

Measuring the impact of conservation on species' populations

Sean Jellesmark

A dissertation submitted for the degree of Doctor of Philosophy

University College London

January 2022

Supervised by:

Prof. Tim M. Blackburn

Prof. Richard D. Gregory

Mike Hoffmann

Louise McRae

Dr. Piero Visconti

I, Sean Jellesmark, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Sean Jellesmark, 31st January 2022

Abstract

The global biodiversity crisis has sparked a rise in conservation actions. However, the impact of conservation actions on species' populations is often not evaluated. In particular, studies of the impact of large-scale conservation, such as area protection or national legislation aimed at safeguarding biodiversity, on population trends remain poorly represented in the scientific literature. The resulting limited knowledge of conservation effectiveness potentially hinders effective evidence-based decision making and leads to suboptimal conservation outcomes.

Here, I collate longitudinal records of vertebrate abundance, conservation interventions and ancillary information to measure the impact of conservation on species' populations across different spatial and temporal scales. Specifically, I estimate the impact of lowland wet grassland reserves on breeding trends of wading birds in the United Kingdom by creating counterfactual reserve trends from national monitoring records collected by volunteers. This demonstrates that lowland wet grassland reserves have a positive impact on targeted species of wading birds. To understand the drivers of breeding abundance within these reserves, I combine local management records with climatic variables and breeding counts and use a Bayesian mixed modelling framework to estimate the association between conservation actions, site-specific conditions and annual breeding abundance for four wading bird species. The resulting estimates show that breeding abundance associates with different factors for the four species and provides new information on conservation effectiveness that can be used to inform local reserve management. Last, I explore the impact of conservation actions on trends in vertebrate populations worldwide. I categorize conservation actions for more than 26,000 populations, create

counterfactual population indices representing how populations may have developed in the absence of conservation, calculate the impact of conservation on a global population index, and estimate how specific conservation actions relate to population changes. I show that conservation benefits targeted vertebrate populations, that in the absence of conservation, a global index of vertebrate abundance could have declined in addition to what is currently observed and that, in particular, conservation through species and land & water management has a positive impact on targeted populations.

The methods developed and applied in this thesis demonstrate how to estimate the impact of conservation actions on species' populations. This work also highlights the potential of longitudinal abundance records for evaluating conservation impact and emphasizes the importance of large-scale monitoring programmes.

Impact statement

To halt and reverse the widespread decline of global biodiversity, understanding the impact of conservation is imperative. This thesis demonstrates how to measure the impact of conservation on vertebrate species' populations using time series population data and counterfactual methodology. A range of different methods are utilized to evaluate the impact of lowland wet grassland reserves on breeding abundance of wading birds in the United Kingdom and more generally, the impact of conservation on global abundance of vertebrate populations. Outputs from this thesis demonstrate the potential of monitoring records for large-scale impact evaluation in conservation science and contribute to the evidence-base of conservation effectiveness. The findings have been used to inform the wider public about the impact of wetland reserves and policy makers about the impact of global conservation and conservations' importance for biodiversity moving forward. Hopefully, this research will encourage researchers and conservation practitioners to utilize monitoring records and, to a larger extent, evaluate conservation impact with an increased focus on factors affecting the reliability of impact estimates such as study design. This has resulted in the following publications:

- Jellesmark S, Blackburn TM, Dove S, Geldmann J, Visconti P, Gregory RD, McRae L, Hoffmann M. 2022. Assessing the global impact of targeted conservation actions on species abundance. Preprint BioRxiv. <https://doi.org/10.1101/2022.01.14.476374>
- Jellesmark S, Ausden M, Blackburn TM, Hoffmann M, McRae L, Visconti P, Gregory RD. 2021. The effect of conservation interventions on the abundance of breeding waders within nature reserves in the United Kingdom. *Ibis*. In review

- Jellesmark, S., Ausden, M., Blackburn, T.M., Gregory, R.D., Hoffmann, M., Massimino, D., McRae, L. and Visconti, P. (2021), A counterfactual approach to measure the impact of wet grassland conservation on U.K. breeding bird populations. *Conservation Biology*, 35: 15751585. <https://doi.org/10.1111/cobi.13692>

This research has additionally contributed to the document for the post-2020 global biodiversity framework “Expert input to the post-2020 global biodiversity framework: Transformative actions on all the drivers of biodiversity loss are urgently required to achieve the global goals by 2050”, and the publication “Dove S, Freeman R, Böhm M, Jellesmark S, Murrell D. A user-friendly guide to using distance measures to compare time series in ecology. *Methods in Ecology and Evolution*. In review“

Acknowledgements

This PhD would not have been possible without the support of so many amazing people. Foremost, I would like to thank my supervisors - I could not have asked for a better group. Tim and Richard, thank you for your continuous support and guidance – you have done more than what could be expected of any supervisor. Without you, the many lockdowns would have deprived me of my sanity. Louise and Mike, thank you for the always positive comments and expert guidance on the many aspects of biodiversity, policy and, in general, navigating through the scientific world that I knew so little about. Piero, the research in this thesis would not have been possible without your ideas and help and I am deeply grateful for that. As a supervisory group, you have been perfect, and I truly believe that I could not have asked for better or more competent guidance.

A special thank you to Malcolm - for sharing your knowledge about waders, tolerating my lack thereof, endless questions, and bringing me along on a wonderful field trip. I can only hope to achieve your level of expertise someday. A broad thanks to the rest of the Royal Society for the Protection of Birds, especially William being ever helpful with data acquisition and any other technical enquiry that I could come up with. I would also like to thank Dario from the British Trust for Ornithology for his technical guidance and all the volunteers that have devoted thousands of hours collecting the data analysed in this thesis. You are doing amazing work.

My thanks to colleagues and friends from CBER and ZSL. Both institutions have helped make this PhD a wonderful journey. A special thanks to my fellow Blackbeard's Shawn, Gonzalo and Bouve. Also, Ben, Adrienne, Iona, Declan, Jo, Jess, Charlie,

Amy and many more with whom I have enjoyed countless interesting conversations, discussions and pints.

I would also like to express my gratitude towards the Inspire4Nature group and the funders, the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant. I have made many friends along this journey, all of whom I am sure will continue to accomplish amazing things.

Finally, I would like to thank my family. Despite the physical distance between us, you have always been supportive and cheered for me. A special thanks to my fiancée who decided to uproot and move with me to London when this PhD opportunity arose. When I lose faith in myself you still believe. For that, and a million other reasons, I love you.

Table of contents

Abstract	3
Impact statement.....	5
Acknowledgements	7
Table of contents.....	9
List of Tables	11
List of Figures.....	12
1 Introduction.....	15
Biodiversity and Conservation	15
A historical perspective on evaluation.....	17
Evaluation and study designs in conservation science	26
Thesis overview	33
References	35
2 A counterfactual approach to measure the impact of wet grassland conservation on U.K. breeding bird populations	49
Abstract	50
Introduction.....	50
Methods.....	55
Data	55
Data analyses	60
Results.....	62
Discussion	65
References	70
Supplementary material.....	77
3 The effect of conservation interventions on the abundance of breeding waders within nature reserves in the United Kingdom	93
Abstract	94
Introduction.....	94
Method.....	96
Datasets.....	96
Site selection and count data	97
Conservation interventions.....	99
Climatic data	99

Other covariates.....	100
Pre - analysis	100
Models	100
Results.....	105
Discussion	109
References	113
Supplementary material.....	117
4 Assessing the global impact of targeted conservation actions on species	
abundance	127
Summary	128
Introduction.....	129
Methods.....	131
The Living Planet Database	131
Database structure.....	133
Data selection, management data and coding	135
Matching populations and calculating trends	137
Calculating the impact of conservation on a global index of species abundance	
.....	142
Mixed model.....	143
Results.....	144
Conservation actions in the LPD	144
Impact of conservation under four different counterfactual scenarios	147
Impact of conservation on global species abundance.....	149
Drivers of population trends estimated from mixed models.....	150
Discussion	151
References	155
Supplementary material.....	163
5 Discussion	176
Overview	176
Advances resulting from this work.....	178
Study limitations and future research	181
References	183

List of Tables

Table 2.1 Variables used to create the counterfactuals.....	58
Table S2.1 Number of sites or grids available for calculating reserve and counterfactual indices.....	78
Table S2.2 Percentage of data points observed from the total number of data points used to calculate reserve and counterfactual trends.....	80
Table S2.3 Number of transect and the percentage of primary habitat types used to calculate counterfactual trends.....	80
Table S2.4 Proposed relationship between wetland management and species' benefit.....	81
Table S3.1 Variable descriptions.....	117
Table 4.1 Types of population data that meets the data standards for tracking trends in the Living Planet Database.....	132
Table 4.2 Variable descriptions.....	141
Table S4.1 Parameter estimates and SE for conservation actions.....	165

List of Figures

Figure 2.1 Hypothesized effect of lowland wet grassland conservation on breeding bird trends.....	59
Figure 2.2 Main breeding trends for the five wading bird species.....	64
Figure 2.3 Counterfactual breeding trends for the five wading bird species.....	65
Figure S2.1 Breeding trends for Black-tailed godwit, Garganey and Shoveler.....	82
Figure S2.2 Breeding trends for Snipe and Yellow Wagtail with and without the Ouse Washes.....	83
Figure S2.3 Counterfactual breeding trends for the five wading bird species including SEs.....	83
Figure S2.4 Main breeding trends for the five wading bird species when excluding the top three sites which contain the most breeding pairs.....	84
Figure S2.5 Relationship between site age and annual change in breeding pair....	85
Figure S2.6 Main breeding trends when excluding sites acquired later than 1994..	86
Figure S2.7 Location of lowland wet grassland reserves and Breeding Bird Survey squares.....	87
Figure 3.1 Study sites under RSPB management.....	98
Figure 3.2 Distributions of breeding pair counts.....	104
Figure 3.3 Posterior parameter estimates and 95% credible intervals for conservation interventions, temperature, rainfall and site size.....	107

Figure 3.4 Posterior estimates and 95% credible intervals for site age.....	108
Figure S3.1 Mean annual number of breeding Lapwing pairs.....	121
Figure S3.2 Mean annual number of breeding Redshank pairs.....	122
Figure S3.3 Mean annual number of breeding Curlew pairs.....	123
Figure S3.4 Mean annual number of breeding Snipe pairs.....	124
Figure S3.5 Lapwing model validation.....	125
Figure S3.6 Redshank model validation.....	126
Figure S3.7 Snipe model validation.....	127
Figure S3.5 Curlew model validation.....	128
Figure 4.1 Location of managed and unmanaged populations.....	146
Figure 4.2 Conservation actions and number of targeted populations.....	147
Figure 4.3 Vertebrate population trends for species subject to conservation actions and potential counterfactual trends in the absence of conservation.....	148
Figure 4.4 Impact of conservation on a global population index.....	149
Figure 4.5 Parameter estimates and SE for conservation actions.....	151
Figure S4.1 Detailed conservation actions and number of targeted populations....	166
Figure S4.2 Time series length for managed populations.....	168
Figure S4.3 First year of observation for managed populations.....	169
Figure S4.4 Sensitivity test for the four counterfactual scenarios.....	170

Figure S4.5 Number of species within taxonomical class for the four counterfactual scenarios.....172

Figure S4.6 Validation plot for link between listed reason for population increases and recorded conservation actions.....173

Figure S4.7 Global population indices with and without the impact of conservation.....174

1 | Introduction

Biodiversity and Conservation

The world is facing a mass extinction, with thousands of species' populations lost over the last century (Barnosky et al. 2011; Ceballos et al. 2015, 2020). Over the last 50 years, populations of vertebrate species, as measured by the Living Planet Index, have declined by 68% on average, with declines of up to 94% for populations in the tropical subregions of the Americas (WWF 2020a). For many, these precipitous declines have ended up with the disappearance of the entire species: extinction rates are tens to thousand time higher than the background rates, and it is estimated that a million more species now face extinction over the course of the next century (Pimm et al. 2014; IPBES 2019).

Conservation actions play a key role in reversing this loss of biodiversity (Hoffmann et al. 2010, 2015; Leclère et al. 2020; Bolam et al. 2020). A plethora of different conservation actions has been developed and applied, from local single-species reintroduction programmes, to multilateral treaties aimed at safeguarding global biodiversity (Salafsky et al. 2008; CBD 2011). The resources available for conserving biodiversity are limited though and so selecting the intervention with the highest impact is imperative to ensure effective conservation (Ferraro & Pattanayak 2006). However, deciding which conservation action is most effective relies on empirical evaluation of former interventions (Ferraro 2009; Baylis et al. 2016). Historically, conservation practice has relied on anecdotal sources, often supported by limited evidence (Sutherland et al. 2004; Margoluis et al. 2009), and thus potentially with reduced effectiveness (Sutherland & Wordley 2017; Christie et al. 2020b).

Recent years have seen a rise in the call for evaluation of conservation programmes and access to conservation evidence (Kleiman et al. 2000; Pullin & Knight 2001; Salafsky et al. 2002; Salafsky & Margoluis 2003; Sutherland et al. 2004; Ferraro & Pattanayak 2006; Ferraro 2009; Sutherland & Wordley 2017). For example, the Conservation Evidence Project (www.conservationevidence.com) has summarized information from thousands of scientific articles on local-scale conservation interventions and provides free and publicly available evidence of their effectiveness. In addition to the increased attention on evaluation efforts, the quality of evidence is receiving attention, such as the importance of study design and credibility of population trends, as an important topic within the scientific community, with widespread implications for conservation practitioners and biodiversity in general (Christie et al. 2019, 2020b, 2020c, 2020a; Wauchope et al. 2019b, 2021).

Measuring the impact of conservation is the main purpose of this thesis. To do so accurately requires methods of causal inference. Causal inference is defined as ‘the leveraging of theory and deep knowledge of institutional details to estimate the impact of events and choices on a given outcome of interest’ (Cunningham 2021). At the core of causal inference lies counterfactual theory – representations of worlds that could have been if conditions were different. Counterfactuals are essential because they allow estimates of treatment effects to be inferred, by comparing an observed outcome with a counterfactual that represents the outcome in the absence of treatment, thus allowing any differences to be attributed to the intervention being evaluated. However, reality dictates that an outcome can be observed only under a single set of conditions and counterfactuals are therefore never observed, but instead must be estimated. How counterfactuals are estimated, and from what, affects the degree to which the counterfactual approximates the truth, and thus the reliability of the estimate. In this

Introduction, I first set the scene by providing a historical perspective on the roots of evaluation, I then describe the evolution of evaluation and study designs in conservation science and, finally, provide a brief overview of each of my chapters and how the thesis is organized.

A historical perspective on evaluation

Arguably, the first documented experimental evaluation is described in “The Book of Daniel” in the Bible. It tells the story of how, in 597 BC, Babylon’s king Nebuchadnezzar wanted the people of his court to eat meat and drink wine exclusively. However, for religious reasons, a young boy of royal blood named Daniel asked that he and three of his friends be given a diet of vegetables instead. To show that the vegetarian diet would not diminish their ability to serve the king, he proposed an experiment: for ten days, Daniel and his friends would follow a strict diet of nothing but legumes and water while another group of children would be fed the king’s meat and wine. After the ten days, the physical state of the children in the two groups were compared. Daniel and his friend appeared better nourished than the group that had eaten nothing but meat and wine and the king permitted them to continue their vegetarian diet (Bhatt 2010; Pearl & Mackenzie 2019). Although this experiment did not follow the scientific rigor of present-day standards, the purpose was similar to what scientists do today – test a hypothesis and produce evidence. Daniel and his three friends received the intervention (vegetarian diet) while the other group served as the control (King’s diet) in order to test the effect of vegetarianism.

The first modern controlled trial on record was performed in 1746, more than 2000 years after Daniel, by the Scottish doctor James Lind. Lind was interested in the treatment of scurvy, a disease that killed an estimated 2 million sailors between 1500

and 1800 AD (Pearl & Mackenzie 2019). He tested the effects of six treatments for 12 patients with scurvy and found a treatment that contained ascorbic acid to be particularly effective. However, his findings had little impact on practice, and it was another 42 years before his proposed treatment for scurvy was introduced by the relevant authorities (Tröhler 2005).

In 1864, the next milestone occurred with the emergence of placebo treatments. A placebo is a type of treatment that has no therapeutic effect, but which can still have a psychological effect. The US physician Austin Flint compared the effectiveness of a placebo drug to an active remedy in treating rheumatic disorder (a condition causing pain in joints or connective tissue). Since wildlife populations do not hold expectations about conservation actions and are therefore not biased in the same way as humans in clinical trials, placebo treatments are rarely necessary in conservation science. However, they are not entirely absent (Cisneros et al. 2015; Neely et al. 2020).

In 1885, not long after the introduction of placebo treatments, randomization appeared in psychophysical experiments with humans (Peirce & Jastrow 1885). In impact evaluation, randomization refers to the process of randomly assigning treatment to units within the eligible population so that the treatment and control group are chosen at random. However, it was not until somewhere between 1925 and 1935 that randomization became the basis for scientific inference, introduced by the (in)famous statistician Ronald Fisher in agricultural experiments (Fisher & Mackenzie 1923; Bhatt 2010; Rubin 2019). In clinical research, the first randomized trial was carried out in 1946, run by the MRC Streptomycin in Tuberculosis Trials Committee to test the effect of streptomycin (an antibiotic) on Pulmonary Tuberculosis. Streptomycin was of limited availability and control subjects therefore did not receive any treatment – scientifically convenient but unethical following today's standards. Several advances have been

made since then, but the overall idea of randomization has remained the same. Randomization was pivotal for experimental designs, and still is to this day, because the randomization procedure excludes potential bias. Random allocation of treatment groups ensures that study units, regardless of what is being studied, have an equal probability of receiving treatment. Treatment allocation is therefore not determined by the practitioners, participants or other observed or unobserved factors, thereby assuring that study participants, regardless of receiving a treatment or serving as control, have similar matched characteristics. Randomization thus ensures that estimates of a treatments' effect can be attributed to the treatment alone, and not to other confounding factors (Stolberg et al. 2004).

Most progress within the practice of experimental evaluation has originated from developments and advances of research designs in agriculture and the medical sciences. Here, evidence-based practice has been particularly important in driving progress. However, much of the theory that guides and underpins modern evaluation methods, such as the theory of why a large sample size improves an estimate, was developed hundreds of years ago in other disciplines. For example, probability theory, the foundation for probability distributions, was developed by French mathematicians around the 1650s. The Swiss mathematician Jakob Bernoulli proved the Law of Large Numbers in 1713, which explains how the average of a large sample is a better approximation of the true population average than a small sample. The central limit theorem was proved by the French Mathematician Pierre-Simon Laplace in 1810 and states that, for a random variable, the sample mean distribution will be normal if the sample size is sufficiently large. These are all fundamental for scientific inference as we know it today.

Early evaluation efforts, such as those described in the history of David and King Nebuchadnezzar and the Scurvy trial by James Lind, were primarily concerned with testing a hypothesis, and less concerned about study designs, or how to estimate the degree of association between an intervention and an outcome of interest. However, this all changed at the end of the 19th century, with the emergence of modern statistics. Initially pioneered by Francis Galton and his biological research on heredity – the degree to which traits manifest from one generation into the next – the discovery of correlation paved the way for much present-day analysis. For example, Galton’s research led to the initial conceptualization of linear regression (Stanton 2001; Aldrich 2017). However, it was not until the ideas of Karl Pearson, William Gosset, Jerzy Neyman and Ronald Fisher emerged that statistics shifted away from being descriptive in nature towards more inference-oriented methods (Lehmann 1993; Rao 2006). Ronald Fisher in particular is credited for this success with his development of maximum likelihood estimation of unknown parameters (Rao 2006). These statistical developments allowed researchers to estimate the level of uncertainty in generalizations made from an observed sample to the population level, and with the introduction of the Chi-squared test for goodness of fit in 1900, and the t-test for drawing inferences from normally distributed means (Franke et al. 2012; Aldrich 2017), statistics was recognized as a separate scientific discipline.

Correlation and statistical tests enable hypothesis testing and estimates of association, but do not identify causality. Correlation does not imply causation is an often-repeated mantra of science that was indisputable until recently. In fact, Pearson insisted on correlation being the only metric of scientific interest and has been said to have ridiculed those who opposed him (Pearl & Mackenzie 2019). He believed that “Statistics may be regarded as ... the study of methods of the reduction of data”

thereby discarding any relevance of causal knowledge about the data-generation process (Pearl & Mackenzie 2019). However, in the late 1920s, Sewall Wright showed that correlation can in fact imply causation (Pearl & Mackenzie 2019). Like many other prominent scientists of his time, Wright studied genetics. More specifically, he studied inheritance of coat colour in Guinea pigs. By solving algebraic equations in conjunction with path diagrams (directional arrows from cause to effect), Wright showed that coat patterns and colour depend not only on the inherited genes, but also on their combination, and where they are expressed and suppressed (Pearl & Mackenzie 2019). While this might sound trivial to some, it demonstrated how correlation coupled with a causal theory based on subject-specific knowledge, illustrated through a path diagram, can turn correlation into causation. In Wright's example that meant measuring and estimating how development factors cause changes in fur colour (Lleras 2004).

Building on the foundation of path analysis, Turing award winner and computer scientist Judea Pearl, has pioneered the modern use of graphical models for causal inference (Pearl 2010). A causal diagram is a visual representation of the hypothesized relationship between an intervention of interest, an outcome that is being evaluated, and other characteristics that should be considered. The direction of cause is depicted through directional arrows flowing from cause to effect, with the cause always preceding the effect. For example, if a wildlife vaccination programme is assumed to impact mortality rates for a group of targeted populations, an arrow would go from the treatment – the vaccination program, to the outcome of interest – the population's mortality rates. In chapter two, I use a specific type of causal diagram - a Directed Acyclic Graph (Figure 2.1) - to visualize how wetlands affect breeding trends of wetland birds, to demonstrate the assumptions that the analysis relies upon, and to

select the variables that I believe are required to reduce confounding bias when inferring the impact of the wetlands being evaluated. Graphical representations of relationships between cause and effect play an important role for robust inferences by visualizing proposed theories of how interventions affect outcomes of interest (often referred to as “theory of change”), which in turn aids researchers in determining what data to collect, and how to obtain unbiased effect estimates (Sills et al. 2017; Adams et al. 2019; Guerra et al. 2019; Schleicher et al. 2019; Grace & Irvine 2020).

Confounding effects are the main concern when analyzing observational data and understanding the causal structure of a problem is therefore particularly important (Staplin et al. 2017). Confounding arises if a variable that affects both the outcome of interest and the mechanism that assigns treatment is not accounted for. This is not a concern in randomized experimental evaluations, because treatment is a randomized process, thereby ensuring that, on average, the characteristics of the groups being evaluated are similar. If the sample size is sufficiently large, this ensures that any differences in outcome are caused by the treatment and not the covariates since these are, on average, similar across trials. However, when estimating the impact of an intervention of interest using observational data, such as the impact of protected areas on bird abundance (Wauchope et al. 2019a; Cazalis et al. 2020, 2021b; Jellesmark et al. 2021), the intervention is not assigned at random (Larsen et al. 2019). For example, if we were to create a study testing the effect of protected areas on the abundance of breeding birds, it would neither be logical nor practically feasible to randomly allocate protected areas across time and space. Therefore, areas will often be protected because of area-specific covariates that influence both the likelihood of receiving protection *and* biodiversity outcomes within. For example, areas consisting of more pristine habitat could be more likely to receive protection than areas of degraded

habitat. If this is the case, and breeding bird abundance between protected and unprotected areas is compared, then strict protection would be erroneously assessed as being effective, whereas high abundance was in fact due to pre-treatment habitat quality.

In 1974, Donald Rubin presented a framework that enables causal inference from observational data (Rubin D. B 1974; Rubin 1990; Sekhon 2008). Rubin's framework developed theory presented in earlier work by Jerzy Neyman (one of the great statisticians of the twentieth century). The theory first presented by Neyman explained how each unit has two outcomes, one if the unit is treated and the other if untreated, with a causal effect being the difference between them. Neyman's theory strictly considered randomized experiments, but Rubin expanded this framework to allow causal inference with observational data (Sekhon 2008; Rubin 2019). Randomization ensures that units in the treatment and control group are exchangeable. Treatment assignment is therefore not associated with any of the potential outcomes: the average outcome in one group serves as the counterfactual for the other, and the treatment effect can be estimated as the average difference between group outcomes. With observational data, inferences can be made by assuming that treatment selection depends on a set of observable covariates and that all units could, in theory, have received treatment. Returning to the earlier example of a hypothetical study examining the effect of protection on bird abundance, habitat quality would be included as a covariate so that abundance was compared between areas that were of similar habitat quality prior to protection. Moreover, I discuss methods that can balance group samples, such as matching, later in this chapter. For such analysis to be feasible and a causal estimate to be valid, data of high quality are required, inferences must be made using appropriate statistical analysis, and most importantly, the causal structure

of the system or problem being evaluated (i.e. the potential covariates) must be thoroughly understood, as the researcher will not be able to identify which covariates to control for otherwise. This requires not only technical knowledge about data analysis but also subject-level expertise of the system (Hernán et al. 2019).

The developments in statistical techniques, study designs and causal inference theory through the 20th century provided tremendous insights into the effects of different treatments (Rubin 2019). However, despite the rise in knowledge gained through these advances, new information does not always find its way into practice or policy. As mentioned earlier, James Lind's findings on scurvy were not formally adopted until 40 years later. In fact, in mature disciplines, such as medicine, anecdotal evidence still guides common healthcare interventions even in modern times (Shah & Chung 2009). However, the last 50 years has seen an increase in the adoption of evidence-based policy and practice. This owes much to the pivotal role of Archie Cochrane who, in 1972, advanced evidence-based decision making within the medical sciences by promoting the need for the systematic collation of scientific evidence and cementing the role of medical research within both practice and policy (Cochrane 1972; Christie 2021).

Parallels have been drawn between medicine and conservation – both are crisis disciplines that often require urgent action in situations with limited information (Pullin & Knight 2001). However, unlike the major advances in the application of evidence for deciding treatment strategies seen in medicine, it was not until the early 2000s that similar approaches started to be promoted in conservation science (Kleiman et al. 2000; Pullin & Knight 2001; Salafsky et al. 2002; Salafsky & Margoluis 2003; Pullin et al. 2004; Sutherland et al. 2004; Ferraro & Pattanayak 2006; Christie 2021). In 2004, Bill Sutherland identified that access to evidence was a serious barrier for practitioners

and policymakers wishing to apply the best possible conservation interventions (Sutherland et al. 2004). Yet, despite more than a decade of advances in the availability of evidence through open-access publications and free repositories, evidence is still not being appropriately utilized, and conservation actions are still not being evaluated accordingly (Sutherland & Wordley 2017).

The slow uptake of evidence can be attributed to a variety of causes (Sutherland et al 2017). For example, a conservation practitioner or policy maker may believe that the knowledge they possess is sufficient, or that the evidence available is not relevant to the conservation task. Another limitation is how to use the existing evidence for evidence-based conservation (Salafsky et al. 2019). The training and background of conservation practitioners is likely very different from conservation scientists. Furthermore, conservation practitioners operate in different settings and environments than clinical researchers, which restricts the degree to which evidence-based frameworks from medicine can translate to the field of conservation. Additionally, because of limited resources, the spatial extent of conservation actions such as protected areas, and the complex and dynamic conditions that conservation projects operate within (Margoluis et al. 2009; Salafsky et al. 2019), conservation actions are often not evaluated. Furthermore, conservation projects often target threatened species or large areas, thus making manipulative experiments potentially unethical or requiring extensive resources. The spatial extent of conservation actions such as protected areas are particularly prohibitive with regard to applying randomized controlled trials, as protection can rarely be randomly allocated in the broader landscape. Furthermore, to secure ecological representation of influential factors, such as habitat composition, biological communities, human pressure and other factors potentially influencing the impact of a conservation action, would require replication to

a degree that is impractical and rarely possible (Ewers et al. 2011). The lack of sufficient spatial replicates could lead to low external validity (the degree to which the findings of a study are valid outside the settings they were obtained from, as opposed to internal validity, which refers to the degree to which the findings obtained from a specific study population represents the truth (Margoluis et al. 2009; Patino & Ferreira 2018)), therefore reducing the relevance to the broader conservation community.

Evaluation and study designs in conservation science

Understanding the impact of past conservation actions is what enables future conservation to be successful in a cost-effective manner, searching as conservationists do, for the biggest 'bang-for-buck'. However, the credibility of evidence from past evaluation efforts depends on the evaluation methods and study designs applied in the individual studies (de Palma et al. 2018; Christie et al. 2019; Ribas et al. 2020). The impact of conservation actions is influenced by social, political, and biological factors, thereby increasing the need for extensive data and risk of model misspecification, and thus complicating how to estimate conservation effectiveness. For example, underlying differences between communities can influence the degree to which individuals comply with conservation actions such as hunting quotas or bans (Andrade & Rhodes 2012). Similarly, the abundance of local predators is likely to influence populations of breeding birds, thus presenting eradication programmes as more efficient in areas with lower predator pressure (Bolton et al. 2007). Without adequate data that enables control of factors such as community differences or initial predator pressure and correctly specified models, estimates of conservation effectiveness can therefore be misleading (Keane et al. 2008). Another issue which

has just recently started to receive attention in conservation science, is the degree to which different study designs bias effect estimates (Eigenbrod et al. 2011; Ranganathan & Aggarwal 2018; Christie et al. 2019, 2020c, 2020a). Study design refers to the systematic collection of data, is an integral part of any evaluation, and largely determines the type of evaluation possible and the credibility of the evidence produced (Christie et al. 2019, 2020a; Larsen et al. 2019).

The estimation error describes the degree to which an estimate approximates the true causal effect (Zhao et al. 2019) with the following equation that elegantly illustrates the three fundamental causes of error in causal estimates:

$$\text{Estimator} - \text{True causal effect} = \text{Hidden bias} + \text{Misspecification bias} + \text{Noise}$$

Where the difference between the estimator and the true causal effect is the estimation error, the hidden bias term is due to poor study design, the misspecification bias is due to parametric modelling, and the noise is random variation due to a finite sample. Increasing the sample size reduces the potential negative effect of the noise term, which is a strong argument for using larger observational datasets. The bias of the study design arises the instant that we decide what data to collect (or include in the case of post-hoc analysis) and cannot be corrected or identified by statistical methods applied later, thus referred to as hidden bias in the equation. For example, if bias is induced by systematic differences in pre-intervention covariates, but only post-intervention information is collected (or included for post-hoc analysis), then there is no method that can correct for this bias. Misspecification is easier to correct for, at least in theory, and occurs when important variables that affect an outcome of interest are not included in a model, irrelevant variables are included, or an erroneous

functional form is specified between a dependent variable and an outcome of interest (Begg & Lagakos 1990; Zhao et al. 2019).

Study designs are an integral part of any evaluation. While the study designs most frequently applied in conservation science are discussed in chapter two, I will briefly summarize the primary components here. Christie et al. (2020a) describes the components linked to varying levels of study design bias as; randomization, sampling before and after the impact of interest occurs, and using a control group. These components can be combined to form study designs that vary in complexity and inherent levels of bias. The simplest study design is the 'After' design, where exposed units are monitored after an impact (e.g forest fire) or intervention of interest (e.g a conservation intervention). This study design can provide information on rates of change but cannot link these changes to an intervention of interest and thus have no causal interpretation (de Palma et al. 2018). If combined with sampling pre-impact or pre-intervention, this becomes a 'Before-After' study design. Returning to the theory of counterfactuals, the pre-intervention outcome is intended to represent the counterfactual outcome for the post-intervention period in the absence of treatment. The 'Before-After' study design has the obvious advantage that pre-intervention outcomes are known, but it still lacks an appropriate control from the post-intervention time period: any inference is therefore confounded if conditions that affect the outcome (e.g. weather or habitat) pre-intervention differ from conditions post-intervention. Another frequently used study design is the 'Control-Impact'. Here, an outcome of interest is monitored post-intervention inside areas that received the intervention (Impact) and inside areas that did not (Control). The control represents the counterfactual outcome for the impact group in the absence of treatment. Unlike the 'Before-After' study design, controls that are different to the actual exposure unit and

from the same time period are used, but pre-intervention conditions are not accounted for, which increases the risk of bias caused by initial differences between the 'Control' and 'Impact' sampling units. The 'Before-After' and 'Control-Impact' study designs can be combined into the 'Before-After-Control-Impact' study design, which improves the reliability of effect estimates inferred by accounting for differences in outcome between control and impact groups pre- and post-intervention.

Experimental evaluations that use RCTs remain particularly rare in conservation, with few notable exceptions evaluating large-scale interventions (Jayachandran et al. 2017; Chaves et al. 2018; Wiik et al. 2019; Weigel et al. 2021). For example, Wiik et al (2019) used randomization to evaluate the impact of a conservation incentive programme on deforestation in the Bolivian Andes. Weigel et al (2021) also used a randomized controlled trial to test a conservation programme's ability to facilitate adoption of conservation practices on rented farmland.

Quasi-experimental methods can be used for post-hoc evaluations of observational data with non-random allocation of treatment groups. They are generally less resource demanding, and so studies applying methods from this category are far more frequent in the conservation literature (Ferraro et al. 2007; Butsic et al. 2017; Wauchope et al. 2019a; Cazalis et al. 2020; Terraube et al. 2020). The potential to use historical data to answer new research questions is especially important. Researchers can utilize advances in technology such as remote sensing and larger datasets to retrospectively evaluate the impact of past actions and answer questions in situations without experimental manipulation (Sagarin & Pauchard 2010; Larsen et al. 2019; Christie et al. 2020a). The primary concern that researchers face with quasi-experimental evaluations is ruling out alternative explanations for observed effects (Shadish et al. 2002). As discussed earlier, the problem arises if groups systematically differ on the

conditions of receiving treatment in ways that could affect the outcome. Several statistical approaches can address such potential confounding and are therefore suitable for balancing samples, thus improving evaluations of conservation impact, in particular in observational settings (Joppa & Pfaff 2010, 2011; Geldmann et al. 2013a, 2015, 2018, 2019; Ferraro & Pressey 2015; Barnes et al. 2016; Ferraro & Miranda 2017; Wichman & Ferraro 2017; Butsic et al. 2017; Wauchope et al. 2019a, 2021; Schleicher et al. 2019; Terraube et al. 2020).

One approach that is particularly useful for quasi-experimental evaluations is statistical matching. This can be used to select treatment and control groups with similar covariate distributions from observational data, thus reducing the likelihood of confounding bias (Stuart 2010). Matching is an umbrella term that covers a set of conceptually similar but technically different approaches. Treatment and control groups are selected from the full sample based on similarity expressed through a unit level distance measure. Here, unit refers to an entity being evaluated that does or does not receive the treatment of interest (e.g. species' populations in this thesis, patients in clinical research). Exact matching is the simplest form. Here, treatment and control groups are selected so that all units are similar given a vector of covariates. Observations with dissimilar covariate values are discarded, thereby ensuring that the treatment and control groups are identical when compared on the selected vector of covariates. This type of matching theoretically secures the highest degree of sample balance but can greatly reduce the sample size of the groups being evaluated, because fewer observations will match stricter conditions. This is particularly true if including continuous covariates, a high number of categorical covariates, or both (Ho et al. 2007). Propensity score matching is an alternative approach: here, the two vectors of covariates are collapsed into scalar propensity scores, often using logistic

regression, representing the probability of receiving treatment given the covariates. The distance between two observations can then be calculated as the scalar difference, and treatment and control groups can be selected so that the minimum distance is obtained.

Matching is increasingly being used to evaluate large-scale conservation interventions, such as the effect of protected areas on outcomes that can either be represented or proxied from remote sensing data (Joppa & Pfaff 2010, 2011; Geldmann et al. 2013a, 2019; Schleicher et al. 2019; Ford et al. 2020), so that changes in the outcome of interest are compared between areas with similar characteristics (e.g. elevation, slope, type of land cover). Other measures of conservation interest, such as population counts from monitoring programmes, can similarly be used to evaluate an intervention of interest by matching conservation targeted populations (e.g. bird counts from protected areas) to similar populations that were not targeted (e.g. bird counts from outside protected areas). The populations without conservation will then be approximations of how the conservation targeted populations could have developed in the absence of conservation, and estimates of impact derived thereof (Wauchope et al. 2019a, 2021; Jellesmark et al. 2021).

Non-experimental evaluations, such as protecting an area and then evaluating its impact by observing the development of wildlife populations within, generally have low credibility. As there is no comparison group, observed changes cannot be attributed to the intervention being evaluated. However, non-experimental designs are not without use and play a key role in day-to-day management. For example, in the Royal Society for the Protection of Birds (RSPB), annual breeding counts are used as one of many indicators to guide management decisions. While not ideal for settings that require a high level of credibility, non-experimental evaluation is a cost-effective

approach that utilizes the available information and is especially useful when combined with the local knowledge of reserve managers.

An evaluation design that does not qualify for any of the above categories but is both interesting and, to my knowledge, unique to conservation science, can be considered under the broader theme of “Inferential approaches” (Grace et al. 2021a). In statistics, inferential methods use sample-level information to generalize about the overall population. However, in conservation science, the term also refers to studies that use expert-knowledge to inform counterfactual scenarios, often representing species’ extinction risk without the impact of conservation. The impact of conservation can then be assessed as the difference between a species extinction risk in the factual scenario (what is currently observed) and its counterfactual representation without conservation (Butchart et al. 2006; Hoffmann et al. 2010, 2015; Young et al. 2014; Bolam et al. 2020; Grace et al. 2021a). While the evaluation theory underlying these studies is reasonably similar – they utilize expert knowledge to assess the extinction risk for species in the absence of conservation – considerable variation still exists between the methods of the different studies and how the counterfactuals are approximated.

For example, Butchart et al (2006) – to my knowledge the first to utilize expert-informed counterfactuals for evaluating the impact of conservation actions – used a counterfactual framework for estimating species extinction risk in the absence of conservation. To measure the impact of conservation interventions on reducing species’ extinction risk, avian experts considered population size, trends, severity of threats and intensity and effectiveness of conservation actions for 27 bird species between 1994-2004. Of these 27 species, 16 were deemed likely to have gone extinct in the absence of conservation. The methods initially pioneered by Butchart et al (2006) were later developed and used to assess the impact of conservation on the

status of the world's vertebrates (Hoffmann et al. 2010), on the world's ungulates (Hoffmann et al. 2015) and to evaluate the impact of a conservation organization (Young et al. 2014). More recently, Bolam et al (2020) used a similar expert-informed method to measure the impact of conservation on bird and mammal species' risk of going extinct. A similar method is being used to assess the impact of past conservation actions in the new IUCN framework: The Green Status of Species (Grace et al. 2021b, 2021a). While it is widely recognized that quantitative assessments would be preferred, the data required to do so are often not available (Butchart et al. 2006; Grace et al. 2021a). Developing quantitative counterfactual methods and thus advancing quantitative assessments of conservation impact was the initial inspiration for this thesis. By using matching and similar quantitative analysis, I set out to combine datasets to create counterfactual population trends, and to evaluate large-scale conservation actions using these trends. I hope that this research will advance the field of conservation science, improve our understanding of how conservation impacts upon species' populations, and inform policy by estimates of large-scale conservation impact. Foremost, I hope that this research will help contribute to the reversal of ongoing global biodiversity loss.

