Quantifying the impact of protected areas on near-global waterbird population trends

PRE-ANALYSIS PLAN

Hannah Wauchope[1*], Julia P G Jones[2], Tatsuya Amano[1,3,4], Jonas Geldmann[1], Daniel Blanco[5], Richard A. Fuller[3], Tom Langendoen[5], Taej Mundkur[5], Benno I. Simmons[1,6], Nagy Szabolcs[5], William J. Sutherland[1].

[1]Conservation Science Group, Department of Zoology, University of Cambridge, David Attenborough Building, Cambridge, UK
[2]College of Environmental Science and Engineering, Bangor University, Bangor, UK
[3]School of Biological Sciences, University of Queensland, Brisbane, Australia
[4]Centre for the Study of Existential Risk, University of Cambridge, Cambridge, UK
[5]Wetlands International, Horapark 9, Ede, Netherlands
[6]Department of Animal and Plant Sciences, University of Sheffield, Western Bank, Sheffield, UK

*Corresponding Author: hsw34@cam.ac.uk.

1. Summary

There is increasing interest in the effectiveness of protected areas (PAs) for supporting populations of wildlife. While there are a number of association studies showing a relationship between protected areas and abundance or trends in wild species, studies with an appropriate counterfactual (what would have happened in the absence of protection) are rare. We use the world's largest database on waterbird counts (covering 587 species at 21,989 sites globally) to answer three questions: 1) Do PAs have a positive impact on waterbird population trends relative to a counterfactual (this includes cases where a PA has lessened, but not halted, a population decline)?; 2) are PAs performing successfully by maintaining or increasing populations? and 3) what factors contribute to PA impact and performance? We selected 15,703 waterbird populations (here defined as a site species combination), consisting of 311 species at 870 protected sites, where PA designation occurred at least 5 years after the first survey date, and 5 years before the last. We will use this to compare trends before PA designation to those afterwards. We then matched these sites to unprotected sites with similar covariates in the years before PA designation, resulting in a matching dataset of 6,451 populations pairs consisting of 39 species at 769 pairs of protected and unprotected sites. We will use this to compare trends both before and after PA designation *and* inside and outside of PAs. Our results will shed light on the impact of PA on hundreds of waterbird species, providing much needed evidence regarding PA effectiveness. As PA performance is a sensitive subject and it is important to develop hypotheses before knowing the results (especially for the relatively complex data analysis used in matching protected and unprotected sites), we present a pre-analysis plan. This will ensure that the final paper's analyses are hypotheses testing, rather than generating, and avoids the risk of, or perception of, data dredging.

2. Introduction

Protected Areas (PAs) are receiving increasing attention, both being held up as the solution to high levels of species' extinctions and being subject to increasing debate over their real-world effectiveness (Barnes *et al*., 2018). The Convention on Biological Diversity (CBD)'s Strategic Plan for Biodiversity 2011-2020 set out Aichi Biodiversity Targets, #11 of which dictated that by 2020 at least 17% of terrestrial and inland water areas will be conserved through effectively managed Protected Areas (Convention on Biological Diversity, 2011). As a result, the global PA network has been expanding, with many countries on track to meet the target (global terrestrial protected areas coverage currently stands at 14.9%; UNEP-WCMC & IUCN and NGS, 2018). However, a number of studies show that some PAs are not managed effectively (Blom *et al*., 2004; Leverington *et al*., 2010; Gill *et al*., 2017; Geldmann *et al*., 2018) and question the ultimate effectiveness of PAs (Baillie *et al*., 2016; Pringle, 2017). Comprehensive data on PA effectiveness is still lacking (Geldmann *et al*., 2013) and calls have been made to establish whether PAs positively impact biodiversity and whether they are achieving what they set out to achieve: preserving habitat and supporting wildlife populations (Watson

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27741v2 | CC BY 4.0 Open Access | rec: 21 Jun 2019, publ: 21 Jun 2019

et al., 2014, 2016). Understanding the effectiveness of one of the most important conservation approaches (Protected Areas) implemented to date is greatly needed as the CBD prepares a post-2020 Biodiversity Framework (IUCN, 2018; CBD, 2019).

Studies assessing the impact of PAs need an estimate of the counterfactual (what would have happened in the absence of the intervention; Ferraro 2009). There are robust, counterfactual studies attesting to the effectiveness of PAs at averting forest clearance (Bruner *et al*., 2001; Nelson & Chomitz, 2009; Scharlemann *et al*., 2010; Joppa & Pfaff, 2011; Geldmann *et al*., 2013, though see Clark *et al*., 2013). However, studies considering the effectiveness of PAs in maintaining populations of species are in their infancy (Geldmann *et al*., 2013). A number of recent large-scale studies have considered wildlife population trends within PAs (Laurance *et al*., 2012; Barnes *et al*., 2016, Amano *et al*., 2018), however studies including appropriate counterfactuals either in comparisons to trends of populations before PAs were designated, or to comparable trends of suitably matched populations outside of PAs are lacking (Geldmann *et al*., 2013). Association studies cannot determine whether a PA directly benefits a population or is simply more likely to be designated at sites where a population is doing well (*de facto* protection).

Ways of describing PA impact and success are explored in Figure 1. A PA has a positive impact if the trend of the protected population is more positive than the counterfactual (blue and green lines, Figure1a,b; to the left of diagonal line Figure 1c). It is performing successfully if the trend of the protected population is stable or positive (green and yellow lines, Figure 1a,b; above the horizontal line, Figure 1c). There can be cases where the PA is having a positive impact, but is not performing successfully (blue in Figure 1) and also cases where the PA has had a negative impact, but is still performing successfully (yellow in Figure 1). Note that by this definition PA impact is defined relative to a counterfactual, while performance does not need counterfactuals to be determined – it is simply a measure of the absolute trend of the protected population.

We use a near-global, though European and North America focused, dataset of waterbird population trends from 1966 – 2014 (Amano *et al*., 2018) to test PA impact and performance. Though these data have been used to assess population trends in PAs at a coarse scale (Amano *et al*., 2018), it was not the main aim of the study and counterfactuals, either before/after or inside/outside, were not established. We have filtered Amano et al's dataset to our final dataset of protected sites consisting of 15,703 waterbird populations (here defined as a site species combination) and 311 species at 870 protected sites, where PA designation occurred at least 5 years after the first survey date, and 5 years before the last.

We aim to answer three questions:

1) Do PAs have a positive impact on waterbird population trends (relative to a counterfactual; this includes cases where a PA has lessened, but not halted, a population decline)?
2) Are PAs performing successfully by maintaining or increasing populations (i.e. population trends are stable or increasing)?
3) What factors contribute to PA impact and performance?

