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

PRE-ANALYSIS PLAN

Hannah Wauchope1*, Julia P G Jones2, Tatsuya Amano1,3,4, Jonas Geldmann1, Daniel Blanco5, Richard A. Fuller3, Tom Langendoen5, Taej Mundkur5, Benno I. Simmons1, Nagy Szabolcs5, William J. Sutherland1.

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

*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 9,650 waterbird populations (here defined as a site species combination), consisting of 262 species at 546 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 3,677 populations consisting of 94 species at 514 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 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; Goldmann 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 (Goldmann 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 (Goldmann 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, Figure 1a,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 9,650 waterbird populations (here defined as a site species combination) and 262 species at 546 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 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:

\[
\log(E(\text{Count}_i)) = \beta \log(e_i)
\]

\[
\text{var}(\text{Count}_i) = v_{NB}(E(\text{Count}_i))
\]

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. \(\beta\) 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 (Coetzee 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 9,650 populations, consisting of 262 species at 546 protected sites (Figure 3, blue points, Table 1).

We next created a dataset of unprotected sites and counts at those sites, by taking only those sites that did not interact with any protected areas in the WDPA (not only our cleaned set of PAs, to avoid including, for example, sites that were protected but without a designation date). We then restricted this dataset to only the species present in the Full Protected dataset. This gave us an unprotected dataset of 182,180 populations consisting of 262 species and 8925 sites (Figure 3, green points, Table 1).
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:

$$md_{(x,y)} = \sqrt{(x - y)^T S^{-1}(x - y)}$$ (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). For each species, these were then combined into a larger distance matrix containing all the sites that 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):

\[
\log(E(\text{Count}_z)) = \alpha + \beta_1 Y_z + \log(e_z) + D_z = IWC
\]

Where \( e_z \) is the values of covariate \( z \) is predicted by the Year, and an effort term if the data was taken from the CBC. Variance is expressed by Eq. 2. The \( \beta_1 \) 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.

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):

\[
d_{\text{cov}} = \frac{T_{\text{cov}} - C_{\text{cov}}}{\sqrt{\frac{\text{var}(T_{\text{cov}}) + \text{var}(C_{\text{cov}})}{2}}}
\]

Where \( T_{\text{cov}} \) is the values of covariate \( cov \) for protected sites (mean from the years before and equal to designation), \( C_{\text{cov}} \) is the same for
unprotected sites, \( \text{var} \) is the variance of each of these and \( d_{\text{cov}} \) 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 94 species, 3677 population pairs and 514 protected sites paired to 1286 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.

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>1999</td>
<td>2002</td>
<td>2004</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>X</td>
<td>2</td>
<td>9</td>
<td>A</td>
<td>B</td>
<td>C</td>
</tr>
<tr>
<td>Y</td>
<td>5</td>
<td>6</td>
<td>X</td>
<td>Y</td>
<td>Z</td>
</tr>
<tr>
<td>Z</td>
<td>4</td>
<td>1</td>
<td>E</td>
<td>F</td>
<td>X</td>
</tr>
</tbody>
</table>

**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=514) and unprotected (green, n=1286) 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.