Thesis overview

In this thesis, I focus on how to measure the impact of conservation actions on species populations using quantitative methods. I demonstrate how to measure the impact of conservation on different spatial scales - local, national and global - from observational data for vertebrates. Informed by counterfactual theory, I estimate the impact of protected areas on breeding populations of wetland birds, and investigate the association between conservation actions within these areas and the breeding

abundance of targeted populations. I then estimate the global impact of conservation on species' populations in general by creating multiple counterfactual population trends, measuring the degree to which conservation has affected a global index of vertebrate abundance, and assessing if certain conservation actions are more successful than others.

In chapter two, I evaluate the impact of lowland wet grassland reserves on breeding populations of wading birds in the United Kingdom. I match population counts from a national survey on key reserve covariates to create counterfactual population trends that represent how reserve populations could have developed in the absence of conservation and use different matching specifications to represent the uncertainty in the counterfactual trends.

In chapter three, I focus on the local drivers of breeding abundance within the reserves evaluated in the previous chapter. I use management information collected from annual reserve reports, paired with climatic data, to estimate the association between specific conservation actions, such as water control and predator control, and annual breeding abundance of wading birds.

In chapter four, I use a global database of vertebrate population time series to estimate the global impact of conservation. I categorize conservation actions on a population level to create counterfactual trends that represent how conservation targeted populations may have developed without conservation, estimate the impact of conservation on a global index of vertebrate abundance, and test the effectiveness of the recorded conservation actions on population trends.

Finally, in chapter five I discuss the conclusions of this work, some of its limitations and how to improve future evaluation efforts.

References

Adams VM, Barnes M, Pressey RL. 2019. Shortfalls in Conservation Evidence: Moving from Ecological Effects of Interventions to Policy Evaluation. *One Earth* 1:62–75.

Aldrich J. 2017. The Origins of Modern Statistics. Page (Hájek A, Hitchcock C, editors) *The Oxford Handbook of Probability and Philosophy*. Oxford University Press. Available from <http://oxfordhandbooks.com/view/10.1093/oxfordhb/9780199607617.001.0001/oxfordhb-9780199607617-e-7> Francis Galton.

Andrade GSM, Rhodes JR. 2012. Protected areas and local communities: An inevitable partnership toward successful conservation strategies? *Ecology and Society* 17.

Barnes MD et al. 2016. Wildlife population trends in protected areas predicted by national socio-economic metrics and body size. *Nature Communications* 7:1–9. Nature Publishing Group.

Barnosky AD et al. 2011. Has the Earth's sixth mass extinction already arrived? *Nature* 471:51–57. Nature Publishing Group.

Baylis K, Honey-Rosés J, Börner J, Corbera E, Ezzine-de-Blas D, Ferraro PJ, Lapeyre R, Persson UM, Pfaff A, Wunder S. 2016. Mainstreaming Impact Evaluation in Nature Conservation. *Conservation Letters* 9:58–64.

Begg MD, Lagakos S. 1990. On the consequences of model misspecification in logistic regression. *Environmental Health Perspectives* 87:69–75.

Bhatt A. 2010. Evolution of clinical research: a history before and beyond james lind. *Perspectives in clinical research* 1:6–10. Available from

www.ncbi.nlm.nih.gov/pubmed/21829774www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3149409.

Bolam FC et al. 2020. How many bird and mammal extinctions has recent conservation action prevented? *Conservation Letters*:e12762. Available from <http://doi.wiley.com/10.1111/conl.12762>.

Bolton M, Tyler G, Smith K, Bamford R. 2007. The impact of predator control on lapwing *Vanellus vanellus* breeding success on wet grassland nature reserves. *Journal of Applied Ecology* 44:534–544.

Butchart SHM, Stattersfield AJ, Collar NJ. 2006. How many bird extinctions have we prevented? ORYXDOI: 10.1017/S0030605306000950.

Butsic V, Lewis DJ, Radeloff VC, Baumann M, Kuemmerle T. 2017. Quasi-experimental methods enable stronger inferences from observational data in ecology. *Basic and Applied Ecology* 19:1–10. Elsevier GmbH. Available from <http://dx.doi.org/10.1016/j.baae.2017.01.005>.

Cazalis V, Belghali S, Rodrigues ASL. 2021. Using a large-scale biodiversity monitoring dataset to test the effectiveness of protected areas at conserving North-American breeding birds. *Peer Community Journal* 1. Available from <https://peercommunityjournal.org/articles/10.24072/pcjournal.5/>.

Cazalis V, Princé K, Mihoub JB, Kelly J, Butchart SHM, Rodrigues ASL. 2020. Effectiveness of protected areas in conserving tropical forest birds. *Nature Communications* 11:1–8. Springer US. Available from <http://dx.doi.org/10.1038/s41467-020-18230-0>.

CBD. 2011. Convention of Biological Diversity: Strategic Plan for Biodiversity 2011-2020. Page Convention of Biological Diversity: Strategic Plan for Biodiversity 2011-2020.

Ceballos G, Ehrlich PR, Barnosky AD, García A, Pringle RM, Palmer TM. 2015. Accelerated modern human-induced species losses: Entering the sixth mass extinction. *Science Advances* 1.

Ceballos G, Ehrlich PR, Raven PH. 2020. Vertebrates on the brink as indicators of biological annihilation and the sixth mass extinction. *Proceedings of the National Academy of Sciences of the United States of America* 117:13596–13602.

Chaves WA, Valle DR, Monroe MC, Wilkie DS, Sieving KE, Sadowsky B. 2018. Changing Wild Meat Consumption: An Experiment in the Central Amazon, Brazil. *Conservation Letters* 11:1–10.

Christie AP et al. 2020a. Quantifying and addressing the prevalence and bias of study designs in the environmental and social sciences. *Nature Communications* 11:6377. Available from <http://www.nature.com/articles/s41467-020-20142-y>.

Christie AP. 2021. Determining the biases and consistencies in the evidence for conservation:238. Available from <https://www.repository.cam.ac.uk/handle/1810/321463?show=full>.

Christie AP, Amano T, Martin PA, Petrovan SO, Shackelford GE, Simmons BI, Smith RK, Williams DR, Wordley CFR, Sutherland WJ. 2020b. Poor availability of context-specific evidence hampers decision-making in conservation. *Biological Conservation* 248:108666. Elsevier. Available from <https://doi.org/10.1016/j.biocon.2020.108666>.

Christie AP, Amano T, Martin PA, Petrovan SO, Shackelford GE, Simmons BI, Smith RK, Williams DR, Wordley CFR, Sutherland WJ. 2020c. The challenge of biased evidence in conservation. *Conservation Biology* DOI: 10.1111/cobi.13577. Available from <https://onlinelibrary.wiley.com/doi/abs/10.1111/cobi.13577>.

Christie AP, Amano T, Martin PA, Shackelford GE, Simmons BI, Sutherland WJ. 2019. Simple study designs in ecology produce inaccurate estimates of biodiversity responses. *Journal of Applied Ecology*:2742–2754.

Cisneros E, Zhou SL, Börner J. 2015. Naming and shaming for conservation: Evidence from the Brazilian Amazon. *PLoS ONE* 10:1–24.

Cochrane AL. 1972. *Effectiveness and Efficiency: Random Reflections on Health Services*. Available from <https://www.bmj.com/lookup/doi/10.1136/bmj.328.7438.529>.

Cunningham S. 2021. *Causal Inference*, 1st edition. Yale University Press.

de Palma A, Sanchez-Ortiz K, Martin PA, Chadwick A, Gilbert G, Bates AE, Börger L, Contu S, Hill SLL, Purvis A. 2018. *Challenges With Inferring How Land-Use Affects Terrestrial Biodiversity: Study Design, Time, Space and Synthesis*. Page Advances in Ecological Research, 1st edition. Elsevier Ltd. Available from <http://dx.doi.org/10.1016/bs.aecr.2017.12.004>.

Eigenbrod F, Hecnar SJ, Fahrig L. 2011. Sub-optimal study design has major impacts on landscape-scale inference. *Biological Conservation* 144:298–305. Elsevier Ltd. Available from <http://dx.doi.org/10.1016/j.biocon.2010.09.007>.

Ewers RM et al. 2011. A large-scale forest fragmentation experiment: The stability of altered forest ecosystems project. *Philosophical Transactions of the Royal Society B: Biological Sciences* 366:3292–3302.

Ferraro PJ. 2009. Counterfactual thinking and impact evaluation in environmental policy. *New Directions for Evaluation* 2009:75–84. Available from <https://onlinelibrary.wiley.com/doi/10.1002/ev.297>.

Ferraro PJ, McIntosh C, Ospina M. 2007. The effectiveness of the US endangered species act: An econometric analysis using matching methods. *Journal of Environmental Economics and Management* 54:245–261.

Ferraro PJ, Miranda JJ. 2017. Panel Data Designs and Estimators as Substitutes for Randomized Controlled Trials in the Evaluation of Public Programs. *Journal of the Association of Environmental and Resource Economists* 4:281–317.

Ferraro PJ, Pattanayak SK. 2006. Money for Nothing? A Call for Empirical Evaluation of Biodiversity Conservation Investments. *Plos biology* 4:482–488.

Ferraro PJ, Pressey RL. 2015. Measuring the difference made by conservation initiatives: Protected areas and their environmental and social impacts. *Philosophical Transactions of the Royal Society B: Biological Sciences* 370:4–8.

Fisher RA, Mackenzie WA. 1923. Studies in crop variation. II. The manorial response of different potato varieties. *J Agr Sci* 13:311–320.

Ford SA, Jepsen MR, Kingston N, Lewis E, Brooks TM, MacSharry B, Mertz O. 2020. Deforestation leakage undermines conservation value of tropical and subtropical forest protected areas. *Global Ecology and Biogeography* 29:2014–2024.

Franke TM, Ho T, Christie CA. 2012. The Chi-Square Test: Often Used and More Often Misinterpreted. *American Journal of Evaluation* 33:448–458.

Geldmann J et al. 2015. Changes in protected area management effectiveness over time: A global analysis. *Biological Conservation* 191:692–699. Elsevier B.V. Available from <http://dx.doi.org/10.1016/j.biocon.2015.08.029>.

Geldmann J et al. 2018. A global analysis of management capacity and ecological outcomes in terrestrial protected areas. *Conservation Letters* 11:1–10.

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. Elsevier Ltd. Available from <http://dx.doi.org/10.1016/j.biocon.2013.02.018>.

Geldmann J, Manica A, Burgess ND, Coad L, Balmford A. 2019. A global-level assessment of the effectiveness of protected areas at resisting anthropogenic pressures. *Proceedings of the National Academy of Sciences of the United States of America* 116:23209–23215. Available from <https://doi.org/10.1073/pnas.1908221116>.

Grace JB, Irvine KM. 2020. Scientist's guide to developing explanatory statistical models using causal analysis principles. *EcologyDOI*: 10.1002/ecy.2962.

Grace MK et al. 2021a. Building robust , practicable counterfactuals and scenarios to evaluate the impact of species conservation interventions using inferential approaches. *Biological Conservation* 261:109259. Elsevier Ltd. Available from <https://doi.org/10.1016/j.biocon.2021.109259>.

Grace MK et al. 2021b. Testing a global standard for quantifying species recovery and assessing conservation impact. *Conservation Biology*:1–17.

Guerra CA et al. 2019. Finding the essential: Improving conservation monitoring across scales. *Global Ecology and Conservation* 18:e00601. Elsevier Ltd. Available from <https://doi.org/10.1016/j.gecco.2019.e00601>.

Hernán MA, Hsu J, Healy B. 2019. A Second Chance to Get Causal Inference Right: A Classification of Data Science Tasks. *CHANCE* 32:42–49. Available from <https://www.tandfonline.com/doi/full/10.1080/09332480.2019.1579578>.

Ho DE, Imai K, King G, Stuart EA. 2007. Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis* 15:199–236.

Hoffmann M et al. 2010. The impact of conservation on the status of the world's vertebrates. *Science* 330:1503–1509.

Hoffmann M, Duckworth JW, Holmes K, Mallon DP, Rodrigues ASL, Stuart SN. 2015. The difference conservation makes to extinction risk of the world's ungulates. *Conservation Biology* 29:1303–1313.

IPBES. 2019. Global Assessment Report on Biodiversity and Ecosystem Services. Page Global Assessment Summary for Policymakers. Available from https://ipbes.net/system/tdf/ipbes_global_assessment_report_summary_for_policymakers.pdf?file=1&type=node&id=35329.

Jayachandran S, de Laat J, Lambin EF, Stanton CY, Audy R, Thomas NE. 2017. Cash for carbon: A randomized trial of payments for ecosystem services to reduce deforestation. *Science* 357:267–273. Available from <https://www.science.org/doi/10.1126/science.aan0568>.

Jellesmark S, Ausden M, Blackburn TM, Gregory RD, Hoffmann M, Massimino D, McRae L, Visconti P. 2021. A counterfactual approach to measure the impact of wet grassland conservation on UK breeding bird populations. *Conservation Biology*:cobi.13692. Available from <https://onlinelibrary.wiley.com/doi/10.1111/cobi.13692>.

Joppa L, Pfaff A. 2010. Reassessing the forest impacts of protection. *Annals of the New York Academy of Sciences* 1185:135–149. Available from <http://doi.wiley.com/10.1111/j.1749-6632.2009.05162.x>.

Joppa LN, Pfaff A. 2011. Global protected area impacts. *Proceedings of the Royal Society B: Biological Sciences* 278:1633–1638. Available from <https://royalsocietypublishing.org/doi/10.1098/rspb.2010.1713>.

Keane A, Jones JPG, Edwards-Jones G, Milner-Gulland EJ. 2008. The sleeping policeman: understanding issues of enforcement and compliance in conservation. *Animal Conservation* 11:75–82. Available from <https://onlinelibrary.wiley.com/doi/10.1111/j.1469-1795.2008.00170.x>.

Kleiman DG, Reading RP, Miller BJ, Clark TW, Scott JM, Robinson J, Wallace RL, Cabin RJ, Felleman F. 2000. Improving the evaluation of conservation programs. *Conservation Biology* 14:356–365.

Larsen AE, Meng K, Kendall BE. 2019. Causal analysis in control–impact ecological studies with observational data. *Methods in Ecology and Evolution* 10:924–934.

Leclère D et al. 2020. Bending the curve of terrestrial biodiversity needs an integrated strategy. *Nature* 2018. Available from <https://doi.org/10.1038/s41586-020-2705-y>
<http://www.nature.com/articles/s41586-020-2705-y>.

Lehmann EL. 1993. The fisher, Neyman–Pearson theories of testing hypotheses: One theory or two? *Journal of the American Statistical Association* 88:1242–1249.

Lleras C. 2004. Path Analysis. *Encyclopedia of Social Measurement* 3:25–30.

Margoluis R, Stem C, Salafsky N, Brown M. 2009. Design alternatives for evaluating the impact of conservation projects. *New Directions for Evaluation* 2009:85–96. Available from <http://doi.wiley.com/10.1002/ev.298>.

Neely KL, Macaulay KA, Hower EK, Dobler MA. 2020. Effectiveness of topical antibiotics in treating corals affected by Stony Coral Tissue Loss Disease. *PeerJ* 2020.

Patino C, Ferreira J. 2018. Research methods knowledge base: Qualitative validity. *J Bras Pneumol.* 44:183.

Pearl J. 2010. An Introduction to Causal Inference. *The international journal of biostatistics* 6:1–59. Available from <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2836213&tool=pmcentrez&rendertype=abstract>.

Pearl J, Mackenzie D. 2019. *The Book of Why*. Penguin Books, Harlow, England.

Peirce CS, Jastrow J. 1885. On Small Differences in Sensation. *Memoirs of the National Academy of Sciences* 3:73–83.

Pimm SL, Jenkins CN, Abell R, Brooks TM, Gittleman JL, Joppa LN, Raven PH, Roberts CM, Sexton JO. 2014. The biodiversity of species and their rates of extinction, distribution, and protection. *Science* 344.

Pullin AS, Knight TM. 2001. Effectiveness in Conservation Practice: Pointers from Medicine and Public Health. *Conservation Biology* 15:50–54. Available from <http://dx.doi.org/10.1111/j.1523-1739.2001.99499.x>.

Pullin AS, Knight TM, Stone DA, Charman K. 2004. Do conservation managers use scientific evidence to support their decision-making?

Ranganathan P, Aggarwal R. 2018. Study designs: Part 1-An overview and classification. *Perspectives in Clinical Research* 9:184–186.

Rao R. 2006. R. A. Fisher: The Founder of Modern Statistics. *Statistical Science* 7:34–48. Available from <https://www.jstor.org/stable/2245989>.

Ribas LG dos S, Pressey RL, Loyola R, Bini LM. 2020. A global comparative analysis of impact evaluation methods in estimating the effectiveness of protected areas. *Biological Conservation* 246:108595. Elsevier. Available from <https://doi.org/10.1016/j.biocon.2020.108595>.

Rubin D. B. 1974. Estimating causal effects of treatment in randomized and nonrandomized studies. *Journal of Educational Psychology* 66:688–701. Available from http://www.fsb.muohio.edu/lij14/420_paper_Rubin74.pdf.

Rubin DB. 1990. Comment: Neyman (1923) and causal inference in experiments and observational studies. *Statistical Science* 5:472–480.

Rubin DB. 2019. Essential concepts of causal inference: a remarkable history and an intriguing future. *Biostatistics and Epidemiology* 3:140–155.

Sagarin R, Pauchard A. 2010. Observational approaches in ecology open new ground in a changing world. *Frontiers in Ecology and the Environment* 8:379–386.

Salafsky N et al. 2008. A standard lexicon for biodiversity conservation: Unified classifications of threats and actions. *Conservation Biology* DOI: 10.1111/j.1523-1739.2008.00937.x.

Salafsky N et al. 2019. Defining and using evidence in conservation practice. *Conservation Science and Practice* 1:e27.

Salafsky N, Margoluis R. 2003. What conservation can learn from other fields about monitoring and evaluation. *BioScience* 53:120–122.

Salafsky N, Margoluis R, Redford KH, Robinson JG. 2002. Improving the practice of conservation: a conceptual agenda for conservation science. *Conservation Biology* 16:1469–1479.

Schleicher J, Eklund J, Barnes M, Geldmann J, Oldekop JA, Jones JPG. 2019. Statistical matching for conservation science. *Conservation Biology*:cobi.13448. Available from <https://onlinelibrary.wiley.com/doi/abs/10.1111/cobi.13448>.

Sekhon JS. 2008. The Neyman-Rubin Model of Causal Inference and Estimation Via Matching Methods. Page The Oxford Handbook of Political Methodology.

Shadish WR, Cook TD, Campbell DT. 2002. Experimental and quasi- experimental designs for generalized causal inference. Houghton Mifflin.

Shah HM, Chung KC. 2009. Archie cochrane and His vision for evidence-based medicine. *Plastic and Reconstructive Surgery* 124:982–988.

Sills EO, de Sassi C, Jagger P, Lawlor K, Miteva DA, Pattanayak SK, Sunderlin WD. 2017. Building the evidence base for REDD+: Study design and methods for evaluating the impacts of conservation interventions on local well-being. *Global Environmental Change* 43:148–160. Elsevier Ltd. Available from <http://dx.doi.org/10.1016/j.gloenvcha.2017.02.002>.

Stanton JM. 2001. Galton, Pearson, and the Peas: A Brief History of Linear Regression for Statistics Instructors. *Journal of Statistics Education* 9.

Staplin N, Herrington WG, Judge PK, Reith CA, Haynes R, Landray MJ, Baigent C, Emberson J. 2017. Use of Causal Diagrams to Inform the Design and Interpretation of Observational Studies: An Example from the Study of Heart and Renal Protection (SHARP). *Clinical Journal of the American Society of Nephrology* 12:546–552. Available from <https://cjasn.asnjournals.org/lookup/doi/10.2215/CJN.02430316>.

Stolberg HO, Norman G, Trop I. 2004. Randomized Controlled Trials. *American Journal of Roentgenology* 183:1539–1544. Available from <http://www.ajronline.org/doi/10.2214/ajr.183.6.01831539>.

Stuart EA. 2010. Matching Methods for Causal Inference: A Review and a Look Forward. *Statistical Science* 25:1–21.

Sutherland WJ, Pullin AS, Dolman PM, Knight TM. 2004. The need for evidence-based conservation.

Sutherland WJ, Wordley CFR. 2017. Evidence complacency hampers conservation. *Nature Ecology and Evolution* 1:1215–1216. Springer US. Available from <http://dx.doi.org/10.1038/s41559-017-0244-1>.

Terraube J, van doninck J, Helle P, Cabeza M. 2020. Assessing the effectiveness of a national protected area network for carnivore conservation. *Nature Communications* 11:1–9. Springer US. Available from <http://dx.doi.org/10.1038/s41467-020-16792-7>.

Tröhler U. 2005. Lind and scurvy: 1747 to 1795. *Journal of the Royal Society of Medicine* 98:519–522.

Wauchope H et al. 2019a. Quantifying the impact of protected areas on near-global waterbird population trends, a pre-analysis plan:1–17. Available from <https://doi.org/10.7287/peerj.preprints.27741v1>.

Wauchope Hannah S, Amano T, Geldmann J, Johnston A, Simmons BI, Sutherland WJ, Jones JPG. 2021. Evaluating Impact Using Time-Series Data. *Trends in Ecology & Evolution* 36:196–205. Available from <https://doi.org/10.1016/j.tree.2020.11.001>.

Wauchope HS, Amano T, Sutherland WJ, Johnston A. 2019b. When can we trust population trends? A method for quantifying the effects of sampling interval and duration. *Methods in Ecology and Evolution* 10:2067–2078.

Weigel C, Harden S, Masuda YJ, Ranjan P, Wardropper CB, Ferraro PJ, Prokopy L, Reddy S. 2021. Using a randomized controlled trial to develop conservation strategies on rented farmlands. *Conservation Letters* 14:1–8.

Wichman CJ, Ferraro PJ. 2017. A cautionary tale on using panel data estimators to measure program impacts. *Economic Letters* 151:82–90.

Wiik E, d'Annunzio R, Pynegar E, Crespo D, Asquith N, Jones JPG. 2019. Experimental evaluation of the impact of a payment for environmental services program on deforestation. *Conservation Science and Practice* 1:1–11.

WWF. 2020. Living planet report 2020 - Bending the curve of Biodiversity loss. Page Almond R.E.A., Grooten M. and Petersen, T. (Eds). WWF, Gland, Switzerland. Available from <https://f.hubspotusercontent20.net/hubfs/4783129/LPR/PDFs/ENGLISH-FULL.pdf>.

Young RP, Hudson MA, Terry AMR, Jones CG, Lewis RE, Tatayah V, Zuël N, Butchart SHM. 2014. Accounting for conservation: Using the IUCN Red List Index to evaluate the impact of a conservation organization. *Biological Conservation* 180:84–96. Elsevier Ltd. Available from <http://dx.doi.org/10.1016/j.biocon.2014.09.039>.

Zhao Q, Keele LJ, Small DS. 2019. Comment: Will competition-winning methods for causal inference also succeed in practice? *Statistical Science* 34:72–76.

2 | A counterfactual approach to measure the impact of wet grassland conservation on U.K. breeding bird populations

This work was conducted in collaboration with my supervisors, Malcolm Ausden from the Royal Society for the Protection of Birds and Dario Massimino from the British Trust for Ornithology and was published in January 2021 as:

Jellesmark, S., Ausden, M., Blackburn, T.M., Gregory, R.D., Hoffmann, M., Massimino, D., McRae, L. and Visconti, P. (2021), A counterfactual approach to measure the impact of wet grassland conservation on U.K. breeding bird populations. *Conservation Biology*, 35: 1575-1585.

Abstract

Wet grassland populations of wading birds in the United Kingdom have declined severely since 1990. To help mitigate these declines, the Royal Society for the Protection of Birds has restored and managed lowland wet grassland nature reserves to benefit these and other species. However, the impact of these reserves on bird population trends has not been evaluated experimentally due to a lack of control populations. We compared population trends from 1994 to 2018 among five bird species of conservation concern that breed on these nature reserves with counterfactual trends created from matched breeding bird survey observations. We compared reserve trends with 3 different counterfactuals based on different scenarios of how reserve populations could have developed in the absence of conservation. Effects of conservation interventions were positive for all four targeted wading bird species: Lapwing (*Vanellus vanellus*), Redshank (*Tringa totanus*), Curlew (*Numenius arquata*), and Snipe (*Gallinago gallinago*). There was no positive effect of conservation interventions on reserves for the passerine, Yellow Wagtail (*Motacilla flava*). Our approach using monitoring data to produce valid counterfactual controls is a broadly applicable method allowing large-scale evaluation of conservation impact.

Introduction

Halting the decline of global biodiversity is currently one of humanity's greatest environmental challenges. Within animal populations, declines have predominantly been attributed to changes in land use, invasive species, exploitation of species and habitats, pollution, and climate change (IPBES, 2019). Different actors, including policy makers, nongovernmental organizations, and conservation practitioners, are addressing this global decline through a range of conservation actions, chiefly habitat

and species management interventions in- and outside protected areas. From 2010 to 2018, £817 million was spent on average each year to promote and protect biodiversity in the United Kingdom alone (Department for Environment, Food & Rural Affairs, 2019). Despite these efforts, biodiversity indicators in the United Kingdom continue to decline (Hayhow et al., 2019). An important question in understanding the impact of conservation interventions on target populations is the extent to which those interventions mitigate or reverse population declines (Hoffmann et al., 2010, 2015). However, limited resources often mean that evaluation efforts do not extend beyond simple measures of association. Population trends are often monitored in protected areas, but appropriate control trends are not. Thus, whether population changes in target species are caused by the management measures or represent changes that would have occurred in the absence of that management remains untested.

To assess the impact of conservation, it is necessary to understand what would have happened in the absence of conservation, that is, the counterfactual conservation outcome (e.g., Baylis et al., 2016; Ferraro, 2009; Ferraro & Pattanayak, 2006). The exact form of the counterfactual can never be known for certain. Ideally, a robust study design, such as a randomized controlled trial (RCT) (random assignment of treatment and control groups), could be used to infer the causal effect of a treatment by approximating the counterfactual outcome. However, RCT designs are rarely used in conservation. This is because randomization is often infeasible. For example, there can be legislative obligations to manage protected sites in ways considered beneficial to conservation, which makes it difficult to include unmanaged controls. In addition, the scale of conservation interventions and sampling units may be too large to allow for sufficient replication (Margoluis et al., 2009; Baylis et al., 2016; Wiik et al., 2019).

Conservation practitioners resort to other evaluation designs because of the financial, practical, and logistical challenges of the RCT design. These include after (A) methods (e.g., increasing or decreasing posttreatment population size), before-after (BA) methods (e.g., pre- treatment population changes are compared with post- treatment population changes), and control-impact (CI) methods (e.g., comparing population densities inside reserves with population densities outside reserves). Such approaches are important in determining the extent to which conservation objectives are being achieved and are a prerequisite for adaptive management. However, if potential biases are not properly addressed, these approaches cannot be used to determine cause and effect with a high level of confidence. The after study design describes the posttreatment rate of change and direction but does not provide insight into whether the change would have differed without the treatment. The before– after study design assumes that temporal variability and confounding factors before and after the intervention are comparable, and control–impact assumes time-for-space substitution and comparability between groups. The validity of such inferences is therefore compromised if a population would have developed similarly regardless of conservation (e.g. A), if the effect of confounding variables is not homogenous across time (e.g., BA), and if local variation is systematically different between impact and control groups (e.g., CI) (e.g., Ferraro & Pressey, 2015; De Palma et al., 2018; Adams et al., 2019). To improve the credibility of an inference, the BA and the CI study design can be combined, forming the before– after–control–impact (BACI) study design (e.g., comparing pretreatment and posttreatment densities in a treated and a control group while accounting for the pretreatment density difference between treated and control group). Using simulated ecological data, the BACI design can be used to estimate the true effect size better than the true effect and direction), CIs (3.2–4.6 times more RCTs

(1.3—1.8 times more likely to estimate $\pm 30\%$ of likely), and A study designs (7.1—10.1 times more likely) (Christie et al., 2019). However, this study design has many of the same limitations as the RCT and is further limited if appropriate controls cannot be identified ex ante (e.g., appropriate controls cannot be selected prior to measuring the outcome of interest if confounders are unknown or poorly understood).

To produce reliable conservation effect estimates, matching techniques are increasingly being used in conservation science (Sills et al., 2017; Schleicher et al., 2019; Sonter et al., 2019). The intent of matching is to create treatment and control groups with similar covariates by creating subsets of treatment and control samples so that comparisons are carried out with groups that have similar characteristics (e.g. comparing the outcome of a treated group to the outcome of a control group where both groups are from the same habitat type, elevation, and country). The post matching control group then represents the counterfactual outcome of the treated group, and the effect of a given treatment can be inferred as the difference between outcomes. For example, Ferraro et al. (2007) tested the effectiveness of U.S. Endangered Species Act listing and funding on species recovery based on matching of a set of observable covariates to account for bias in the listing and funding process. They found listing is effective only when accompanied by adequate funding. Geldmann et al. (2019) assessed whether protected areas (PAs) reduce anthropogenic pressure. They used 10 variables linked to PA selection to match PAs to similar unprotected areas and found that, on average, PAs do not reduce human pressure. Nevertheless, although the theoretical potential of these methods has been highlighted, examples of their application remain scarce (Ferraro & Pattanayak, 2006; Margoluis et al., 2009; Joppa & Pfaff, 2010).

We adopted a matching approach to explore the impact of specific conservation interventions on a particular habitat of conservation concern in Europe: lowland wet grassland (Franks et al., 2018). Conversion to other habitat types, changes in grazing regimes, drainage, and agricultural intensification have adversely affected these grasslands (Wilson et al., 2004). In particular, wetland bird species that use this habitat to breed and overwinter, such as wading birds (Charadriiformes), have exhibited severe breeding population declines as a result of these habitat changes (Wilson et al., 2005; Boatman et al., 2007; Colhoun et al., 2017). For example, Lapwing (*Vanellus vanellus*) populations, once abundant in the countryside of the United Kingdom, declined by 42% from 1995 to 2017 (Harris et al., 2019). To help mitigate these declines, the Royal Society for the Protection of Birds (RSPB) has allocated resources to purchasing, restoring, and managing reserves in lowland wet grassland habitats to benefit breeding wading birds in the United Kingdom. Conservation interventions, such as raising and manipulating water levels, beneficial stock grazing regimes, control and exclusion of generalist predators, and mechanical vegetation control, are implemented on these reserves (Ausden et al., 2019). Conservation efforts of this type are associated with increasing wading bird populations (Ausden & Hirons, 2002; Malpas et al., 2013; Smart et al., 2014). However, a central problem is whether the conservation actions result in positive benefits to the target populations: is the population performance better than would have occurred in the absence of these interventions? We tested this by comparing breeding trends on the reserves with matched counterfactual trends that represent how the trends may have developed in the absence of reserve-based conservation interventions. This is, to our knowledge, the first time post hoc evaluation of conservation interventions using quasi-experimental after-control-impact ACI analyses has been carried out for conservation

interventions in the United Kingdom. We used trends after intervention thus after and matching reserve trends to counterfactual controls, therefore control–impact.

Methods

Data

We used bird counts from RSPB lowland wet grassland reserves and from the U.K. Breeding Bird Survey (Harris et al., 2019) for the period 1994–2018. The RSPB manages over 200 reserves across the United Kingdom; 47 of these contain lowland wet grassland (Figure S2.7). Most of these reserves are in England (35); the rest are in Scotland (7), Wales (3), and Northern Ireland (2). We chose lowland wet grassland (i.e., periodically flooded grasslands below approximately 250 m elevation [Jefferson & Grice, 1998]) because this is a habitat in which considerable resources have been invested in habitat restoration and creation in recent decades. The area of lowland wet grassland on individual reserves varies from 18 to 1,300 ha (mean site area = 95 ha [SD 144]). Some reserves consist of two or more noncontiguous blocks of lowland wet grassland, which we refer to as sites. We used a total of 101 sites in the 47 reserves. We treated new acquisitions of land as separate sites. The RSPB reserves are managed in accordance with the biological requirements of priority species selected for that reserve. The number of breeding pairs of priority bird species are counted three times annually between April and June at each site with standard methods described in Gilbert et al. (1998) (Appendix 2.1).

The focal wetland species were Garganey (*Anas querquedula*), Shoveler (*A. clypeata*), Black-tailed Godwit (*Limosa limosa*), Lapwing, Curlew (*Numenius arquata*), Snipe (*Gallinago gallinago*), Redshank (*Tringa totanus*), and Yellow Wagtail (*Motacilla*

flava). Analysis concentrated on the latter five abundant species. These species were chosen for practical reasons. First, populations breed on reserves; second, they are currently RSPB priority species and have been monitored both on reserves and in the wider countryside (see below); and third, and most importantly, conservation interventions are designed to closely match their biological breeding requirements, making the number of breeding birds a natural response to the conservation type we evaluated (Table S2.4).

In the case of Snipe and Yellow Wagtail, a large proportion of their breeding reserve population (59% and 90%, respectively, at the start of the period analysed) occurred at a single reserve, the Ouse Washes in Norfolk and Cambridgeshire. This site is atypical because breeding birds are sometimes disrupted by flooding during the breeding season; the site is designed to temporarily store floodwater. This flooding is outside the control of the reserve management and explains population declines for Black-tailed Godwit (Ratcliffe et al., 2005). We therefore carried out analyses with and without the Ouse Washes for Snipe and Yellow Wagtail.

We obtained matching data to compute counterfactual population trends from the U.K. Breeding Bird Survey (BBS), managed by the British Trust for Ornithology. This scheme was started in 1994 and monitors changes in the national breeding trends of more than a hundred common and widespread bird species (Gregory et al., 2000; Harris et al., 2019). Surveying is performed in 1 × 1 km grids, each consisting of 10 transects. The type of habitat is recorded in a separate visit prior to 2 annual bird counts between April and June (Appendix 2.1). We used the habitat data recorded in the BBS and elevation data from the OS terrain 50 data set and the USGS EROS Archive – Digital Elevation (SRTM) 1 Arc-Second Global to calculate mean elevations.

We selected observations from lowland wet grassland sites and target species to create 1 reserve sample (i.e., treated sites) and matched the BBS data exactly on covariates affecting reserve selection and breeding trend (Table 2.1) to create the counterfactual sample (i.e. the control sites) for each species. We call this our benchmark counterfactual, as opposed to 2 other variants introduced to test sensitivity of the results (see below and Table 2.1). The counterfactuals were created by selecting observations from BBS grids containing certain habitats (Table 2.1) because we believe these are the best approximations of how reserve land would have developed without reserve conservation. We did not set a minimum proportion of the selected habitats or the exact mix of habitats that a grid had to contain to be included in the counterfactual sample. In the BBS, birds are counted in transect of 200 m and habitat is determined similarly. This also means that both bird numbers and habitat distinction come with some uncertainty regarding exactly where habitat changes and birds are observed. To account for this uncertainty, we operated on 1-km grid level.

We used a directed acyclic graph (DAG) to present our hypothesis for how wetland conservation affects breeding trends and to select matching covariates (Fig. 2.1) (e.g., Stuart, 2010; Pearl & Mackenzie, 2018; Hernan & Robins, 2020). Lowland wet grassland conservation is a cause of change in habitat quality (habitat, hydrology, food availability, and predator pressure) (Smart & Coutts, 2004; Verhulst et al., 2007; Eglington et al., 2008; Acreman et al., 2010; Ausden & Bolton, 2012; Smart et al., 2014), which then causes a change in the breeding trend. Habitat quality is improved by converting or forming the habitat from other habitat types to grassland, by changing the hydrological conditions using water control structures and land forming, by maintaining a habitable sward through grazing by domestic livestock and mowing; by

mechanically removing shrubs and trees to remove perches for avian predators; and by reducing the impact of predation by controlling or excluding generalist predators.

We excluded counts from the matched control sample if they originated from grids spatially overlapping with the chosen reserves (see “Stable Unit Treatment Value Assumption” in Rubin [1980]). Transect counts were summarized for each grid, excluding transect counts with >10 individuals as birds on passage because it is not likely the study species breed in such high densities (Field & Gregory, 1999). The maximum annual grid count for each species was used, and grids surveyed only once over the entire period were excluded. Furthermore, to avoid uncertain trend estimates, we excluded all BBS species that were observed in <30 grids annually (Newson et al. 2009). Pre-analysis data manipulation and graphics were done with the tidyverse packages (Wickham et al., 2019) and DAGs with the dagitty package (Textor et al., 2016). All analysis, visualization, and manipulation were implemented using R version 3.5.1 (R Core team, 2019).

Table 2.1 *The variables used in the creation of the benchmark, liberal and stringent counterfactuals.*

Benchmark	Liberal	Stringent
<ul style="list-style-type: none"> • Elevation < 250m • From 1994-2018 • From the UK • Contains the target species • Grids* containing semi-natural 	<ul style="list-style-type: none"> • From 1994-2018 • From the UK • Contains the target species 	<ul style="list-style-type: none"> • Elevation < 250m • From 1994-2018 • From the UK • Contains the target species • Grids* containing semi-natural grassland types more similar to wet

<p>grassland / marsh (chalk downland, grass moor, grass moor mixed with heather, machair other dry grassland, water- meadow/grazing marsh, reed swamp, other open marsh or saltmarsh)</p>		<p>grassland (dry grasslands, water meadows/grazing marsh, reed swamp or other open marsh).</p>
---	--	---

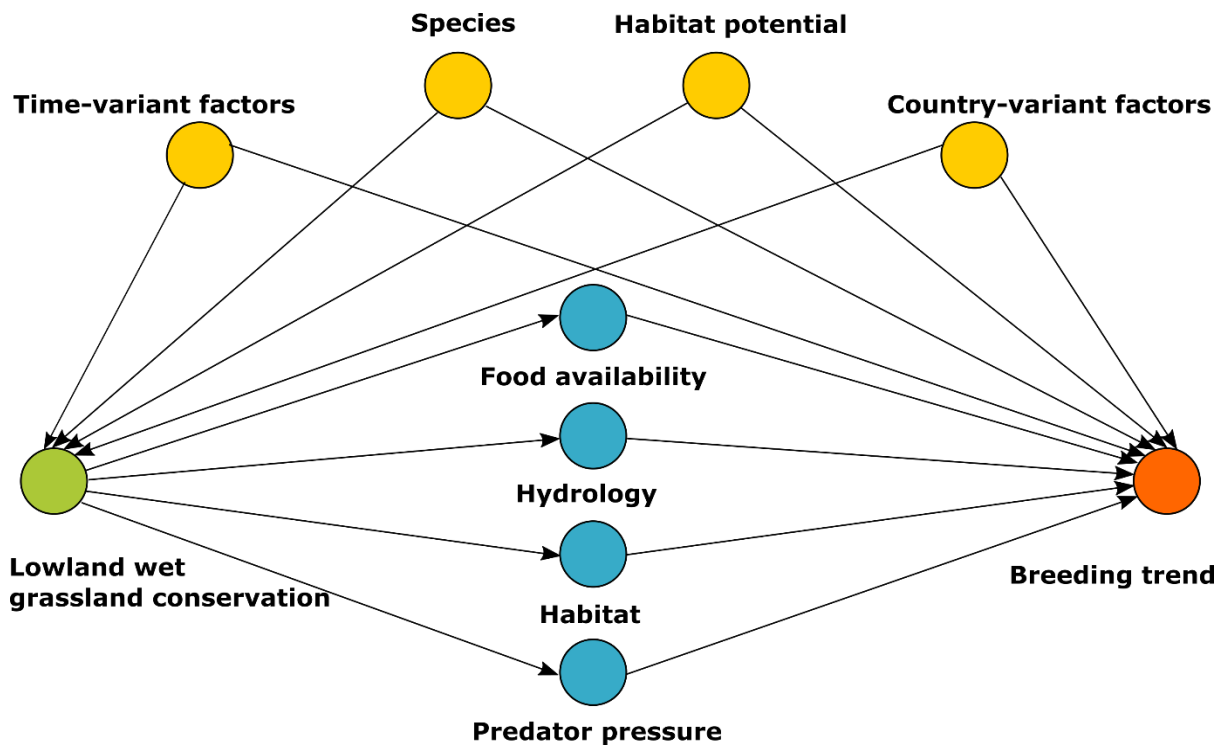


Figure 2.1 Hypothesized effect of management of habitat, hydrology, food availability, and predators in lowland wet grasslands on breeding bird trends (yellow circles represent confounding factors that may affect reserve selection and the breeding trends).