To answer Question 1 we use two approaches to estimate the counterfactual. We will compare trends before and after designation for all populations in the dataset (Before After Analysis), and, where appropriate matches between protected and unprotected sites are available, compare trends inside and outside PAs, as well as before and after designation (Before After Control Intervention [BACI] analysis, McDonald *et al*., 2000). Each analysis has advantages and disadvantages. The Before After analysis makes use of the full dataset while the BACI dataset is biased towards common species as it was not possible to find matches for all species at all protected sites. The benefit of a BACI study is that it allows us to account for counterfactuals in both space and time, while with just a Before After comparison, we cannot know if the change was due to wider changes in the population generally (although the fact that designation date of PA varied substantially across the dataset is helpful here). By combining space and time counterfactuals, we can come close to knowing the true impact of PA designation. To answer

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27741v2 | CC BY 4.0 Open Access | rec: 21 Jun 2019, publ: 21 Jun 2019

Question 2 we need only the trends of protected populations in the years after designation.

To answer Question 3, we will see how the impact and performance of PAs, derived from Questions 1 & 2, varies with covariates that might influence PA effectiveness. We make three hypotheses. First, that well-managed, large PAs in countries with better governance will have a more positive impact/perform better (the latter being found by association in Amano *et al.*, 2018). Second that non-migrant, large-bodied species will respond best to PAs, the former because they are not affected by outside influences and the latter because they are more susceptible to exploitation outside PAs (Barnes *et al.*, 2016). Third, that PAs in more remote regions will benefit less from protection as they are likely to be experiencing less human pressure.

Here, we present a pre-analysis plan. There is growing interest in improving the quality of impact evaluations in conservation (Baylis *et al.*, 2016), and in reducing the impacts of human bias on post-hoc statistical analysis. A risk in all such analyses is HARKing (Hypothesizing After Results are Known; Kerr, 1998) and it is therefore important to draw a clear distinction between generating hypotheses with existing observations and testing hypotheses with new observations. This is especially important in matching studies: there is high variability in matching procedures (Stuart, 2010) and it is important to avoid the risk of trying many and selecting the one which gives a more desired result. Preregistering of analysis has been proposed as a way to address this (Hardwicke & Ioannidis, 2018; Nosek *et al.*, 2018). We have therefore divided our analysis into two stages: Stage 1) obtaining a dataset, selecting protected sites and matching these to unprotected sites, and Stage 2) analysis of PA impact and performance on population trends (Qs 1 & 2) and factors affecting these trends (Q3). Stage 1 is complete and we report results from it, but these simply tell us the quality of the datasets to be used for Stage 2. Stage 2 has not been carried out, and we propose our methods for how Qs 1-3 will be assessed.

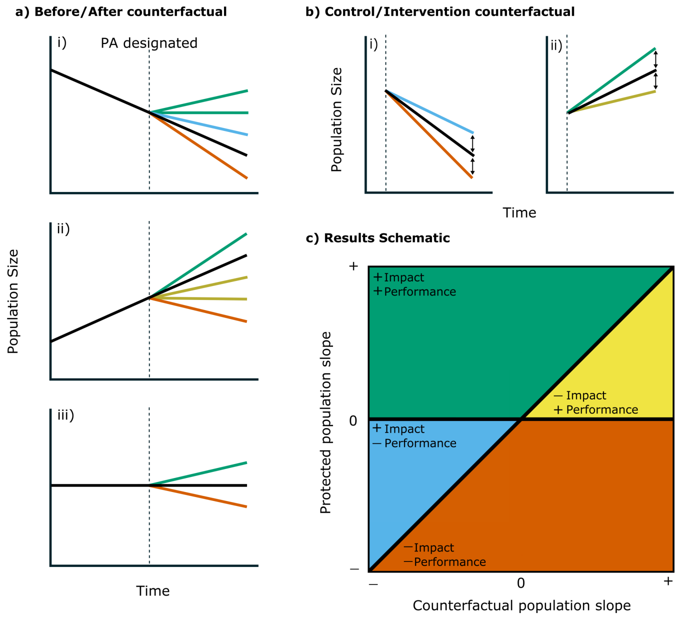


Figure 1. The impact of a PA (vertical dotted line) is assessed relative to a counterfactual while performance refers simply to the absolute trend. **a)** Comparing trends after protection to before protection (a Before After counterfactual) can show a positive impact and positive performance (green lines), a negative impact but positive performance (yellow lines) a positive impact but negative performance (blue lines), a negative impact and negative performance (orange lines) **b)** Comparing trends inside PAs with matched, non-protected sites (a BACI counterfactual). **c)** The categories when comparing the protected population slope (y axis) to a counterfactual population slope (x axis).

3. Methods Stage 1

So far we have completed the following steps:

- Obtained count data for waterbirds from sites across the world
- Determined which sites are inside PAs
- Established a clean dataset of populations (i.e. site-species combinations) within PAs.
    - We term this the Full Protected dataset, and it allows us to do before-after comparisons of PA effectiveness
- Matched each population from the Before After dataset to a population from the unprotected sites that was of the same species, surveyed over the same period, displayed a similar trend and was similar in habitat according to a number of climate, land-type and human impact criteria (Table 2). This created a second dataset consisting of pairs of matched protected and unprotected populations.
    - We term this the "BACI" (before/after/control/intervention) dataset, as it allows us to do before-after-control-intervention comparisons of PA effectiveness.

3.1 Count Data

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 (see Amano et al., 2018 for the full methods). Our initial dataset consisted of 587 species at 21,989 sites. We removed populations (i.e. site-species combinations) with zero counts in all years of the time series. We then restricted our data to only sites surveyed in December to February, that covered at least 10 survey years (but did not require every year in that period to be surveyed, in accordance with Wauchope *et al*., 2019).

As Christmas Bird Count data is not standardized for effort, we required that species taken from CBC data showed a log-linear relationship with effort (i.e. the rate of new individuals detected 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 (Venables & Ripley, 2002), using all available CBC data for that species:

$$\log\left(E(Count_i)\right) = \beta\log(e_i) \qquad (1)$$

$$var(Count_i) = v_{NB}\left(E(Count_i)\right) \qquad (2)$$

Where *Count* is all counts of a species and e_i is the number of survey hours for each count. The variance of the counts is negative binomial (Eq. 2). We retained CBC data for all species where there was a significant positive relationship between count and effort (i.e. β was significant and greater than 0).

3.2 Protected (and Unprotected) Area Data

We took our protected area data from the World Database on Protected Areas (UNEP-WCMC & IUCN, 2019). We downloaded the full dataset of all protected areas globally, and overlaid our sites to determine which fell in protected areas. We removed any sites where the PA designation status was proposed, and any UNESCO biosphere reserves as these are often not afforded formal protection (Coetzer *et al*., 2014). We next removed any sites where there was no information about the designation date of the PA. In some cases, there were multiple PA data for a site, in these cases we took the earliest designation year given. Finally, we reduced the dataset to only those cases where the designation date of the PA occurred at least 5 years after the first survey date of the population, and at least 5 years before the last survey date, with data taken in at least 3 years before and after (Wauchope et al., 2018). This gave us a Full Protected dataset of counts for 15,703 populations, consisting of 311 species at 870 protected sites (Figure 3, blue points, Table 1).