<table>
<thead>
<tr>
<th>Order</th>
<th>Family</th>
<th>Genera</th>
<th>Species</th>
<th>Protected Sites</th>
<th>Unprotected Sites</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anseriformes</td>
<td>Anatidae</td>
<td>40 (32)</td>
<td>130 (76)</td>
<td>528 (468)</td>
<td>8501 (889)</td>
</tr>
<tr>
<td>Charadriiformes</td>
<td>Alcidae</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>73</td>
</tr>
<tr>
<td></td>
<td>Charadriidae</td>
<td>8 (6)</td>
<td>44 (12)</td>
<td>169 (122)</td>
<td>4054 (167)</td>
</tr>
<tr>
<td></td>
<td>Haematopodidae</td>
<td>2 (2)</td>
<td>12 (2)</td>
<td>65 (47)</td>
<td>624 (47)</td>
</tr>
<tr>
<td></td>
<td>Jacanidae</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>392</td>
</tr>
<tr>
<td></td>
<td>Laridae</td>
<td>18 (2)</td>
<td>70 (16)</td>
<td>196 (106)</td>
<td>5286 (117)</td>
</tr>
<tr>
<td></td>
<td>Recurvirostrida</td>
<td>6 (2)</td>
<td>10 (2)</td>
<td>52 (25)</td>
<td>1305 (25)</td>
</tr>
<tr>
<td></td>
<td>Scolopacidae</td>
<td>22 (16)</td>
<td>84 (34)</td>
<td>177 (131)</td>
<td>4119 (228)</td>
</tr>
<tr>
<td></td>
<td>Stercorariidae</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td>84</td>
</tr>
<tr>
<td>Ciconiiformes</td>
<td>Ciconidae</td>
<td>8</td>
<td>14</td>
<td>14</td>
<td>858</td>
</tr>
<tr>
<td>Gaviiformes</td>
<td>Gaviidae</td>
<td>2 (2)</td>
<td>6 (2)</td>
<td>19 (6)</td>
<td>1211 (6)</td>
</tr>
<tr>
<td>Gruiformes</td>
<td>Aramidae</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>93</td>
</tr>
<tr>
<td></td>
<td>Gruidae</td>
<td>6</td>
<td>8</td>
<td>6</td>
<td>220</td>
</tr>
<tr>
<td></td>
<td>Rallidae</td>
<td>20 (8)</td>
<td>28 (10)</td>
<td>421 (232)</td>
<td>6246 (256)</td>
</tr>
<tr>
<td>Pelicaniformes</td>
<td>Ardeidae</td>
<td>18 (10)</td>
<td>36 (16)</td>
<td>316 (167)</td>
<td>6284 (219)</td>
</tr>
<tr>
<td></td>
<td>Pelecanidae</td>
<td>2</td>
<td>10</td>
<td>16</td>
<td>797</td>
</tr>
<tr>
<td></td>
<td>Threskiornithida</td>
<td>8</td>
<td>14</td>
<td>18</td>
<td>958</td>
</tr>
<tr>
<td>Phoenicopteriformes</td>
<td>Phoenicopterida</td>
<td>4</td>
<td>4</td>
<td>8</td>
<td>280</td>
</tr>
<tr>
<td>Podicipediformes</td>
<td>Podicipedidae</td>
<td>10 (6)</td>
<td>20 (12)</td>
<td>374 (177)</td>
<td>5751 (202)</td>
</tr>
<tr>
<td>Suliformes</td>
<td>Anhingidae</td>
<td>2</td>
<td>4</td>
<td>5</td>
<td>496</td>
</tr>
<tr>
<td></td>
<td>Phalacrocoracida</td>
<td>4 (2)</td>
<td>16 (4)</td>
<td>218 (22)</td>
<td>4973 (22)</td>
</tr>
<tr>
<td></td>
<td>Sulidae</td>
<td>2 (2)</td>
<td>2 (2)</td>
<td>8 (3)</td>
<td>169 (3)</td>
</tr>
<tr>
<td><strong>All Orders</strong></td>
<td><strong>All Families</strong></td>
<td><strong>96 (45)</strong></td>
<td><strong>261 (94)</strong></td>
<td><strong>546 (514)</strong></td>
<td><strong>8925 (1286)</strong></td>
</tr>
</tbody>
</table>
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.