Data analyses

We used imputed counts to calculate the species totals used to create reserve and counterfactual trend indices. Imputed means that if a given site (BBS grid or reserve site) in a given year has been monitored, then the observed count is used; otherwise, the missing count is estimated (Appendix 2.3). Missing population counts were estimated separately for each species x reserve or counterfactual combination with a loglinear model with Poisson error terms. Each count was modeled as a function of site and year effects (Eq. 1) with the rtrim package (Bogart et al., 2020). The SE was adjusted for overdispersion and temporal autocorrelation (Bogart et al., 2020; Pannekoek et al., 2018).

$$\text{Ln}Y_{ij} = \alpha_i + \beta_j \quad (1)$$

where Y_{ij} is the estimated count for site i at time j , α_i is the average log-count of site i , and β_j is the average log-count deviation at time j across all sites.

We used indices to reflect relative changes in breeding pairs through time. The indices were calculated by dividing each annual total imputed count by a reference value that was set as the total count in the first time point (year 1994). Each set of indices was

then tested against its counterfactual to examine whether the two sets of indices were different based on a Welch 2-sample t test. If any difference could be statistically substantiated ($p < 0.05$), the effect size was assessed as the mean trend of the counterfactual indices subtracted from the corresponding annual reserve indices.

A concern with quasi-experimental inferences is whether the correct variables have been included in the matching process (Stuart, 2010). We therefore created two alternative counterfactuals, imposing different matching requirements (Table 2.1). We created a liberal counterfactual that imposes only exact species as a covariate restriction. The liberal counterfactual relaxes the criterion to define like for like in control populations but has the potential advantage of increasing the number of control populations. This counterfactual assumes that, on average, the reserve populations would have developed like any other population in the United Kingdom. We also created a stringent counterfactual that matches on exact species observations and has a subset of the habitat types used in the benchmark that is closer to the lowland wet grasslands in RSPB reserves. That is, matching grids were lowland (mean elevation below 250 m) and contained transects of either dry grassland; water meadows or grazing marsh; reed swamp; or open marshland. The stringent counterfactual thus assumes that, for each species, the average reserve trend would have developed like that of an average primarily lowland wet habitat regardless of conservation action. The increase in similarity requirements of matching populations comes at the cost of further limiting their numbers, thus potentially reducing the statistical power of the analyses. However, it might better describe the effect of conservation by reducing confounding effects. We assessed whether the results were robust to the counterfactual used by comparing the t-test results from both the liberal and stringent counterfactual (each one tested separately against the reserve indices)

with the t-test results of the benchmark counterfactual (benchmark indices tested against reserve indices). We also examined the relationship between site age and changes in breeding counts and whether reserve trends were sensitive to exclusion of sites with large breeding counts (Figure S2.4).

Results

Shoveler, Garganey, and Black-tailed Godwit were not sufficiently represented in the BBS data to create valid benchmark counterfactuals but showed either stable or increasing trends on reserves (Figure S2.1). The distribution of the remaining target species across lowland wet grassland reserve sites varied considerably. Lapwing and Curlew were present on most reserve sites and BBS grids, and Yellow Wagtail and Redshank were consistently rarer than other species, regardless of the counterfactual approach used (Table 2.2 and Table S2.2). The BBS grids used for the benchmark counterfactuals consisted primarily of farmland (45.5%), wet grassland transects (seminatural grassland types used in the stringent counterfactual in Table 2.1) (19.9%), and other semi-natural grassland transects (remaining semi-natural grassland types) (12.7%), whereas the liberal counterfactuals consisted primarily of farmland (67.3%) and other habitat types (24.7%). The stringent counterfactuals consisted primarily of wet grassland transects (27.6%) and farmland (47.4%) (Table S2.3). The largest relative increase in breeding pairs occurred within the first 10 years of reserve creation (Figure S2.5).

The breeding indices for Snipe and Yellow Wagtail across all lowland wet grassland reserves could not be statistically distinguished from their benchmark counterfactuals (Snipe: $t = 1.9$, $df = 40$, $p = 0.07$; Yellow Wagtail: $t = -0.3$, $df = 39$, $p = 0.79$). However, when the Ouse Washes was excluded from the reserve data set (because its spring

flooding is known to negate the effect of wetland management), the Snipe indices became more positive than its benchmark counterfactual (Fig. 2.2 & Figure S2.2) ($t = 4$, $df = 47$, $p = 0.0002$). The indices for Yellow Wagtail were unchanged by this exclusion (Figure S2.2).

Indices of Lapwing ($t = 7.6$, $df = 40$, $p < 0.0001$), Redshank ($t = 9.4$, $df = 45$, $p < 0.0001$), and Curlew ($t = 5.3$, $df = 35$, $p < 0.0001$) were all more positive on reserves. The mean annual trend difference represented an improvement of around 2.4% for Lapwing, 4.5% for Redshank, 1.5% for Snipe (Ouse Washes excluded) and 1.4% for Curlew. Thus, from 1994 to 2018 on lowland wet grassland reserves, Snipe populations increased by 36%, whereas the benchmark counterfactual remained stable around 1, suggesting that conservation interventions on these reserves were responsible for that increase. Curlew populations decreased by 23% compared with a 55% decline on the benchmark counterfactual, implying a 33% improvement caused by conservation interventions on reserves. From 1994 to 2018, Lapwing populations increased by 13%, but the benchmark counterfactual suggested they would have decreased by 44% without conservation interventions, resulting in a 57% index improvement by conservation. Redshank populations on reserves increased by 51%, whereas the benchmark counterfactual decreased by 57% without conservation, attributing a relative improvement of 108% to conservation interventions.

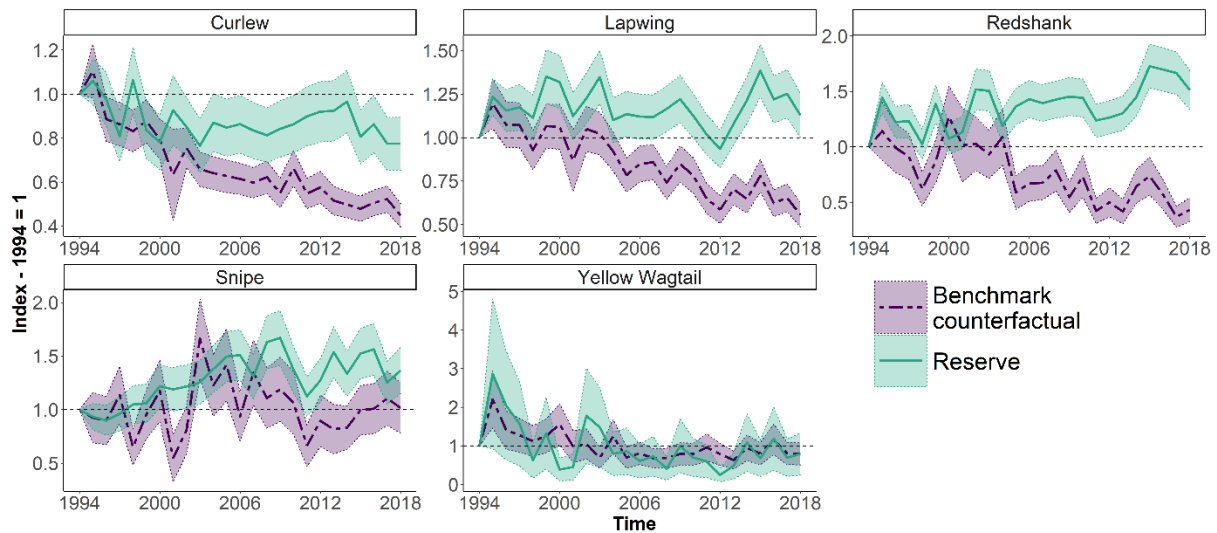


Figure 2.2 Breeding trends from 1994 to 2018 for the five target bird species inside reserves (solid line) and the benchmark counterfactual trends (dashed line) (shading, SE). The Ouse washes reserve was excluded for Snipe. Indices are calculated using imputed counts from loglinear models (see Methods).

Regardless of which counterfactual we compared with, the reserve indices were more positive for the four wading bird species and similar for Yellow Wagtail (Figure 2.3). The three counterfactuals were very similar for Lapwing, Redshank and Yellow Wagtail but more dissimilar for Curlew and Snipe. The difference between the Curlew reserve indices and its liberal counterfactual became less pronounced (Figure 2.3; $t = 2.4$, $df = 39$, $p = 0.02$) than when the reserve indices were compared with the benchmark scenario, whereas the reserve indices differed more from their stringent counterfactuals for both Curlew ($t = 5.1$, $df = 32$, $p < 0.0001$) and Snipe ($t = 10.2$, $df = 48$, $p < 0.0001$).

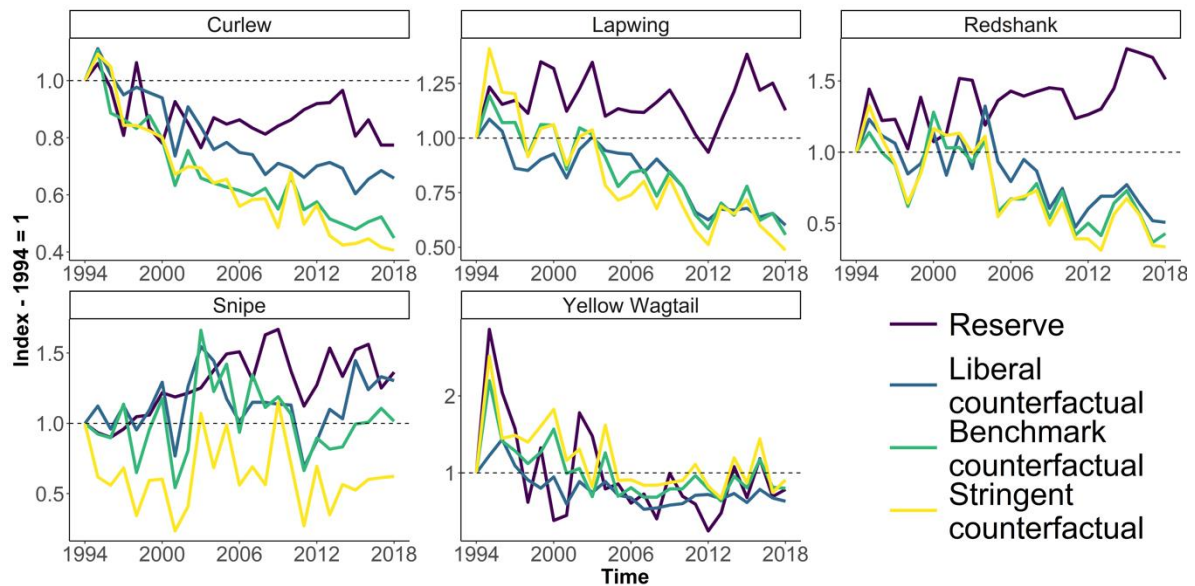


Figure 2.3 Breeding trends from 1994 to 2018 based on the liberal, benchmark, and the stringent matching settings. The Ouse washes reserve was excluded for Snipe. Indices are calculated using imputed counts from loglinear models (see Methods). To ease visualization, SEs are excluded (see Figure S2.3 for figure with SEs).

Discussion

We used a quasi-experimental approach to demonstrate how long-term population monitoring data can be used to evaluate the impact of conservation. We found that lowland wet grassland conservation has benefitted Lapwing, Redshank, and Curlew populations and, if an atypical site is excluded, that it also benefitted Snipe. We found no reserve effect for Yellow Wagtail and were not able to compare breeding populations of three other species (Black-tailed Godwit, Garganey, and Shoveler) because they were too rare outside of nature reserves, although they showed either stable or increasing trends on reserves. Based on the benchmark counterfactual trends, Snipe (Ouse Washes excluded), Lapwing, and Redshank populations all increased on reserves, but would have decreased or remained stable without this

conservation, whereas Curlew populations decreased much less on reserves than they would otherwise have done. For the four wading bird species, the reserve indices were higher than their counterfactuals regardless of which counterfactual they were compared with; positive effects of reserve conservation were strong in all cases. However, different counterfactuals can produce different results, here illustrated by the different counterfactual trends in each species (Figure 2.3). The effect of reserve conservation became less pronounced for Curlew under the liberal counterfactual, suggesting that this species may be faring slightly better in habitats other than wet grassland. Nevertheless, the differences in the three counterfactual trends for Curlew were small (Figure 2.3). Overall, our findings concur with others (Ausden et al., 2019; Verhulst et al., 2007) in substantiating the positive effects of conservation actions on target breeding wetland bird populations.

The target wading bird species in our study should theoretically benefit from lowland wet grassland conservation, but not necessarily in equal measure. European grassland-breeding wading birds display species-specific responses to different types of grassland conservation (Franks et al., 2018). Wetland conservation management incorporates a range of different intervention types – from the conversion of, for example, ex-arable land to grassland, to changes in hydrology and grazing and mowing regimes. The degree to which each intervention type provides suitable conditions for the different study species may therefore differ. For example, Ausden et al. (2019) suggest that limiting livestock grazing in spring, which aims to reduce trampling of wading birds' nests, could also reduce habitat quality for Yellow Wagtail because they often feed in close association with domestic livestock. While Yellow Wagtail breed in wetland habitats, it has not been a priority species until recently and has not been actively targeted by management. This species is also the only long-

distance migrant among the study species, and changes on its wintering grounds in Africa and migration paths may also affect its breeding population (Wood, 1992; Newton, 2006), thereby rendering conservation efforts in the breeding range less effective or redundant.

There are also multiple reserve specific conditions we did not account for. For example, because of improved breeding conditions, new sites recruit breeding pairs faster than older reserve sites (Figure S2.5). Further research is needed to explore why reserve effects differ across study species (e.g. the declining reserve trend for Curlew in contrast to the increasing reserve trends for Redshank, Lapwing, and Snipe) and in particular how population responses relate to site-specific interventions, reserve age and size, and finer-scale abiotic and habitat covariates.

We created separate reserve and counterfactual indices for each species based on the total annual number of breeding pairs. Because of the method used, a large decline on one reserve and stable or slightly increasing breeding numbers in all other reserves could still produce a decreasing trend if the total number of breeding birds declined overall. This can potentially mask the individual reserves' conservation success, as illustrated when excluding the Ouse Washes from the analysis of Snipe populations. However, our results were largely robust to exclusion of sites with large proportions of breeding numbers (Figure S2.4).

The method we used provides several benefits over other evaluation methods for conservation impacts. It allows the use of population monitoring data sets to emulate a robust ex post study design. The interpretation of the results is intuitive (diverging lines in Figure 2.2 mean that the observed scenario differs from its counterfactual), and results are easily communicated to an audience without statistical knowledge.

Although our method is marginally more complex than study designs such as the “After”, it does not require more resources. European monitoring data, such as the BBS data, are often freely available.

This method also allows a more detailed analysis of impacts than other study designs. For example, using the “After” method, which examines the reserve trend after the establishment of the reserve exclusively, Redshank and Snipe would be the only species with a clear increasing trend. Assessing whether reserve conservation works exclusively based on whether a population trend is increasing implicitly assumes that the population would remain stable in the absence of conservation, which is far from the reality of ongoing population declines outside reserves (Harris et al., 2019). If the assessment had been done using a classical land-use control-impact study design, where the number of birds in each reserve would have been counted at one point in time, we would be able to compare densities but not trends. Our method (after-control-impact) ex post compares trends and depicts the dynamic development of populations through time, whereas control-impact studies provide only a temporal snapshot. The dynamic element is advantageous because it allows identification of divergent mechanisms through time and shows visually how adding new reserves affects the overall reserve trend.

Matching is increasingly being used in combination with regression techniques to assess the effect of conservation initiatives (Terraube et al., 2020). However, matching alone does not necessarily improve effect inferences and, because of reductions in sample size, may not have the same power to detect effects as regression techniques (Brazauskas & Logan, 2016). The RSPB reserve and BBS data sets we used covered long periods (>20 years) and included breeding bird counts derived from robust study designs. Such data sets are not common, and a quasi-experimental evaluation design

like ours will not necessarily be applicable or appropriate elsewhere (see Walker et al., [2018] for alternative impact evaluation using BBS monitoring data). Furthermore, for matching to be appropriate, it requires a clear theory of how the treatment changes the outcome (Figure 2.1 and Data section) and careful selection of matching variables and methods accordingly (Schleicher et al., 2019). Using exact matching, we were able to retain sufficiently large sample sizes to run the loglinear models for five out of eight species. Other quasi-experimental designs with fewer data or higher covariate complexity (higher number of covariates or continuous covariates) will either be impractical or require other matching methods (Iacus et al., 2019).

Reserves and BBS grids are surveyed using different survey protocols. Some of these differences could potentially lead to larger uncertainty and year-on-year variance; however, we do not believe this is the case. Each grid or site is surveyed with consistent effort each year, which means that a potential bias is also consistent and accounted for by using indices. Additionally, the counterfactuals created from the BBS are generally based on a relatively large number of annual observations. For further discussion see Appendix 2.2. One way to create credible counterfactuals is through well-monitored control areas. This should reduce the likelihood of a mis-specified control group and enhance the credibility of the inference, but in order to make this possible, monitoring of control sites must be a priority, with a further emphasis on consistent survey method. This may be difficult for the reasons described in the Introduction. Our results nonetheless suggest that dedicated conservation efforts have benefited target lowland wet grassland bird species and that monitoring programs can be used to evaluate the impact of conservation interventions by creating credible counterfactuals through matching approaches.

References

Acreman M C., Mountford J O, Gowing DJ, Stratford C, Durrel S, Duenas M A., Rispen, E., Youngs E, & Moy IL. 2010. Restoration of wet grasslands through re-instatement of surface grips. Extended Final report to Department for Environment, Food and Rural Affairs. Centre for Ecology and Hydrology, Wallingford.

Adams V M, Barnes M, & Pressey R L. 2019. Shortfalls in Conservation Evidence: Moving from Ecological Effects of Interventions to Policy Evaluation. *One Earth*, 1(1), 62–75.

Ausden M & Hirons G J M. 2002. Grassland nature reserves for breeding wading birds in England and the implications for the ESA agri-environment scheme. *Biological Conservation*. [https://doi.org/10.1016/S0006-3207\(01\)00254-3](https://doi.org/10.1016/S0006-3207(01)00254-3)

Ausden M & Bolton M. 2012. Breeding waders on wet grassland: Factors influencing habitat suitability. In R. Fuller (Ed.), *Birds and Habitat: Relationships in Changing Landscapes* (Ecological Reviews, pp. 278-306). Cambridge: Cambridge University Press.

Ausden M, Hirons G, & White G. 2019. Wetland restoration by the RSPB – what has it achieved for birds? *British Birds*, 112 (June), 315–336.

Baylis K, Honey-Rosés J, Börner J, Corbera E, Ezzine-de-Blas D, Ferraro P J, Lapeyre R, Persson U M, Pfaff A & Wunder, S. 2016. Mainstreaming Impact Evaluation in Nature Conservation. *Conservation Letters*, 9(1), 58–64.

Brazauskas, R & Logan B R. 2016. Observational Studies: Matching or Regression? *Conservation Letters*, 9(3), 289–313.

Boatman N D, Parry H R, Bishop J D & Cuthbertson A G S. 2007. Impacts of agricultural change on farmland biodiversity in the UK. *Environmental Science and Technology* 25. <https://doi.org/10.1039/9781847557650-00001>

Bogart P, van der Loo M & Pannekoek, J. 2020. rtrim: Trends and Indices for Monitoring Data. R package version 2.1.1. <https://CRAN.R-project.org/package=rtrim>

Christie A P, Amano T, Martin P A, Shackelford G E, Simmons B I & Sutherland W J. 2019. Simple study designs in ecology produce inaccurate estimates of biodiversity responses. *Journal of Applied Ecology* 5: 2742–2754.

Crick H Q P. 1992. A bird-habitat coding system for use in Britain and Ireland incorporating aspects of land-management and human activity, *Bird Study* 39:1, 1-12,

Colhoun K, Mawhinney K, McLaughlin M, Barnett C, McDevitt A M, Bradbury R B & Peach W. 2017. Agri-environment scheme enhances breeding populations of some priority farmland birds in Northern Ireland. *Bird Study* 64(4), 545–556.

De Palma A, Sanchez-Ortiz K., Martin P A, Chadwick A, Gilbert G, Bates A. E, Börger L, Contu S, Hill S L L & Purvis A. 2018. Challenges With Inferring How Land-Use Affects Terrestrial Biodiversity: Study Design, Time, Space and Synthesis. In *Advances in Ecological Research* 58(2). <https://doi.org/10.1016/bs.aecr.2017.12.004>

Department for Environment, Food and Rural Affairs, UK. 2019. UK Biodiversity Indicators 2019.

Eaton M A, Gibbons D W, Evans A, Gregory R D, Brown A F, Noble D G, Musgrove A J, Hearn R D & Aebischer N J. 2015. Birds of conservation concern 4: the population status of birds in the united kingdom, channel islands and isle of Man. *British Birds*, 108(6), 708–746.

Eglington S M, Gill J A, Bolton M, Smart M A, Sutherland W J & Watkinson A R. 2008. Restoration of wet features for breeding waders on lowland grassland. *Journal of Applied Ecology*, 45, 305–314.

Ferraro P J. 2009. Counterfactual thinking and impact evaluation in environmental policy. *New Directions for Evaluation*, 122, 75- 84.

Ferraro P J, McIntosh C & Ospina M. 2007. The effectiveness of the US endangered species act: An econometric analysis using matching methods. *Journal of Environmental Economics and Management*, 54(3), 245–261.

Ferraro P J & Pattanayak S K. 2006. Money for Nothing? A Call for Empirical Evaluation of Biodiversity Conservation Investments, *Plos biology* 4(4), 482–488.

Ferraro P J & Pressey R L. 2015. Measuring the difference made by conservation initiatives: Protected areas and their environmental and social impacts. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370(1681), 4–8.

Field, R H. & Gregory R D. 1999. Measuring population changes from the Breeding Bird Survey. BTO Research Report No. 217. The Nunery, Thetford, Norfolk.

Franks S E, Roodbergen M, Teunissen W, Carrington Cotton A & Pearce-Higgins J W. 2018. Evaluating the effectiveness of conservation measures for European grassland-breeding waders. *Ecology and Evolution*, 8(21), 10555–10568.

Geldmann J, Manica A, Burgess N D, Coad L & Balmford A. 2019. A global-level assessment of the effectiveness of protected areas at resisting anthropogenic pressures. *Proceedings of the National Academy of Sciences of the United States of America*, 116(46), 23209–23215. <https://doi.org/10.1073/pnas.1908221116>

Gilbert G, Gibbons D W, Evans J. 1998. *Bird Monitoring Methods*. RSPB, Sandy.

Gregory R, Baillie S, Bashford R. 2000. Monitoring breeding birds in the United Kingdom. *Bird Census News* 13, 101–112.

Harris S J, Massimino D, Eaton M A, Gillings S, Noble D G, Balmer D E, Pearce-Higgins J W & Woodcock P. 2019. The Breeding Bird Survey 2018. BTO Research Report 717. British Trust for Ornithology, Thetford.

Hayhow D B, et al. 2017. The state of the UK's birds 2017. The RSPB, BTO, WWT, DAERA, JNCC, NE and NRW, Sandy, Bedfordshire

Hayhow D B, et al. 2019. The State of Nature 2019. The State of Nature partnership.

Hernan M A, Robins J M. 2020. *Causal Inference: What If*. Boca Raton: Chapman & Hall/CRC.

Hoffmann M, Duckworth J W, Holmes K, Mallon D P, Rodrigues A S L & Stuart S N. 2015. The difference conservation makes to extinction risk of the world's ungulates. *Conservation Biology*, 29(5), 1303–1313.

Hoffmann M, et al. 2010. The impact of conservation on the status of the world's vertebrates. *Science*, 330(6010), 1503–1509.

Iacus S M, King G & Porro G. 2019. A theory of statistical inference for matching methods in causal research. *Polit. Anal.* 27, 46–68.

IPBES (2019): Summary for policymakers of the global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. S Díaz, et al (eds.). IPBES secretariat, Bonn, Germany. 56 pages.

Jefferson R G & Grice PV. 1998. The conservation of lowland wet grassland in England. In C.B. Joyce & P.M. Wade (eds) *European Wet Grasslands: Biodiversity, Management and Restoration*: 31– 48. Chichester: Wiley

Joppa L & Pfaff A. 2010. Reassessing the forest impacts of protection. *Annals of the New York Academy of Sciences*, 1185(1), 135–149.

Malpas L R, Kennerley R J, Hirons G J M, Sheldon R D, Ausden M, Gilbert J C. & Smart J. 2013. The use of predator-exclusion fencing as a management tool improves the breeding success of waders on lowland wet grassland. *J. Nat. Conserv.* 21, 37–47 (2013).

Margoluis R, Stem C, Salafsky N & Brown M. 2009. Design alternatives for evaluating the impact of conservation projects. *New Directions for Evaluation*, 122, 85–96.

Newton I. 2006. Can conditions experienced during migration limit population levels of birds. *J Ornithol.* 147, 146-166.

Pannekoek J, Bogaart P, van der Loo M. 2018. *Models and statistical methods in rtrim*. Statistic Netherlands, Haga

R Core Team. 2019. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

Pearl J & Mackenzie D. 2018. *The Book of Why*. Basic Books, New York.

Ratcliffe N, Schmitt S & Whiffin M. 2005. Sink or swim? Viability of a black-tailed godwit population in relation to flooding. *Journal of Applied Ecology*, 42(5), 834–843.

Rubin R D. 1980. Discussion of “Randomization Analysis of Experimental Data in the Fisher Randomization Test,” by D. Basu, *Journal of the American Statistical Association*, 75, 591–593

Schleicher J, Eklund J, Barnes M, Geldmann J, Oldekop J A & Jones J P G. 2019. Statistical matching for conservation science. *Conservation Biology*, 34, 538-549.

Sills E O, de Sassi C, Jagger P, Lawlor K, Miteva D A, Pattanayak S K and Sunderlin W D. 2017. Building the evidence base for REDD+: study design and methods for evaluating the impacts of conservation interventions on local well-being *Glob. Environ. Change* 43, 148–60

Rubin DB: Randomization analysis of experimental data: the fisher randomization test. *J Am Stat Assoc* 1980, 75:591-593

Smart J, Wotton S R, Dillon I A, Cooke A I, Diack I, Drewitt A L, Grice P V & Gregory R D. 2014. Synergies between site protection and agri-environment schemes for the conservation of waders on lowland wet grasslands. *Ibis*, 156(3), 576–590.

Smart M & Coutts K. 2004. Footdrain management to enhance habitat for breeding waders on lowland wet grassland at Buckenham and Cantley Marshes, Mid-Yare RSPB Reserve. *Conservation Evidence*, 1, 16–19.

Sonter LJ, Barnes M, Matthews JW, Maron M. 2019. Quantifying habitat losses and gains made by U.S. Species Conservation Banks to improve compensation policies and avoid perverse outcomes. *Conservation Letters*:1–9.

Stuart E A. 2010. Matching Methods for Causal Inference: A Review and a Look Forward. *Statistical Science*, 25(1), 1–21.

Terraube J, Van doninck J, Helle P, Cabeza M. 2020. Assessing the effectiveness of a national protected area network for carnivore conservation. *Nature Communications* 11:1–9.

Textor J, van der Zander B, Gilthorpe M S, Liškiewicz M & Ellison G T. 2016. Robust causal inference using directed acyclic graphs: The R package “dagitty.” *International Journal of Epidemiology*, 45(6), 1887–1894.

Verhulst J, Kleijn D & Berendse F. 2007. Direct and indirect effects of the most widely implemented Dutch agri-environment schemes on breeding waders. *Journal of Applied Ecology*, 44, 70-80

Walker L K, Morris A J, Cristinacce A, Dadam D, Grice P V & Peach W J. 2018. Effects of higher-tier agri-environment scheme on the abundance of priority farmland birds. *Animal Conservation*, 21(3), 183–192.

Wickham et al. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43), 1686, <https://doi.org/10.21105/joss.01686>

Wiik E, d’Annunzio R, Pynegar E, Crespo D, Asquith N, Jones J P G. 2019. Experimental evaluation of the impact of a payment for environmental services program on deforestation. *Conservation Science and Practice* 1:1–11.

Wilson A M, Ausden M & Milsom T P. 2004. Changes in breeding wader populations on lowland wet grasslands in England and Wales: Causes and potential solutions. *Ibis*, 146(suppl. 2), 32–40.

Wilson A, Pendlebury C & Vickery J. 2005. Changes in Lowland Wet Grassland Breeding Wader Number: The Influence of Site Designation. BTO Research Report 365. British Trust for Ornithology, Thetford.

Wood B. 1992. Yellow Wagtail *Motacilla flava* migration from West Africa to Europe: pointers towards a conservation strategy for migrants on passage. *Ibis*, 134 (suppl 1), 66–76.

Supplementary material

Table S1.1 *Number of sites or grids per year available in the reserve and counterfactual indices. R = number of reserve sites, B = number of grids in the benchmark counterfactual, L = number of grids in the liberal counterfactual, S = number of grids in the stringent counterfactual. Shown for each species.*

Year	Black-tailed godwit				Curlew				Garganey				Lapwing				Redshank				Shoveler				Snipe				Yellow wagtail			
	R	B	L	S	R	B	L	S	R	B	L	S	R	B	L	S	R	B	L	S	R	B	L	S	R	B	L	S	R	B	L	S
1994	0	5	14	3	13	80	700	51	0	1	10	1	26	105	976	69	24	31	185	20	0	14	55	12	22	59	321	37	1	23	376	16
1995	2	6	15	5	11	89	757	59	14	2	10	2	23	114	1078	81	21	32	202	23	9	14	61	13	16	60	359	37	5	26	417	18
1996	2	9	24	6	16	97	851	61	14	2	10	2	31	119	1188	84	28	39	221	26	9	18	69	17	25	68	420	43	5	28	447	20
1997	2	8	22	7	16	110	961	76	14	2	11	2	33	139	1360	103	29	42	236	31	9	20	74	19	26	79	437	51	5	36	533	26
1998	2	6	22	5	16	120	994	84	14	1	11	1	34	146	1398	110	30	50	255	39	8	21	82	20	26	87	461	55	5	36	519	28
1999	2	8	21	7	16	125	1028	93	14	1	11	1	37	147	1449	115	33	50	257	39	8	23	79	22	27	82	448	53	5	36	541	27
2000	2	9	23	8	16	117	979	89	14	3	11	3	40	136	1390	102	37	50	262	41	11	24	79	23	30	70	426	47	6	35	516	27
2001	2	4	9	3	6	17	165	14	15	0	3	0	34	30	328	23	31	15	63	13	12	9	35	8	21	12	78	10	6	8	130	8
2002	3	8	26	7	16	105	924	75	18	2	8	2	44	128	1311	91	41	44	235	35	13	19	76	18	31	66	414	44	6	26	460	22
2003	3	9	26	8	18	120	969	88	20	3	12	3	51	146	1386	111	45	52	252	41	15	27	88	26	35	78	441	54	8	29	491	24
2004	3	9	24	8	19	135	1050	96	23	3	13	3	54	165	1500	125	46	52	266	39	18	26	83	25	38	85	462	55	11	42	527	34
2005	3	14	30	13	20	152	1137	107	25	4	17	4	62	195	1725	148	53	56	307	45	27	29	102	27	40	93	506	66	11	44	600	36
2006	3	13	37	10	20	152	1324	106	28	5	17	5	69	194	1988	143	59	62	377	49	29	32	104	28	42	86	628	57	16	51	673	37
2007	3	19	38	16	22	164	1610	110	30	4	15	4	75	217	2221	159	66	68	421	51	33	34	105	30	47	98	817	62	18	53	714	36
2008	3	15	35	12	22	145	1466	98	32	5	17	5	83	189	2026	139	73	58	390	44	38	32	93	30	48	83	747	54	21	44	639	30
2009	3	17	39	14	22	127	1439	85	34	5	14	5	86	173	2009	122	77	51	389	38	40	29	97	28	50	72	743	45	22	41	634	27
2010	3	13	38	10	22	130	1398	87	34	5	15	5	87	169	1957	127	76	46	377	36	43	26	92	25	50	62	715	39	23	40	601	28
2011	3	15	37	13	22	120	1177	82	35	4	16	4	90	155	1735	117	80	44	325	32	43	23	87	22	52	64	557	41	23	36	587	26

2012	3	13	35	10	22	119	1475	77	36	5	17	5	91	156	2017	112	82	42	387	30	44	24	102	23	51	61	735	40	26	41	628	30
2013	3	13	36	10	23	135	1566	90	38	5	18	5	95	171	2091	126	85	50	391	36	46	26	98	25	52	70	774	43	29	42	643	32
2014	3	13	34	11	23	165	1487	111	37	5	16	5	94	214	2078	158	83	67	354	48	46	31	100	29	51	99	646	61	26	48	692	38
2015	3	15	34	13	22	181	1465	121	38	6	17	6	96	236	2098	175	85	71	356	51	47	35	108	32	53	109	624	67	28	52	686	40
2016	3	16	32	13	22	187	1488	120	38	6	16	6	97	250	2116	180	87	74	349	52	46	37	105	34	53	112	652	67	27	59	720	43
2017	3	16	29	13	23	178	1481	114	38	6	17	6	97	257	2106	185	87	74	343	50	47	36	106	33	53	113	661	70	27	57	730	39
2018	3	15	29	13	23	173	1448	107	38	6	16	6	97	243	2020	172	87	72	356	49	46	36	99	33	53	108	662	69	27	50	671	34

0 **Table S2.2** Percentage of data points (site or grid x year) observed and the total
 1 number of estimated and observed data points used to create the reserve and
 2 counterfactual trends.

3

Species	Reserve		Benchmark counterfactual		Liberal counterfactual		Stringent counterfactual	
	Observed	Total	Observed	Total	Observed	Total	Observed	Total
Curlew	81.9%	575	35%	9275	51.6%	80575	32.4%	9500
Lapwing	67%	2425	34.4%	12175	47.3%	61925	33%	6675
Redshank	66.4%	2175	58.5%	2208	51.3%	14725	35.5%	2700
Snipe	74.8%	1325	34.8%	5675	45.3%	30300	34.5%	3675
Yellow wagtail	53.3%	725	37.4%	2625	55.6%	25475	35.4%	2050

4

5 **Table S2.2** The percentage of transects from all five target species used in the
 6 counterfactuals which contains the grassland types “Other dry grassland”, “Water
 7 meadows/ grazing marsh”, “Reed swamp” and “Other open marsh” (Wet grassland),
 8 remaining semi-natural grassland in at least one of the two primary habitat
 9 categories but none of the wet grassland habitat types (Grassland), farmland but not
 10 grassland (Farmland), and the percentage of transects which contain neither of the
 11 above habitat types (Other).

Counterfactual	Wet grassland	Grassland	Farmland	Other	Number of transects
Benchmark	19.9%	12.7%	45.5%	21.9%	67123
Liberal	2.8%	5.2%	67.3%	24.7%	612340
Stringent	27.6%	2.7%	47.4%	22.3%	47918

12

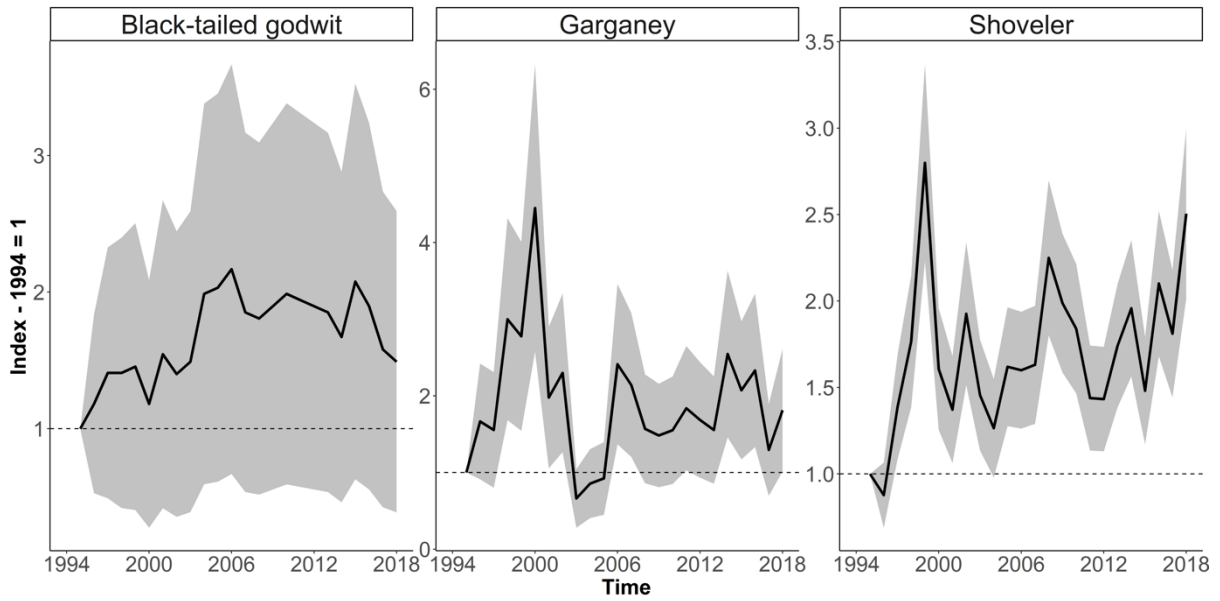
13 **Table S2.3** Proposed relationship between four categories of wetland management

14 and the species that benefit from these (X = benefit).