We next created a dataset of counts at unprotected sites, by only including sites that were at least 1km from a protected area, to avoid any confounding of results from spill-over effects (Ament & Cumming, 2016). We then restricted this dataset to only the species present in the Full Protected dataset. This gave us an unprotected dataset of 103,979 populations consisting of 311 species and 5081 sites (Figure 3, green points, Table 1)

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27741v2 | CC BY 4.0 Open Access | rec: 21 Jun 2019, publ: 21 Jun 2019

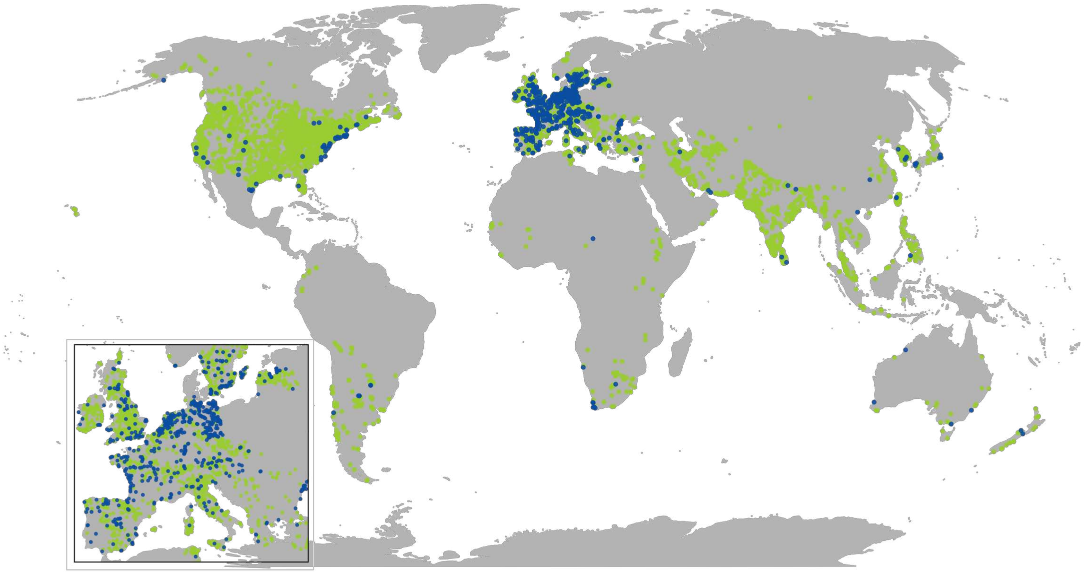


Figure 3. The Full Protected dataset of 870 protected sites in blue, as well as the 5081 unprotected sites that were available for matching in green, with zoomed inset for Western Europe.

3.3 Site Matching

As we also want to compare trends of populations inside and outside of protected areas (the BACI analysis), we needed to establish a reasonable counterfactual through which to compare protected and unprotected sites. If the sites we compare are not well matched in terms of other covariates that influence protected area designation and population trends, then it is not possible to fairly compare them (Gelman & Hill, 2006). In other words, as closely as possible, sites should differ only in whether they have PAs or not, to test the effect of PA designation *per se*.

We use statistical matching to achieve this, by selecting control sites that have similar covariate (Table 2) distributions to treated sites in the years before treatment (See Stuart, 2010 for a comprehensive summary of matching methods). For this analysis, it was necessary to develop a novel matching method because we needed to match on covariates relating only to the years prior to designation for each protected area (see https://github.com/hannahchoppie/PAImpact for code). The covariates we used for matching, how we prepared them and justification for their use are given in Table 2. We removed highly correlated variables by first calculating the variance inflation factor (using the usdm package in R; Naimi *et al*., 2014) of all covariates, and iteratively removing variables with a VIF greater than four until none were over four (Salmerón Gómez *et al*., 2016). We next removed variables with a Pearson's Correlation Coefficient of over 0.7. The reduced set of covariates is denoted in Table 2 as any covariates with at least one asterisk (*).

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 (Gu & Rosenbaum, 1993; Stuart 2010), recent criticisms of the most commonly used matching method, Propensity Score Matching (King & Nielsen 2019), meant we were interested to test other options and found Mahalanobis distance matching to perform markedly better in comparisons.

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

$$md_{(x,y)} = \sqrt{(\boldsymbol{x}-\boldsymbol{y})^T\boldsymbol{S}^{-1}(\boldsymbol{x}-\boldsymbol{y})} \qquad (3)$$

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.

We needed to match only on covariates in the years prior to designation, as we would expect and hope for protected and unprotected sites to vary in the years after PA designation, especially in covariates related to human impact. Further, mahalanobis distance requires each site to have one value for each covariate, whereas in our dataset we have values for each year, meaning we need to take means of the values for the years pre-designation. Because designation year varies, we created a series of mahalanobis distance matrices for each year that protected areas were designated, finding the distance, $md_{(unprotected\ site,\ protected\ site)}$, between protected and unprotected sites using covariate means from the years before the relevant designation year (See Figure 4i, ii 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 PA in our dataset was designated. This resulted in a loss of only 9 sites.

For each species, these mahalanobis distance matrixes were then combined into a larger distance matrix containing all the sites across all designation years that the species occurred in (Figure 4iii).

We then required that sites were exactly matched on a number of criteria. For each protected site, we removed unprotected sites not of the same anthrome category and geographic region (which roughly equates to continent) and only retained those where the PA was designated at least 5 years after the first year of surveys and at least 5 years before the last year, with counts taken in at least 3 years before and after (Wauchope *et al*., 2019) (Figure 4iv). To satisfy assumptions in the proposed models in Stage 2, we required potential matches to have the same population trend in pre-designation years (Figure 4iv). We calculated trends using a simple negative binomial glm (glm.nb, R package MASS, Venables & Ripley, 2002):

$$\log\left(E(Count_z)\right) = \alpha + \beta_1 Y_z \begin{cases} & D_z = IWC \\ + \log(e_z) & D_z = CBC \end{cases} \tag{4}$$

Where the count of the population in year z is predicted by the Year, and an effort term if the data was taken from the CBC. Variance is expressed by Eq. 2. The β_1 coefficient gives the slope of the population. If over 6 years of data were available, the population trend was classified as positive or negative according to the slope as this has an 80% chance of being a good estimate of the 10 year slope of the population even if not significant (Wauchope et al., 2019, supp. material), if under 6 years of data were available the slope was classified as either stable or, if the slope was significant at $p < 0.05$, positive or negative. Sites with zero counts in all years were excluded.

If no unprotected sites met the exact match criteria, the protected site did not have a match and was excluded (e.g. Figure 4iv, Site E).