<table>
<thead>
<tr>
<th>Category and reason for inclusion</th>
<th>Variable</th>
<th>Used in Matching</th>
<th>Data source</th>
<th>Resolution</th>
<th>Data transformation</th>
</tr>
</thead>
<tbody>
<tr>
<td>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.</td>
<td>Total annual precipitation (mm)</td>
<td>*</td>
<td>CRU TS4.01 (Harris et al., 2014)</td>
<td>0.5°, monthly (1961-2016)</td>
<td>Yearly sum of Jan-Dec</td>
</tr>
<tr>
<td></td>
<td>Total precipitation December – February (mm)</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mean annual temperature (°C)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Minimum annual temperature (°C)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Maximum annual temperature (°C)</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mean temperature December – February (°C)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Minimum temperature December – February (°C)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Maximum temperature December – February (°C)</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>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.</td>
<td>Nitrogen (g N/m² cropland/yr)</td>
<td>*</td>
<td>Lu &amp; Tian, 2017</td>
<td>0.5°, yearly (1961–2013)</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>Phosphorous (g P/m² cropland/yr)</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>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 &amp; Pfaff, 2009).</td>
<td>Anthrome (categorical)</td>
<td>+</td>
<td>HYDE 3.2.001 (Hurt et al., 2011)</td>
<td>5', centennial (10,000BC-1600AD) decadal (1700-2000), yearly (2001-2016)</td>
<td>Pre-2000 data taken from nearest decade</td>
</tr>
<tr>
<td></td>
<td>Grazing land (km²/gridcell)</td>
<td></td>
<td></td>
<td></td>
<td>Temporal linear interpolation to obtain yearly data between decades of 1960-2000</td>
</tr>
<tr>
<td></td>
<td>Irrigated land (not rice; km²/gridcell)</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Irrigated land (rice; km²/gridcell)</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pasture land (km²/gridcell)</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Rangeland (km²/gridcell)</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Rainfed crop land (no rice; km²/gridcell)</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Rainfed crop land (rice; km²/gridcell)</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Human presence.</td>
<td>Human population density (inhabitants/km² pergridcell)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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).

<table>
<thead>
<tr>
<th>Protected areas are more likely to be designated in areas far from humans (Joppa &amp; Pfaff, 2009), and human presence can also affect waterbird numbers either directly through hunting or through habitat degradation (ref).</th>
</tr>
</thead>
</table>

**Total built up area (km² per gridcell)**

**Human population count** *(inhabitants/gridcell)*

**Travel time to nearest city**

**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)**

**Surface water (presence/absence)**

**Elevation**

**Region (categorical)**

**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.

**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 taken across all years because data is only available from 1996. Therefore just one value per site for all years.**

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)

**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)**

**Surface water (presence/absence)**

**Elevation**

**Region (categorical)**

**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.
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?

Not all protected populations had suitable matches and so the BACI dataset contains fewer protected populations than the Full Protected dataset 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>
<thead>
<tr>
<th>Question 1 (Do PAs have a positive impact on population trends)</th>
<th>Dataset</th>
<th>Analysis Summary</th>
<th>Report Section</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full Protected</td>
<td>Model quantifying change in population trend directions from before PA designation to after.</td>
<td>4.1.3</td>
<td></td>
</tr>
<tr>
<td>BACI</td>
<td>Model quantifying change in population trend directions from before PA designation to after, when compared to unprotected populations.</td>
<td>4.1.2</td>
<td></td>
</tr>
</tbody>
</table>

| 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:

\[
\log(E(\text{Count}_{ij})) = \alpha + \beta_1 Y_{ij} + \beta_2 Y_{ij} \cdot BA_{ij} + \beta_{5:22} CovA_{ij} + (1|\text{Species}) + (1|\text{Site}) + \log(\varepsilon_i)
\]

Where the count of species \(i\) in site \(j\) in year \(z\) is predicted by the intercept \((\alpha)\), 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.

A positive $\beta_2$ 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 $\beta_4$ 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 $\beta_4$ 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:

$$\log\left(E(\text{Count}_{ijz})\right) = \alpha + \beta_1 Y_{ijz} + \beta_2 Y_{ijz} BA_{ijz} + \beta_3 Y_{ijz} CI_{ijz} + \beta_4 Y_{ijz} BA_{ijz} CI_{ijz} + \beta_{5:22} CovA_{ijz} + (1|\text{Species}) + (1|\text{Site}) + \log(\epsilon_i)$$

Where the count of species $i$ in site $j$ in year $z$ is predicted by the intercept ($\alpha$), 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 ($CovA$, 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 and/or collinearity tests.

As before, a positive $\beta_2$ indicates that slopes are more positive after protection than before. A positive $\beta_3$ indicates more positive trends inside PAs than outside. And a positive $\beta_4$ 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. $\beta_{1:4}$ 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.