Species	Predator control	Vegetation management	Water control structures	Wet features
Black-tailed godwit	X	X	X	X
Curlew	X	X	X	
Garganey	X		X	X
Lapwing	X	X	X	X
Redshank	X	X	X	X
Shoveler	X		X	X
Snipe	X	X	X	X
Yellow Wagtail		X	X	X
Mechanism	Reduces levels of nest and chick predation	Used to provide suitable sward height and structure for nesting and feeding birds	Helps maintain optimal water levels for the relevant bird species	Increases the area of shallow water/mud for birds to feed in
Reference	Malpas et al 2013; Ausden et al 2019	Smart & Coutts 2004; Ausden et al 2019	Smart & Coutts 2004; Eglington et al 2008; Ausden et al 2019	Smart & Coutts 2004; Eglington et al 2008; Ausden et al 2019

15

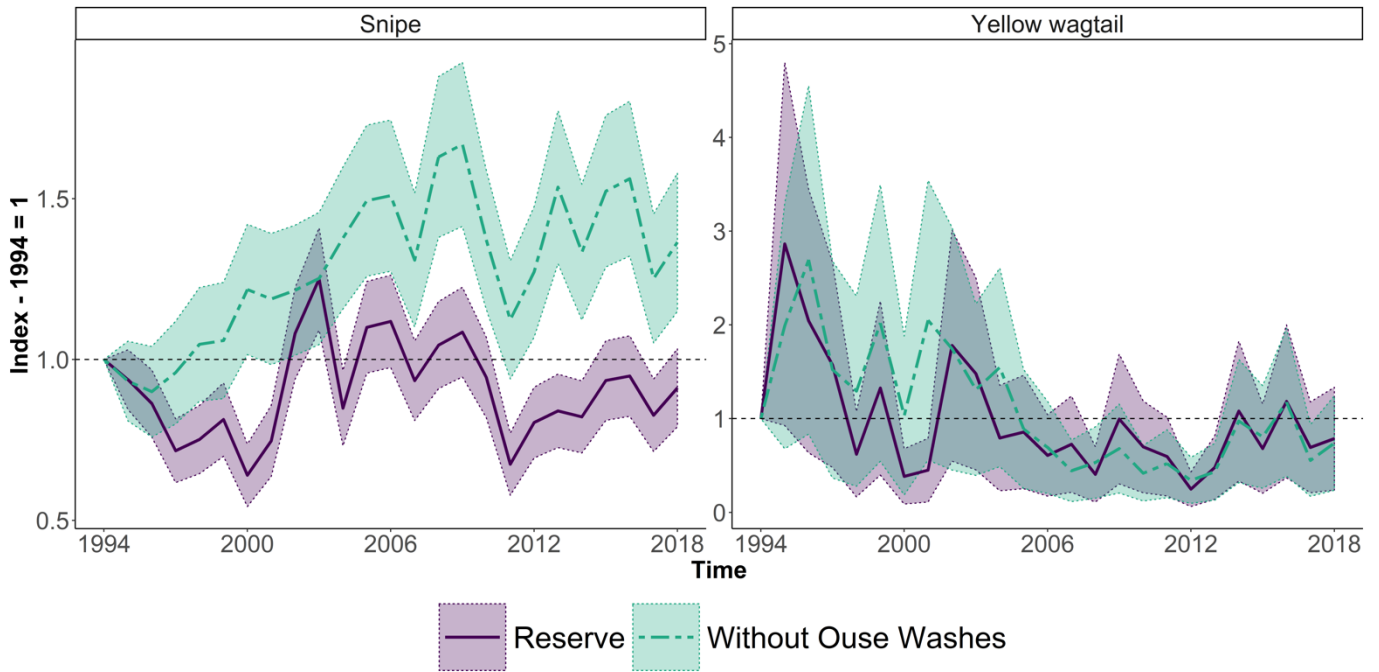
16



17

18 **Figure S2.1** Reserve trends from 1994-2018 for Black-tailed godwit, Garganey and
19 Shoveler. Indices have been calculated using imputed counts from loglinear models
20 (as described in the method section). The shaded area delineates the standard
21 error.

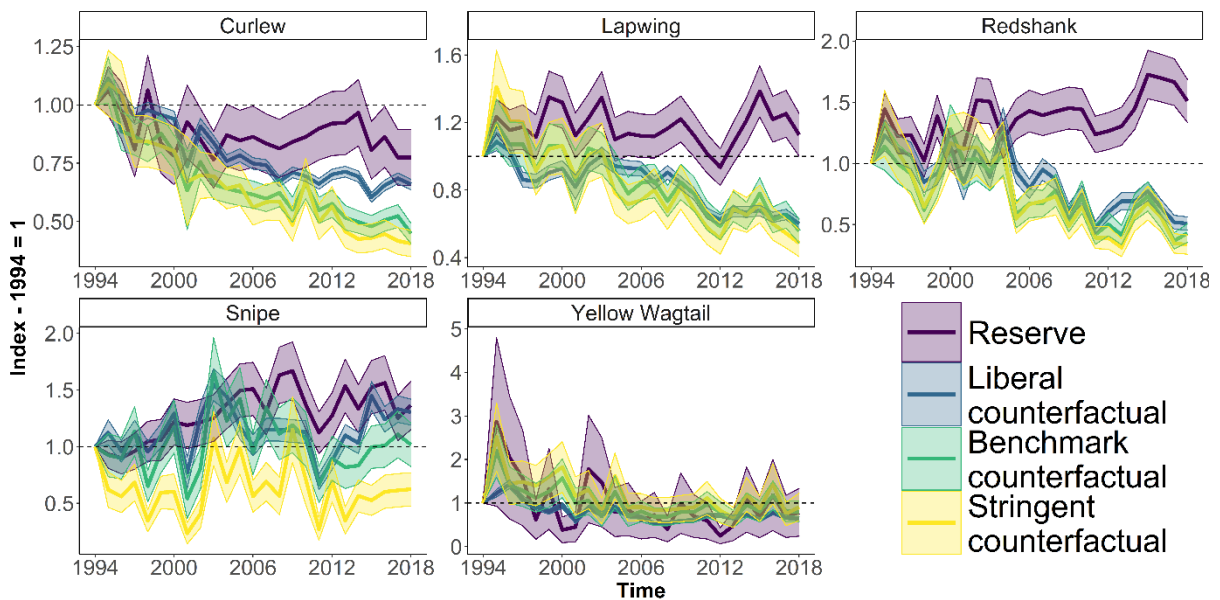
22



23

24 **Figure S2.2** The Snipe and Yellow Wagtail reserve trend with and without the Ouse
 25 Washes. Indices have been calculated using imputed counts from loglinear models
 26 (as described in the method section). The shaded area delineates the standard error.

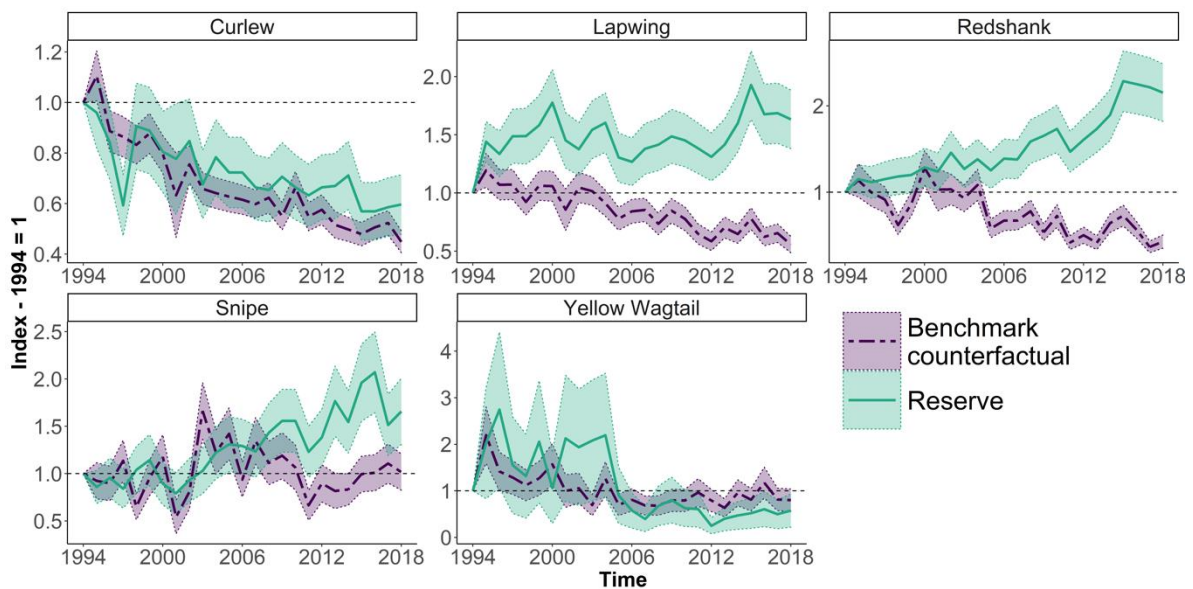
27



28

29 **Figure S2.3** Reserve and counterfactual breeding trends from 1994 - 2018 using the
 30 liberal, benchmark and the stringent matching settings as in earlier analysis. The Ouse
 31 washes reserve was excluded for Snipe. Indices were calculated using imputed counts
 32 from loglinear models (as described in the method section). The shaded area
 33 delineates the standard error.

34

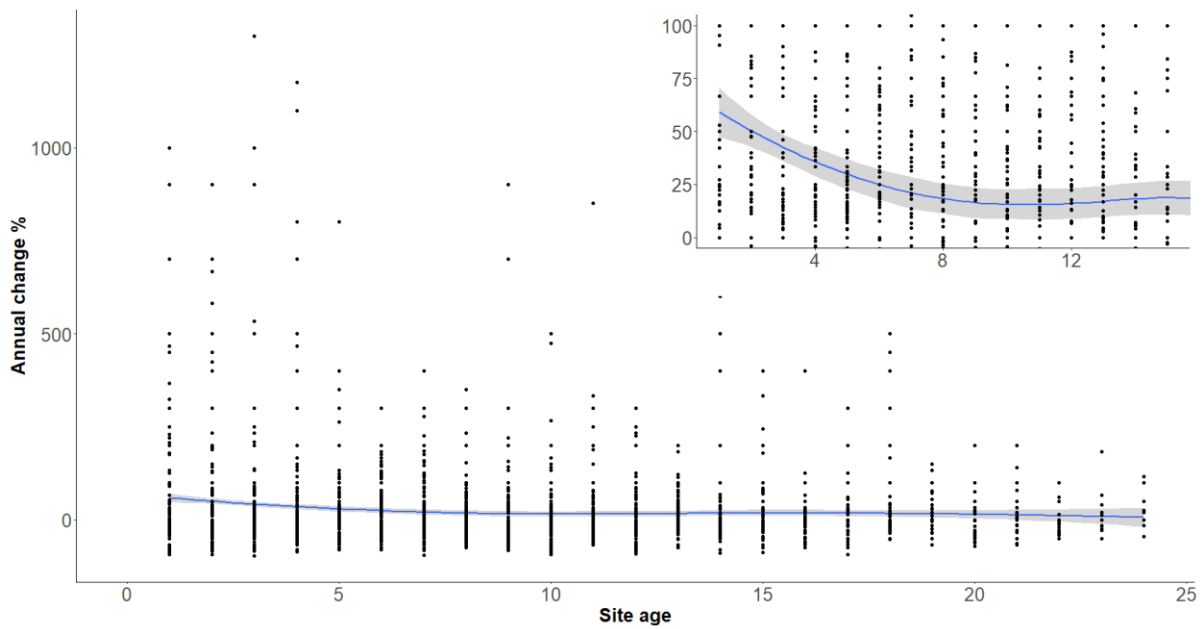


35

36 **Figure S2.4** Reserve and counterfactual breeding trends from 1994 – 2018. For each
 37 species, the top three reserve sites which contain the most breeding pairs have been
 38 excluded. Indices were calculated using imputed counts from loglinear models (as
 39 described in the method section). The shaded area delineates the standard error.

40

41

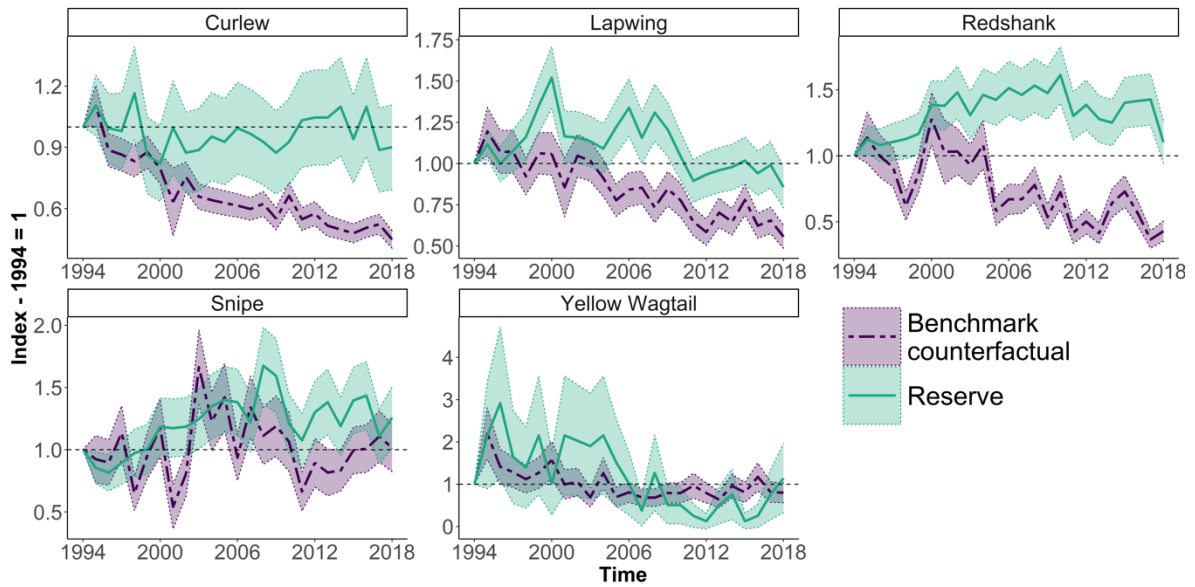


42

43

44 **Figure S2.5** Relationship between site age (using first site count as year 0) and annual
 45 change % in breeding pairs and corresponding 95% CI (grey shaded area). Fitted
 46 using a Loess regression for all sites acquired after 1993 using Snipe, Redshank,
 47 Lapwing, Curlew and Yellow Wagtail counts ($N = 3990$). To avoid infinite values, we
 48 added 1 to all breeding counts. The right corner of the graph shows the fitted
 49 relationship between annual change % restricted between 0 – 100 % in year 1- 15.

50



51

52 **Figure S2.6** Reserve and counterfactual breeding trends from 1994 - 2018 using
 53 reserve sites which were under RSPB management in 1994 and excluding Ouse
 54 Washes (Number of sites: Curlew = 10; Lapwing = 22; Redshank = 22; Snipe = 19;
 55 Yellow Wagtail = 4). Indices have been calculated using imputed counts from loglinear
 56 models (as described in the method section). The shaded area delineates the standard
 57 error.



Figure S2.7 Location of Lowland wet grassland reserves (green bird symbol) and BBS grids (red dots) used in the analysis.

Appendix 2.1

RSPB reserves

The RSPB has created lowland wet grassland reserves in low-lying habitats in the UK, mainly on drained grassland (78 sites), ex-arable land (16 sites), mixed grass and arable land (5 sites) and on a small number of ex-mineral extraction sites (2 sites) (Ausden et al. 2019). These areas are selected to be reserves because they either already support valuable assemblages of species, or are considered to have the potential to do so in the future, based on factors such as their water supply and potential to form part of large, contiguous blocks of wetland habitat. Reserve management follows a species-based conservation approach. This means that each reserve is managed in accordance with the biological requirements of the priority species selected for that reserve. Priority species are selected based on: 1) the species being of national conservation concern (Eaton et al. 2015); 2) a significant proportion of their breeding population being found on reserves; and 3) whether the conservation actions required to aid population recovery are considered known (RSPB 2010; Ausden et al. 2019).

RSPB reserves survey

Lowland breeding wader populations are surveyed using the standard protocol for surveying lowland sites (Gilbert et al. 1998). Breeding birds are counted three times a year, between mid-April and the end of June, with a minimum of one week between each visit. Survey sites are defined as the amount of ground that can be covered in one visit and can consist of separate fields. Each field is surveyed from 100 m and the

numbers of pairs are recorded. The maximum counts are used for Lapwing and Snipe whereas the mean number of birds are used for Curlew and Redshank following standard practice (Gilbert et al. 1998).

Breeding Bird Survey UK

The BTO/JNCC/RSPB Breeding Bird Survey is a national survey used to monitor breeding bird populations in the United Kingdom since 1994 (Gregory et al. 2000; Harris et al., 2019). The scheme employs a stratified random sample of 1-km squares, where squares are stratified regionally to match potential observer availability. The number of squares that have been surveyed in each year has gradually increased from around 1500 in 1994 to over 4000 in 2020. Birds are counted twice a year, once between April and mid-May (early visit), and once between mid-May and the end of June (late visit). In each square, the surveyor walks along two 1-km transects which are divided into 200-m sections. In each section, all birds are recorded and classified into three distance bands (0–25 m, 25–100 m, 100 m+) or in an ‘in flight’ category if they are only seen flying. In addition to birds, surveyors also record habitat (usually during a reconnaissance visit made before the bird breeding season). Each transect section is categorised using a hierarchical system, with two primary habitat types for each section and up to four detail levels further specifying each habitat categorisation (Crick, 1992).

Appendix 2.2

Reserves which were not under reserve management in 1994 and which were converted from a non-favourable breeding habitat, such as ex-mineral extraction sites or highly intensive arable land, into wet grassland are most likely underestimated. Population counts prior to reserve creation (1994 – the first year that the site is

surveyed as a reserve) have been estimated and will therefore contribute to the 1994 reference count. However, as these sites were unfavourable prior to reserve management it is likely that these reserves did not support any breeding pairs in the reference year. This effectively means that the relative increase these sites contribute to the overall trend is less than if they had contributed with a zero count in the reference year. Furthermore, an unknown proportion of the data contributing to the counterfactual trends may have been from nature reserves managed by other organisations and individuals and/or from land subject to other conservation interventions through, for example, agri-environment schemes. In addition, our target species breed in a variety of habitats in the UK and might be subject to different pressures in these different habitats. Without more closely matching the counterfactual data to the reserve data, we would therefore be comparing, for example in the case of Redshank, their trend on lowland wet grassland reserves with their trend in a combination of saltmarsh, upland habitats and lowland freshwater wetland habitats. We addressed this uncertainty by examining whether results were consistent across different counterfactuals. While this provides a degree of sensitivity testing, to which our results were largely robust, the correct counterfactual is still unknown.

Appendix 2.3

Producing reserve trends:

Step 1 – Creating the reserve data: Select observations that match conditions. For example, to create the Lowland wet grassland reserve indices for Lapwing, we select all Lapwing counts from Lowland wet grassland reserve sites from 1994 - 2018. The RSPB data is already summarized to one observation per species x year x site.

Step 2: Impute missing counts using equation 1 (model 3 in the Rtrim framework), summarize the total imputed count for each year and create indices and SEs from the time totals (using observed counts where available and estimated where counts are missing).

Producing counterfactual trends:

Step 1 – Creating the counterfactual data: Select observations that match conditions (table 1). For example, to create the stringent counterfactual indices for Lapwing, we first select all 1x1 km grids that contain at least one transect of the selected set of semi-natural grassland habitat types (table 1), with a mean altitude equal to or below 250m. For each of the selected grids, we then summarize each transects breeding number (as breeding pairs are recorded in multiple distance bands within each transect) and exclude transect counts above 10. We then summarize transect counts into grid counts (an early and a late grid count for each year) and select the maximum annual grid count. We exclude grids without any positive Lapwing observations or only one positive observation. The result is one maximum annual count of breeding Lapwing per grid that has been surveyed.

Step 2: Impute missing counts using equation 1 (model 3 in the Rtrim framework), summarize the total imputed count for each year and create indices and SEs from the time totals (using observed counts where available and estimated where counts are missing).

Testing whether reserve indices are different from the benchmark counterfactuals:

Select a species' reserve and counterfactual indices. For example, if testing whether Lapwing reserve indices are different than the indices from the Lapwing benchmark

counterfactual, we select the indices for Lapwing (reserve and counterfactual indices consisting of 25 indices each which have been calculated from the species total, both starting in 1994 with a value of 1). We then test whether the 25 reserve indices are different than the 25 benchmark counterfactual indices using a Welch two-sample t test.

Testing whether a different counterfactual would have led to a different conclusion:

Select a species' reserve and alternative counterfactual indices. For example, if testing whether having used the stringent counterfactual as the preferred counterfactual, would have led us to another conclusion about the effectiveness of reserve conservation, we select the indices for Lapwing (reserve and stringent counterfactual indices consisting of 25 indices each which have been calculated from the species total, both starting in 1994 with a value of 1). We then test whether the indices are different using a Welch two-sample t-test. We compare the outcome of this test to the outcome of the benchmark test and evaluate whether using a different counterfactual would had led to a different conclusion about the effect of reserves on the target species, than when using the benchmark counterfactual.

3 | The effect of conservation interventions on the abundance of breeding waders within nature reserves in the United Kingdom

This work was conducted in collaboration with my supervisors and Malcolm Ausden from the Royal Society for the Protection of Birds and was submitted to *Ibis* for publication.

Abstract

Breeding populations of many wading birds have declined globally, primarily caused by habitat degradation and loss. In the United Kingdom, population declines have been particularly notable on lowland wet grasslands. In response, lowland wet grasslands have been restored and receive ongoing management to improve the breeding conditions of target species. Here, we assess the efficacy of management measures using a Bayesian framework and controlling for confounding factors. We focus on four wader species, Northern Lapwing (*Vanellus vanellus*), Eurasian Curlew (*Numenius arquata*), Common Snipe (*Gallinago gallinago*) and Common Redshank (*Tringa totanus*), that breed in numbers on wet grassland reserve sites in the UK. We collated annual site-specific climate, management information (e.g. the creation of wet features and predator control measures) and bird counts between 1994-2018. We found the effects of conservation actions varied between intervention types and species. For lapwing and redshank, excluding predators by predator-exclusion fencing, especially in combination with fox control, and improvements in water control structures and water surface features, were generally associated with higher breeding counts. For all study species, sites with longer histories of management were associated with higher breeding numbers, with the effect of site age being particularly notable for management on former arable land. Our findings support the effectiveness of targeted conservation actions to achieve high numbers of breeding waders on lowland wet grassland reserves and highlight the value of consistent and reliable monitoring data.

Introduction

Wetlands are among the world's most biodiverse ecosystems and play a key role in helping to mitigate climate change, providing essential ecosystem service benefits,

and contributing to people's livelihoods (Ramsar Convention on Wetlands 2018). However, due to intensification and expansion of agriculture, water extraction and drainage, these habitats and the species that depend on them are disappearing at alarming rates (WWF 2020b). In Europe, population declines have been particularly notable for breeding wading birds (Order Charadriiformes) on lowland wet grasslands (Franks et al. 2018; Hayhow et al. 2019).

Considerable resources have been invested into halting these declines. For example, in the United Kingdom (UK), between 1993 and 2018, the Royal Society for the Protection of Birds (RSPB) acquired and restored over 7000 ha of land at more than 80 lowland wet grassland nature reserve sites. Land at these sites has been converted from drained grassland, arable land, or former mineral extraction sites, into wet grasslands by raising water levels, excavating pools, scrapes and foot drains (shallow, water-filled ditches) and, in the case of ex-arable sites, establishing a grass sward. Subsequent ongoing management has involved manipulating water levels, grazing and mechanical mowing, and reducing the impacts of generalist predators on nesting birds using predator-exclusion fencing and lethal control (Ausden et al. 2019).

The positive effects that these conservation measures have on waders has been well documented (Ausden et al. 2001; Ausden & Hirons 2002; Smart & Coutts 2004; Wilson et al. 2004; Smart et al. 2006; Eglington et al. 2008; Malpas et al. 2013; Franks et al. 2018). For example, Malpas et al (2013) tested the effectiveness of predator-exclusion fencing on nest survival at 10 lowland wet grassland reserves from 2004 – 2011. They found that nest survival and overall productivity increased within predator-exclusion fenced areas. However, for a variety of reasons, including limited person power and resources, studies have often tested a single intervention type or category of protected area as a uniform treatment, therefore not accounting for potential differences between

management across study sites. Studies have furthermore often been limited to a few selected sites that have been monitored over a relatively short period. This increases the risk that observed effects are due to confounding factors. Considerable efforts have recently been made to document and summarise the effectiveness of conservation interventions in various contexts (Sutherland et al. 2019) but large-scale evaluations in general remain rare.

In this study, we collated and combined breeding counts of wading birds and conservation management information from nature reserves to assess how different site-based conservation interventions relate to the abundance of breeding waders on managed lowland wet grassland reserves across the UK. Specifically, we tested for the effects of interventions that target the manipulation of site hydrology and control of predators. We also tested whether the duration of site management (hereafter referred to as site age) and former land use affected breeding abundance when accounting for other site-specific conservation actions and climatic conditions. Unlike previous studies, this dataset allowed us to estimate the effect of multiple conservation actions while accounting for potential confounding effects from other ongoing management actions and abiotic factors.

Method

Datasets

We extracted data from RSPB reserve management reports (building on work done in chapter two), and online repositories of weather records, to create a national, spatially explicit dataset of annual numbers of breeding bird pairs, wetland conservation interventions and climatic conditions, at a spatial resolution of 1km² (Table S3.1).

Site selection and count data

We collated breeding bird counts for four wading bird species – Northern Lapwing (*Vanellus vanellus*), Eurasian Curlew (*Numenius arquata*), Common Snipe (*Gallinago gallinago*) and Common Redshank (*Tringa totanus*) – for the period 1994-2018, from RSPB-managed lowland wet grassland nature reserves in the UK. We chose these species because they are all of high conservation concern in the UK and RSPB reserve management actions are designed to benefit their populations (Ausden et al. 2019; Jellesmark et al. 2021). We defined sites as blocks of lowland wet grassland habitat acquired by the RSPB in the same year. A small number of sites where it was not always possible to differentiate between numbers of breeding waders on blocks of land acquired at different times were excluded from the analysis. These sites made up about 8% of the current total area of lowland wet grassland on RSPB reserves. Our analyses are thus based on annual breeding pair counts from 5781 ha of lowland wet grassland across 79 sites acquired between 1993 - 2018 (mean site area = 73.5 ha \pm 67.9 SD; Figure 3.1). The number of breeding pairs was estimated using standard lowland wader survey methods described in Gilbert et al (1998). In 2018, the study sites had on average been under reserve management for 16 years.

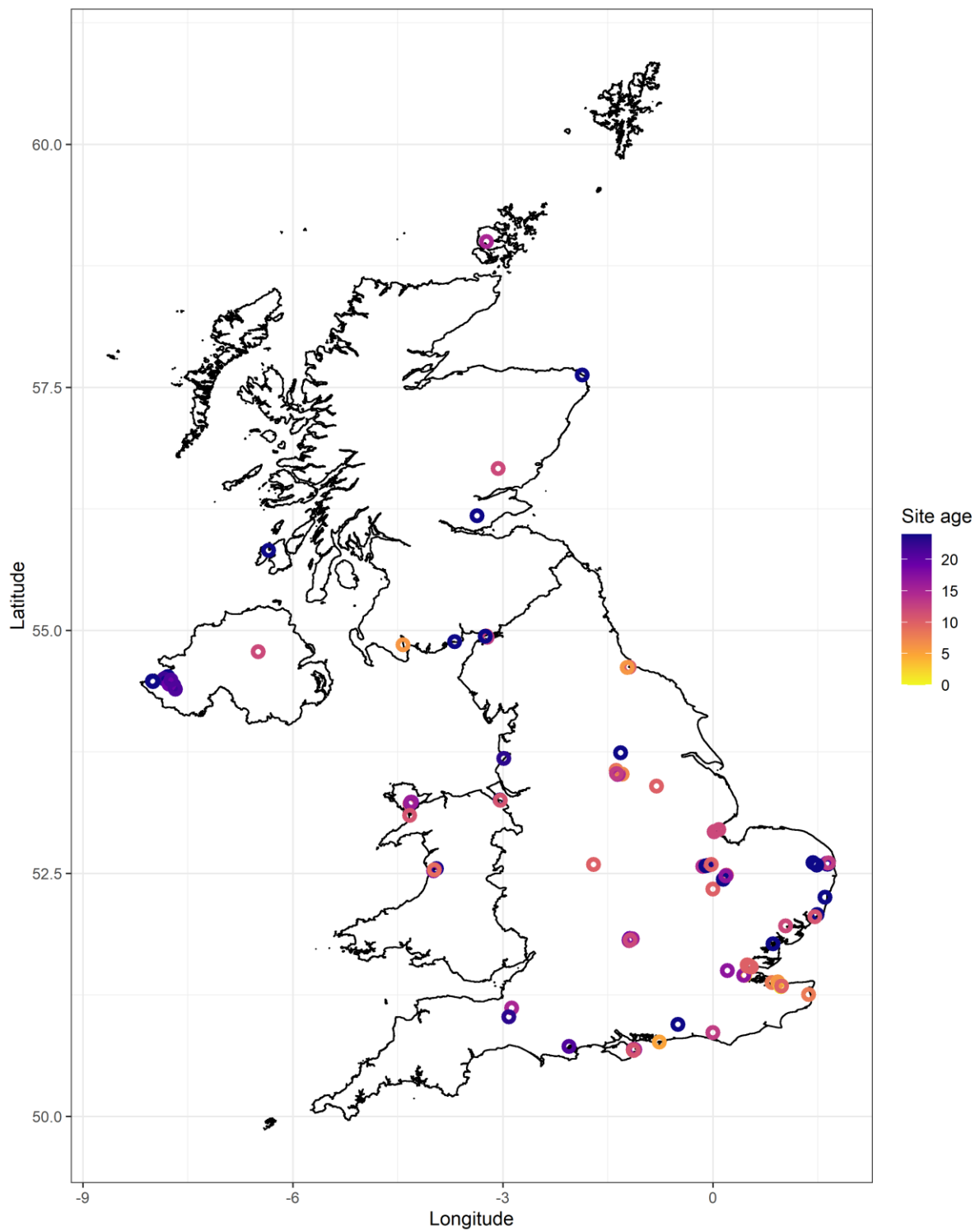


Figure 3.1 Site age for UK lowland wet grassland reserve sites acquired in 1993 or after ($n=79$) under RSPB management.

Conservation interventions

We gathered data on conservation interventions and site management activity from annual reserve reports for the period 1993-2018. Annual reports contain prescribed management information and are submitted annually to the RSPB headquarters. In a small number of cases, older annual reports were missing, and the relevant information was instead collated using information from other centralised RSPB sources in liaison with staff. The management and conservation information contained in reserve annual reports usually included: vegetation management (e.g., the type of livestock used for grazing, the period grazed and if mechanical vegetation removal was used); predator control (e.g., whether nests were protected from predation by Red Fox *Vulpes vulpes* and Badger *Meles meles* using predator-exclusion fencing (Malpas et al. 2013), whether Red Foxes or Carrion Crows *Corvus corone* were killed to protect wader eggs and chicks, whether predator-exclusion fencing was applied in combination with lethal control, or whether neither predator-exclusion fencing or lethal control were applied); and manipulation of site hydrology (e.g., installation of water control features, or excavation to create pools, scrapes or foot drains). Predator control was recorded as being active when at least one individual of Red Fox or Carrion Crow, or one Carrion Crow nest, was removed within the reporting period (hereafter referred to as foxes and crows) (Table S3.1).

Climatic data

We created seasonal climatic variables for each year using temperature and rainfall observations from the HadUK 1km grid monthly climatic data (Hollis et al. 2019). The autumn/winter season temperature variable was created as the mean monthly temperature between October and March. The autumn/winter season rainfall variable

was created by summing monthly precipitation from October until March. These variables provide a measure of, in particular, wetness during the winter before waders settle to breed from March onwards, as well as of the severity of the winter cold prior to the breeding season. The spring/summer variables were created similarly using monthly temperature and precipitation data from April, May and June and provide a measure of climatic conditions during the wader breeding season. We paired the seasonal climatic data to reserve sites by selecting the 1km climate grid that overlapped with each site's centre point (British National Grid projection).

Other covariates

We recorded the area of each site, the former habitat type and date of land acquisition (used to calculate site age) by the RSPB, all of which are held on a central database.

Pre - analysis

To avoid statistical problems in the primary analysis we explored the data for each species prior to specifying the explanatory models (Zuur et al. 2010). We checked for correlation between explanatory variables using Pearson's correlation values (excluding variables if Pearson correlation $r > 0.7$) and examined for collinearity between variables using Generalized Variance Inflation Factor values (excluding collinear variable when the Variance Inflation Factor exceeded 3) (Zuur & Ieno 2018).

Models

We used zero inflated negative binomial (ZINB) and Poisson models to explain the effect that site specific factors and management actions have on the number of breeding pairs on lowland wet grassland reserve sites. We used a combination of linear and nonlinear effects of the covariates and fitted all models in a Bayesian

framework with Integrated Nested Laplace Approximation (INLA) (Rue et al. 2009). Each model was initially specified using a zero inflated negative binomial distribution and the Watanabe Akaike information criterion (WAIC) value was obtained and used to compare with less complex models, such as Poisson and negative binomial. Based on the WAIC values, we selected the ZINB model for redshank, lapwing and snipe. The breeding pair counts for these three species contain a large proportion of zeros (Fig. 3.2; lapwing = 18.7%; redshank = 19.7%; snipe = 38.3%) and counts vary substantially between sites. Curlew was fitted using a Poisson distribution, as the more complex models did not improve the WAIC value. Curlew breeding pair counts contain a relatively low proportion of zeros (Figure 3.2; 13.8%) and vary less than the other three species.

The zero-inflation part of the models was fitted using an intercept only. For the count (Negative Binomial and Poisson) part of the models, site was used as a random intercept in all models. We furthermore considered random intercepts for reserve identity and climate district. Spatial dependency between proximate sites was included with the stochastic partial differential equation approach (SPDE) (Lindgren et al. 2011; Lindgren & Rue 2015). We used default priors for the fixed effects and non-informative priors for the random effects (Carroll et al. 2015). The priors for the second order random walk functions were penalized complexity priors with parameter values $U = 1$ and $a = 0.01$ (Simpson et al. 2017).

A stepwise model selection approach, based on WAIC values, was used to select the best fitting combination of random effects, and to determine whether to include the spatial term. We considered a model improved when the WAIC value decreased by at least 3 and selected the most parsimonious model. Sites without observations were excluded from our analysis. We specified a separate model for each species as we

expected different responses to conservation actions between the study species (Franks et al. 2018; Ausden et al. 2019; Jellesmark et al. 2021).

The breeding count for each species for each combination of site and year was modelled as a function of site size, fox and crow control, predator-exclusion fencing, precipitation and temperature in the winter and spring season, water control features and water surface features. Foot drains, excavation of ponds and scrapes and other fixed structures, such as earth bunds, were aggregated and included as a single water surface features variable, reflecting improvements in a site's ability to retain surface water during the wader breeding season. If improvements were completed within a site prior to the breeding season in any of the three water surface categories (fixed structures, foot drain or excavation of pools and scrapes), the variable increases by 1 for each improved category, otherwise the value of the previous year carries over. Water control structures were modelled similarly but were restricted to a maximum annual improvement of 1. Most predator control requires appropriate permission from the relevant statutory conservation agency. Crow control is performed immediately before and during the wader breeding period, and virtually all fox control during January to March. Lethal control of predators can increase the number of breeding pairs by increasing chick survival and thus the number of birds in the following year (Niemczynowicz et al. 2017; Laidlaw et al. 2020). We modelled this as a lagged relationship so that control efforts after the breeding season in year t_{-1} but prior and during the breeding season in year t_0 were assessed relative to the breeding counts in year t_1 . Predator-exclusion fencing was assumed to be associated with higher breeding numbers through reduced predation risk (Fontaine & Martin 2006) and therefore modelled relative to the current years' breeding numbers. We included interaction terms between fox control and predator-exclusion fencing, between crow

control and predator-exclusion fencing, and between adjustable water control structures and surface water features.

To examine how site age relates to the breeding numbers for the four target species, we fitted site age as a smoothed function using a second-order random walk process. The second-order random walk produces a smoothed term based on the second order differences that allows us to identify whether a pattern exists between site age and the breeding numbers while accounting for the other explanatory variables. If a pattern exists, and thus a changing effect of site age, the random walk trend diverges from a horizontal line of no change. To allow for a different temporal trend between sites of different former habitats (i.e., drained grassland, arable land former mineral extraction sites and mixed arable and grassland), site age was specified individually for each former habitat type using a dummy variable. Site age was included to capture the temporal effect of factors such as ceasing conventional agricultural practices, reseeded former arable land, introducing beneficial grazing regimes, and the overall gradual effect that continuous site management through time is expected to have on populations of breeding birds (Ausden & Bolton 2012).

We assessed two different mechanisms by which rainfall during spring/summer can affect the breeding numbers within a site. Large amounts of precipitation during the spring/summer in year t_0 can flood a site, causing pairs to breed in sub-optimal habitats outside reserves (Ratcliffe et al. 2005). Precipitation in year t_0 thereby affects the number of breeding birds in year t_0 . Additionally, flooded sites cause breeding pairs to nest on adjacent suboptimal land, leading to reduced breeding productivity and therefore a lower number of birds in year t_1 . Each of these potential mechanisms were tested in separate models.

Posterior parameter mean estimates and 95% credible Intervals above 0 on the log scale were considered to show a positive effect on breeding counts, while estimates lower than 0 were considered to show a negative effect. All continuous explanatory variables with linear effects were standardized to make regression coefficients comparable.

All analyses were done in R version 4.0.2. We used the tidyverse packages for data manipulation (Wickham et al 2019) and the ncd4 package (Pierce 2019) for climatic data in nc format. All models were fitted in INLA (Rue et al. 2009). All code used is available at <https://github.com/seanjellesmark>.

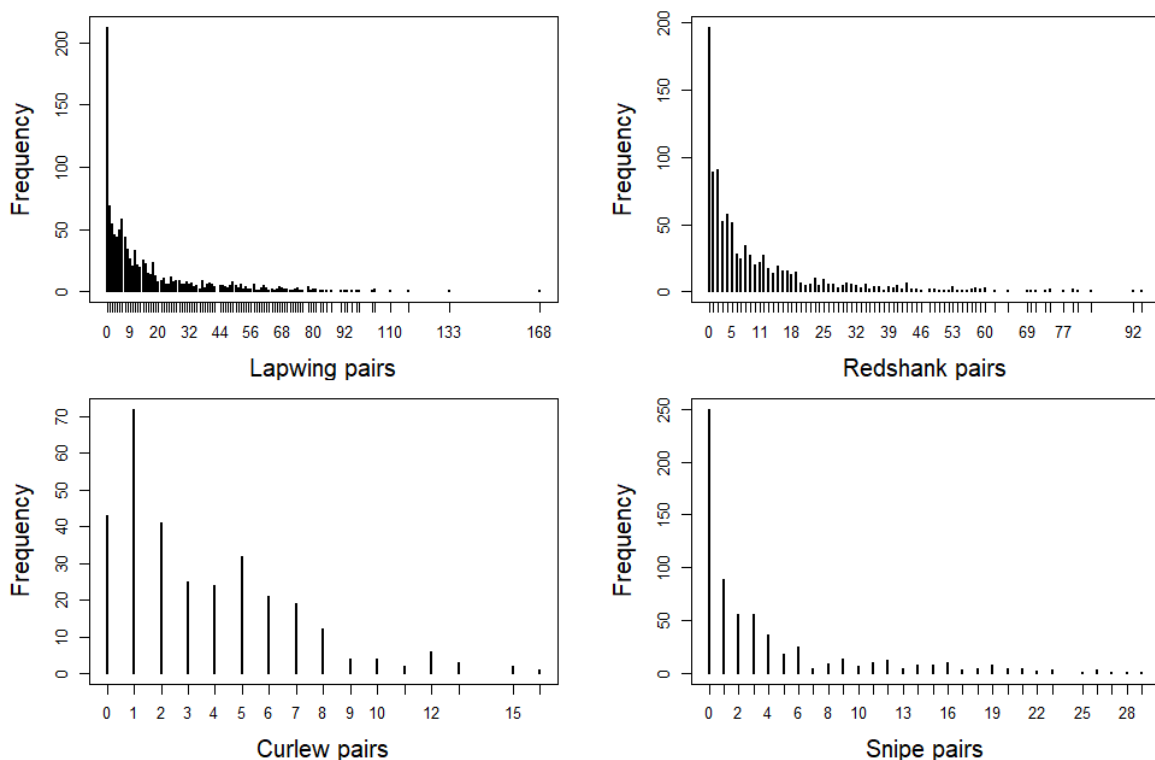


Figure 3.2 The frequency distribution of breeding pair counts between 1994 - 2018 for Lapwing ($N = 1132$), Redshank ($N = 995$), Curlew ($N = 311$) and Snipe ($N = 625$) across 79 study sites.

