

# Reductions in global biodiversity loss predicted from conservation spending

Anthony Waldron<sup>1,2</sup>, Daniel C. Miller<sup>2</sup>, Dave Redding<sup>3</sup>, Arne Mooers<sup>4</sup>, Tyler S. Kuhn<sup>5</sup>, Nate Nibbelink<sup>6</sup>, J. Timmons Roberts<sup>7</sup>, Joseph A. Tobias<sup>1,8</sup> & John L. Gittleman<sup>9</sup>

**Halting global biodiversity loss is central to the Convention on Biological Diversity and United Nations Sustainable Development Goals<sup>1,2</sup>, but success to date has been very limited<sup>3–5</sup>. A critical determinant of success in achieving these goals is the financing that is committed to maintaining biodiversity<sup>6–9</sup>; however, financing decisions are hindered by considerable uncertainty over the likely impact of any conservation investment<sup>6–9</sup>. For greater effectiveness, we need an evidence-based model<sup>10–12</sup> that shows how conservation spending quantitatively reduces the rate of biodiversity loss. Here we demonstrate such a model, and empirically quantify how conservation investment between 1996 and 2008 reduced biodiversity loss in 109 countries (signatories to the Convention on Biological Diversity and Sustainable Development Goals), by a median average of 29% per country. We also show that biodiversity changes in signatory countries can be predicted with high accuracy, using a dual model that balances the effects of conservation investment against those of economic, agricultural and population growth (human development pressures)<sup>13–18</sup>. Decision-makers can use this model to forecast the improvement that any proposed biodiversity budget would achieve under various scenarios of human development pressure, and then compare these forecasts to any chosen policy target. We find that the impact of spending decreases as human development pressures grow, which implies that funding may need to increase over time. The model offers a flexible tool for balancing the Sustainable Development Goals of human development and maintaining biodiversity, by predicting the dynamic changes in conservation finance that will be needed as human development proceeds.**

The rapid loss of global biodiversity has major consequences for human wellbeing<sup>5,19</sup> and consequently, governments worldwide have committed to international agreements aimed at reducing these losses; these include the Convention on Biological Diversity (CBD) and United Nations Sustainable Development Goals (SDGs)<sup>1,2</sup>. However, outcomes to date have been poor: the 2010 CBD target was not achieved and it also seems probable that the 2020 Aichi biodiversity targets will not be accomplished<sup>3,4</sup>. One of the most important determinants of policy success is our ability to correctly identify (and secure) the level of financing needed to adequately fund biodiversity-conservation strategies (as outlined in Aichi target 20 and SDG 17)<sup>1,2,6–8</sup>. A second way to substantially improve outcomes is to adopt an evidence-based approach, in which decision making is guided by reliable evaluations of past successes and failures ('conservation impact assessments')<sup>10–12</sup>. In many fields, the financing of strategic goals is already evidence-based, analysing previous spending outcomes to guide current budget decisions<sup>20,21</sup>. Surprisingly, however, no study has yet tested whether global conservation investment has