<table>
<thead>
<tr>
<th></th>
<th>Before designation</th>
<th>After designation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unprotected</td>
<td>$\beta_1$</td>
<td>$\beta_1 + \beta_2$</td>
</tr>
<tr>
<td>Protected</td>
<td>$\beta_1 + \beta_3$</td>
<td>$\beta_1 + \beta_2 + \beta_3 + \beta_4$</td>
</tr>
</tbody>
</table>

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(\text{Count}_z)) = \alpha + \beta_1Y_z + \log(e_z) \quad (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):

\[ E(\beta_{zi}) = \alpha + \beta_{z1,\text{Species}} \text{CovB} + (1|\text{Species}) \]
\[ + (1|\text{Country}) \]
\[ + (1|\text{Country}:\text{Site}) \]

Where the \( \beta_{zi} \) is the value of impact or performance (either \( \beta_1, \beta_2 \) or \( \beta_4 \) 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:

\log(E(\text{Count}_z)) = \alpha + \beta_1Y_{iz} + \beta_2Y_{ijz}\text{BA}_{iz} \quad \text{model 8} \]
\[ + \log(e_z) \]

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 \( \beta_2 \) coefficient is positive, trends are more positive after PA designation. This \( \beta_2 \) 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:

\log(E(\text{Count}_z)) = \alpha + \beta_1Y_{iz} + \beta_2Y_{ijz}\text{BA}_{iz} \quad \text{model 9} \]
\[ + \beta_3Y_{ijz}\text{Cl}_{iz} + \beta_4Y_{ijz}\text{BA}_{ijz}\text{Cl}_{iz} \]
\[ + \log(e_z) \]

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 (Cl; 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 \( \beta_4 \) coefficient is positive, trends are more positive after PA designation, inside the PA. This \( \beta_4 \) 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 \( \beta_1 \) value from this will be used in model 9.
Table 5. Covariates used to assess what factors affect PA effectiveness

<table>
<thead>
<tr>
<th>Category</th>
<th>Variable and Reason for Inclusion</th>
<th>Category/Levels</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Species</td>
<td>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.</td>
<td>Continuous</td>
<td>Birdlife.org</td>
</tr>
<tr>
<td></td>
<td>Taxonomic group. Different taxonomic groups may respond differently and so we intend to investigate this.</td>
<td>Categorical: Family</td>
<td>Birdlife.org</td>
</tr>
<tr>
<td>Species</td>
<td>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 consider migratory networks). Some species are migrants in parts of their range and non-migrant in others, so we will categorize each population at each site separately.</td>
<td>Categorical: Non-migrant, Migrant</td>
<td>Birdlife.org</td>
</tr>
<tr>
<td>Site (nested within Site)</td>
<td>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 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.</td>
<td>Categorical: Urban, Village, Croplands, Rangeland, Semi-natural, Wild</td>
<td>HYDE (Hurtt et al., 2011; see Table 1)</td>
</tr>
<tr>
<td></td>
<td>PA Size. We expected larger PAs will perform better, because of reduced edge effects.</td>
<td>Continuous</td>
<td>World Database on Protected Areas (UNEP-WCMC &amp; IUCN, 2019)</td>
</tr>
<tr>
<td></td>
<td>PA Management. This is not available for all PAs, but we will assess correlations for the PAs for which there are data. We expect that better managed PAs will perform better, as they will succeed more in reducing hunting and habitat degradation. We will assess the performance of PAs using METT questions 12 -19, which relate to Capacity and Resources (see Geldmann et al., 2018).</td>
<td>Continuous</td>
<td>Global Database of Protected Area Management Effectiveness (UNEP-WCMC &amp; IUCN, 2019), Management Effectiveness Tracking Tool (METT; WWF International, 2007).</td>
</tr>
<tr>
<td>Country</td>
<td>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).</td>
<td>Continuous</td>
<td>World Bank (Kaufmann &amp; Kraay, 2019; see Table 1)</td>
</tr>
</tbody>
</table>
6. References


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


Watson JEM, Darling ES, Venter O et al. (2016)