Results

Between 1994 – 2018, a sum total of 31,665 pairs of the four study species were counted breeding on reserve sites acquired within the study period (17,456 lapwing pairs; 10,578 redshank pairs; 1,095 curlew pairs; 2,536 snipe pairs). Lapwings were found on 75 sites, redshank on 65 sites, snipe on 39 sites and curlew on 17 sites. Lapwing and redshank were widely distributed across all the lowland wet grassland reserve sites, and on average in higher breeding numbers than snipe and curlew (Figure 3.2). Curlew occurred primarily in Northern Ireland and Scotland, while snipe were almost entirely absent from reserves located on the southern and south-eastern coast of England (Figure S3.1:3.S4).

The models that included site as the only random effect produced the lowest WAIC values for lapwing, snipe and curlew. Reserve identity was included as an additional random effect for redshank as it reduced the WAIC value beyond the predefined threshold value. None of the models were improved by including a spatial term. Model validation indicated that the models for lapwing, redshank and curlew were slightly under-dispersed with a small number of outliers (Figure S3.5:S3.8).

Larger sites were associated with higher breeding counts for all wader species except curlew, although the 95% credible interval for snipe overlapped zero (Figure 3.3). Sites with predator-exclusion fencing were associated with higher breeding counts of lapwing, redshank and curlew, although the 95% credible intervals of this effect overlapped zero for all species. Snipe were negatively associated with predator-exclusion fencing. Higher breeding counts for lapwing and redshank were generally associated with a combination of fox control and predator-exclusion fencing. The breeding abundance of redshank and lapwing was higher on sites with more

adjustable water control structures and surface water improvements, but the opposite was true for curlew and snipe. However, the credible intervals overlapped zero for all species except curlew. Neither fox nor crow control on their own had a clear effect on the breeding abundance of the target species, other than that numbers of breeding snipe were negatively associated with crow control (Figure 3.3). High amounts of precipitation in the spring and early summer were negatively associated with breeding numbers for redshank and lapwing, whereas warmer winter temperatures were positively associated with higher breeding abundance for these two species the following spring, with the credible intervals overlapping zero. Similar positive associations were estimated between winter rainfall and breeding numbers for redshank, lapwing and curlew but with zero within the credible intervals. There was no clear association between the amount of rainfall in year t_0 and the breeding number in year t_1 (mean \pm 95% credible interval; lapwing -0.003 ± 0.06 ; redshank -0.003 ± 0.06 ; snipe 0.06 ± 0.08 ; curlew -0.007 ± 0.08)

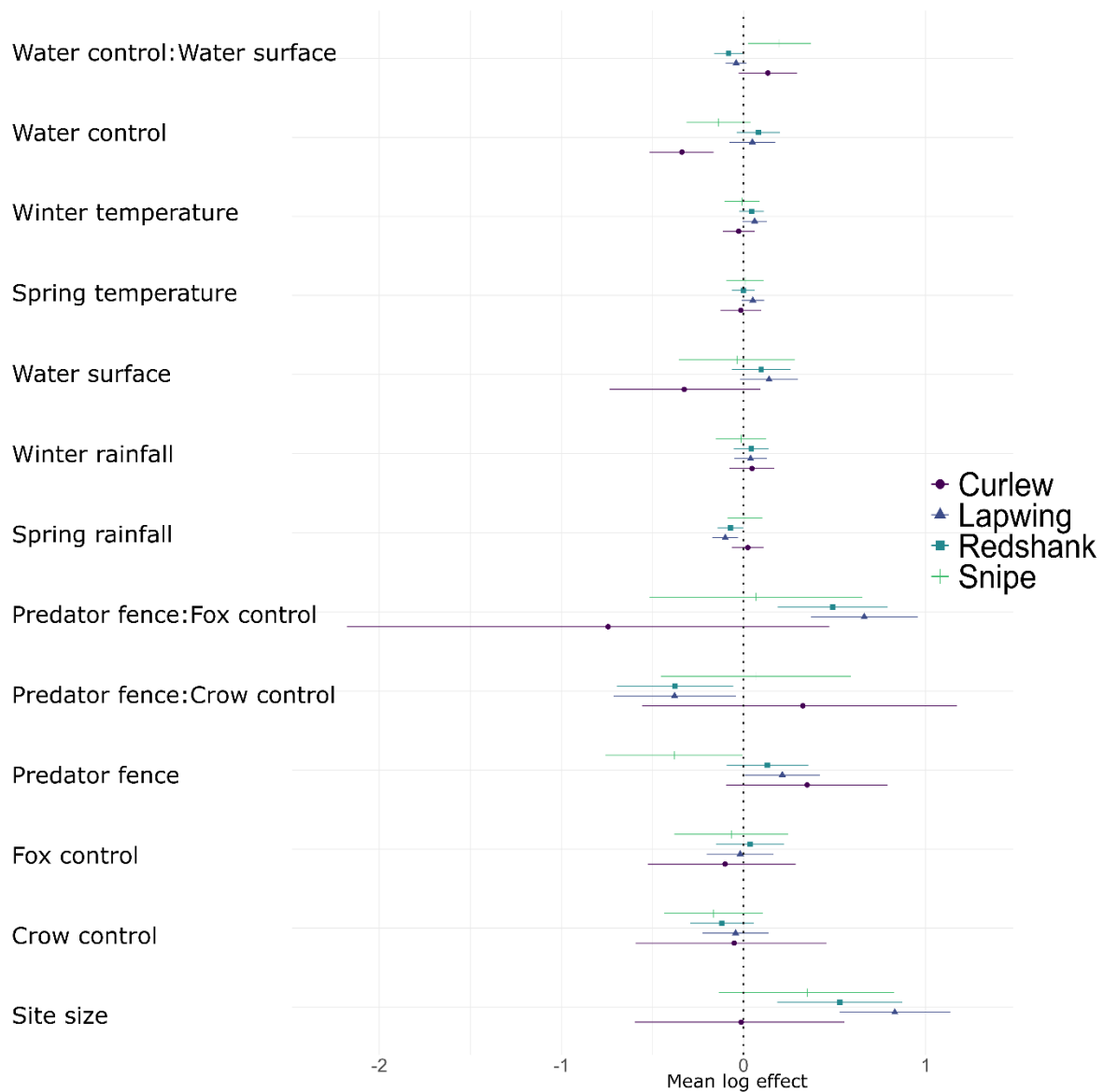


Figure 3.3 Posterior parameter estimates and 95% credible intervals for log mean effects of conservation interventions, temperature, rainfall and site size.

The relationships between site age and breeding pairs were largely similar for all species within each former land type (Figure 3.4). The number of breeding pairs on former drained grasslands increased initially for all the study species but decreased slightly thereafter for lapwing and curlew (although these subsequent declines are based on a very small number of sites), whereas both redshank and snipe displayed

a second, slight increase after around 12 years. For lapwing, redshank and snipe, the effect of site age was more prominent on former arable sites.

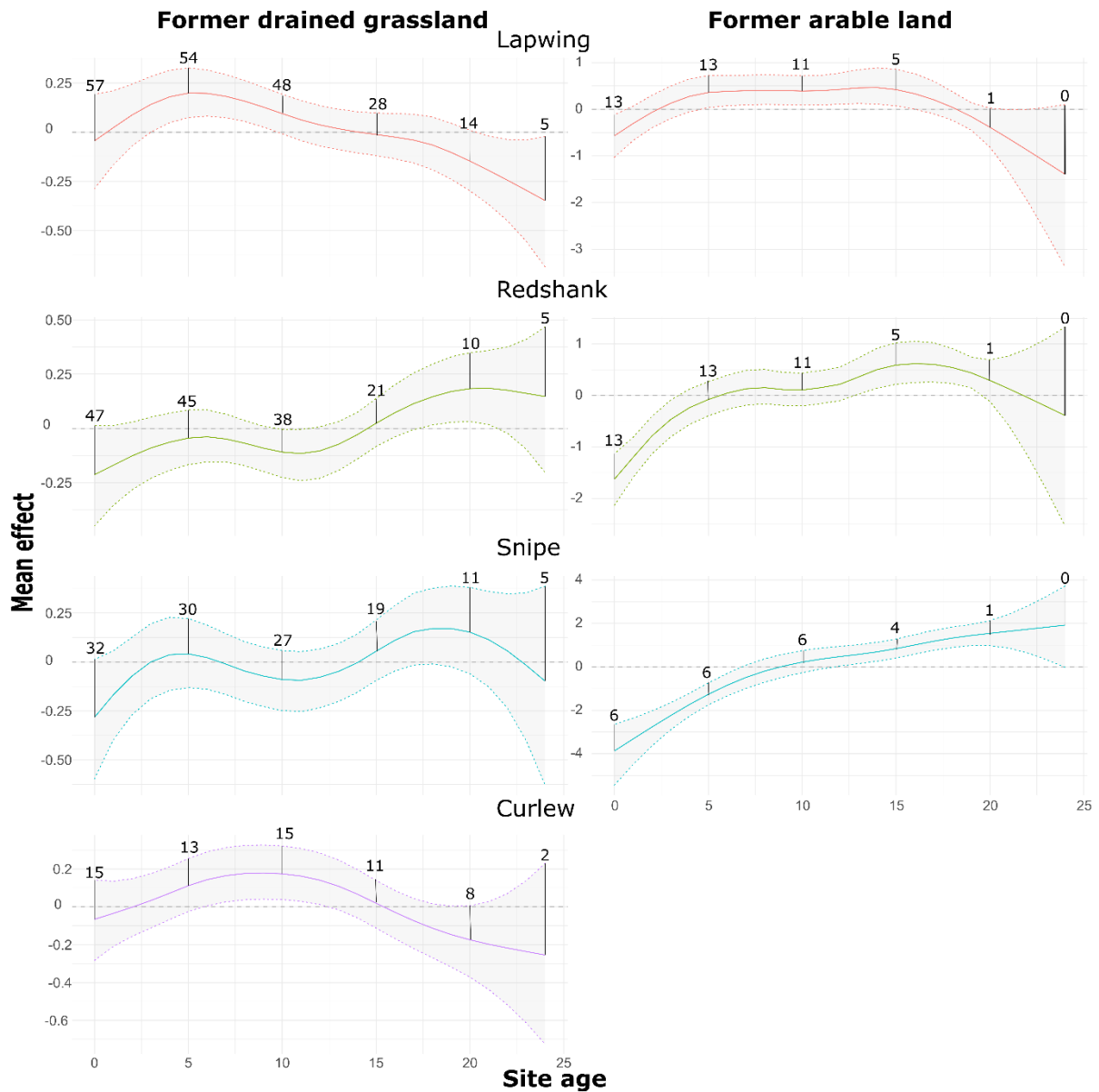


Figure 3.4 Estimated effect of site age on breeding pairs. The x axis shows years since acquisition on former grassland (left) and former arable (right) sites. The number of sites for each species x former land type are shown in five-year intervals. The estimated relationships are presented as smoothed functions with 95% credible intervals. Curlew on former arable land is not included as the data only contains a

single site. Former mineral extraction and mixed arable and grassland sites are not included as they were too scarce to produce reliable estimates.

Discussion

In this study, we carried out a detailed assessment of the impact of different site-based conservation actions on the number of breeding waders on lowland wet grassland nature reserve sites in the UK. Combining information from annual management reports with climatic data and site counts allowed us to evaluate multiple site-based interventions and the effect of site age on a national scale while controlling for potential confounding variables. We found that excluding foxes and badgers using predator fencing is an effective measure associated with high breeding abundance of lapwing and redshank. The largest relative increases in breeding numbers were observed in the years immediately after site acquisition. The breeding populations generally increased more on former arable land than on former drained grasslands, which is unsurprising as arable land undergoes larger habitat changes than drained grassland when transformed into wet grassland.

Our analyses concur with previous studies (Smith et al. 2011; Malpas et al. 2013) that predator fencing is effective and strongly associated with higher numbers of breeding waders, especially for lapwing and redshank. However, fencing a site for predator exclusion is costly and resource intensive and requires ongoing maintenance. For organisations with a limited budget, such as the RSPB, this means that fences are most often installed on sites that already support high numbers of target species or have the potential to do so. The negative association between predator fencing and breeding populations of snipe probably reflects the fact that most breeding snipe on RSPB wet grassland reserves occur on a small number of sites where it has not been

practical to install predator-exclusion fencing or on an island where there are no foxes or badgers, and not that snipe prefer to breed outside fenced areas. Similarly, curlew only breed at a small number of sites where predator-exclusion fencing has been installed. Hence, we would not necessarily expect a positive relationship between the abundance of breeding curlew or snipe pairs and predator-exclusion fencing.

We did not find a positive association between lethal control of foxes or crows and breeding abundance in the following year. These results are in accordance with Bolton et al (2007), who found no overall effect of predator control on lapwing population trends. Similarly, Franks et al (2018) found no increase in the likelihood of conservation success for populations of curlew and lapwing targeted by predator control. However, our analysis shows that predator fencing leads to higher numbers of redshank and lapwing, and that this effect is further increased by additional lethal control of foxes i.e. carrying out fox control on its own did not appear to increase wader productivity to a high enough level to allow their population to increase. A possible explanation for this is that individual animals removed by lethal control locally may be replaced rapidly from a regional pool, such as removal of territorial crows leading to an influx of other non-breeding crows (Smart et al. 2006; Bolton et al. 2007; Eglington et al. 2008; Fletcher et al. 2010). The initial predator density on sites and on surrounding land therefore influences local effectiveness of predator control. Furthermore, the method and intensity of lethal control also influences its effectiveness. For example, Fletcher et al (2010) found a positive effect of intensive predator control on breeding numbers and breeding success for lapwing and curlew and Bolton et al (2007) found that fox control increased nest survival for lapwing on sites with high initial fox densities. We did not account for initial predator densities,

predator densities on land surrounding sites or intensity of predator control, which could have led to different estimates of predator control effectiveness.

Sites with more surface water features were positively associated with breeding abundance for lapwing and redshank, similar to findings in other breeding wader studies in lowland habitats (Smart et al. 2006; Eglinton et al. 2008). However, the opposite was true for curlew and snipe, and none of the associations were strong, with the credible intervals overlapping zero. The opportunistic data collection process potentially explains why water control structures, water surface features and local predator control did not show a clear positive effect on breeding abundance. Data were collected from annual management reports. These reports are written by site managers to describe and monitor local management progress and effectiveness, but not necessarily using similar terminology or level of detail. We accounted for this lack of detail in our data by creating variables of a more general nature. For example, predator control was modelled as a binary variable, either active or inactive relative to the breeding number in a given year, and the water control variable was essentially modelled as a counter that reflects the number of years a site had improvements made. A more accurate representation of the variables, such as the proportion of shallow water area relative to site size, or other more fine-grained measures of habitat improvements, may have demonstrated the efficacy of the conservation actions. However, for such assessments to be feasible on a national scale, as in this study, requires consistent, standardized, and fine grain monitoring which for many conservation NGOs (working on limited budgets) would divert limited resources away from delivering conservation on the ground.

We found evidence that management duration (herein our site age variable) is an important factor affecting the number of breeding pairs on reserve sites, and that the

effect is different between former land types and species. As a site is acquired and restoration is initiated, breeding populations increase until a local carrying capacity is reached. It should be noted that the decrease we show here after year 20 for lapwing and redshank is driven by a limited number of observations, which is reflected in the larger credible intervals and limited number of sites (Figure 3.4). Nonetheless, the results support other findings showing that the number of years since land has been transformed into a reserve, and management initiated, are important factors to consider when designing studies and testing site-based conservation measures (Ausden et al. 2019). Considering the impact of interventions through time is important not only for conservation practitioners aiming to understand how to maximise the impact of new conservation measures and land acquisition, but also for scientists evaluating other conservation actions on sites that are being managed to protect species. Likewise, our results show the importance of considering the former management history of reserve sites. If the aim is strictly to increase the breeding abundance of wetland bird species on a national scale, converting arable land may be a better option than converting drained grasslands, as the latter habitat type already serves as breeding habitat for some waders, albeit being suboptimal for most. Other conservation priorities such as preserving populations of invertebrates or plants on existing areas of wet grassland can justify acquisition and restoration of grasslands over arable land. The costs of those two options differ considerably and for conservation organisations that operate under a constrained budget, the cost-benefits of each intervention and potential land acquisition and subsequent management would need to be evaluated.

Site age was the only variable modelled as a non-linear effect. We took this approach because we did not expect a linear effect to adequately capture how suitable habitat

develops at a site upon acquisition and subsequent restoration. Rather, we expected an initial strong effect of time, as the effects of ceasing former management manifest themselves, followed by a smaller increase between later years as populations reach carrying capacity. Our results support these assumptions.

Our results confirm the importance of installing and maintaining wet features, excluding foxes and badgers using predator-exclusion fencing while ideally also controlling foxes outside these fences for breeding populations of lapwing and redshank. Despite reserve management for snipe and curlew being successful (Jellesmark et al. 2021), the effects of specific interventions appear more subtle, potentially influenced by pressures on wintering grounds (Cook et al. 2021).

References

Ausden M, Bolton M. 2012. Breeding waders on wet grassland: factors influencing habitat suitability. Pages 278–306 in R. J. Fuller, editor. *Birds and Habitat*. Cambridge University Press, Cambridge. Available from https://www.cambridge.org/core/product/identifier/CBO9781139021654A019/type/book_part.

Ausden M, Hirons G, White G. 2019. Wetland restoration by the RSPB – what has it achieved for birds? *British Birds* 112:315–336.

Ausden M, Hirons GJM. 2002. Grassland nature reserves for breeding wading birds in England and the implications for the ESA agri-environment scheme. *Biological Conservation*.

Ausden M, Sutherland WJ, James R. 2001. The effects of flooding lowland wet grassland on soil macroinvertebrate prey of breeding wading birds. *Journal of Applied Ecology* 38:320–338.

Bolton M, Tyler G, Smith K, Bamford R. 2007. The impact of predator control on lapwing *Vanellus vanellus* breeding success on wet grassland nature reserves. *Journal of Applied Ecology* 44:534–544.

Carroll R, Lawson AB, Faes C, Kirby RS, Aregay M, Watjou K. 2015. Comparing INLA and OpenBUGS for hierarchical Poisson modeling in disease mapping. *Spatial and Spatio-temporal Epidemiology* 14–15:45–54. Available from <https://linkinghub.elsevier.com/retrieve/pii/S1877584515000350>.

Cook ASCP, Burton NHK, Dodd SG, Foster S, Pell RJ, Ward RM, Wright LJ, Robinson RA. 2021. Temperature and density influence survival in a rapidly declining migratory shorebird. *Biological Conservation* 260:109198. Elsevier Ltd. Available from <https://doi.org/10.1016/j.biocon.2021.109198>.

Eglington SM, Gill JA, Bolton M, Smart MA, Sutherland WJ, Watkinson AR. 2008. Restoration of wet features for breeding waders on lowland grassland. *Journal of Applied Ecology* 45:305–314.

Fletcher K, Aebischer NJ, Baines D, Foster R, Hoodless AN. 2010. Changes in breeding success and abundance of ground-nesting moorland birds in relation to the experimental deployment of legal predator control. *Journal of Applied Ecology* 47:263–272.

Fontaine JJ, Martin TE. 2006. Habitat Selection Responses of Parents to Offspring Predation Risk: An Experimental Test. *The American Naturalist* 168:811–818. Available from <http://www.journals.uchicago.edu/doi/10.1086/508297>.

Franks SE, Roodbergen M, Teunissen W, Carrington Cotton A, Pearce-Higgins JW. 2018. Evaluating the effectiveness of conservation measures for European grassland-breeding waders. *Ecology and Evolution* 8:10555–10568.

Hayhow D et al. 2019. *State of Nature 2019*.

Hollis D, McCarthy M, Kendon M, Legg T, Simpson I. 2019. HadUK-Grid—A new UK dataset of gridded climate observations. *Geoscience Data Journal* 6:151–159.

Jellesmark S, Ausden M, Blackburn TM, Gregory RD, Hoffmann M, Massimino D, McRae L, Visconti P. 2021. A counterfactual approach to measure the impact of wet grassland conservation on UK breeding bird populations. *Conservation Biology*:cobi.13692. Available from <https://onlinelibrary.wiley.com/doi/10.1111/cobi.13692>.

Laidlaw RA et al. 2020. Predator management for breeding waders: a review of current evidence and priority knowledge gaps. *Wader Study* 128:1–12.

Lindgren F, Rue H. 2015. Bayesian spatial modelling with R-INLA. *Journal of Statistical Software* 63:1–25.

Lindgren F, Rue H, Lindström J. 2011. An explicit link between gaussian fields and gaussian markov random fields: The stochastic partial differential equation approach. *Journal of the Royal Statistical Society. Series B: Statistical Methodology* 73:423–498.

Malpas LR, Kennerley RJ, Hirons GJM, Sheldon RD, Ausden M, Gilbert JC, Smart J. 2013. The use of predator-exclusion fencing as a management tool improves the breeding success of waders on lowland wet grassland. *Journal for Nature Conservation* 21:37–47. Elsevier GmbH. Available from <http://dx.doi.org/10.1016/j.jnc.2012.09.002>.

Niemczynowicz A, Świętochowski P, Brzeziński M, Zalewski A. 2017. Non-native predator control increases the nesting success of birds: American mink preying on wader nests. *Biological Conservation* 212:86–95.

Ramsar Convention on Wetlands. 2018. Ramsar Convention on Wetlands. Available from <https://www.global-wetland-outlook.ramsar.org/outlook>.

Ratcliffe N, Schmitt S, Whiffin M. 2005. Sink or swim? Viability of a black-tailed godwit population in relation to flooding. *Journal of Applied Ecology* 42:834–843.

Rue H, Martino S, Chopin N. 2009. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 71:319–392. Available from <http://doi.wiley.com/10.1111/j.1467-9868.2008.00700.x>.

Simpson D, Rue H, Riebler A, Martins TG, Sørbye SH. 2017. Penalising model component complexity: A principled, practical approach to constructing priors. *Statistical Science* 32:1–28.

Smart J, Gill JA, Sutherland WJ, Watkinson AR. 2006. Grassland-breeding waders: Identifying key habitat requirements for management. *Journal of Applied Ecology* 43:454–463.

Smart M, Coutts K. 2004. Footdrain management to enhance habitat for breeding waders on lowland wet grassland at Buckenham and Cantley Marshes , Mid-Yare RSPB Reserve ,. *Conservation Evidence* 1:16–19.

Sutherland WJ, Dicks L v, Ockendon N, Petrovan SO, Smith RK. 2019. What Works in Conservation 2019. Page (Sutherland WJ, Dicks L v., Ockendon N, Petrovan SO,

Smith RK, editors). Open Book Publishers. Available from <https://www.openbookpublishers.com/product/1031>.

Wilson AM, Ausden M, Milsom TP. 2004. Changes in breeding wader populations on lowland wet grasslands in England and Wales: Causes and potential solutions. *Ibis* 146:32–40.

WWF. 2020. Living Planet Report 2020. Bending the curve of biodiversity loss: a deep dive into freshwater. Gland, Switzerland.

Zuur AF, Ieno EN. 2018. GAM and Zero-Inflated Models. Highland Statistics Ltd, Newburgh.

Zuur AF, Ieno EN, Elphick CS. 2010. A protocol for data exploration to avoid common statistical problems. *Methods in Ecology and Evolution* 1:3–14.

Supplementary material

Table S3.1 Variables used in the analysis.

Category	Variable and mechanism affecting breeding wader numbers	Variable description	Source
Conservation	Water control features Mainly affects feeding conditions for chicks and adults	Discrete. Linear fixed effect. The sum of years with new water control improvements added on a given site. This variable increases by 1 if any number of new water control structures are added prior to the breeding season in a given year. If no new features are added, then last year's water control structure variable value is used. The features were: Water pumps,	Annual reports

		sluices and dams, tilting weirs and other similar features that allow water control within a site.	
Conservation	Water surface features Mainly affects feeding conditions for chicks and adults	Discrete. Linear fixed effect. A composite measure that reflects the sum of years with new water surface improvements added on a given site. This variable increases by 1 for any number of water surface improvements added prior to the breeding season in a given year within each of the 3 water surface categories. If no new features are added, then last year's water surface feature variable value is used. The 3 different water surface feature categories are 1; fixed water features (Bunds, changes in landscape elevation to retain site water, connecting a site to existing wet features using regular pipes, blocking drainage and similar actions), 2; foot drains and 3; excavation of ponds and scrapes.	Annual reports
Conservation	Lethal fox control (by shooting) Increases breeding success	Binary (active control/no control). Fixed effect. Control in previous year relative to current years breeding number	Annual reports
Conservation	Lethal crow control (mainly using Larsen traps) Increases breeding success	Binary (active control/no control). Fixed effect. Control in previous year relative to current years breeding number	Annual reports
Conservation	Predator-exclusion fencing Increases breeding success by	Binary (active control/no control). Fixed effect. Reflects if a site had predator exclusion	Annual reports

	excluding foxes and badgers	fencing prior to and during the current years breeding number	
Climate	Precipitation October - March Can affect breeding pairs both positively and negatively by flooding breeding areas or by meaning the grassland is not sufficiently wet during the breeding season	Continuous. Total precipitation amount (mm). Linear fixed effect. Sum of precipitation during October to March prior to the breeding season	HadUK 1km Grid (Hollis et al 2019)
Climate	Precipitation April - June Can affect breeding pairs both positively and negatively by flooding breeding areas or by meaning that the grassland is not sufficiently wet during the breeding season	Continuous. Total precipitation amount (mm). Linear fixed effect. Sum of precipitation from April - June	HadUK 1km Grid (Hollis et al 2019)
Climate	Temperature October - March Cold winters are assumed to reduce population sizes thus affecting breeding numbers	Continuous. Mean temp (°C). Linear fixed effect. Mean temperature from October to March prior to the breeding season	HadUK 1km Grid (Hollis et al 2019)
Climate	Temperature April - June Cold springs can reduce population size	Continuous. Mean temp (°C). Linear fixed effect. Mean temperature from April - June	HadUK 1km Grid (Hollis et al 2019)
Other	Area of lowland wet grassland habitat Larger areas can support higher breeding numbers	Continuous. Size (ha). Linear fixed effect. Area of site	Reserve dataset
Other	Habitat prior to acquisition Habitat upon acquisition is expected to affect the number of breeding pairs because some habitats are more suitable than others	Categorical (Grassland, Arable, Ex-mineral extraction site or mixed arable and grassland). Site habitat type before being managed as a reserve.	Reserve dataset
Other	Site age The age of a site is expected to affect breeding abundance by ceasing drainage thus	Discrete. Smoothed 2 nd order random walk function. Number of years since the RSPB	Reserve dataset

	increasing surface water area, ceasing former management and gradually converting the land from sub-optimal habitat such as arable land, into suitable breeding habitat.	started managing the site	

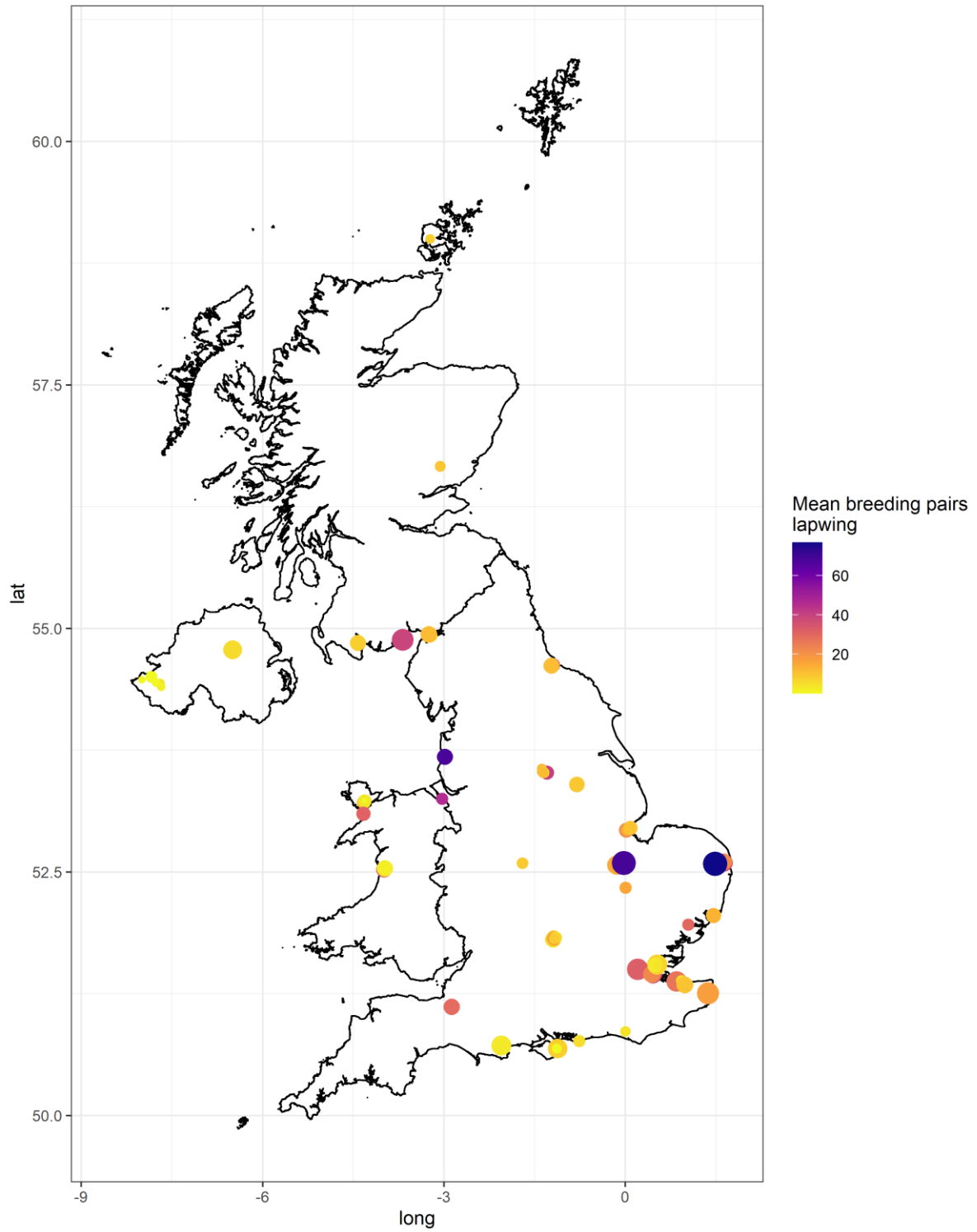


Figure S3.1 Mean annual number of breeding Lapwing pairs. The point size corresponds to site size.

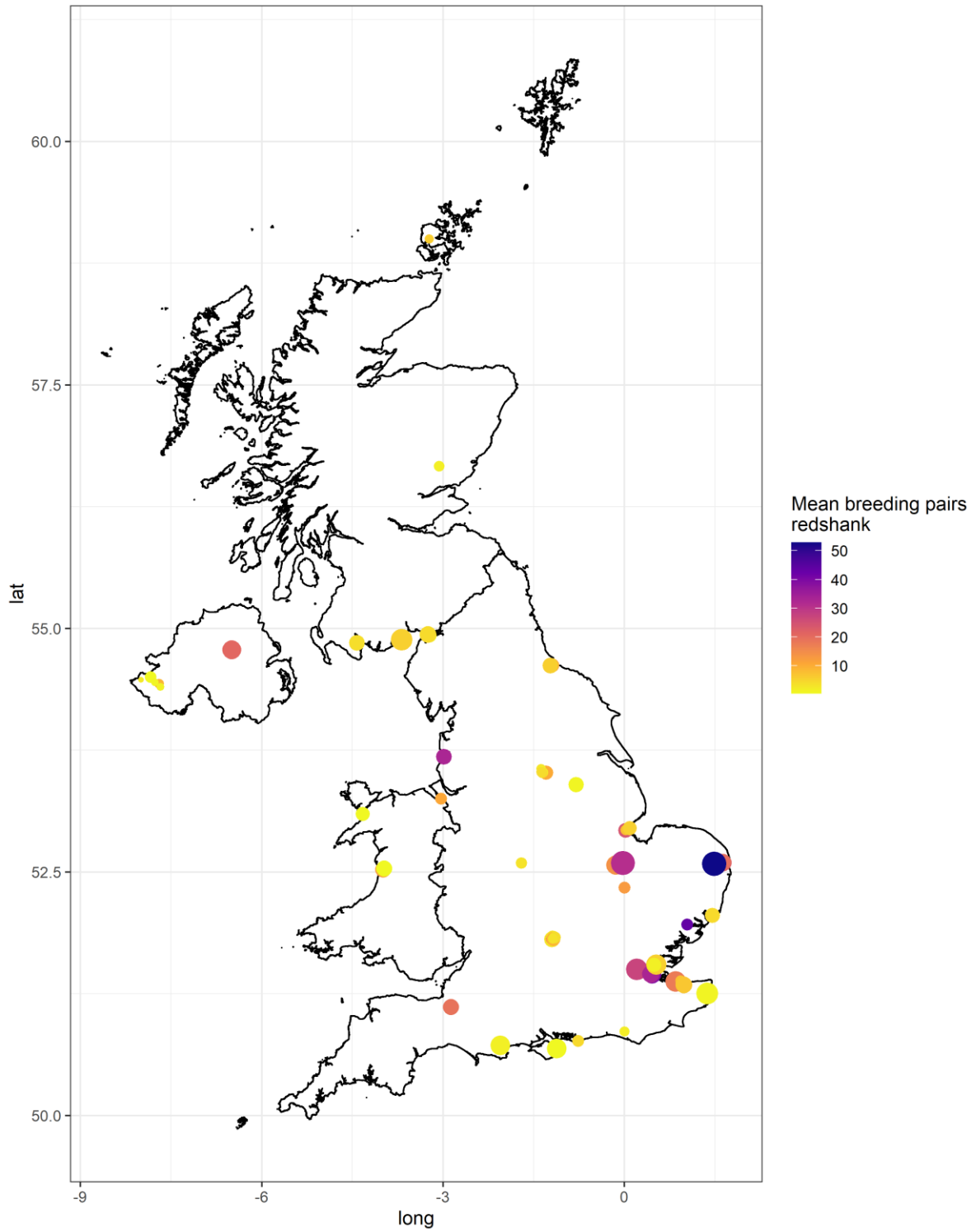


Figure S3.2 Mean annual number of breeding Redshank pairs. The point size relates to site size.

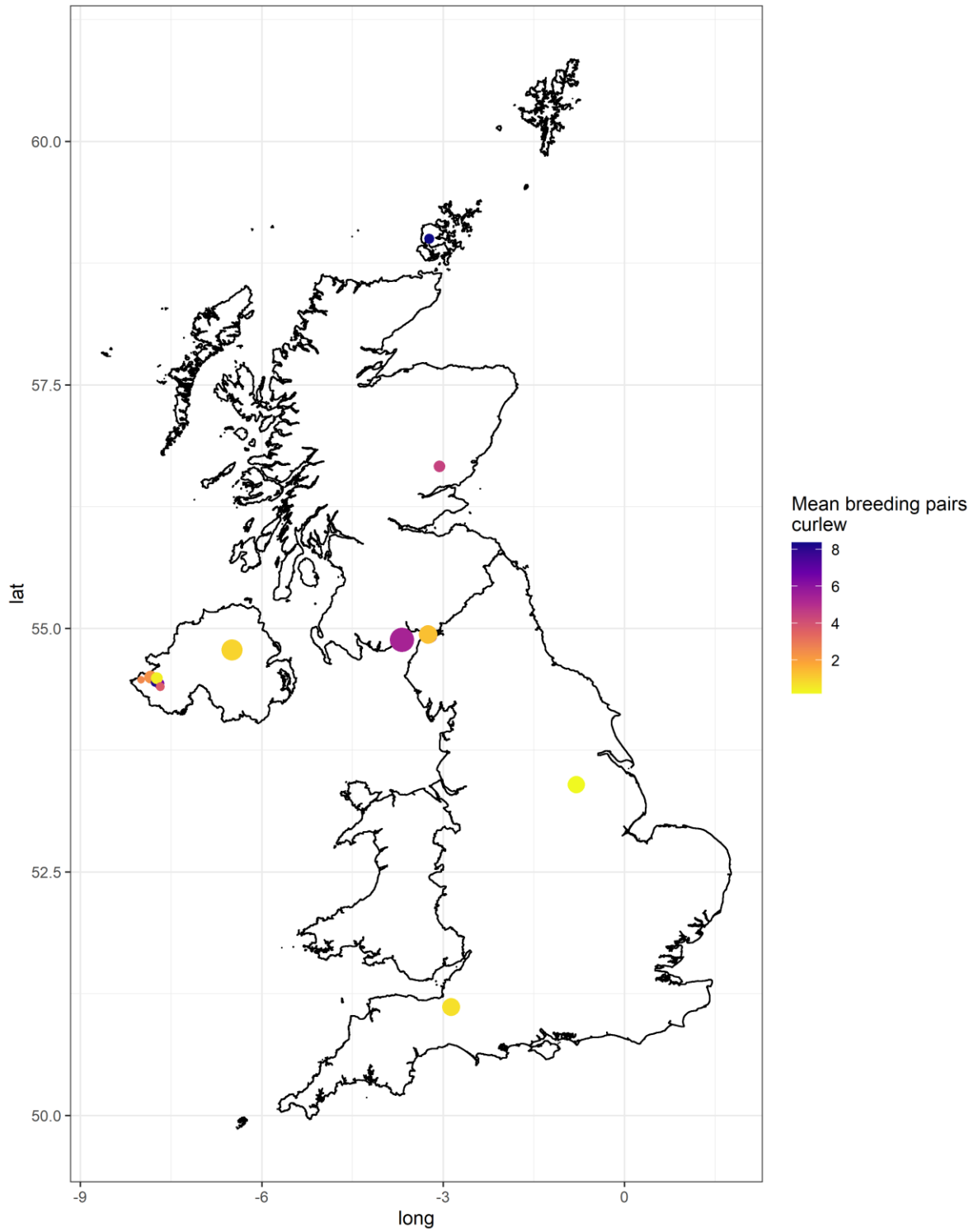


Figure S3.3 Mean annual number of breeding Curlew pairs. The point size relates to site size.