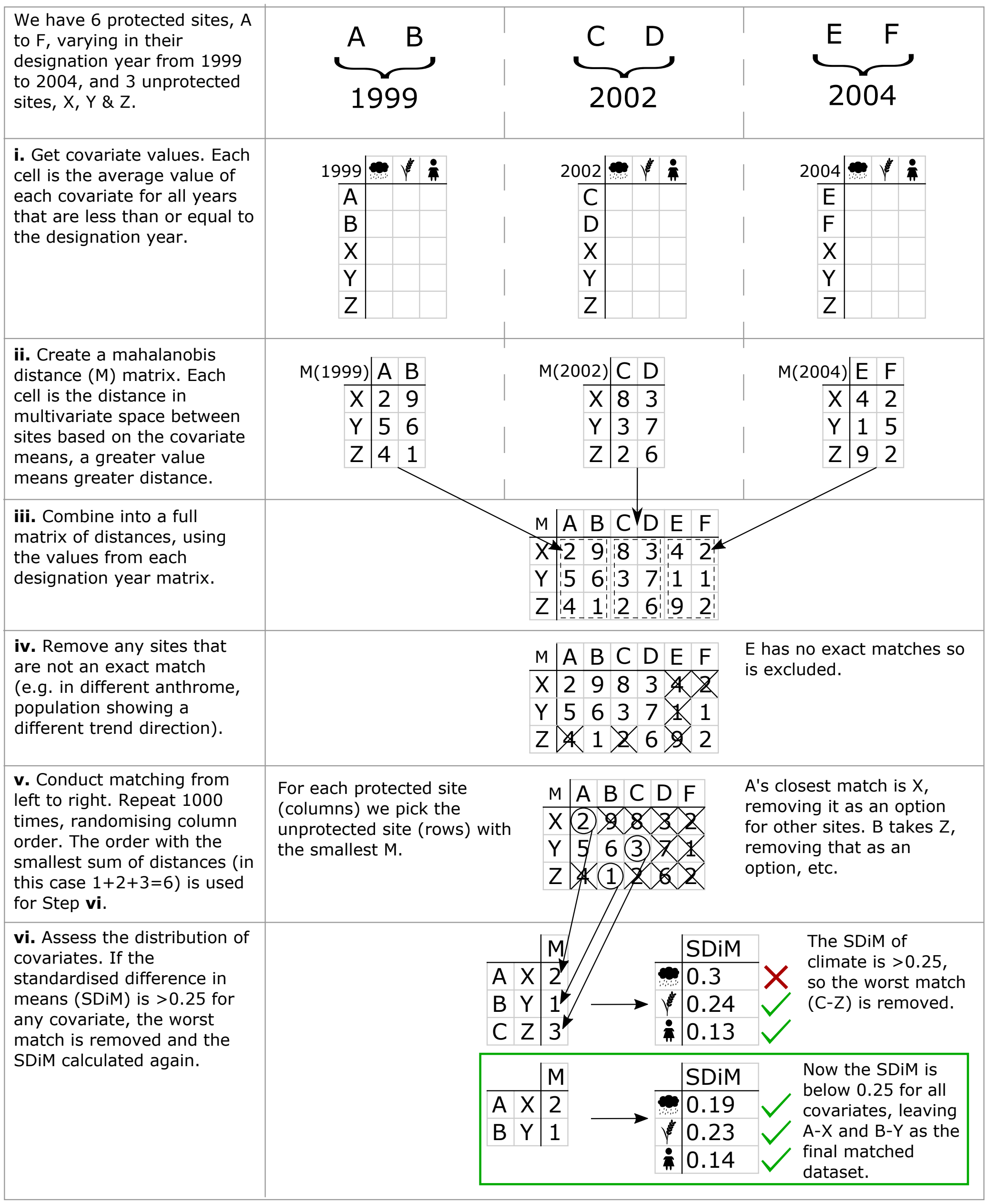
Next, we ran an optimized greedy nearest-neighbour algorithm to select, from the exact matched subset, the unprotected site with the closest distribution of covariates to each protected site. We ran this without replacement, meaning each protected site could be matched to only one protected site, to 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 (Simmons et al., 2019), 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 (Figure 4v, e.g. with randomisations in the figure a smaller global distance would be detected, with the column order FBACD).

Once our matched sets were obtained for each species, we finally 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 (Austin, 2009):

$$d_{cov} = \frac{\bar{\boldsymbol{T}}_{cov} - \bar{\boldsymbol{C}}_{cov}}{\sqrt{\frac{var(\boldsymbol{T}_{cov}) - var(\boldsymbol{C}_{cov})}{2}}} \quad (5)$$

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 was below 0.25 for all covariates (Rubin, 2001; Stuart, 2010). 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, the remaining matched pairs were considered the 'final' matched dataset for that species (Figure 4vi). If this point was never reached, the species was removed.

The final matched dataset contains 39 species, 6451 population pairs and 769 protected sites paired to 1247 unprotected sites (Figure 5, Table 2). There are more unprotected than protected sites as matching was done by species and a protected site may contain multiple species each matched to different unprotected sites.



We have 6 protected sites, A to F, varying in their designation year from 1999 to 2004, and 3 unprotected sites, X, Y & Z.

i. Get covariate values. Each cell is the average value of each covariate for all years that are less than or equal to the designation year.

1999	🌧	🌾	🧍
A			
B			
X			
Y			
Z			

2002	🌧	🌾	🧍
C			
D			
X			
Y			
Z			

2004	🌧	🌾	🧍
E			
F			
X			
Y			
Z			

ii. Create a mahalanobis distance (M) matrix. Each cell is the distance in multivariate space between sites based on the covariate means, a greater value means greater distance.

M(1999)	A	B
X	2	9
Y	5	6
Z	4	1

M(2002)	C	D
X	8	3
Y	3	7
Z	2	6

M(2004)	E	F
X	4	2
Y	1	5
Z	9	2

iii. Combine into a full matrix of distances, using the values from each designation year matrix.

M	A	B	C	D	E	F
X	2	9	8	3	4	2
Y	5	6	3	7	1	1
Z	4	1	2	6	9	2

iv. Remove any sites that are not an exact match (e.g. in different anthrome, population showing a different trend direction).

M	A	B	C	D	E	F
X	2	9	8	3	~~4~~	~~2~~
Y	5	6	3	7	~~1~~	1
Z	~~4~~	1	~~2~~	6	~~9~~	2

E has no exact matches so is excluded.

v. Conduct matching from left to right. Repeat 1000 times, randomising column order. The order with the smallest sum of distances (in this case 1+2+3=6) is used for Step **vi**.

For each protected site (columns) we pick the unprotected site (rows) with the smallest M.

M	A	B	C	D	F
X	②	~~9~~	~~8~~	~~3~~	~~2~~
Y	~~5~~	6	③	~~7~~	~~1~~
Z	~~4~~	①	~~2~~	~~6~~	~~2~~

A's closest match is X, removing it as an option for other sites. B takes Z, removing that as an option, etc.

vi. Assess the distribution of covariates. If the standardised difference in means (SDiM) is >0.25 for any covariate, the worst match is removed and the SDiM calculated again.

		M
A	X	2
B	Y	1
C	Z	3

	SDiM	
🌧	0.3	✗
🌾	0.24	✓
🧍	0.13	✓

The SDiM of climate is >0.25, so the worst match (C-Z) is removed.