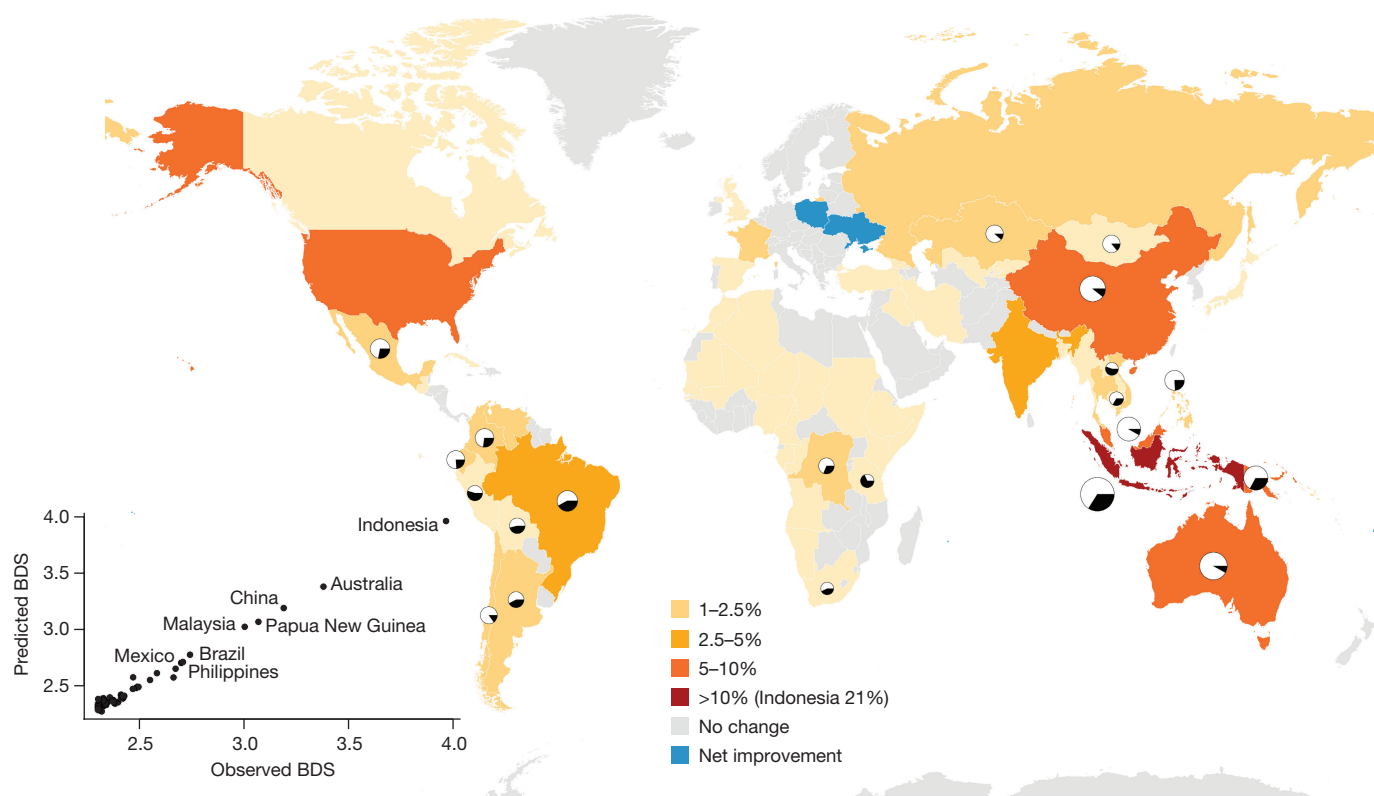
actually reduced biodiversity decline across CBD signatory countries, nor quantified the differential impacts of different funding levels.

A second key policy need is for models that reliably predict biodiversity decline, so that future losses can be forecast and timely action taken<sup>15,22</sup> (as already occurs with climate change<sup>23</sup>). In bio-political science, predictive models typically quantify the ways in which biodiversity loss is driven by human socioeconomic pressures, such as economic or agricultural expansion<sup>14–16,24</sup>. To date, evidence-based impact assessments and predictive decline models have largely been developed separately. However, the ultimate rate of biodiversity change depends on the simultaneous effects of human pressures and conservation impacts<sup>3,25</sup>. To make accurate predictions, we therefore need unified models that combine both aspects: 'pressures-and-conservation-impact' (PACI) models. Indeed, one of the core challenges for the SDGs is to balance the often-conflicting goals of human development (for example, SDGs 1, 2 and 8) and biodiversity conservation (SDG 15)<sup>2,14–18,24</sup>; models for policy use therefore need to include both. Finally, such models need to apply at the level of sovereign countries, the key geopolitical decision-making scale for the CBD and SDGs. This demands a finer geographic resolution than the planet-scale approaches commonly used in assessing global biodiversity change<sup>3,7</sup>.