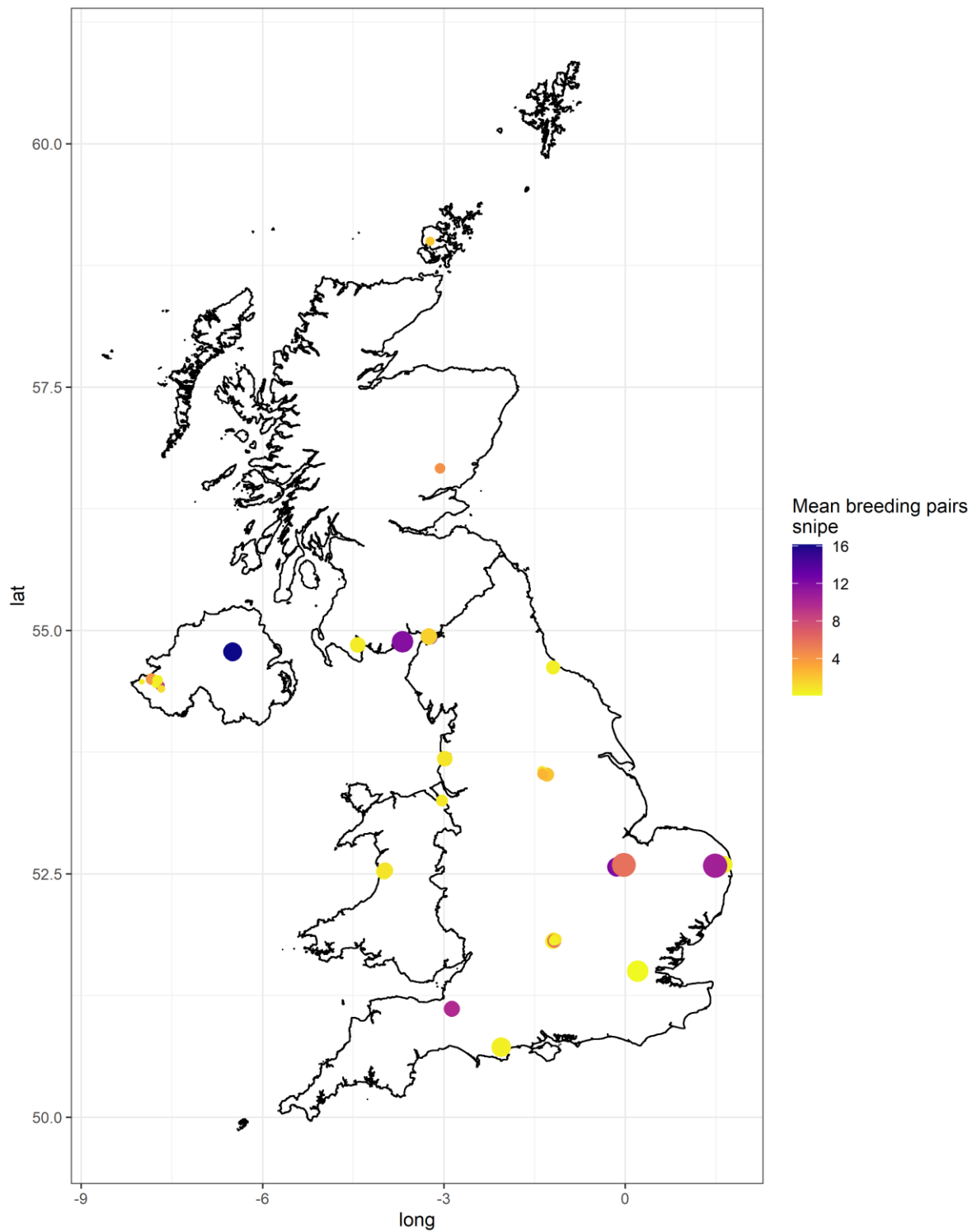


Figure S3.4 Mean annual number of breeding Snipe pairs. The point size relates to site size.

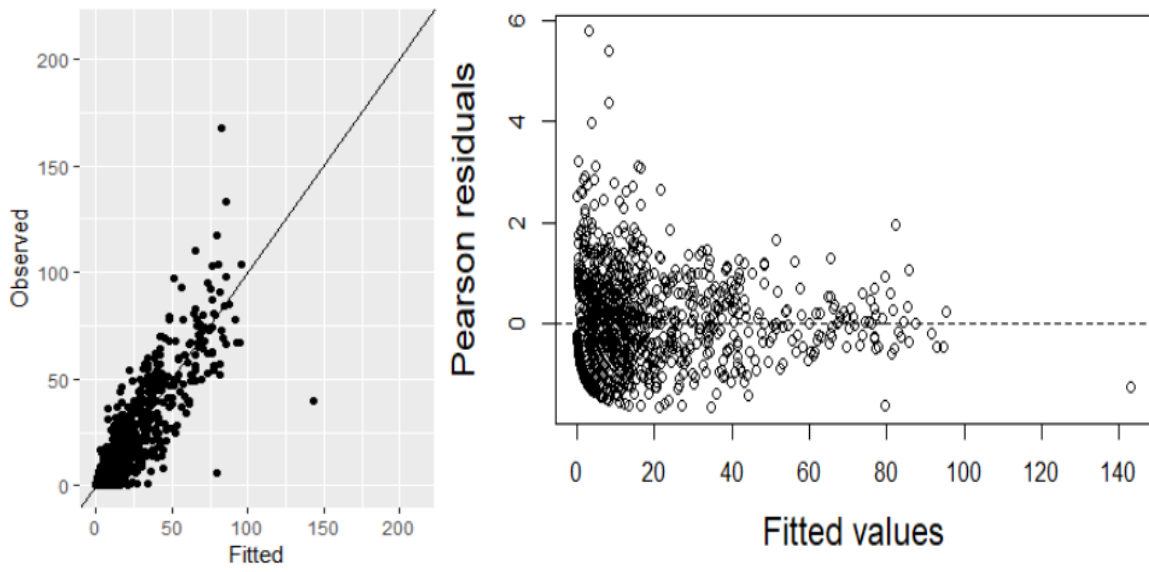


Figure S3.5 Lapwing validation. Observed vs fitted (left) and Pearson residuals vs fitted (right)

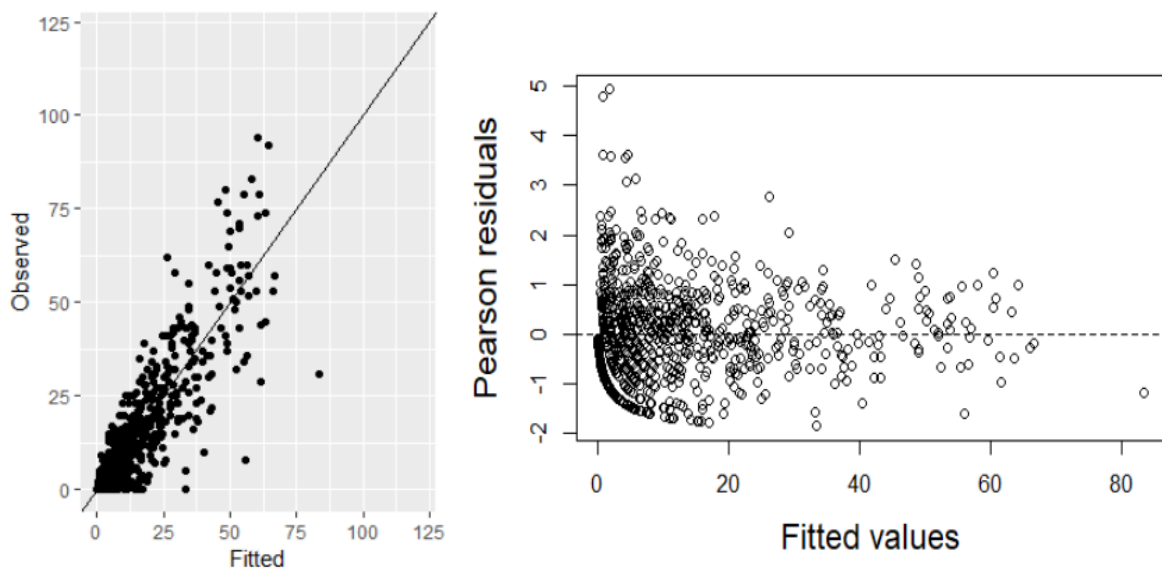


Figure S3.6 Redshank validation. Observed vs fitted (left) and Pearson residuals vs fitted (right)

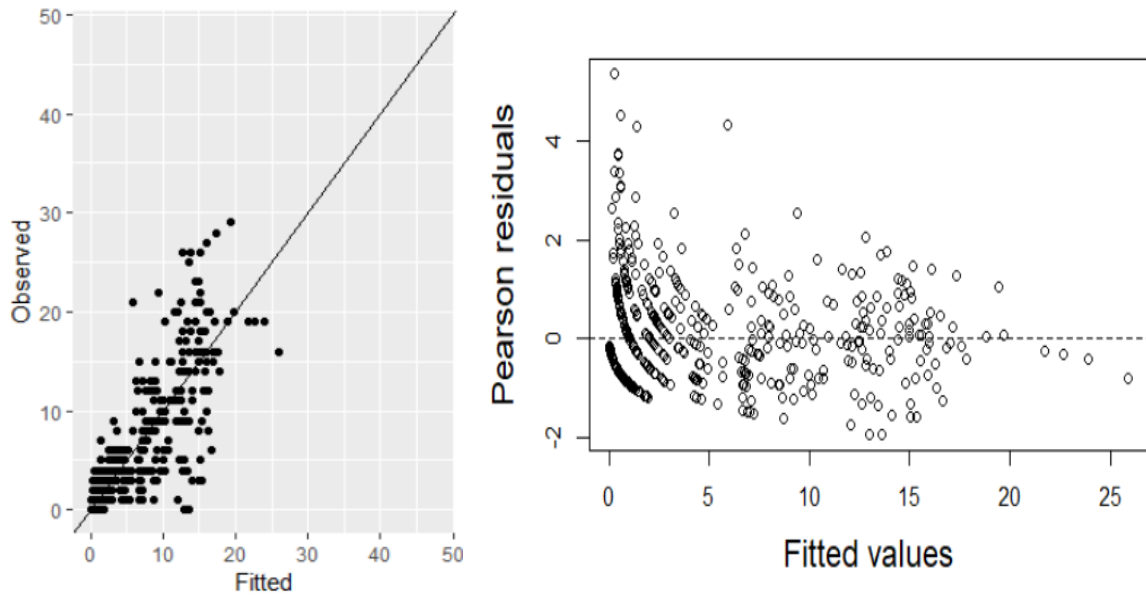


Figure S3.7 Snipe validation. Observed vs fitted (left) and Pearson residuals vs fitted (right)

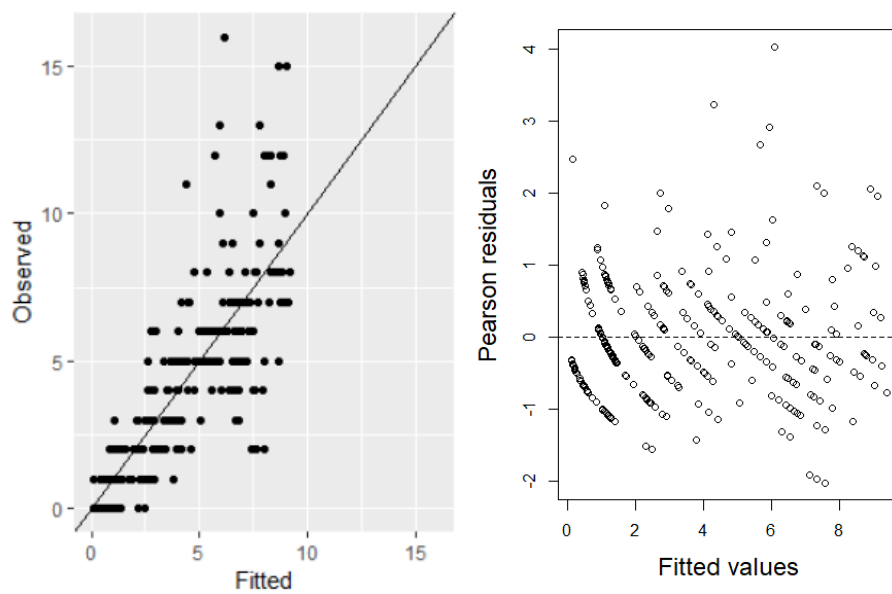


Figure S3.8 Curlew validation. Observed vs fitted (left) and standardized residuals vs fitted (right)

4 | Assessing the global impact of targeted conservation actions on species abundance

This work has been conducted in collaboration with my supervisors, my colleague Shawn Dove, and Jonas Geldmann from University of Cambridge and University of Copenhagen and is published as a preprint as:

Jellesmark S, Blackburn TM, Dove S, Geldmann J, Visconti P, Gregory RD, McRae L, Hoffmann M. 2022. Assessing the global impact of targeted conservation actions on species abundance. Preprint BioRxiv.

Summary

In recent years, vertebrate population abundance has declined at unprecedented rates (WWF 2020a). In response, targeted conservation measures – such as breeding programs or species-specific habitat management – have been applied to halt population declines, aid population recovery, and reduce and reverse the loss of biodiversity (Salafsky et al. 2008; Bolam et al. 2020). Until now, assessments of conservation actions have focused on the extent to which they reduce extinction risk, impact populations within protected areas, or increase the global area of land under protection (Hoffmann et al. 2010, 2015; Barnes et al. 2016; Maxwell et al. 2020; Bolam et al. 2020; Grace et al. 2021b). Here, we record and analyze conservation actions for 26,904 vertebrate populations from 4,629 species in the Living Planet Database, to estimate the association between targeted conservation and vertebrate abundance. Using a counterfactual approach to represent population trends in the absence of conservation, we demonstrate larger population increases in conservation targeted populations thus suggesting that targeted actions have delivered substantial positive effects on the abundance of recipient vertebrate populations worldwide. We show that, in the absence of conservation, a global indicator of vertebrate abundance could have declined even more. Positive population trends were associated with vertebrate populations subject to species or habitat management. Our results suggest that targeted conservation actions can help to reverse global biodiversity loss. However, substantial improvements in global biodiversity data are required to develop our knowledge of global conservation impact – an important step towards reversing biodiversity declines.

Introduction

Alterations to global ecosystems have caused widespread declines in biodiversity worldwide (Díaz et al. 2019; WWF 2020a), captured by global indicators of the state of biodiversity such as the Red List Index (Butchart et al. 2010), Living Planet Index (LPI) (WWF 2020a), and the Biodiversity Intactness Index (Biggs & Scholes 2005). Numerous conservation responses have been implemented to try to halt these declines, from local species-specific efforts, such as ex-situ breeding programs and conservation translocations, to more general large-scale measures, such as the designation of protected areas and international legislation aimed at protecting species (Salafsky et al. 2008; Maxwell et al. 2020; Bolam et al. 2020). Understanding the efficacy of conservation interventions (i.e., to what extent they have contributed to safeguarding biodiversity) is a prerequisite for effective decision-making in conservation and a key priority for researchers, policy and decision makers (Ferraro & Pattanayak 2006; Rose et al. 2018).

Ideally, the impacts of conservation interventions would be assessed using experimental designs that account for potential confounding effects through random allocation of treatment and control groups, such as randomized controlled trials (RCT). If control groups are carefully selected to mimic the treatments in all but the intervention being studied, RCTs offer an experimentally robust approach to estimate the impact of said treatment (Margoluis et al. 2009). However, while experimental designs are possible in certain conservation settings (Wiik et al. 2020), capacity, ethical considerations, and the spatial extent of actions, such as large protected areas, limit the ability to apply experimental evaluations (de Palma et al. 2018; Pynegar et al. 2019). When randomized experiments are not feasible, quasi-experimental designs,

based on statistical methods such as matching, can be used instead (Stuart 2010; Joppa & Pfaff 2011; Butsic et al. 2017; Geldmann et al. 2019; Schleicher et al. 2019). For example, annual population counts carried out within and outside protected areas can be matched on observable covariates, using the matched counts to determine how protection relates to population changes (Wauchope et al. 2019a, 2021; Jellesmark et al. 2021). Alternatively, inferential approaches that use logical arguments (Grace et al. 2021a) or expert knowledge and elicitation to inform counterfactual scenarios can be used to estimate conservation impact (Butchart et al. 2006; Hoffmann et al. 2010; Bolam et al. 2020). For example, Bolam et al (2020) used expert elicitation to estimate the impact of recent conservation actions in averting species extinction and found that, between 1993-2020, conservation may have prevented 21-32 bird and 7-16 mammal extinctions. These studies have advanced the field of counterfactual impact evaluation in conservation science but, until now, we have lacked assessments describing the global impact of conservation actions and responses on species abundance across taxonomic classes.

In this study, we explored the association between global species-targeted conservation actions and trends in abundance using population data from the Living Planet Database (LPD)(LPD 2020). Our study has four aims, namely to:

- 1) Describe targeted conservation actions for species populations in the LPD
- 2) Assess the impact of conservation actions through a counterfactual approach comparing how population indices differ between conservation targeted and non-targeted populations in the LPD
- 3) Measure the impact of conservation on a global population index given different counterfactual scenarios for conservation targeted populations, and

- 4) Test if specific conservation actions or responses are associated with species' population trends.

To achieve this, we (1) categorized conservation actions for each managed population in the LPD, (2) created four scenarios to compare trends from conservation targeted populations against matched counterfactuals, (3) created composite population indices in the absence of conservation impact for a subset of populations, and (4) estimated the impact of seven different types of conservation actions on species trends. This allowed us to present a global overview of targeted conservation actions, assess the impact of these actions by approximating how targeted populations were likely to have developed in the absence of conservation, measure the impact of these conservation actions on a global population index, and provide estimates of how each of these conservation actions affects population trends.

Methods

The Living Planet Database

The LPD is one of the largest global databases for population time series (WWF 2020). Since 1998, the LPD has provided the vertebrate population abundance data used to estimate the Living Planet Index (Collen et al. 2009; McRae et al. 2017; McRae et al. 2020), one of the key global indicators for biodiversity, and a measure adopted and used by the Convention of Biological Diversity to track progress towards halting the global decline of biodiversity (Butchart et al. 2010; Tittensor et al. 2014; WWF 2020a). Today, the LPD is managed by the Zoological Society of London in a collaborative partnership with the World Wildlife Fund, and is continually populated with primary data on vertebrate population abundance, that underpins research in global

biodiversity change and is used for indicators to inform both policy makers and the public. The database currently contains information on more than 27,000 populations from almost 5,000 species. These populations are distributed across 11 taxonomic classes (Actinopteri, Coelacanthi, Dipneusti, Elasmobranchii, Holocephali, Myxini, Petromyzonti, Amphibia, Aves, Mammalia, Reptilia). The majority of populations belong to Aves (birds, npops (number of populations) = 10,143), Actinopteri (ray-finned fish, npops = 9,571) and Mammalia (mammals, npops = 5,117) predominantly from North America (npops = 9,692), Europe (npops = 4,997), Latin America and the Caribbean (npops = 4,166) and Asia (npops = 3,835).

Population time-series data are added to the LPD from published or unpublished data if certain data standards are met. First, the data must be for a single species monitored at a defined location over time. Additionally, the species must be a vertebrate (mammal, bird, fish, reptile or amphibian). Several types of abundance data are accepted (Table 4.1). For example, full population counts are acceptable units of abundance whereas survival rates are not. A minimum of two years of abundance data is required. If a data source contains multiple annual measures, these are converted into a single annual figure using a mean, the peak count, or selecting the most consistently monitored season or month. Besides population data, the source must contain information on the geographic location of the population and the monitoring method. A variety of data sources are accepted given that these are referenced and traceable. This includes peer-reviewed articles from scientific journals, books, reports, online databases and grey literature.

Table 4.1 *Types of population abundance data that meets the data standards for tracking trends in the abundance of species populations (Accepted) and that does not meet the data standards (Not accepted)*

Accepted population abundance data	Not accepted population abundance data
Full population counts	Occupancy (unless it is used specifically as a proxy for abundance)
Estimates (e.g. population size estimated from measured parameters)	Data from experimental observations
Densities (including converted camera trap data)	Survival rates
Indices	Recruitment data e.g. number of eggs or young
Proxies (e.g. breeding pairs, nests, tracks)	Catch or hunting data with no measure of effort
Measures per unit effort (e.g. fish caught per net per hour)	Data where method has changed (unless corrected for)
Biomass (e.g. spawning stock biomass)	Opportunistic sighting data
Samples (e.g. where a proportion of the population is regularly monitored)	

Database structure

In the LPD, each population is stored with a unique ID and contains annual population data alongside additional information that covers eight broad categories relating to the species or the population. The first category is 'Base information', which contains information about the source, the year that the source was published or accessed, the reason for data collection, if the data overlaps with other populations, and the reference for the data. The second category is 'Taxonomy'. Here, taxonomic information is stored such as the common and Latin species name, Class, Order, Family and Genus. The taxonomic authorities used are:

- *Mammals* – Wilson, D. E. and Reeder, D. M. (2005) *Mammal species of the world: a taxonomic and geographic reference* (Third Edition). Johns Hopkins University Press, 2,142 pp.
- *Birds* – BirdLife International or IUCN Red List. This is largely consistent with the standard taxonomy for birds (Sibley, C.G. and Monroe, B.L. (1993) *Distribution and taxonomy of the birds of the world*. Yale University Press: New Haven, USA).
- *Amphibians* – Frost, D. R. (2005) *Amphibian Species of the World: an Online Reference* (Version 3.0). American Museum of Natural History: New York, USA.

- *Fishes* – Catalog of Fishes

researcharchive.calacademy.org/research/ichthyology/catalog/fishcatmain.as

[p](#)

- *Reptiles* – The Reptiles Database www.reptile-database.org

The third category, 'Geography', contains a brief description of a population's location, latitude, longitude, country and political region. The fourth, 'Ecology' category, contains information on the realm and biome in which a population occurs, the habitat type, whether the species is resident in the location, native or alien, and if it is invasive then what impact it has. The fifth category covers population data. Here, population units are recorded along with the sampling method, if the population data have been transformed, the proportion that the population represents of the global population, the annual population value, and if the population increased then additional information about the reason for population increase (reasons can be: Introduction, Recolonisation, Recruitment, Removal of threat, Rural to urban migration, Reintroduction, Range shift, Legal protection, Management, Other, Unknown).

The sixth category contains 'Protected Area' information that describes if the population is inside of a protected area, and the type of protected area if so. The seventh category covers 'Management' aspects, which indicates if the population is managed, the type of management (see Table 4.2 for management details), utilized status, CITES and CMS listing. The last category contains 'Threat' information, which

describes if the population is threatened, the types of threats (threats can be: Habitat Loss, Habitat degradation/Change, Invasive species/genes, Climate change, Pollution, Disease, Exploitation) and whether or not the population is exploited (exploitation can be: Caught and used, Pet trade, Sport hunting, Persecuted as pest, Indirect killing). To ensure consistency, trained personnel record the 'Management', 'Threat' and 'Reasons for increase' information from the original sources using a set of guidelines. This aims to reduce potential bias, but there is likely still to be some individual interpretation of the information in the data source.

Data selection, management data and coding

We extracted data for every population in the LPD (LPD 2020)– 26,904 populations representing 4,629 species from 11 taxonomic classes. For each of these populations we used the additional data stored in the LPD indicating whether a population was managed, utilised, located inside a protected area, or likely benefitting from conservation action (Table 4.2). Conservation actions were categorized by extracting all populations in the LPD with management recorded. We first excluded populations where management was unknown (coded as: unmanaged (npops = 9,296); managed (npops = 5,362); or unknown (npops = 12,246). The managed and unmanaged sample contained populations from 182 countries, with 136 countries represented in the managed sample and 169 in the unmanaged sample (Figure 4.1). For the populations with management records, we determined if the recorded management qualified as

conservation by extracting the management comments from each of the 5,362 managed populations (Table 4.2). Using these comments on management interventions from the LPD, along with original sources for the population data and additional information provided online (e.g. webpage for specific species recovery projects), we removed populations where the management did not qualify as conservation or research. For the remaining populations we documented conservation and research actions according to the relevant IUCN-CMP conservation actions and research actions classification schemes (Salafsky et al. 2008). These schemes record actions in a hierarchical structure: For example, in the conservation scheme, actions are first divided into seven primary categories (Land & water protection, Land & water management, Species management, Education & awareness, Law & policy, Livelihood Economic & other incentives, External capacity building), and then further divided into detailed sub-categories, such as invasive/problematic species control or species re-introduction (see the IUCN-CMP conservation actions and research actions classification scheme for detailed categories and Appendix 4.1 for further information and discussion of management information). Populations that could not be categorized in terms of conservation or research actions were removed from the final analysis (npops = 223). Conservation actions were assessed for a total of 14,329 populations in the LPD (53.3% of all), of which 5,243 populations from 1,207 species were recorded as potentially targeted by conservation, and 9,086 populations from 3,106

species were recorded as without targeted conservation. Of these populations, conservation actions were categorized for 4,347 populations, with 41 populations being solely targeted by research actions.

Matching populations and calculating trends

To explore the impact of species-targeted conservation actions (hereafter, conservation), we used statistical matching to select the populations used to calculate trends for the populations targeted for conservation and the counterfactual comparison groups, using the MatchIt package (Ho et al. 2021). Matching addresses potential biases between treatment and control populations that could influence the estimates of conservation impacts. This is critical in stochastic environments where the outcome of interest is affected by multiple drivers of change – as is the case for measurements of abundance across ecosystems – and counterfactuals can take many forms (Bull et al. 2020). To characterize this uncertainty, we created four counterfactual scenarios to represent different ways conservation targeted populations could have developed in the absence of conservation.

Scenario one assumed that, in the absence of conservation, populations would have developed similarly to other non-targeted populations from the same genus and political region. Thus, this scenario uses exact matching on genus and region to compare treated populations to non-treated populations from that genus and found in the same region. Exact matching describes similarity between populations using a

distance D given a vector X of covariates, where $D = 0$ for individual i and j if $X_i = X_j$, and $D = \infty$ if $X_j \neq X_i$. Therefore, each targeted population was matched to all possible control populations with the exact same covariates (Stuart 2010). This creates subcategories based on unique combinations of the selected covariates, assigning populations with similar covariates to the same subcategories. Populations within subcategories that lack either treatment or control are removed. Exact matching allows a single population to have multiple matches, which we use to represent multiple realizations of how a targeted population could have developed.

Scenario two assumed that, in the absence of conservation, a conservation targeted population would show similar trends in abundance to any other population from the same species and country. Relative to the first scenario, exact matching on country and species reduces the sample size but imposes stricter conditions in terms of similarity between the treatment and control populations.

Scenario three assumed that, given no conservation, a targeted population would have developed similarly to a non-targeted population from the same taxonomic class and political region. In addition, we matched each population on the populations' year of first record and time series length, so that each was matched to a single non-targeted population from the same class and region, and with similar time-series characteristics. We did this using a combination of exact and propensity score matching. Propensity score matching uses logistic regression to predict a probability

of receiving treatment, which in this case is targeted conservation, given a set of observed predictors (Williamson & Forbes 2014). Taxonomic class and global region were included as exact matches, whereas the first year of observation and time series length were included as continuous covariates using one-to-one covariate matching without replacement based on the propensity score. This ensured that the targeted sample and the counterfactual contained the same number of populations, each population compared to its closest match given the observed covariates. Including the year of the time-series was to address any overarching changes within the regions that might mean that comparing populations from different periods of time would be problematic.

In the fourth scenario, we made no assumptions about how the targeted populations would have developed without conservation. We thus compared the full sample of conservation targeted populations for the conservation group with the full sample of populations without conservation as the counterfactual group.

For each of the four scenarios, we created multi-species indices of relative abundance using the *rpi* package (Freeman et al. 2020). Here, annual population growth rates are modelled using the chain method for populations with fewer than six data points, and a Generalized Additive Model (GAM) for populations with six or more observations (Collen et al. 2009; McRae et al. 2017). For species with multiple populations, the estimated annual trends were averaged into a single annual trend following $\bar{d}_t =$

$\frac{1}{n} \sum_{i=1}^{n_t} d_{it}$ where n_t is the number of populations and d_{it} is the annual population change rate in a given year. The rate of change is given by $d_t = \log_{10} \frac{N_t}{N_{t-1}}$ where N is the population estimate and t is the year. Annual log growth rates were capped between -1 and 1. Indices were created based on a geometric mean approach using the log-transformed growth rates where the index year $I_0 = 1$ and the following indices are calculated as $I_t = I_{t-1} * 10^{\bar{d}_t}$ (McRae et al. 2017). The 95% confidence intervals were generated with 10,000 bootstrap replicates across species-level annual growth rates (Collen et al. 2009).

Our approach is similar to that used to create the LPI, except that no taxonomic or other weighting was applied, so that each species was weighted equally. The LPI aims to characterize global trends in vertebrate populations in a balanced fashion, and therefore applies weighting to account for taxonomic and geographical inequalities in the sampled data (McRae et al. 2017). However, we set out to estimate the impact of conservation actions on target populations using a matching approach which similarly corrects for bias. Furthermore, we created the four counterfactuals from matched subsamples of the LPD. The weighted approach is not suitable because these samples are much smaller than the actual LPD, and not randomly selected. Therefore, we did not apply the LPI weighting, as this could potentially exacerbate the effect of any selection biases in ways that would be difficult to interpret.

Clusters of populations with extreme abundance changes and time series length have been shown to influence population trend indicators (Wauchope et al. 2019b; Leung et al. 2020). We therefore tested the sensitivity of our indices by recreating them without the top and bottom 1 % quantiles of species with increasing and decreasing populations, and by restricting populations to those with time series spanning a minimum of 5 and 10 years.

Table 4.2 Variables extracted from the LPD which are used for categorizing conservation actions and analysis. The reasons for population increases that have been used to calculate the global impact of conservation on the unweighted LPI are underlined.

Variable	Description
Species	Taxonomic information according to the latest authority for that species
Genus	Taxonomic information according to the latest authority for that genus
Class	Taxonomic information according to the latest authority for that class. Coding: 'Actinopteri', 'Coelacanthi', 'Dipneusti', 'Elasmobranchii', 'Holocephali', 'Myxini', 'Petromyzonti', 'Amphibia', 'Aves', 'Mammalia', 'Reptilia'.
Country	The country (or countries) that the population occurs in from the list. Marine data are allocated a country if it is within its EEZ, or as International Waters. Multiple countries are selected in order of proportion of the population it represents starting with the greatest.
Region	The political region a country is assigned to. Coding: 'Africa', 'Antarctic', 'Asia', 'Europe', 'International Waters', 'Latin America and Caribbean', 'North America', 'Oceania'
First year of observation	The first recorded year with an abundance estimate for a given population
Time series length	The number of years from first to last population abundance estimate
Managed	A population that receives targeted management (some of which involves sustainable use). This is usually to promote recovery in a population or can incentivise it's use for conservation. It can include measures to stem 'unsustainable' population growth. Coding: 'Yes', 'No', 'Unknown'
Utilised	A population that is intentionally regularly or systematically utilised, either individuals or eggs. This may be sustainable or unsustainable, and the population does not necessarily have to be threatened by use or overexploited. This refers to consumptive use whereby individuals or parts of individuals are removed from the wild. Coding: 'Yes', 'No', 'Unknown'
Targeted conservation actions	A population that is intentionally targeted by conservation. Coding of conservation actions follows categories in Salafsky et al 2008
Reason for population increase	Indicates the reasons given by the data source for any increase in the population. Coding: ' <u>Introduction</u> ', ' <u>Recolonisation</u> ', ' <u>Recruitment</u> ', ' <u>Removal of threat</u> ', 'Rural to urban migration', ' <u>Reintroduction</u> ', ' <u>Range shift</u> ', ' <u>Legal protection</u> ', ' <u>Management</u> ', 'Other', 'Unknown'.
Protected area status	Indicates if the population is inside a protected area. Coding: 'Both', 'No', 'No (area surrounding PA)', 'No (Large survey area)', 'Unknown', 'Yes'.

Calculating the impact of conservation on a global index of species abundance

To explore the wider global impact of conservation on populations, we used a quantitative method that built on Hoffmann et al (2010). First, we calculated a global vertebrate population index using all trend data in the LPD. Then, to evaluate how conservation actions have affected this particular global species abundance metric, we calculated the index under alternative assumptions into three counterfactual indices. The first was a simple population index excluding all conservation targeted populations. This index was calculated by removing the populations with records of targeted conservation but otherwise using all the available LPD population data.

We calculated a second population index where the impact of targeted conservation actions was excluded by assuming stable trends for otherwise increasing populations with records of conservation actions. For this index, we first identified those populations exhibiting an observed increase over the time series, and then selected all populations categorized as ‘conservation targeted’ and for which information about the reason for population increase was recorded. From this sample, we selected increasing populations with a plausible link between the observed change and a conservation intervention (see Figure S4.6 and Appendix 4.1 for the full method of validating this link). The reasons for population increase that we selected are underlined in Table 4.2. For this subsample, we replaced the observed abundance estimates with a constant, thus assuming that trends for these populations would have at least remained stable without conservation. The index was then calculated from all of the available LPD population data, but with constant annual abundance estimates for the selected subset of populations.

Finally, we calculated a population index where the impacts of targeted and collateral conservation were excluded (Hoffmann et al. 2015). This population index, excluding the impact of both targeted and collateral conservation, was calculated similar to the second index but in addition excluding the impact of collateral conservation. By collateral, we mean that a population could have benefitted from conservation without being specifically targeted, which we defined as having increasing population trends within a protected area, while not being specifically chosen for any of the targeted actions. This index was therefore calculated using the full LPD data but assuming stable trends for the same populations as in the second index and additionally populations without targeted conservation, but which were inside a protected area and had a reason for population increase recorded. Population indices were calculated using the `rlpi` package (Freeman et al. 2020) without applying taxonomic or geographical weighting. To visualize the impact of conservation, we plotted the difference between the reference population index, calculated using all of the population trend data in the LPD, and the three potential scenarios that represent the reference index without the likely impact of conservation.

Mixed model

We compared the effects of the seven primary targeted conservation actions on abundance trends (the log of the summed rate of population change) using a mixed model framework (Mcrae et al. 2020). The `rlpi` package generates a matrix of annual rates of population change for each population. We summed these rates into a logged value of total change for each population. To test the effects of conservation actions in general, and of the different types of interventions, two separate models were specified.

In the first, conservation actions were aggregated into a single fixed effect binary variable (1 = targeted or 0 = not targeted by conservation) to estimate the overall effect of actions regardless of action type.

In the second, we specified a binary variable for each of the seven main conservation actions. This allowed us to estimate the effect of conservation actions (model one) and then disentangle the individual effects of specific types of actions (model two). We included time series length, taxonomic class and the utilization status of each population as fixed effects, as these have been shown to affect abundance changes (Wauchope et al. 2019b; McRae et al. 2020). We specified similar random effect structures as in McRae et al (2020), including Family, Genus and population location to account for taxonomic and site specific effects (Model one: $\text{sum_lambda} \sim 0 + \text{TS_length} + \text{Utilised} + \text{Conservation} + \text{Class} + (1|\text{Family/Binomial}) + (1|\text{Location})$;
Model two: $\text{lambda_sum} \sim 0 + \text{Utilised} + \text{ts_length} + \text{land_water_protection} + \text{land_water_management} + \text{species_management} + \text{education_awareness} + \text{law_policy} + \text{incentives} + \text{external_capacity} + \text{research} + \text{Class} + (1|\text{Family/Binomial}) + (1|\text{Location})$).

Results

Conservation actions in the LPD

Mammals had the highest number of managed populations in the LPD, albeit across a relatively low number of species (nspp (number of species) = 244, npops = 2,200), followed by fish (nspp = 548, npops = 1,994), birds (nspp = 305, npops = 756), reptiles (nspp = 58, npops = 220) and amphibians (nspp = 52, npops = 73). The taxonomic classes included in the fish, mammal and bird groups, maps of the starting year (Figure

S4.2) and the length (Figure S4.3) of the population time series, are all given in the Supplementary material.

Species management was the most frequent conservation action ($n = 2,937$), followed by land & water management ($n = 1,095$), and law & policy actions ($n = 467$) (Figure 4.2; see Figure S4.1 for detailed categories). Conservation actions differed between classes, with species management being the most abundant action for mammals and fishes, while land & water management was more frequently applied for birds.

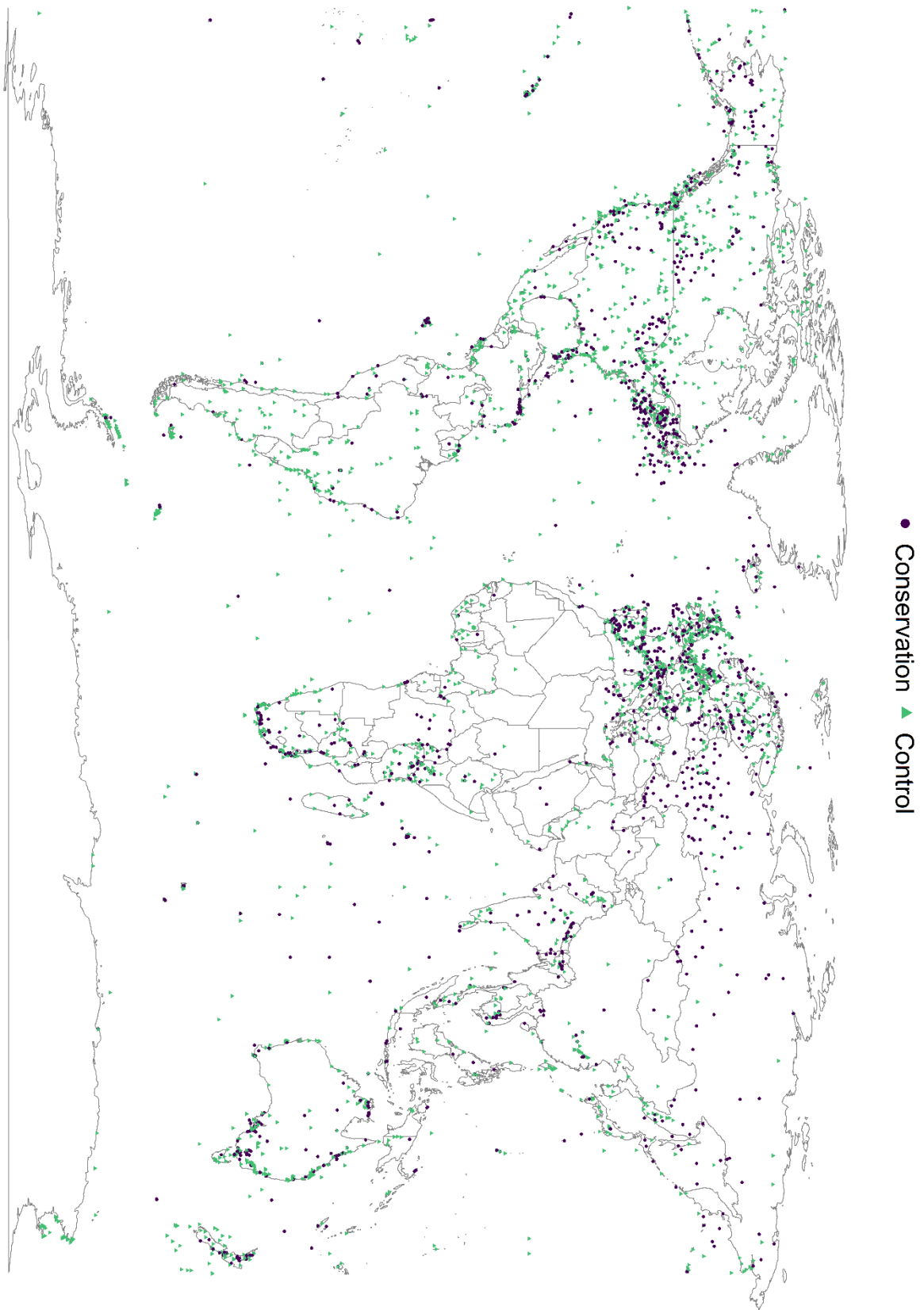


Figure 4.1 Locations of managed ($n = 5,243$) and unmanaged populations ($n = 9,086$).