		M
A	X	2
B	Y	1

	SDiM	
🌧	0.19	✓
🌾	0.23	✓
🧍	0.14	✓

Now the SDiM is below 0.25 for all covariates, leaving A-X and B-Y as the final matched dataset.

Figure 4. 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).

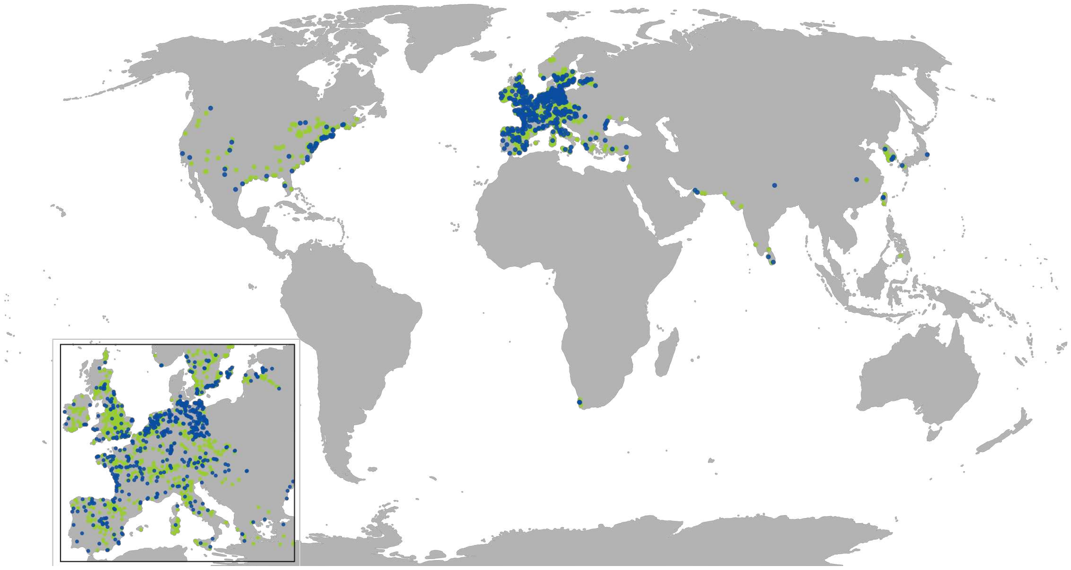


Figure 5. BACI dataset containing protected (blue, n=769) and unprotected (green, n=1247) sites, with zoomed inset for Western Europe. Note that there are more unprotected than protected sites as matching was done by species and a protected site may contain multiple species each matched to different unprotected sites.

Table 1. Taxonomic distribution for our Before/After data (with the subset used in the BACI analysis in brackets). Where families have no species present in the BACI dataset the row is greyed out. Note that the number of 'lost' protected sites in the BACI dataset appear to be much higher than reported, but this is because in many cases some taxa at a protected site were unable to be appropriately matched, while others were, retaining the site in the overall dataset but not for that particular taxa.

Order	**Family**	**Genera**	**Species**	**Protected Sites**	**Unprotected Sites**
Anseriformes	Anatidae	44 (20)	148 (34)	825 (727)	4634 (1062)
	Anhimidae	2	2	3	14
Charadriiformes	Alcidae	2	2	1	53
	Charadriidae	8 (6)	46 (10)	247 (115)	2428 (121)
	Haematopodidae	2 (2)	12 (2)	80 (10)	168 (10)
	Jacanidae	6	6	7	246
	Laridae	22 (2)	82 (8)	294 (215)	2842 (253)
	Recurvirostridae	4	8	75	732
	Rostratulidae	2	2	1	3
	Scolopacidae	24 (6)	97 (6)	260 (100)	2484 (100)
	Stercorariidae	2	4	11	74
Ciconiiformes	Ciconiidae	10	18	25	500
Gaviiformes	Gaviidae	2	6	42	942
Gruiformes	Aramidae	2	2	4	51
	Gruidae	6	10	9	112
	Rallidae	24 (4)	50 (4)	646 (493)	3321 (519)
Pelecaniformes	Ardeidae	22 (2)	46 (6)	471 (344)	3473 (353)
	Pelecanidae	2	10	35	563
	Threskiornithidae	10	14	27	504
Phoenicopteriformes	Phoenicopteridae	6	8	13	187
Podicipediformes	Podicipedidae	10 (4)	24 (6)	561 (389)	3006 (426)
Procellariiformes	Procellariidae	1	1	0	17
Suliformes	Anhingidae	2	4	13	271

	Phalacrocoracidae	4 (2)	16 (2)	334 (161)	2447 (161)
	Sulidae	2	2	29	93
All Orders	**All Families**	**111 (24)**	**311 (39)**	**861 (769)**	**5081 (1246)**

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27741v2 | CC BY 4.0 Open Access | rec: 21 Jun 2019, publ: 21 Jun 2019

Table 2. Covariates used to perform site matching. First, the two categorical variables (anthrome and region) were used for exact matching (shown by a + sign). Next, all continuous variables were assessed for collinearity and highly collinear variables were removed. Those with an asterisk were retained, and were used in matching.

Category and reason for inclusion	**Variable**	**Used in Matching**	**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 (Harris *et al*., 2014)	0.5°, 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 (°C)				Mean, min, max of months Jan-Dec
	Minimum annual temperature (°C)				
	Maximum annual temperature (°C)	*			
	Mean temperature December – February (°C)				Mean, min, max of Dec previous year and Jan & Feb current year
	Minimum temperature December – February (°C)				
	Maximum temperature December – February (°C)	*			
Fertiliser input. Eutrophication can affect waterbird populations (Lehikoinen *et al*., 2016), and can also 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, 2017	0.5°, 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 PA designation – PAs are less likely to be designated in areas suitable for agriculture and farming (Joppa & Pfaff, 2009).	Anthrome (categorical)	+	HYDE 3.2.001 (Hurtt *et al*., 2011)	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)				

Protected areas are more likely to be designated in areas far from humans (Joppa & Pfaff, 2009), and human presence can also affect waterbird numbers either directly through hunting or through habitat degradation (ref).	Total built up area (km^2 per gridcell)				
	Human population count (inhabitants/gridcell)	*			
	Travel time to nearest city	*	WorldPop (Lloyd *et al*., 2017)	1km, yearly	Spatial bilinear interpolation to 5' grid cells
Governance. Governance in a country is a significant predictor of PA effectiveness (Amano *et al*., 2018), meaning it is important we compare PAs 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 (Kaufmann & Kraay, 2019)	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*., 2016)	30m, 1985-2005	Converted to 5' gridcells by taking sum of 'presence' 30m^2 cells in each
Elevation. PAs are biased towards where they can least prevent land conversion (Joppa & Pfaff, 2009) 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 PAs or not.	Elevation	*	WorldPop (Lloyd *et al*., 2017)	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.	Region (categorical)	+	TM World Borders (Sandvik, 2009)	NA	NA

4. Methods Stage 2 (Proposal)

We aim to answer three questions: 1) Do PAs have a positive impact on waterbird population trends relative to a counterfactual (this includes cases where a PA has lessened, but not halted, a population decline)?; 2) are PAs performing successfully by maintaining or increasing populations? and 3) what factors contribute to PA impact and performance?

