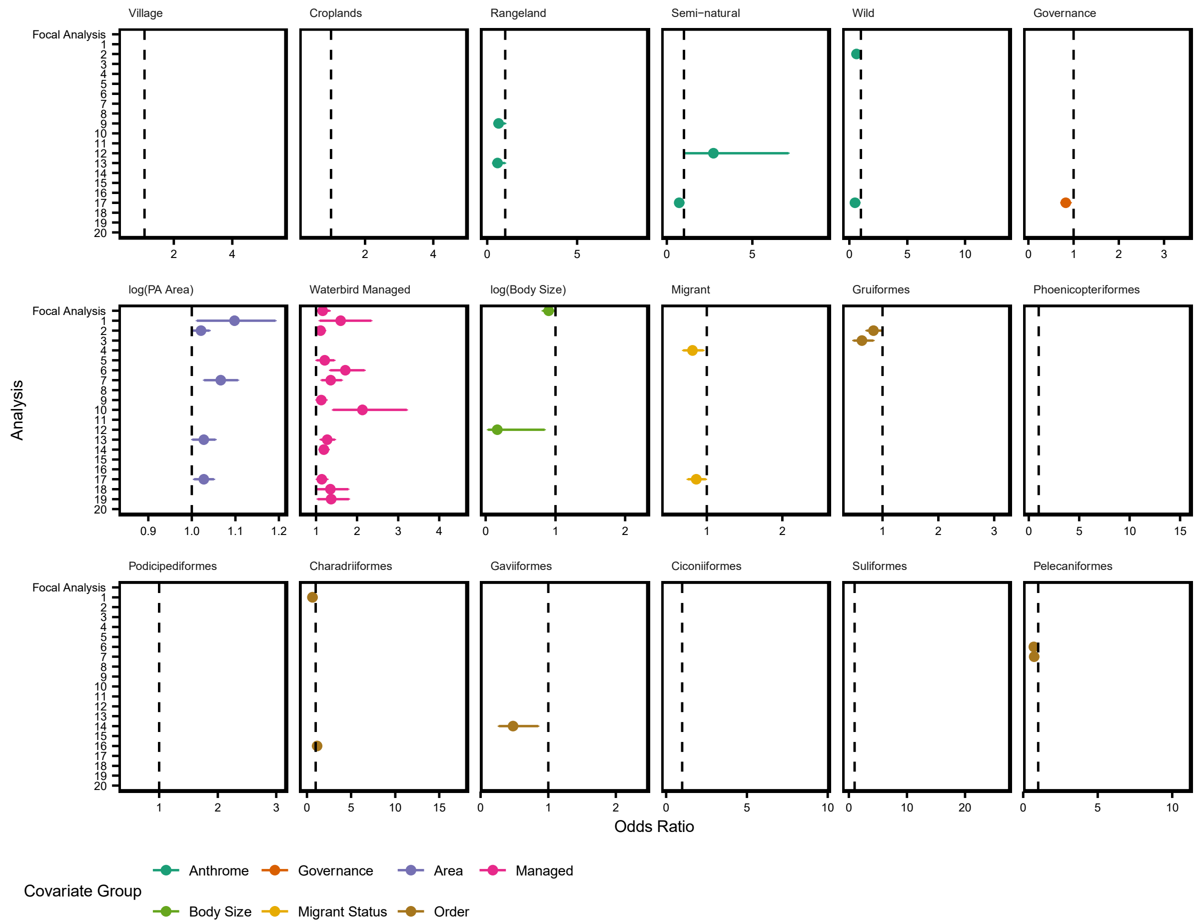
Extended Data Figure 2. Estimates of protected area impact under a BACI study design, for all analyses. Percentage of populations that have been positively, negatively or not impacted by protected areas, by each analysis (n=21; 20 full parameter analyses, plus one focal analysis). Each point is an analysis, with boxplots showing distribution (box bounded by 25th and 75th percentiles, centre shows 50th percentile, whiskers extend to 1.5*IQR above 75th percentile, for maxima, or below 25th percentile, for minima). Large points show estimates from focal analysis. Panels show estimates under BACI (a), Before-After (b) or Control-Intervention (c) frameworks.