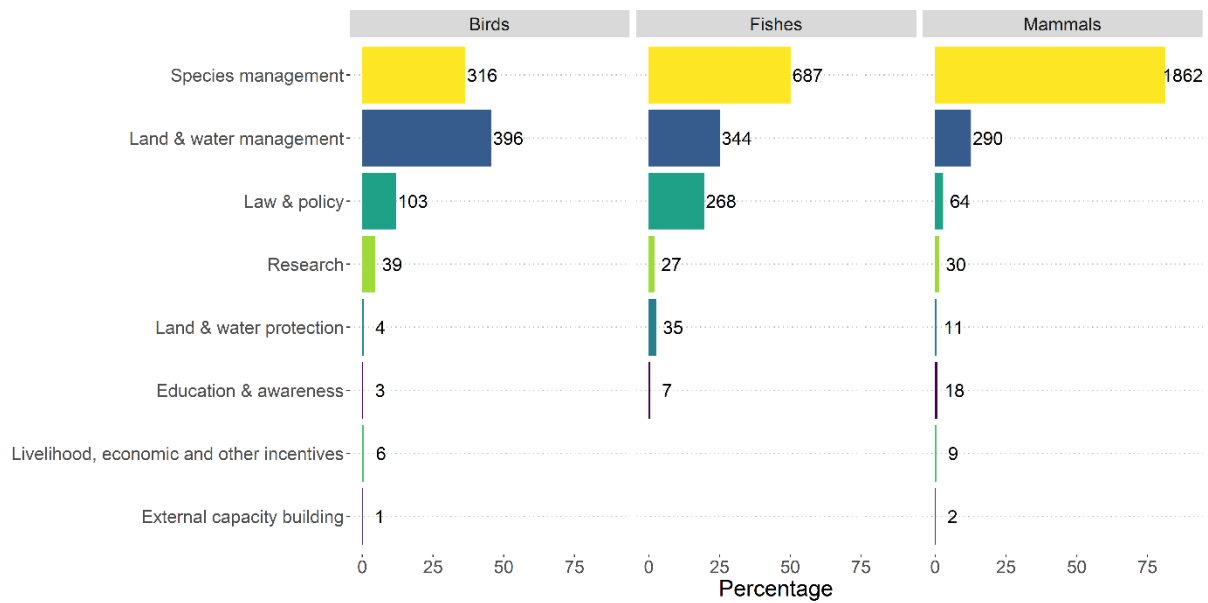


Figure 4.2 Number of targeted populations and the relative percentage of conservation actions for fish, birds and mammals. For each of the three groups with targeted conservation actions, the x-axis shows the percentage of populations targeted by the seven primary conservation actions and research (Salafsky et al. 2008). The number of targeted populations is shown for each bar.

Impact of conservation under four different counterfactual scenarios

Trends for populations targeted by conservation actions increased consistently and strongly when compared with counterfactuals (Figure 4.3). The largest difference was observed when comparing populations of the same species within the same country (scenario 2). Here, the index for the conservation targeted sample increased from 1 to 3.36 (234% increase), whereas the counterfactual sample increased to just 1.01 (1% increase). The smallest difference between the indices was observed in scenario 3 (matching on taxonomic class, region, time series length and starting year) where the index for the conservation targeted sample increased to 1.6 (60% increase) while the counterfactual decreased to 0.79 (21% decrease). Sensitivity tests showed that the conservation targeted population indices remained higher than the counterfactual in

all cases (Figure S4.4. See Figure S4.5 for the number of species in each class within each scenario).

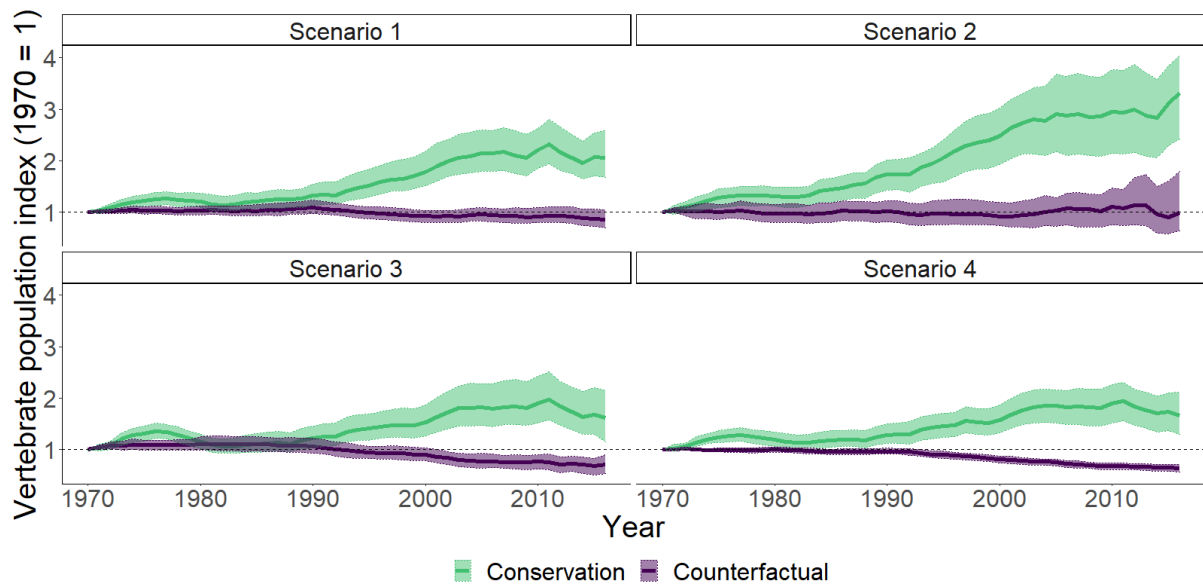


Figure 4.3 Vertebrate population trends for species subject to conservation actions or responses (in green – upper lines) and not targeted by conservation responses (in purple - lower lines) representing counterfactual species trends. Shaded areas show 95% confidence intervals. Dashed line equals index 1. Scenario 1 – Genus + Region; Conservation + 103%; nspp = 785, npop = 3,377, Counterfactual - 30%; nspp = 1,001, npop = 3,463: Scenario 2 – Species + Country; Conservation + 234%; nspp = 348, npop = 1483, Counterfactual + 1%; nspp = 347, npop = 895: Scenario 3 – Class + Region + Time series length + Start year; Conservation + 60%; nspp = 1,010, npop = 2,929, Counterfactual - 21%; nspp = 1,449, npop = 2,929: Scenario 4 – Full sample; Conservation + 75%; nspp = 1,207, npop = 5,243, Counterfactual - 35%; nspp = 3,099, npop = 9,071

Impact of conservation on indices of global species abundance

Our global vertebrate population index decreased by 24% between 1970-2016 but could have declined by 31% (7% points more) if conservation targeted populations had remained stable, or by 32% (8% points more) if both conservation targeted populations and populations affected by collateral conservation remained stable (Figure 4.4, Figure S4.7).

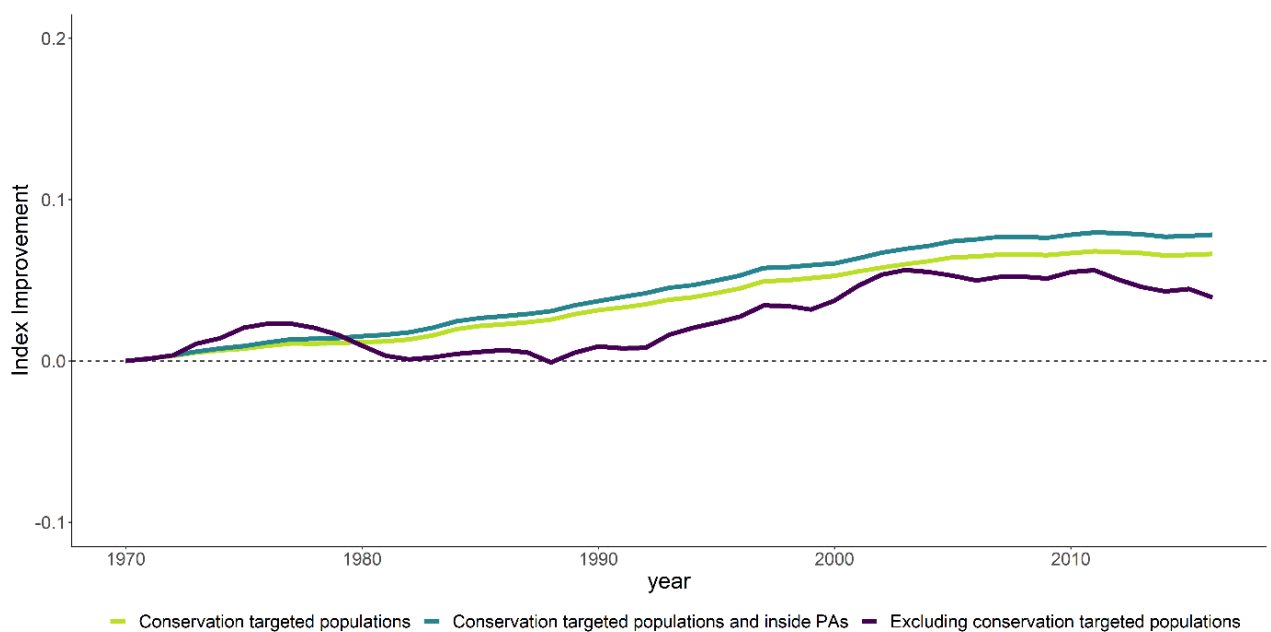


Figure 4.4 Improvements in vertebrate population trends when assuming stable trends for populations with increasing population trends attributable to conservation (light green – middle line), when assuming stable trends for increasing conservation targeted populations and populations inside PAs attributable to conservation (turquoise – top line) and when excluding populations targeted by conservation actions (purple – bottom line). Improvements are calculated by subtracting each of the three counterfactual population trends from the global reference trend. The dashed line

represents no difference between the reference trend and any of the three alternative trends. See Figure S4.7 for the original trends.

Drivers of population trends estimated from mixed models

Conservation actions had a positive effect on targeted populations (Estimate = 0.12, Std Error = 0.02, t value = 5.8). Land & water management, species management and land & water protection actions for species were highly associated with population increases, suggesting a particularly strong effect of actions within these three conservation categories (Figure 4.5). Research actions were negatively associated with population trends. Longer population time series were more likely to have increased, while utilized populations did not display a clear trend.

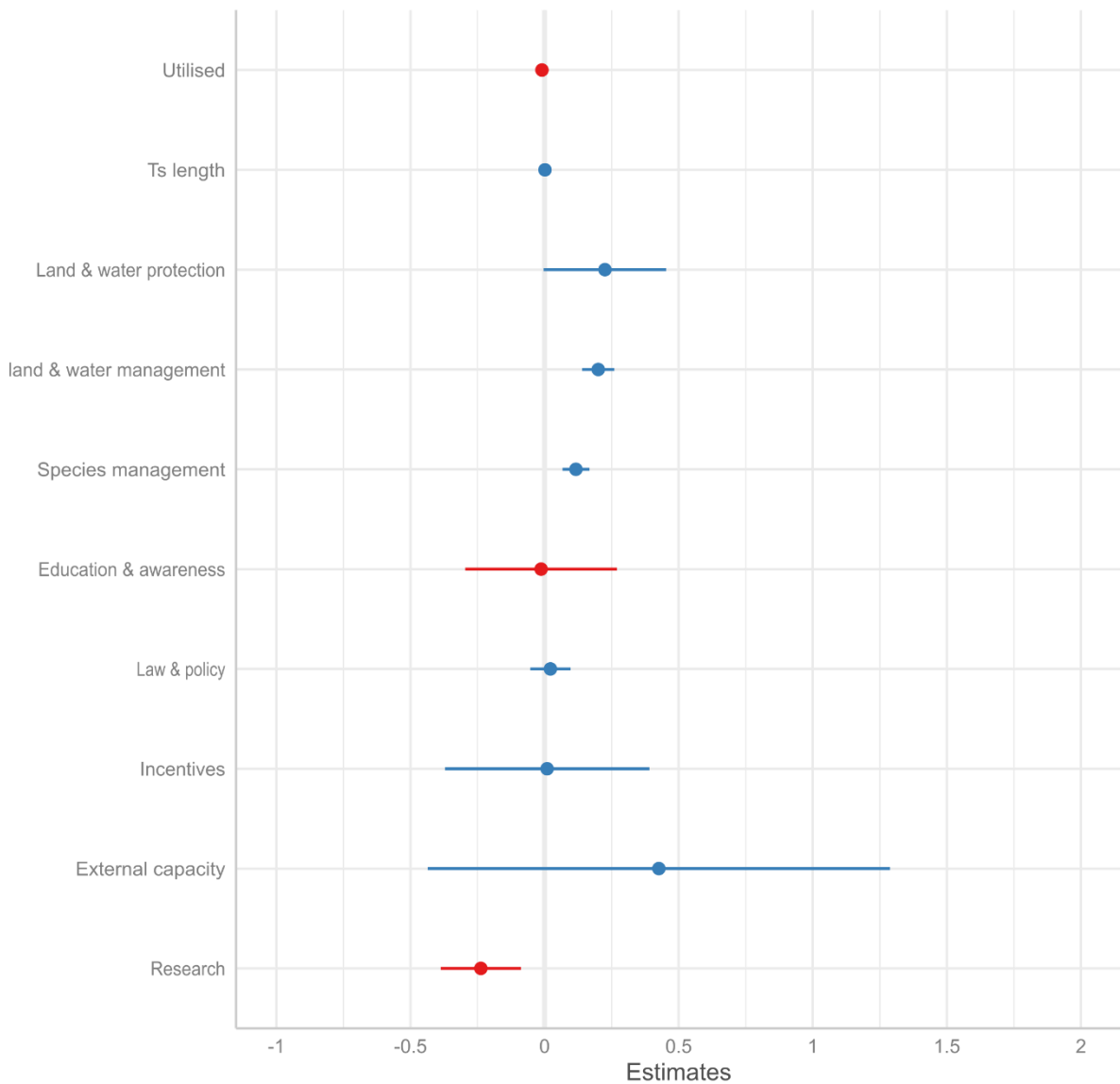


Figure 4.5 Parameter estimates (estimated total change on the log scale) for the seven primary types of conservation actions, research, utilization status and time series length. See Table S1 for parameter estimates, standard errors and *t* values.

Discussion

Our analyses, using one of the largest global datasets of population time-series, suggest that conservation actions have had a positive influence on global vertebrate populations. Our results were consistent and robust, with a positive impact of

conservation detected for all scenarios of counterfactual population developments tested, substantiated by a marked difference between global indices including and excluding the impact of conservation. Furthermore, we saw an effect of conservation not only when comparing the relative difference between treatments and counterfactuals: increasing population trends in absolute terms were also more likely for populations targeted by conservation actions, especially land & water management and species management. Our findings therefore suggest that conservation has delivered substantial benefits to targeted populations.

Our analysis demonstrates the importance of conservation actions that are less frequently evaluated, and thus expands on previous large-scale evaluation efforts within conservation science. Previous efforts have, to a large degree, focused on protected areas (Geldmann et al. 2013a, 2013b, 2019; Butchart et al. 2015; Barnes et al. 2016; Wauchope et al. 2019a; Cazalis et al. 2020, 2021a; Maxwell et al. 2020; Terraube et al. 2020). However, protected areas are under a wide variety of management regimes, with large differences in management effectiveness (Geldmann et al. 2015). This means that the percentage of area protected by itself is inadequate to measure conservation effectiveness (Visconti et al. 2019; Rodrigues & Cazalis 2020), and potentially conceals the effectiveness of certain conservation interventions. By identifying the effects of specific conservation actions targeting populations of vertebrate species, we demonstrate the positive impact of conservation efforts on vertebrate populations globally.

Conservation actions without any immediate effect on population trends, such as education & awareness (Figure 4.5) can still provide important conservation benefits. We assessed the impact of conservation on population abundance, but there are many alternative outcome metrics which could have illustrated the effectiveness of these

actions. For example, we found no effect of awareness & education as this category of conservation might not cause population increases directly. Instead, awareness & education can work indirectly by providing funding for conservation organizations and by giving mandate and support for legal protection to governments.

Accurate assessments of conservation impact depend on accessible and representative data across different aspects of biodiversity. Likewise, appropriately specified counterfactuals require information about the covariates that affect the sample. Currently, global biodiversity data are taxonomically and geographically biased (McRae et al. 2017; Troudet et al. 2017; Hughes et al. 2021). Additionally, fine-scale data and contextual information for conservation targeted populations, such as the type or duration of conservation management, are extremely limited. For example, we could not distinguish between population trends pre and post intervention, as such data do not readily exist on a global level. Similarly, because of the limited data, we specified the counterfactual control groups using a limited set of covariates which means that treatment and control groups can still differ on important covariates. Advances in remote sensing are already reducing this knowledge gap somewhat, with data derived from remote sensing images widely used to inform where to target conservation actions and evaluate the impact of environmental policy (Chen et al. 2019; Runting et al. 2020). However, many measures cannot be proxied by land cover (for example reintroduction programmes) and must therefore be complemented by in-situ monitoring. Here, we show the relevance of retaining and standardizing such local information, but also that widespread systematic monitoring is required to improve evaluation efforts, especially outcome-based assessments, and to determine the progress towards future biodiversity targets. To explore patterns of causation between conservation actions and wildlife populations requires substantial improvements of

global biodiversity data. For example, future monitoring records could be standardized to capture when monitored populations were targeted by conservation actions, and the associated costs. Additional information, such as the temporal exposure to a conservation intervention, would allow effect estimates to be derived with greater confidence from more reliable study designs (de Palma et al. 2018; Christie et al. 2019; Wauchope et al. 2021). However, substantial investments are required for these improvements. We did not account for any potential publication bias in the original sources of the LPD. While studies reporting negative or mixed results provide important knowledge, positive findings are often prioritised (Wood 2020). Large databases, such as the LPD, could therefore also be biased similar to the input sources. An increased focus on reporting negative findings and large-scale data collation can help to address such concerns in future analysis.

Our analysis calculating global indices with and without the impact of conservation likely underestimates the impact of conservation. First, we assumed that, in the absence of conservation, conservation targeted populations that increased would have remained stable. This is in stark contrast to the general pattern of global declines (WWF 2020a). Second, because of data limitations, only a small subset of the total LPD has a reason for increase recorded. Furthermore, effective conservation is not conditional on population increase. Instead, for conservation to be effective requires only that the outcome of interest is improved by conservation, relative to a scenario without conservation.

While populations in the LPD only represent a fraction of global biodiversity, our results offer a glimmer of hope and underpin the importance of conservation efforts in halting the global loss of biodiversity. As the global parties to the Convention on Biological Diversity are preparing the Post-2020 Global Biodiversity Framework, it is relevant to

reflect on conservation progress made over the last decade, as well as how to measure progress towards achieving these targets. One of the most prominent elements of the CBDs Strategic Plan for biodiversity 2011-2020 was focused on the establishment of effective and representative networks of protected areas (Sustainable Development Goals – Aichi Target 11), and post-2020 targets set out more ambitious targets in this regard, currently suggesting that 30% of global terrestrial area should be placed under formal protection (First draft of the post-2020 global biodiversity framework). It is important to recognize the progress made towards increasing the global coverage of protected area, with marked increases observed both on land, in freshwater environments and in the marine realm (Maxwell et al. 2020; UNEP-WCMC & IUCN 2021). However, vertebrate populations continue to decline (WWF 2020a). Similarly, without increased conservation efforts, global biodiversity projections predict continuous declines in the future (Leclère et al. 2020), highlighting the need for effective conservation actions and outcome-based targets (Butchart et al. 2015; Visconti et al. 2019; Pressey et al. 2021). We show that such targeted conservation interventions can be highly effective.

References

Barnes MD et al. 2016. Wildlife population trends in protected areas predicted by national socio-economic metrics and body size. *Nature Communications* 7:1–9. Nature Publishing Group.

Biggs R, Scholes RJ. 2005. A biodiversity intactness index. *Nature* 434:45–49. Available from <http://go.galegroup.com/ps/i.do?id=GALE%7CA185471773&v=2.1&u=ntu&it=r&p=AONE&sw=w>.

Bolam FC et al. 2020. How many bird and mammal extinctions has recent conservation action prevented? *Conservation Letters*:e12762. Available from <http://doi.wiley.com/10.1111/conl.12762>.

Bull JW, Strange N, Smith RJ, Gordon A. 2020. Reconciling multiple counterfactuals when evaluating biodiversity conservation impact in social-ecological systems. *Conservation Biology*:cobi.13570. Available from <https://onlinelibrary.wiley.com/doi/abs/10.1111/cobi.13570>.

Butchart SHM et al. 2010. Global biodiversity: Indicators of recent declines. *Science* 328:1164–1168.

Butchart SHM et al. 2015. Shortfalls and Solutions for Meeting National and Global Conservation Area Targets. *Conservation Letters* 8:329–337.

Butchart SHM, Stattersfield AJ, Collar NJ. 2006. How many bird extinctions have we prevented? ORYXDOI: 10.1017/S0030605306000950.

Butsic V, Lewis DJ, Radeloff VC, Baumann M, Kuemmerle T. 2017. Quasi-experimental methods enable stronger inferences from observational data in ecology. *Basic and Applied Ecology* 19:1–10. Elsevier GmbH. Available from <http://dx.doi.org/10.1016/j.baae.2017.01.005>.

Cazalis V, Barnes MD, Johnston A, Watson JEM, Şekercioğlu CH, Rodrigues ASL. 2021. Mismatch between bird species sensitivity and the protection of intact habitats across the Americas. *Ecology Letters*:ele.13859. Available from <https://onlinelibrary.wiley.com/doi/10.1111/ele.13859>.

Cazalis V, Princé K, Mihoub JB, Kelly J, Butchart SHM, Rodrigues ASL. 2020. Effectiveness of protected areas in conserving tropical forest birds. *Nature*

Communications 11:1–8. Springer US. Available from <http://dx.doi.org/10.1038/s41467-020-18230-0>.

Chen C et al. 2019. China and India lead in greening of the world through land-use management. *Nature Sustainability* 2:122–129. Springer US. Available from <http://dx.doi.org/10.1038/s41893-019-0220-7>.

Christie AP, Amano T, Martin PA, Shackelford GE, Simmons BI, Sutherland WJ. 2019. Simple study designs in ecology produce inaccurate estimates of biodiversity responses. *Journal of Applied Ecology*:2742–2754.

Collen B, Loh J, Whitmee S, McRae L, Amin R, Baillie JEM. 2009. Monitoring change in vertebrate abundance: the living planet index. *Conservation biology : the journal of the Society for Conservation Biology* 23:317–27. Available from <http://doi.wiley.com/10.1111/j.1523-1739.2008.01117.x>.

de Palma A, Sanchez-Ortiz K, Martin PA, Chadwick A, Gilbert G, Bates AE, Börger L, Contu S, Hill SLL, Purvis A. 2018. Challenges With Inferring How Land-Use Affects Terrestrial Biodiversity: Study Design, Time, Space and Synthesis. Page *Advances in Ecological Research*, 1st edition. Elsevier Ltd. Available from <http://dx.doi.org/10.1016/bs.aecr.2017.12.004>.

Díaz S et al. 2019. Pervasive human-driven decline of life on Earth points to the need for transformative change. *Science* 366.

Ferraro PJ. 2009. Counterfactual thinking and impact evaluation in environmental policy. *New Directions for Evaluation* 2009:75–84. Available from <https://onlinelibrary.wiley.com/doi/10.1002/ev.297>.

Ferraro PJ, Pattanayak SK. 2006. Money for Nothing? A Call for Empirical Evaluation of Biodiversity Conservation Investments. *Plos biology* 4:482–488.

Freeman R, McRae L, Deinet S, Amin R, Collen B. 2020. rlp: Tools for calculating indices using the Living Planet Index method. Available from https://github.com/Zoological-Society-of-London/living_planet_index.

Geldmann J et al. 2015. Changes in protected area management effectiveness over time: A global analysis. *Biological Conservation* 191:692–699. Elsevier B.V. Available from <http://dx.doi.org/10.1016/j.biocon.2015.08.029>.

Geldmann J, Barnes M, Coad L, Craigie ID, Hockings M, Burgess ND. 2013a. Effectiveness of terrestrial protected areas in reducing habitat loss and population declines. *Biological Conservation* 161:230–238. Elsevier Ltd. Available from <http://dx.doi.org/10.1016/j.biocon.2013.02.018>.

Geldmann J, Barnes M, Coad L, Craigie ID, Hockings M, Burgess ND. 2013b. Effectiveness of terrestrial protected areas in reducing habitat loss and population declines. *Biological Conservation* 161:230–238. Available from <https://linkinghub.elsevier.com/retrieve/pii/S0006320713000670>.

Geldmann J, Manica A, Burgess ND, Coad L, Balmford A. 2019. A global-level assessment of the effectiveness of protected areas at resisting anthropogenic pressures. *Proceedings of the National Academy of Sciences of the United States of America* 116:23209–23215. Available from <https://doi.org/10.1073/pnas.1908221116>.

Grace MK et al. 2021a. Testing a global standard for quantifying species recovery and assessing conservation impact. *Conservation Biology*:1–17.

Grace MK et al. 2021b. Building robust , practicable counterfactuals and scenarios to evaluate the impact of species conservation interventions using inferential approaches. *Biological Conservation* 261:109259. Elsevier Ltd. Available from <https://doi.org/10.1016/j.biocon.2021.109259>.

Ho D, Imai K, King G, Stuart E, Whitworth A, Greifer N. 2021. Package ‘ MatchIt .’ Available from <https://github.com/kosukeimai/MatchIt>.

Hoffmann M et al. 2010. The impact of conservation on the status of the world's vertebrates. *Science* 330:1503–1509.

Hoffmann M, Duckworth JW, Holmes K, Mallon DP, Rodrigues ASL, Stuart SN. 2015. The difference conservation makes to extinction risk of the world's ungulates. *Conservation Biology* 29:1303–1313.

Hughes AC, Orr MC, Ma K, Costello MJ, Waller J, Provoost P, Yang Q, Zhu C, Qiao H. 2021. Sampling biases shape our view of the natural world. *Ecography* 44:1259–1269.

Jellesmark S, Ausden M, Blackburn TM, Gregory RD, Hoffmann M, Massimino D, McRae L, Visconti P. 2021. A counterfactual approach to measure the impact of wet grassland conservation on UK breeding bird populations. *Conservation Biology*:cobi.13692. Available from <https://onlinelibrary.wiley.com/doi/10.1111/cobi.13692>.

Joppa LN, Pfaff A. 2011. Global protected area impacts. *Proceedings of the Royal Society B: Biological Sciences* 278:1633–1638. Available from <https://royalsocietypublishing.org/doi/10.1098/rspb.2010.1713>.

Leclère D et al. 2020. Bending the curve of terrestrial biodiversity needs an integrated strategy. *Nature* 2018. Available from <https://doi.org/10.1038/s41586-020-2705-y>
<http://www.nature.com/articles/s41586-020-2705-y>.

Leung B, Hargreaves AL, Greenberg DA, Mcgill B, Dornelas M, Freeman R. 2020. Clustered versus catastrophic global vertebrate declines. *Nature* DOI: 10.1038/s41586-020-2920-6. Springer US. Available from <http://dx.doi.org/10.1038/s41586-020-2920-6>.

LPD. 2020. Living Planet Index database. Available from <https://livingplanetindex.org> (accessed November 16, 2020).

Margoluis R, Stem C, Salafsky N, Brown M. 2009. Design alternatives for evaluating the impact of conservation projects. *New Directions for Evaluation* 2009:85–96. Available from <http://doi.wiley.com/10.1002/ev.298>.

Maxwell SL et al. 2020. Area-based conservation in the twenty-first century. *Nature* 586:217–227. Springer US. Available from <http://dx.doi.org/10.1038/s41586-020-2773-z>.

McRae L, Deinet S, Freeman R. 2017. The diversity-weighted living planet index: Controlling for taxonomic bias in a global biodiversity indicator. *PLoS ONE* 12.

McRae L, Freeman R, Geldmann J, Moss GB, Kjær-hansen L, Burgess D. 2020. A global indicator of utilised wildlife populations: regional trends and the impact of management DOI: 10.1101/2020.11.02.365031.

Pressey RL, Visconti P, McKinnon MC, Gurney GG, Barnes MD, Glew L, Maron M. 2021. The mismeasure of conservation. *Trends in Ecology & Evolution* 36:808–821. Elsevier Ltd. Available from <https://doi.org/10.1016/j.tree.2021.06.008>.

Pynegar EL, Gibbons JM, Asquith NM, Jones JPG. 2019. What role should randomized control trials play in providing the evidence base for conservation? *Oryx* 55:235–244. Available from https://www.cambridge.org/core/product/identifier/S0030605319000188/type/journal_article.

Rodrigues ASL, Cazalis V. 2020. The multifaceted challenge of evaluating protected area effectiveness. *Nature Communications* 11:1–4. Springer US. Available from <http://dx.doi.org/10.1038/s41467-020-18989-2>.

Rose DC et al. 2018. The major barriers to evidence-informed conservation policy and possible solutions. *Conservation Letters* 11:1–12.

Runting RK, Phinn S, Xie Z, Venter O, Watson JEM. 2020. Opportunities for big data in conservation and sustainability. *Nature Communications* 11:1–4. Springer US. Available from <http://dx.doi.org/10.1038/s41467-020-15870-0>.

Salafsky N et al. 2008. A standard lexicon for biodiversity conservation: Unified classifications of threats and actions. *Conservation Biology* DOI: 10.1111/j.1523-1739.2008.00937.x.

Schleicher J, Eklund J, Barnes M, Geldmann J, Oldekop JA, Jones JPG. 2019. Statistical matching for conservation science. *Conservation Biology*:cobi.13448. Available from <https://onlinelibrary.wiley.com/doi/abs/10.1111/cobi.13448>.

Stuart EA. 2010. Matching Methods for Causal Inference: A Review and a Look Forward. *Statistical Science* 25:1–21.

Terraube J, van doninck J, Helle P, Cabeza M. 2020. Assessing the effectiveness of a national protected area network for carnivore conservation. *Nature Communications* 11:1–9. Springer US. Available from <http://dx.doi.org/10.1038/s41467-020-16792-7>.

Tittensor DP et al. 2014. A mid-term analysis of progress toward international biodiversity targets. *Science* 346:241–244.

Troudet J, Grandcolas P, Blin A, Vignes-Lebbe R, Legendre F. 2017. Taxonomic bias in biodiversity data and societal preferences. *Scientific Reports* 7:1–14.

UNEP-WCMC, IUCN. 2021. Protected Planet Report 2020. Available from <https://livereport.protectedplanet.net/>.

Visconti P, Butchart SHM, Brooks TM, Langhammer PF, Marnewick D, Vergara S, Yanosky A, Watson JEM. 2019. Protected area targets post-2020. *Science* 364:239–242. Available from <http://www.sciencemag.org/lookup/doi/10.1126/science.aav6886>.

Wauchope H et al. 2019a. Quantifying the impact of protected areas on near-global waterbird population trends, a pre-analysis plan:1–17. Available from <https://doi.org/10.7287/peerj.preprints.27741v1>.

Wauchope HannahS, Amano T, Geldmann J, Johnston A, Simmons BI, Sutherland WJ, Jones JPG. 2020. Evaluating impact using time-series data. *Trends in Ecology & Evolution In Review*:1–10. The Authors. Available from <https://doi.org/10.1016/j.tree.2020.11.001>.

Wauchope HS, Amano T, Sutherland WJ, Johnston A. 2019b. When can we trust population trends? A method for quantifying the effects of sampling interval and duration. *Methods in Ecology and Evolution* 10:2067–2078.

Wiik E, Jones JPG, Pynegar E, Bottazzi P, Asquith N, Gibbons J, Kontoleon A. 2020. Mechanisms and impacts of an incentive-based conservation program with evidence from a randomized control trial. *Conservation Biology* 34:1076–1088.

Williamson EJ, Forbes A. 2014. Introduction to propensity scores. *Respirology* 19:625–635.

Wood KA. 2020. Negative results provide valuable evidence for conservation. *Perspectives in Ecology and Conservation* 18:235–237. Available from <https://doi.org/10.1016/j.pecon.2020.10.007>.

WWF. 2020. Living planet report 2020 - Bending the curve of Biodiversity loss. Page Almond R.E.A., Grooten M. and Petersen, T. (Eds). WWF, Gland, Switzerland. Available from <https://f.hubspotusercontent20.net/hubfs/4783129/LPR/PDFs/ENGLISH-FULL.pdf>.

Supplementary material

Appendix 4.1

Taxonomic classes categorized into groups

We categorized Actinopteri, Coelacanthi, Dipneusti, Elasmobranchii, Holocephali, Myxini and Petromyzonti together as ‘fish’, Aves as ‘birds’, Mammalia as ‘mammals’, Reptilia as ‘reptiles’ and Amphibia as ‘amphibians’.

Caveats

The management actions used in this analysis are recorded from information in the LPD and, in few instances, from external sources online. SJ and MH have made every effort to correctly code the actions based on information available in the LPD, original sources and additional online material. However, some errors may remain as

management actions have not been coded according to consistent categories. This means that management actions may be in place for some populations but are missing from the LPD (omission error). Also, management actions may be documented in place, but are relatively trivial / inconsequential to the population (commission error) – our analysis treats all actions as equal importance. Furthermore, management information may be documented, but may not reflect all actions in place for that population. Additionally, conservation actions apply only to the population during that time-series time-period.

Validating the link between conservation actions and reasons for population increase

For the selected subsample used to estimate the impact of conservation on the global index, we assumed a causal relationship between conservation actions and population increases. However, a potential concern was whether the reasons for population increases listed in the original sources and the conservation actions that we categorized matched. A mismatch between the selected populations' reasons for population increase and the targeted conservation actions would either mean that (i) the recorded conservation actions did not cause the population increase or (ii), that conservation actions were missing from the original sources or not categorized accordingly. We verified whether the targeted conservation actions and reasons for increase aligned, effectively meaning that conservation was likely caused the population increase, by comparing the frequency at which conservation actions and different reasons for population increase were recorded for populations (Fig S6). If population increases were caused by conservation actions, we would expect a higher frequency at which specific reason for increase logically caused by a certain conservation action and that specific conservation action occurred. For example, we

would expect increasing populations targeted by effective legislation to have legal protection listed as the reason for increase. For each population, unique combinations of the selected reasons for population increase and primary conservation type were expanded into separate rows, each row representing a unique combination of a single reason listed as the cause of population increase and a single recorded conservation action. The frequency of unique combinations was then summarized into a frequency table which we used to visualize the frequency at which each of the primary conservation actions occurred relative to each of the selected reasons for increases.

Table S4.1 *Parameter estimates for the seven primary types of conservation actions, research, utilization status and time series length.*

	Estimate	Std. Error	t value
Utilised	-0.0091880	0.0112007	-0.820
Ts length	0.0014400	0.0006965	2.067
Land water protection	0.2253907	0.1165584	1.934
Land water management	0.2003847	0.0306416	6.540
Species management	0.1153678	0.0254073	4.541
Education awareness	-0.0118926	0.1442860	-0.082
Law policy	0.0209303	0.0383110	0.546
Incentives	0.0094938	0.1945584	0.049
External capacity	0.4269375	0.4397941	0.971
Research	-0.2371960	0.0762102	-3.112

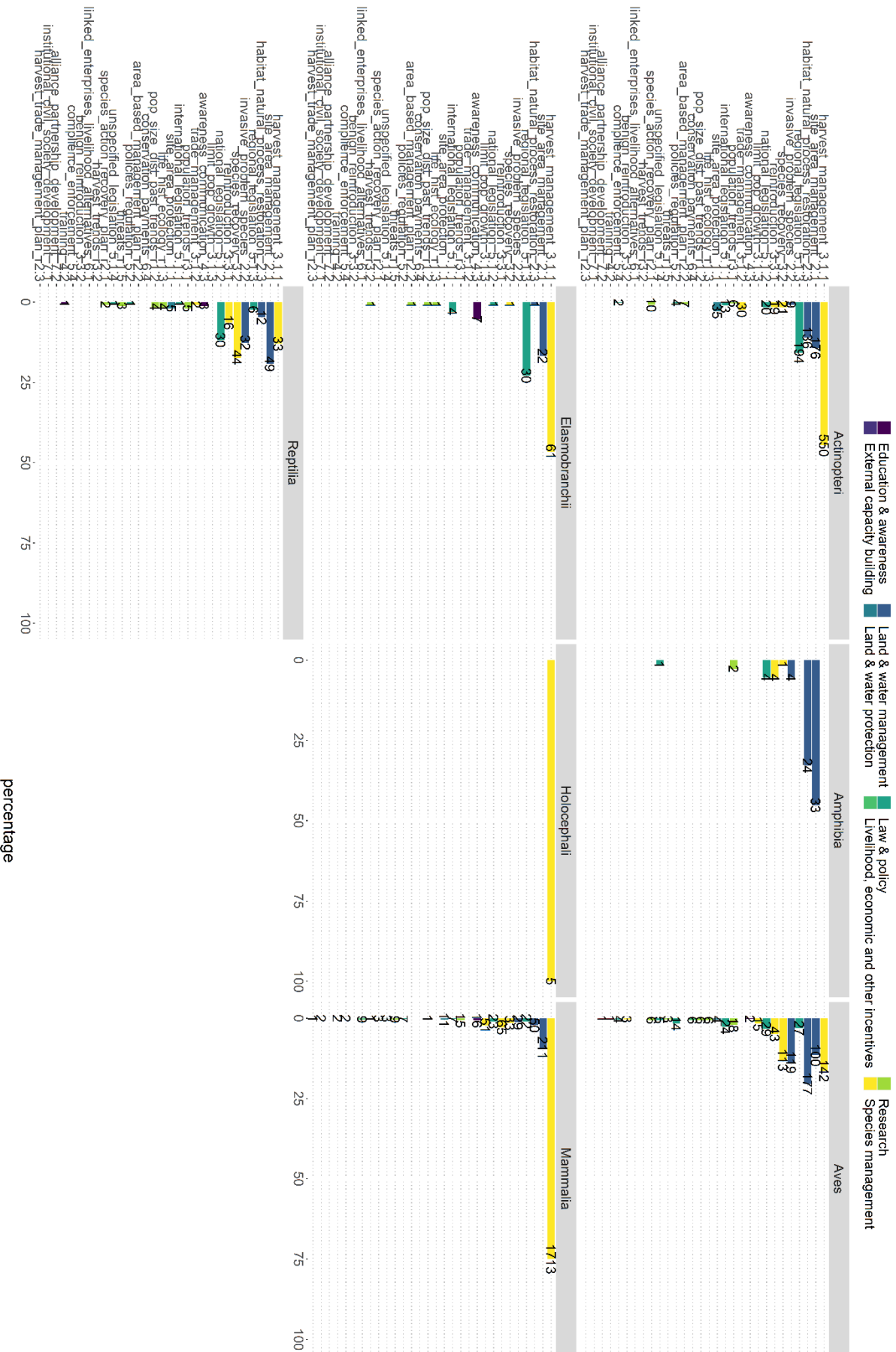


Figure S4.1 Number of targeted populations and the relative percentage of detailed conservation actions for taxonomic classes. For each of the taxonomic classes with targeted conservation actions, the x-axis shows the percentage of populations targeted by the detailed conservation actions and research (Salafsky et al. 2008). The number of targeted populations is shown for each bar.