Though we found suitable matches for at least one species in nearly all protected sites, not all protected species had suitable matches and so the BACI dataset contains a relatively small subset of all the protected species and is biased towards common species (e.g. see Table 1 – taxa occurring at few sites were less likely to be included in the BACI dataset). As a result, the BACI dataset can only estimate PA impact for *common* species. To also assess PA impact on all species we run all analyses on both the Full Protected and the BACI datasets. A brief summary of the analysis flow is give in Table 3.

Table 3. A brief summary of proposed analyses

	Dataset	**Analysis Summary**	**Report Section**
Question 1 (Do PAs have a positive impact on population trends)	Full Protected	Model quantifying change in population trend directions from before PA designation to after.	4.1.3
	BACI	Model quantifying change in population trend directions from before PA designation to after, when compared to unprotected populations.	4.1.2
Question 2 (Are PAs performing successfully by maintaining or increasing populations)	Full Protected	Model quantifying trend of populations after designation of PAs.	4.2
Question 3 (what factors contribute to PA impact and performance?)	Q1 output (Full Protected)	How species and PA characteristics correlate to PA impact (when quantified comparing trends before and after designation).	4.3.1
	Q1 output (BACI)	How species and PA characteristics correlate to PA impact (when quantified comparing trends before and after designation to those inside and outside PAs).	4.3.2
	Q2 output	How species and PA characteristics correlate to PA performance (absolute post designation trends).	4.3.3

4.1 Question 1. Do PAs have a positive impact on waterbird population trends?

4.1.3 Full Protected Dataset

We will run mixed effects models that consider the effect of protected areas on population trends, first using the full dataset and accounting for the 'before' counterfactual. The models will be run using glmer.nb function from the lme4 (Bates *et al.*, 2015) package in R according to the following formula:

$$\log\left(E\left(Count_{ijz}\right)\right) = \alpha + \beta_1 Y_{ijz} + \beta_2 Y_{ijz} BA_{ijz} + \boldsymbol{\beta}_{5:22} \boldsymbol{CovA}_{ijz} + (1|Species) + (1|Site) + \log(e_i) \qquad (6)$$

Where the count of species *i* in site *j* in year *z* is predicted by the intercept (α), year (*Y*), the interaction term between year and the before/after term (*BA*; 0 in years before protection and 1 in years after protection), the covariates used for matching (***CovA***, shown by asterisks in Table 12), crossed random effects for species and site, and finally the log of the effort term (1 for IWC populations and 1 or >1 for CBC populations; see 'Count Data' in Methods 1). Variance is defined by Equation 2. Depending on model complexity, we may undertake some model selection, comparing AICs between models using different covariate combinations tested. We may also need to rescale variables, or recalculate collinearity between variables if there is higher collinearity in data subsets. All test combinations, rescaling or collinearity issues will be reported. Finally, we may instead choose to run these models individually by species (removing the species random effect) if output is too complex.

A positive β_2 indicates that slopes are more positive after protection than before (as displayed in Fig 1, more positive doesn't necessarily imply absolute positive, just more positive than before protection). The random effects in the model allow for there to be different intercepts for each species and each site (i.e. they can have different abundances), however they do not allow for each species/site to have a different slope and therefore assumes that each population behaves in the same way. A model allowing differing trends would be too complex for this analysis, so we will subset the data into three groups, those where populations are increasing, decreasing or stable before protection, and run the model on each. Interpretation of the results will then depend on the group. For PAs to be having a positive impact, we would expect:

- Where a population trend was non-significant or increasing before designation; anything but a significant negative β_4 coefficient. This is because the goal of a PA is not to always increase species populations, and so to maintain a population, or to do anything but cause a decline in a stable or increasing population, the PA has had a positive impact.
- A significant positive β_4 coefficient for species decreasing before designation.

4.1.2 BACI Dataset

Next, to also account for control/intervention counterfactuals we will run an updated version of model 6 using the matched BACI dataset:

$$\log\left(E\left(Count_{ijz}\right)\right) = \alpha + \beta_1 Y_{ijz} + \beta_2 Y_{ijz} BA_{ijz} + \beta_3 Y_{ijz} CI_{ijz} + \beta_4 Y_{ijz} BA_{ijz} CI_{ijz} + \boldsymbol{\beta}_{5:22} \boldsymbol{CovA}_{ijz} + (1|Species) + (1|Site) + \log(e_i) \quad (7)$$

Where the count of species i in site j in year z is predicted by the intercept (α), year (Y), the interaction term between year and the before/after term (BA; 0 in years before protection and 1 in years after protection; unprotected sites assigned values according to the designation date of their matched protected site), the interaction term between year and the control/intervention term (CI; 0 for unprotected sites and 1 for protected sites), the three way interaction between year, before/after and control/intervention, the covariates used for matching ($\boldsymbol{CovA}$, shown by asterisks in Table 12), crossed random effects for species and site, and finally the log of the effort term (1 for IWC populations and 1 or >1 for CBC populations). Variance is defined by Equation 2. Again, we may undertake some model selection, rescaling collinearity tests, and/or run models individually by species.

As before, a positive β_2 indicates that slopes are more positive after protection than before. A positive β_3 indicates more positive trends inside PAs than outside. And a positive β_4 indicates more positive trends after protection AND inside protected areas. The year term is included in all interaction terms so that they represent slope rather than abundance. $\beta_{1\text{-}4}$ can also tell us about the estimated trend of populations, see Table 4. Also as before, we will subset the data into three groups, those where populations are increasing, decreasing or stable before protection, and run the model on each

Table 4. How to estimate trends of populations in various states of protection, using the coefficients output from Equation 9.

	Before designation	**After designation**
Unprotected	β_1	$\beta_1+\beta_2$
Protected	$\beta_1+\beta_3$	$\beta_1+\beta_2+\beta_3+\beta_4$

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27741v2 | CC BY 4.0 Open Access | rec: 21 Jun 2019, publ: 21 Jun 2019

4.2 Question 2. Are PAs performing successfully by maintaining or increasing populations?

The Question 1 analysis has the ability to detect whether protected areas are having a positive impact, but not whether they are performing successfully by maintaining populations (i.e. if the post protection trend is stable or increasing). For this reason, we will conduct individual population models in the Full Protected dataset to ascertain the trend of each population after PA designation. The models will be simple, structured as follows (using function glm.nb, R package MASS, Venables & Ripley, 2002)

$$\log(E(Count_z)) = \alpha + \beta_1 Y_z + \log(e_z) \tag{8}$$

Where Count of the population in year z is determined by the intercept term, the Year and the log of the effort term (1 for IWC populations and 1 or >1 for CBC populations). Variance is defined by Equation 2. As in the matching methods, and following from the results of Wauchope et al (2019), we will consider the slope of any data of 6 years or more, and significant slopes from data of under 6 years. We will sensitivity check this by also assessing only significant slopes.