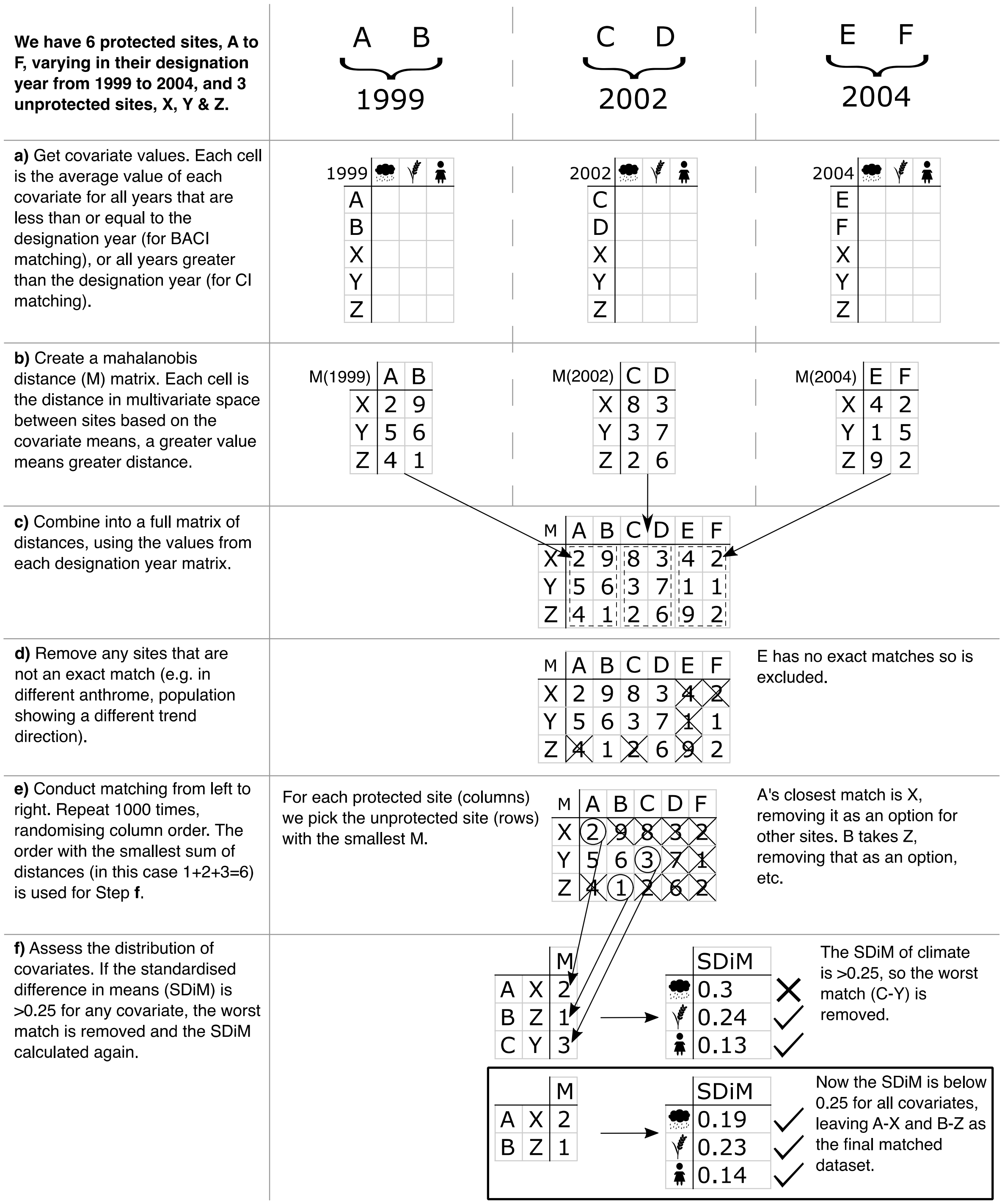
Extended Data Figure 3. Estimates of protected area impact under a BA study design. Proportion of populations (n=6263) showing various responses to protection, per site (a; n=860) and species (b; n=66), when response to protection is calculated in a BA framework. Each species/site is one bar, with the proportion of their populations in each category shown on the y axis. Bar width is scaled to the number of populations of that species/site in the dataset, log scaled in the case of species, with a wider bar meaning the species/site has more populations. Each colour represents a different way a population can respond to protection, and an example of each is shown at the bottom.



Extended Data Figure 4. Estimates of protected area impact under a CI study design. Proportion of populations (n=3783) showing various responses to protection, per site (a; n=698) and per species (b; n=32), when response to protection is calculated in a CI framework. Each species/site is one bar, with the proportion of their populations in each category shown on the y axis. Bar width is scaled to the number of populations of that species/site in the dataset, log scaled in the case of species, with a wider bar meaning the species/site has more populations. Each colour represents a different way a population can respond to protection, and an example of each is shown at the bottom.



Extended Data Figure 5. Predictors of protected area impact, with odds ratios and confidence intervals. Odds ratios for covariates predicting protected area (PA) effectiveness under a BACI framework. Estimated using cumulative link mixed models, points show model estimates, tails show 95% confidence intervals, and significance is indicated by bold colours ($p<0.05$). Dashed line given at an odds ratio of one (ratios above one indicate a positive relationship, and below one a negative relationship). Y axis shows all analyses (20 full parameter analyses, plus one focal analysis, with the focal analysis given in the first row). Colours show covariate grouping. Orders are measured relative to Anseriformes, and Anthromes relative to Urban. Note that we expect continuous variables (PA Area, Body Size, Governance) to have smaller coefficients as they express odds ratios per unit increment.



Extended Data Figure 6. Schematic demonstrating matching procedure. Example of the matching procedure for one species, using a toy dataset of 6 protected sites (A to F) and 3 unprotected sites (X, Y and Z), with three dummy example covariates, climate (cloud), land use (wheat) and human population (person). See methods section "matching" for more detailed step by step walk through of this process.