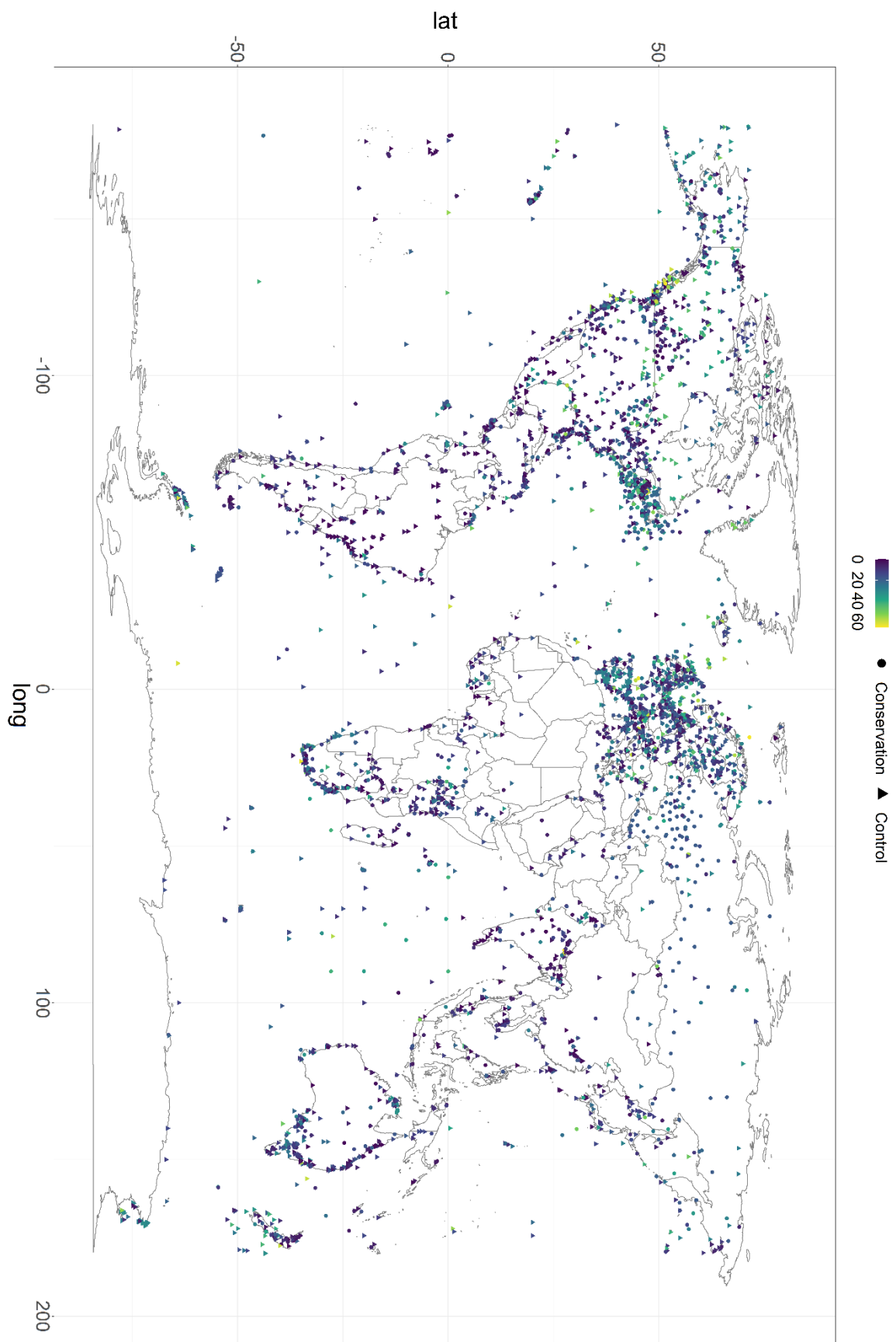


Figure S4.2 Length of each managed population.

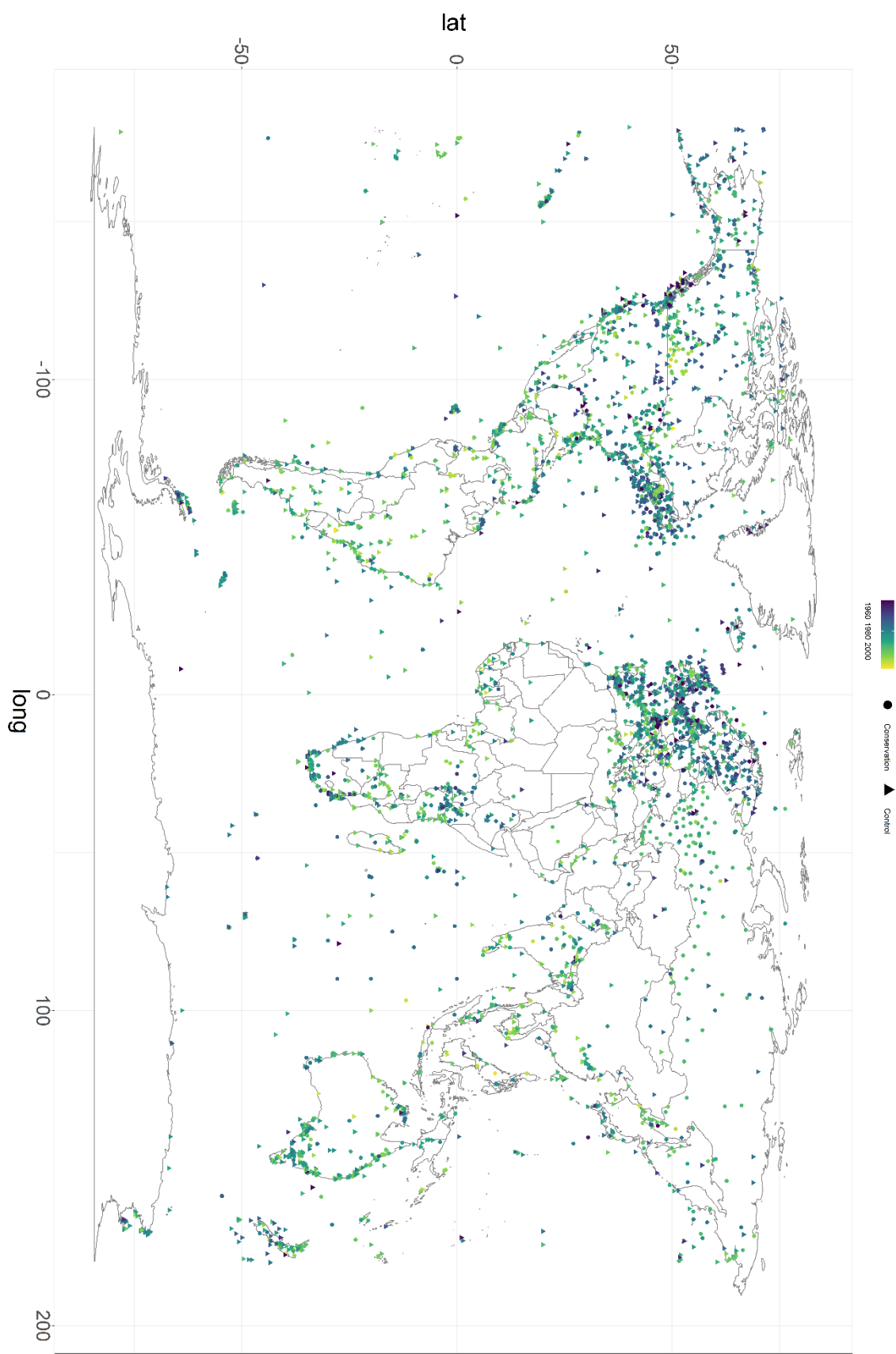
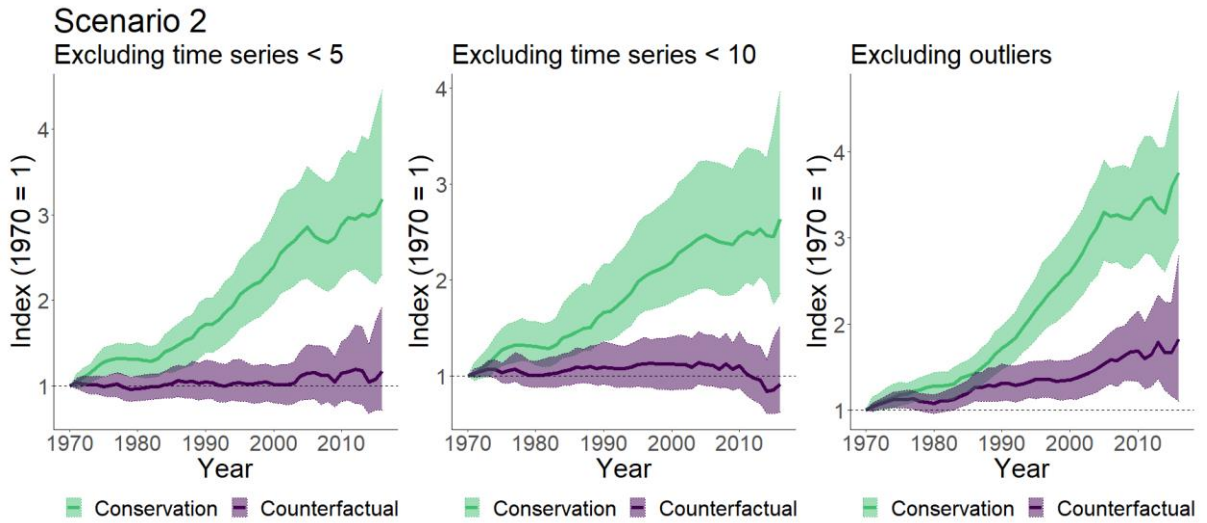
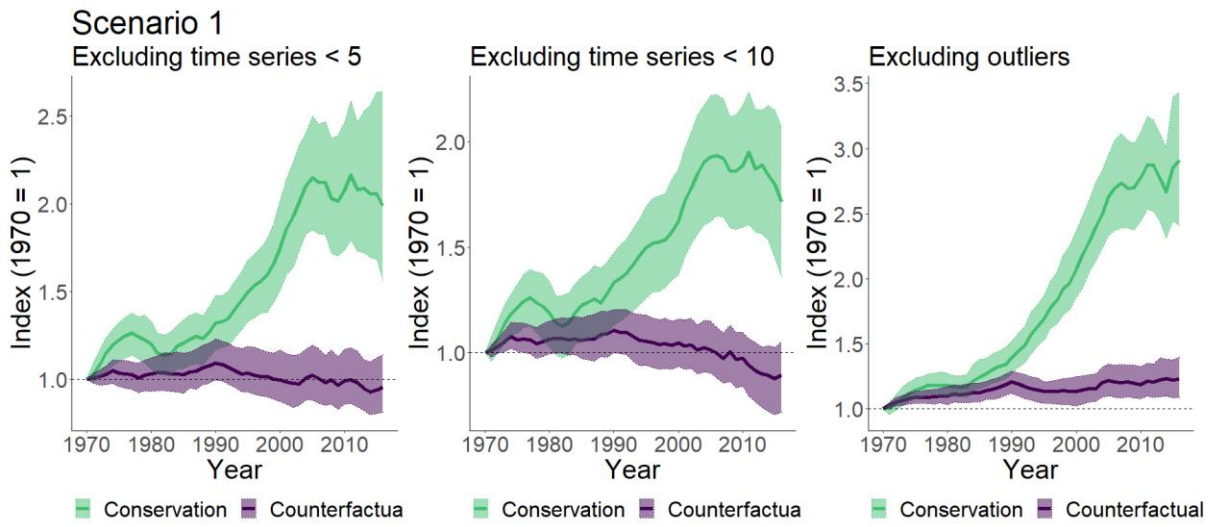


Figure S4.3 Starting year of each managed population.



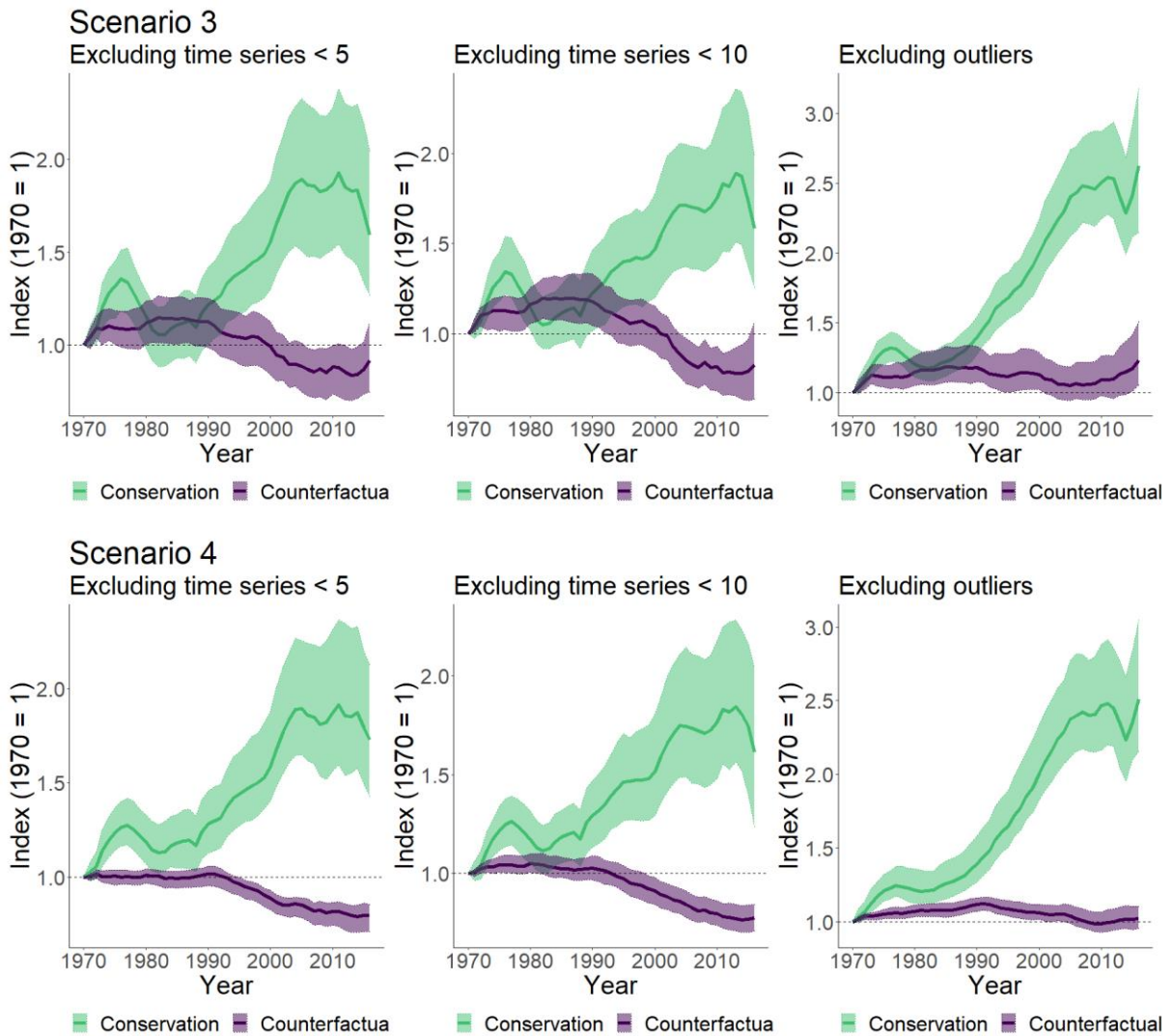


Figure S4.4 Sensitivity test for the four scenarios. Trends were calculated excluding populations covering less than five and ten years and excluding species with populations in the top and bottom 1% annual population change quantiles.

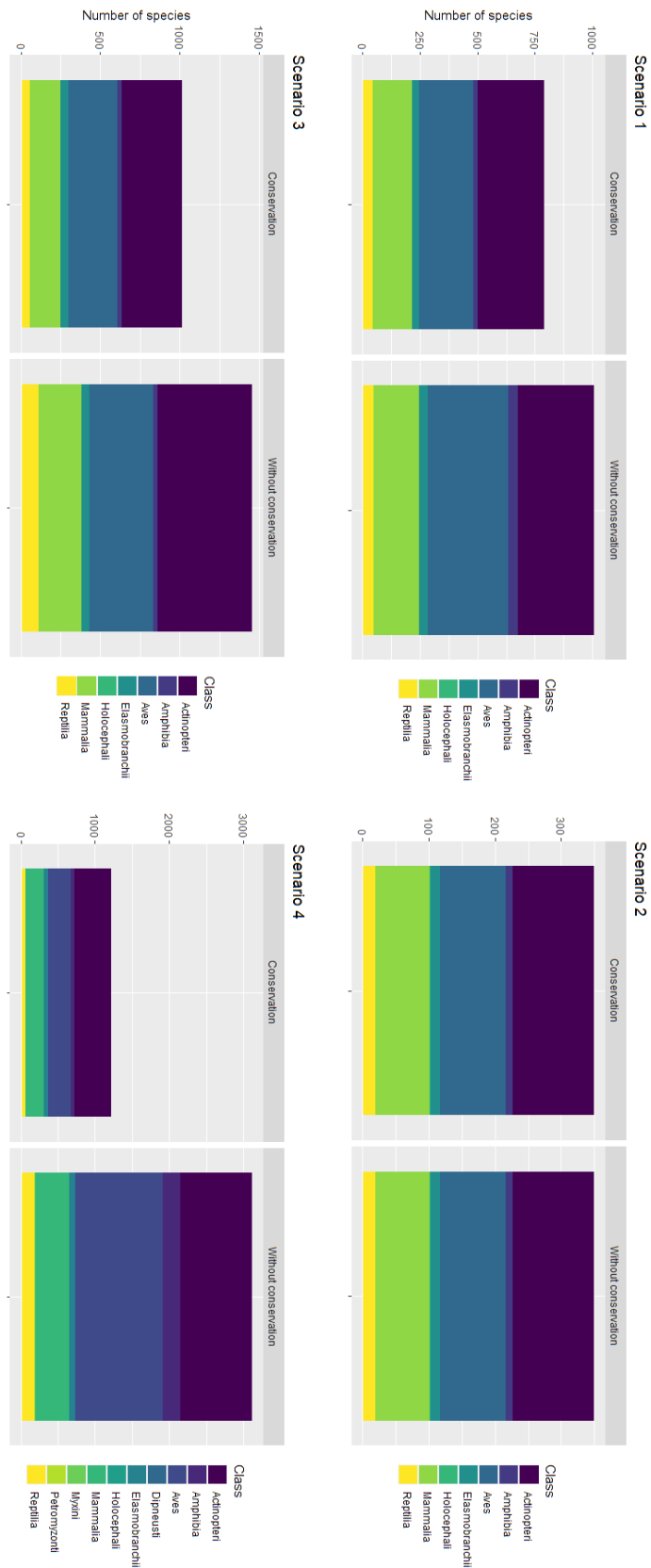


Figure S4.5 Number of species within each taxonomical class for each scenario.

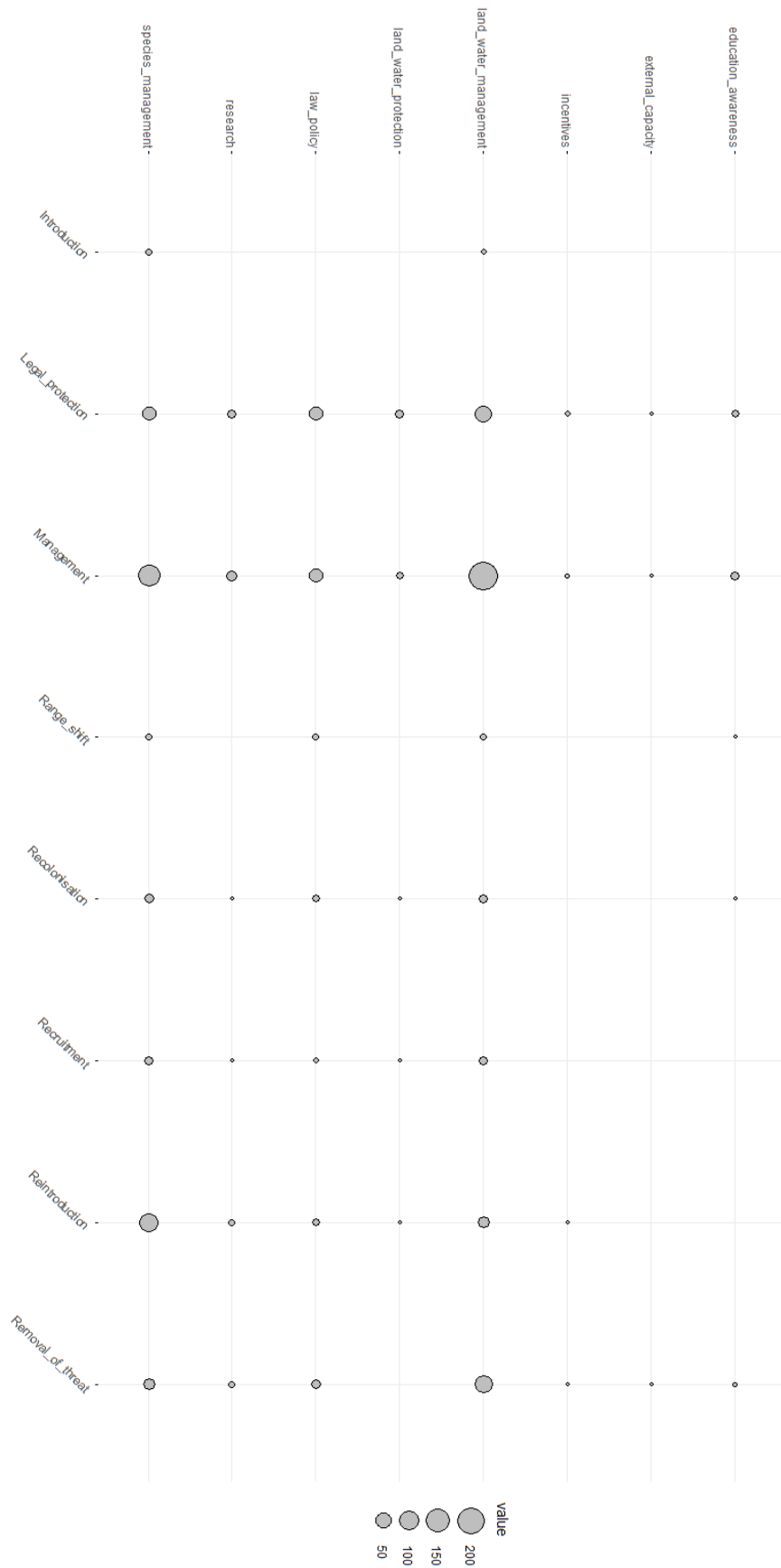


Figure S4.6 Combination frequencies for conservation actions (y-axis) and reasons for increase listed in the LPD (x-axis).

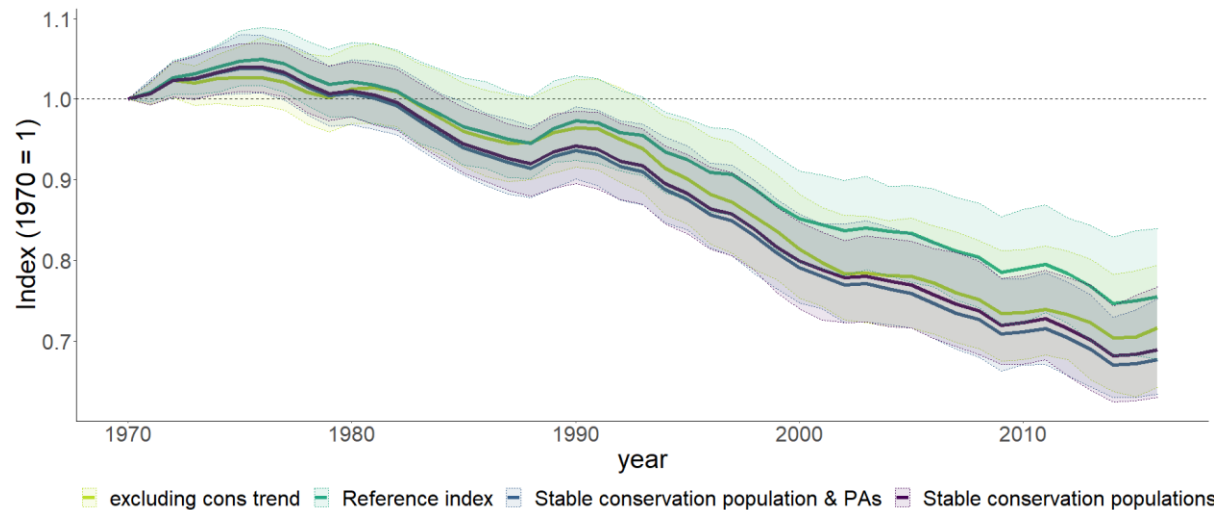


Figure S4.7 A reference global vertebrate population index (top dark green line: -24%; $nspp = 4,622$, $npop = 26,871$): next, a similar index but excluding all conservation targeted populations (second green line from the top: -28%; $nspp = 4226$, $npop = 21628$); population index assuming stable populations for conservation targeted populations with conservation as the reason for increase (second purple line from the bottom: -31%; $nspp = 286$, $npop = 519$): finally, a population index assuming stable populations for conservation targeted populations and populations inside PAs with conservation given as the reason for increase (blue bottom: -32%; $nspp = 329$, $npop = 600$). Note that, the number of species and populations listed for the two last indices refers to the number of species and populations assumed stable in the absence of conservation. The number of species and populations used to create the indices are similar to the full reference index.

5 | Discussion

Overview

The research reported in this thesis has measured the impact of conservation on species' populations using records of vertebrate abundance. By combining abundance records with ancillary habitat and conservation data, each chapter has focused on estimating the impact of conservation at a distinct spatial scale – national in chapter two, local in chapter three, and global in chapter four. These findings demonstrate that abundance records, such as those collected from dedicated volunteers in national monitoring schemes, allow retrospective evaluation of past conservation efforts, and underpin the great importance of consistent and widespread citizen science monitoring. Besides methodological advances, this research adds new information to the evidence base of conservation effectiveness, and thus improves our knowledge of conservation impacts for large scale actions in the conservation literature. At the core of this thesis is the application of a counterfactual approach, enabling comparisons of observed outcomes to likely outcomes in the absence of conservation, and thus allowing causal attribution of observed differences to an intervention of interest. The degree to which a counterfactual approximates the truth depends on its specification, the data available and knowledge of the topic being studied. Therefore, counterfactuals developed from inadequate data or with limited knowledge of a study system increase the probability of an erroneous counterfactual and thus a biased effect estimate.

In chapter two, I measured the impact of lowland wet grassland reserves managed by the Royal Society for the Protection of Birds on breeding trends of priority wading birds,

and calculated counterfactual trends from annual breeding bird counts derived from the Breeding Bird Survey (Harris et al. 2019). This analysis demonstrated that lowland wet grassland reserves have benefitted populations of target breeding waders. In chapter three, I disentangled the drivers of breeding abundance within such reserves to provide information that could be used to guide and fine-tune local reserve management. I collated 25 years of recorded management data for the lowland wet grassland reserves evaluated in the second chapter, and combined these data with local climatic information, to estimate the association between site-specific conservation actions and annual abundance of breeding waders. This chapter showed that conservation actions such as intensive predator control are associated with higher breeding abundance for two of the more abundant species of breeding waders. In the fourth chapter, I categorised conservation actions for vertebrate populations recorded in the Living Planet Index Database to estimate the potential global impact of conservation on species' population trends and on a global index of population abundance. Additionally, I estimated the association between primary conservation intervention categories and vertebrate population trends. This research suggests that conservation has benefitted global abundance of vertebrate populations demonstrably and cements the role and importance of conservation actions of different kinds moving forward.

In this final chapter, I summarize and reflect on the findings of chapters two to four, where I developed and applied methods to measure the impact of conservation on species' populations. I discuss how this research has advanced conservation science and its implications for conservation science, conservation practitioners and policy makers. Finally, I discuss the limitations of the research in this thesis, and the future of impact evaluation in conservation science.

Advances resulting from this work

Effective conservation depends on the availability of reliable and relevant evidence (Christie et al. 2020b). However, rigorous evaluations that demonstrate the impact of large-scale conservation, such as area protection or legislative changes, are often constrained by limited resources or legal obligations to manage habitats or species in a certain way (Margoluis et al. 2009; Baylis et al. 2016; Jellesmark et al. 2021). In chapter two and four, I demonstrated a method for measuring large-scale impact of conservation using longitudinal population data. I showed that evaluation of conservation interventions, such as area protection, is possible using observational data derived from monitoring programmes, and how to emulate more robust study designs post-hoc with statistical balancing methods such as matching. Applying these methods in chapter two showed that breeding populations of lapwing, redshank, snipe and curlew were better off than they would have been without conservation. Similar evaluation methods as those applied here can easily be used to evaluate the impact of other conservation projects, especially as more open access data become available. This applies to interventions on different spatial scales. For example, to evaluate the impact of a local meadow restoration on butterfly abundance, counterfactual population abundance could be calculated from site level records from a butterfly monitoring scheme, such as the United Kingdom Butterfly Monitoring Scheme (UKBMS), based on site-specific covariates. National evaluations could follow a similar approach but aggregating site level counts into general trends, as demonstrated in the second and fourth chapter. Evaluating studies that would otherwise not have been assessed will hopefully help guide conservation investments and ensure that limited resources make as much impact as possible on biodiversity outcomes.

The findings from the second chapter serve as a strong argument for the substantial annual investments directed towards managing nature reserves. In the light of resource scarcity and an increased demand for accountability from financial donors, documenting the effectiveness of conservation management is important (Jeffries et al. 2019; Stephenson et al. 2020; Grace et al. 2021a), especially for organisations that rely heavily on donations from external sources such as the broader public.

By creating counterfactual breeding trends that represent population development in the absence of conservation, the second chapter demonstrated that lowland wet grassland reserves in the United Kingdom greatly benefit breeding waders. However, management of these reserves involves a broad array of conservation actions. For example, hydrological management is improved through the implementation of water control structures, while predator pressure is regulated through predator exclusion and lethal control. Reserves are subjected to different management regimes, and the exact conservation mechanisms driving abundance was therefore still unknown until addressed in the third chapter; though previous studies had looked at individual conservation actions in isolation. We used 25 years of management data along annual climatic records and site-specific information to estimate the drivers of breeding abundance in a Bayesian framework. The results showed that the drivers of breeding abundance differed between the four species evaluated, reflecting their ecologies and their interplay with management actions. The two most abundant species, lapwing and redshank, benefitted from the same types of interventions, namely intensive predator control, whereas the breeding abundance for curlew and snipe did not associate strongly with any of the interventions evaluated. By incorporating a non-linear temporal effect through a second order random walk function, I tried to incorporate the effect of site development in the years following reserve creation. This showed a general

pattern of higher breeding abundance as a site matures, particularly for sites on former arable land. These results show that monitoring data, even if not originally intended for evaluation, can hold great value if utilized accordingly, and provide insights that further advances conservation. Additionally, it expanded on previous studies by using data from multiple sites across the United Kingdom, therefore providing more general estimates of the factors that associate with breeding abundance inside lowland wet grassland reserves.

The successful use of abundance records to create counterfactual reserve trends in the second chapter, combined with the collation of management data from annual reports to estimate the impact of individual conservation actions, inspired the analysis in the fourth chapter. Here, I expanded on these methods to categorize broadly conservation actions captured within the Living Planet Index Database and measure their impact across more than 26,000 vertebrate populations. This described the conservation actions specifically targeting species globally, showed that conservation appears to have delivered substantial benefits to recipient populations, and that global trends of vertebrate abundance would have decreased even further in the absence of conservation. Additionally, this analysis provided new information on the extent to which different broad classes of conservation actions affect population trends.

Other studies have assessed the impact of conservation on vertebrate species (Hoffmann et al. 2010, 2015; Bolam et al. 2020; Grace et al. 2021a). However, the use of longitudinal population data, the combination of abundance records with ancillary management and conservation information, and the depiction of impact using counterfactual trends, demonstrate different approaches to measure the impact of conservation, and provide new information to the evidence base of conservation effectiveness.

This thesis reiterates the value of well-designed general monitoring schemes. For instance, the Breeding Bird Survey and the Living Planet Database all rely on wider data, often collected by volunteers. Had it not been for the tremendous efforts of highly skilled citizen scientists that volunteer on a frequent basis, none of the research in this thesis would have been possible. The importance of monitoring schemes and volunteers extends far beyond this thesis, providing valuable information about global biodiversity demonstrated by indicators such as the Living Planet Index.

Protected areas are at the very core of modern conservation practice and will likely remain so. For example, an area-based target of 30% protected area coverage is suggested by 2030 (Waldron et al. 2020). However, area protection alone does not secure protection of biodiversity (Visconti et al. 2019; Wauchope et al. 2022). To secure effective conservation, area-based targets must be supported by alternative targets, such as population improvements likely caused by conservation, as shown in this thesis.

Study limitations and future research

In this thesis, I have demonstrated methods to measure the impact of conservation, applied these methods to estimate the impact of conservation on breeding waders in the United Kingdom and global vertebrate population abundance, and highlighted the potential value of population data for retrospective evaluations. However, it is important to remember that these evaluations have, at best, been quasi-experimental and can therefore still be biased by unobserved covariates. Additionally, the available data did not allow the application of study designs that also account for pre-intervention conditions, such as the Before-After-Control-Impact design.

Drawing causal inference from observational data comes with certain caveats and can be subject to biases from selection, confounding and measurement. In this thesis, statistical matching was used to reduce such potential biases. However, limited data still restricts the degree to which these estimates approximate true causal effects. For example, a more detailed set of covariates and pre-intervention breeding counts in Chapter 2 would have greatly improved the estimates of lowland wet grassland reserves' impact. In Chapter 3, several conservation actions were tested using regression-based analysis without any causal theory per se, therefore limiting any causal interpretation. In the fourth chapter, global population data were used. Such data are currently extremely coarse and are not intended for purposes of causal inference. However, important associations can still be inferred and thus pave the way for future analysis and guidance of data collection, in order to allow and improve future evaluation efforts. Developments on how to measure conservation impact with time series data, such as that in this study, are receiving attention in the conservation community, with great examples using waterbird abundance data (Wauchope et al. 2019a, 2021). However, the technical aspects of analysis can only be improved so much, and thus cannot surpass the intrinsic limitations imposed by the nature and quality of data.

For example, estimates of conservation impact derived from large samples of experimental data are less likely to be biased than estimates from observational data. While randomly sampled experimental data allows impact estimates unaffected by the shortcomings associated with those derived from non-experimental data, such assessments are often not feasible. Additionally, retrospective evaluations are restricted in regard to sampling method and the units recorded, while differences between sampling methods and protocols can render records non-comparable

between datasets. This can limit how such data can be utilized for alternative analysis. For the research reported here, pre-intervention abundance records were not available for the vast majority of cases, thus excluding the possibility of applying more rigorous study designs. As for the development of impact evaluation using time series data, the importance of study design is receiving increased attention (Christie et al. 2019, 2020c; Christie 2021). However, advances in the analysis of time series data for impact evaluation, and improved knowledge about the bias related to different study designs, are futile without similar developments in the accessible data. Initiatives such as EuropaBON (<https://europabon.org>) are working to overcome shortcomings in the existing data by designing an EU-wide framework for monitoring of biodiversity and ecosystem services. Identifying current gaps in monitoring efforts and further improving on these will result in improved monitoring data and allow better evaluations similar to those in this thesis, that can support improved local, national and global decision making for biodiversity and people.

References

Baylis K, Honey-Rosés J, Börner J, Corbera E, Ezzine-de-Blas D, Ferraro PJ, Lapeyre R, Persson UM, Pfaff A, Wunder S. 2016. Mainstreaming Impact Evaluation in Nature Conservation. *Conservation Letters* 9:58–64.

Bolam FC et al. 2020. How many bird and mammal extinctions has recent conservation action prevented? *Conservation Letters*:e12762. Available from <http://doi.wiley.com/10.1111/conl.12762>.

Christie AP. 2021. Determining the biases and consistencies in the evidence for conservation. Available from <https://www.repository.cam.ac.uk/handle/1810/321463?show=full>.

Christie AP, Amano T, Martin PA, Petrovan SO, Shackelford GE, Simmons BI, Smith RK, Williams DR, Wordley CFR, Sutherland WJ. 2020a. Poor availability of context-specific evidence hampers decision-making in conservation. *Biological Conservation* 248:108666. Elsevier. Available from <https://doi.org/10.1016/j.biocon.2020.108666>.

Christie AP, Amano T, Martin PA, Petrovan SO, Shackelford GE, Simmons BI, Smith RK, Williams DR, Wordley CFR, Sutherland WJ. 2020b. The challenge of biased evidence in conservation. *Conservation Biology*. Available from <https://onlinelibrary.wiley.com/doi/abs/10.1111/cobi.13577>.

Christie AP, Amano T, Martin PA, Shackelford GE, Simmons BI, Sutherland WJ. 2019. Simple study designs in ecology produce inaccurate estimates of biodiversity responses. *Journal of Applied Ecology*:2742–2754.

Grace MK et al. 2021. Building robust, practicable counterfactuals and scenarios to evaluate the impact of species conservation interventions using inferential approaches. *Biological Conservation* 261:109259. Elsevier Ltd. Available from <https://doi.org/10.1016/j.biocon.2021.109259>.

Harris SJ, Massimino D, Eaton MA, Gillings S, Noble DG, Balmer DE, Pearce-Higgins JW, Woodcock P. 2019. The Breeding Bird Survey 2018. Page BTO Research Report 717. Available from www.bto.org.

Hoffmann M et al. 2010. The impact of conservation on the status of the world's vertebrates. *Science* 330:1503–1509.

Hoffmann M, Duckworth JW, Holmes K, Mallon DP, Rodrigues ASL, Stuart SN. 2015. The difference conservation makes to extinction risk of the world's ungulates. *Conservation Biology* 29:1303–1313.

Jeffries G, Withers O, Barichievy C, Gordon C. 2019. The rhino impact investment project—a new, outcomes-based finance mechanism for selected AfRSG-rated ‘key’ black rhino populations. *Pachyderm* 2019:88–95.

Jellesmark S, Ausden M, Blackburn TM, Gregory RD, Hoffmann M, Massimino D, McRae L, Visconti P. 2021. A counterfactual approach to measure the impact of wet grassland conservation on UK breeding bird populations. *Conservation Biology*: Available from <https://onlinelibrary.wiley.com/doi/10.1111/cobi.13692>.

Margoluis R, Stem C, Salafsky N, Brown M. 2009. Design alternatives for evaluating the impact of conservation projects. *New Directions for Evaluation* 2009:85–96. Available from <http://doi.wiley.com/10.1002/ev.298>.

Stephenson PJ, Workman C, Grace MK, Long B. 2020. Testing the IUCN Green List of Species. *Oryx* 54:10–11.

Waldron A et al. 2020. Protecting 30% of the planet for nature: costs, benefits and economic implications (Campaign for Nature). Available from https://www.conservation.cam.ac.uk/files/waldron_report_30_by_30_publish.pdf.

Wauchope H et al. 2019. Quantifying the impact of protected areas on near-global waterbird population trends, a pre-analysis plan:1–17. Available from <https://doi.org/10.7287/peerj.preprints.27741v1>.

Wauchope HS, Amano T, Geldmann J, Johnston A, Simmons BI, Sutherland WJ, Jones JPG. 2021. Evaluating Impact Using Time-Series Data. *Trends in Ecology & Evolution* 36:196–205. Available from <https://doi.org/10.1016/j.tree.2020.11.001>.

Wauchope HS et al. 2022. Protected areas have a mixed impact on waterbirds, but management helps. *Nature* 605:103–107. Springer US.