We do not need to run this analysis on the BACI dataset as counterfactuals are not required.

4.3 Question 3. What factors correlate to protected area impact and performance?

Finally, we will examine what factors correlate to protected area impact and performance. To do this, we require a value of impact/performance for every population (i.e. site species combination), however for impact we so far only have only one overall value (from Equations 6/7). How we obtain these values is detailed in the sections below. Then, using these population specific values we will run the following mixed effects model (using function lmer, R package lme4, Bates *et al*., 2015):

$$\begin{aligned} E(\beta_{xi}) = \alpha + \boldsymbol{\beta}_{5:10}\boldsymbol{CovB} + (1|Species) \\ + (1|Country) \\ + (1|Country{:}Site) \end{aligned} \tag{9}$$

Where the β_{xi} is the value of impact or performance (either β_1, β_2 or β_4, as detailed below) for protected population i, and is predicted by the covariates in Table 5 (***CovB***), with a random effect of site nested within country, crossed with a random effect for species. We will check for any collinearity between predictor variables before proceeding.

4.3.1 PA Impact by population – Full Protected Dataset

To obtain a value of PA impact for every population, when quantified only according to an 'after' counterfactual, will run the following model for each population in the Full Protected Dataset:

$$\begin{aligned} \log(E(Count_z)) = \alpha + \beta_1 Y_z + \beta_2 Y_z BA_z \\ + \log(e_z) \end{aligned} \tag{10}$$

Where the count of the population in year z is modelled by year (Y) and the interaction between year and the before/after term (BA; 0s in years before protection and 1s in years after), plus the log of the effort term (1 for IWC populations and 1 or >1 for CBC populations). Variance is defined by Equation 2. If the β_2 coefficient is positive, trends are more positive after PA designation. This β_2 coefficient will then be used in model 9.

4.3.2 PA Impact by population - BACI Dataset

To obtain a value of PA impact for every population, when quantified according to an 'after' and 'control' counterfactuals, will run the following model on each matched population pair in the BACI dataset:

$$\begin{aligned} \log(E(Count_z)) = \alpha + \beta_1 Y_{jz} + \beta_2 Yj_z BA_{jz} \\ + \beta_3 Y_{jz} CI_{jz} + \beta_4 Y_{jz} BA_{jz} CI_{jz} \\ + \log(e_z) \end{aligned} \tag{11}$$

Where the count of the population in site j (either the protected or the unprotected site) in year z is modelled by year (Y) and the interaction between year and the before/after term (BA; 0s in years before protection and 1s in years after), the interaction term between year and the control/intervention term (CI; 0 for unprotected sites and 1 for protected sites) and the three way interaction between year, before/after and control/intervention), plus the log of the effort term (1 for IWC populations and 1 or >1 for CBC populations). Variance is defined by Equation 2. If the β_4 coefficient is positive, trends are more

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27741v2 | CC BY 4.0 Open Access | rec: 21 Jun 2019, publ: 21 Jun 2019

positive after PA designation, inside the PA. This β_4 coefficient will then be used in model 9.

4.3.3 PA Performance by population

This is already quantified in Question 2 (section 4.2, model 8). The β_1 value from this will be used in model 9

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27741v2 | CC BY 4.0 Open Access | rec: 21 Jun 2019, publ: 21 Jun 2019

Table 5. Covariates used to assess what factors affect PA effectiveness

Category	**Variable and Reason for Inclusion**	**Category/Levels**	**Source**
Species	Body Mass (g) We expect larger species to respond better to PAs (Barnes et al). This is due to the fact that larger bodied species are more vulnerable to hunting	Continuous	Birdlife.org
	Taxonomic group Different taxonomic groups may respond differently and so we intent to investigate this.	Categorical: Family	Birdlife.org
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 PAs (and it beyond the scope of this study to conside migratory networks). Some species are migrants in parts of their range and non-migrant in others, so we will categorise each population at each site separately.	Categorical: Non-migrant, Migrant	Birdlife.org
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. However, we also expect that populations in these regions will generally be more stable/increasing (i.e. above the horizontal line of Fig 1b) because of increased distance from possible impacts of mismanagement such as poaching and habitat degradation.	Categorical: Urban, Village, Croplands, Rangeland, Semi-natural, Wild	HYDE (Hurtt *et al*., 2011; see Table 1)
	PA Size. We expected larger PAs will perform better, because of reduced edge effects.	Continuous	World Database on Protected Areas (UNEP-WCMC & IUCN, 2019)
	~~PA Management~~. Unfortunately the Management Effectiveness Tracking Tool (METT) is not available for all Protected Areas, and is biased away from Europe and the USA, unlike our dataset. Only 7 of the protected areas in our dataset are included in the METT, therefore we do not feel we can do a fair assessment of management effectiveness.	~~Continuous~~	~~Global Database of Protected Area Management Effectiveness (UNEP-WCMC & IUCN, 2019), Management Effectiveness Tracking Tool (METT; WWF International, 2007).~~
Country	Governance. We expect sites in better governed areas to respond better to PAs (this was suggested in the study by Amano et al which was based on associations).	Continuous	World Bank (Kaufmann & Kraay, 2019; see Table 1)

6. References

Ament JM, Cumming GS (2016) Scale dependency in effectiveness, isolation, and social-ecological spillover of protected areas. *Conservation Biology*, **30**, 846:855.

Amano T, Székely T, Sandel B et al. (2018) Successful conservation of global waterbird populations depends on effective governance. *Nature*, **553**, 199-202.

Austin PC (2009) Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. *Statistics in medicine*, **28**, 3083–3107.

Baillie JEM, Joppa L, Robinson JG (2016) Introduction: Do Protected Areas Safeguard Biodiversity? *Protected Areas: Are They Safeguarding Biodiversity?*, 1–10.

Barnes MD, Craigie ID, Harrison LB et al. (2016) Wildlife population trends in protected areas predicted by national socio-economic metrics and body size. *Nature Communications*, **7**, 12747.

Barnes MD, Glew L, Wyborn C, Craigie ID (2018) Prevent perverse outcomes from global protected area policy. *Nature Ecology & Evolution*, **2**, 759–762.

Bates D, Maechler M, Bolker B, Walker S (2015) Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, **67**, 1.

Baylis K, Honey-Rosés J, Börner J et al. (2016) Mainstreaming Impact Evaluation in Nature Conservation. *Conservation Letters*, **9**, 58–64.

Blom A, Yamindou J, Prins HHT (2004) Status of the protected areas of the Central African Republic. *Biological Conservation*, **118**, 479–487.

Bruner AG, Gullison RE, Rice RE, da Fonseca GAB (2001) Effectiveness of Parks in Protecting Tropical Biodiversity. *Science*, **291**, 125–128.

Buuren S van, Groothuis-Oudshoorn K (2011) mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software*, **45**, 1–67.

CBD (2019) *Synthesis of views of parties and observers on the scope and content of the post-2020 global biodiversity framework*.

Clark NE, Boakes EH, McGowan PJK, Mace GM, Fuller RA (2013) Protected Areas in South Asia Have Not Prevented Habitat Loss: A Study Using Historical Models of Land-Use Change. *PLOS ONE*, **8**, 1–7.

Coetzer KL, Witkowski ETF, Erasmus BFN (2014) Reviewing Biosphere Reserves globally: effective conservation action or bureaucratic label? *Biological Reviews*, **89**, 82–104.

Convention on Biological Diversity (2011) TARGET 11 - Technical Rationale extended.

Ferraro PJ (2009) Counterfactual thinking and impact evaluation in environmental policy. In Birnbaum M & Mickwitz P (Eds) *Environmental program and policy evaluation: Addressing methodological challenges*. *New Directions for Evaluation*. **122**, 75-84.

Geldmann J, Barnes M, Coad L, Craigie ID, Hockings M, Burgess ND (2013) Effectiveness of terrestrial protected areas in reducing habitat loss and population declines. *Biological Conservation*, **161**, 230–238.

Geldmann J, Coad L, Barnes MD et al. (2018) A global analysis of management capacity and ecological outcomes in terrestrial protected areas. *Conservation Letters*, **11**, e12434.

Gelman A, Hill J (2006) *Data analysis using regression and multilevel/hierarchical models*. Cambridge university press.

Gill DA, Mascia MB, Ahmadia GN et al. (2017) Capacity shortfalls hinder the performance of marine protected areas globally. *Nature*, **543**, 665.

Gu XS, Rosenbaum PR (1993) Comparison of Multivariate Matching Methods: Structures, Distances, and Algorithms. *Journal of Computational and Graphical Statistics*, **2**, 405–420.

Hardwicke TE, Ioannidis JPA (2018) Mapping the universe of registered reports. *Nature Human Behaviour*, **2**, 793–796.

Harris I, Jones PD, Osborn TJ, Lister DH (2014) Updated high-resolution grids of monthly climatic observations - the CRU TS3.10 Dataset. *International Journal of Climatology*, **34**, 623–642.

Ho D, Imai K, King G, Stuart EA (2011) MatchIt: nonparametric preprocessing for parametric causal inference. *Journal of Statistical Software*, **42**, 1–28.

Hurtt GC, Chini LP, Frolking S et al. (2011) 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.

IUCN (2018) *IUCN views on the preparation, scope and content of the Post-2020 global biodiversity framework*.

Joppa LN, Pfaff A (2009) High and Far: Biases in the Location of Protected Areas. *PLOS ONE*, **4**, e8273.

Joppa LN, Pfaff A (2011) Global protected area impacts. *Proceedings of the Royal Society B: Biological Sciences*, **278**, 1633–1638.

Kaufmann D, Kraay A (2019) Worldwide Governance Indicators (WGI). *The World Bank Group*.

Kerr NL (1998) HARKing: Hypothesizing After the Results are Known. *Personality and Social Psychology Review*, **2**, 196–217.

King G, Nielsen R *Why Propensity Scores should not be used for matching*.

Laurance WF, Carolina Useche D, Rendeiro J et al. (2012) Averting biodiversity collapse in tropical forest protected areas. *Nature*, **489**, 290.

Lehikoinen A, Rintala J, Lammi E, Pöysä H (2016) Habitat-specific population trajectories in boreal waterbirds: Alarming trends and bioindicators for wetlands. *Animal Conservation*, **19**, 88–95.

Leverington F, Costa KL, Pavese H, Lisle A, Hockings M (2010) A Global Analysis of Protected Area Management Effectiveness. *Environmental Management*, **46**, 685–698.

Lloyd CT, Sorichetta A, Tatem AJ (2017) Data Descriptor: High resolution global gridded data for use in population studies Background & Summary. 1–17.

Lu C, Tian H (2017) 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.

McDonald TL, Erickson WP, McDonald LL (2000) Analysis of count data from before-after control-impact studies. *Journal of Agricultural, Biological, and Environmental Statistics*, 262–279.

Naimi B, Hamm N a S, Groen T a., Skidmore AK, Toxopeus AG (2014) Where is positional uncertainty a problem for species distribution modelling? *Ecography*, **37**, 191–203.

Nelson A, Chomitz KM (2009) *Protected area effectiveness in reducing tropical deforestation*. Washington D. C.

Nosek BA, Ebersole CR, DeHaven AC, Mellor DT (2018) The preregistration revolution. *Proceedings of the National Academy of Sciences*, **115**, 2600–2606.

Pekel J-F, Cottam A, Gorelick N, Belward AS (2016) High-resolution mapping of global surface water and its long-term changes. *Nature*, **540**, 418–422.

Pringle RM (2017) Upgrading protected areas to conserve wild biodiversity. *Nature*, **546**, 91.

Rosenbaum PR, Rubin DB (1983) The central role of the propensity score in observational studies for causal effects. *Biometrika*, **70**, 41–55.

Rubin DB (2001) Using Propensity Scores to Help Design Observational Studies: Application to the Tobacco Litigation. *Health Services and Outcomes Research Methodology*, **2**, 169–188.

Salmerón Gómez R, García Pérez J, López Martín MDM, García CG (2016) Collinearity diagnostic applied in ridge estimation through the variance inflation factor. *Journal of Applied Statistics*, **43**, 1831–1849.

Sandvik B (2009) World Borders Dataset. *Thematic Mapping*.

Scharlemann JPW, Kapos V, Campbell A et al. (2010) Securing tropical forest carbon: the contribution of protected areas to REDD. *Oryx*, **44**, 352–357.

Simmons BI, Hoeppke C, Sutherland WJ (2019) Beware greedy algorithms. *Journal of Animal Ecology*, **88**, 804-807.

Stuart EA (2010) 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.

UNEP-WCMC, IUCN (2019) The World Database on Protected Areas (WDPA)/The Global Database on Protected Areas Management Effectiveness (GD-PAME).

UNEP-WCMC, IUCN and NGS (2018) *Protected Planet Report 2018*. Cambridge, UK; Gland, Switzerland; and Washington D.C., USA.

Venables WN, Ripley BD (2002) *Modern Applied Statistics with S*, 4th edn. Springer, New York.

Watson JEM, Dudley N, Segan DB, Hockings M (2014) The performance and potential of protected areas. *Nature*, **515**, 67.

Watson JEM, Darling ES, Venter O et al. (2016) Bolder science needed now for protected areas. *Conservation Biology*, **30**, 243–248.

Wauchope HS, Johnston A, Amano T, Sutherland WJ (2019) When can we trust population trends? A method for quantifying the effects of sampling interval and duration. *bioRxiv*, 498170.

WWF International (2007) *Management Effectiveness Tracking Tool*. Gland, Switzerland.