Here, we develop a unified PACI model, at the sovereign country scale, that statistically quantifies how change in human pressures drives biodiversity decline and conservation spending reduces it. The model therefore informs policymakers not only about expected biodiversity losses but also, more constructively, about the ways in which changes in conservation resourcing can reduce the otherwise-expected losses<sup>3</sup>. We also show how the effects of spending and pressures predictably depend on national socioeconomic contexts, and thus how they may change over time.

A standard policy measure of biodiversity change is the planet-scale sum of all changes in the IUCN (International Union for Conservation of Nature) Red List status of each individual species, using well-known taxa as a proxy for biodiversity<sup>3,26</sup>. To calculate biodiversity change at the scale of sovereign signatory countries (the 'biodiversity decline score' (BDS) for each country), we took Red List status changes for all bird and mammal species for the period 1996–2008 (see Methods for details) and portioned them out among all the countries in which each species is found (treating the few status improvements as negative fractions). We then summed all the decline fractions for each country to calculate its BDS<sup>8,26</sup> (Fig. 1, Supplementary Table 1). Sixty per cent of the total global BDS was ascribable to only seven countries: Indonesia, Malaysia, Papua New Guinea, China, India, Australia and the USA (principally Hawaii). Seven countries had net biodiversity improvements (a negative BDS): Mauritius, Seychelles, Fiji, Samoa, Tonga, Poland and Ukraine (see Extended Data Fig. 1 for mean BDS per species).

<sup>1</sup>Edward Grey Institute, Department of Zoology, Oxford University, Oxford OX1 3PS, UK. <sup>2</sup>Department of Natural Resources and Environmental Sciences, University of Illinois, Urbana-Champaign, Illinois 61801, USA. <sup>3</sup>Department of Genetics, Evolution and Environment, University College London, London WC1E 6BT, UK. <sup>4</sup>Biology Department, Simon Fraser University, Burnaby, British Columbia V5A 1S6, Canada. <sup>5</sup>Scimitar Scientific, Whitehorse, Yukon Y1A 6V6, Canada. <sup>6</sup>Warnell School of Forestry & Natural Resources, University of Georgia, Athens, Georgia 30602, USA. <sup>7</sup>Institute at Brown for Environment and Society, Brown University, Providence, Rhode Island 02912, USA. <sup>8</sup>Department of Life Sciences, Imperial College London, Silwood Park, Buckhurst Road, Ascot, Berkshire SL5 7PY, UK. <sup>9</sup>Odom School of Ecology, University of Georgia, Athens, Georgia 30602, USA.



**Figure 1 | Global biodiversity declines and the effects of conservation spending.** Colours show percentage of all global declines (total BDS) associated with each country. Pie charts show the predicted reduction in decline (in black) if spending had been \$5 million higher (for selected countries); pie size represents the square root of the BDS.

Inset shows predicted versus observed BDS (log-transformed) for the continuous model (see also Extended Data Fig. 4). Country outlines supplied by *esri\_dm* (<https://www.arcgis.com/home/item.html?id=d86e32ea12a64727b9e94d6f820123a2#overview>).

To be useful in policymaking, models of biodiversity change need to have simple generality and demonstrated forecasting accuracy. Therefore, we first built PACI regression models to predict known BDSs, using national-level data on strict-sense conservation spending (annualized, see Methods) combined with the broad socioeconomic pressures of growth in the gross domestic product (GDP), agricultural expansion (and its relationship to forest loss), human population growth and changing governance quality (Extended Data Table 1, Supplementary Table 2). We then tested forecasting accuracy by using cross-validation, which repeatedly presents the model with data it has not seen and asks it to predict a known outcome (see Methods). BDS data were continuous and zero-inflated as a result of the presence of multiple species-poor countries with no status changes, so we used two-part models<sup>27</sup> in which the ‘continuous’ part ( $n = 50$ ) models BDS after truncating the long tail of zeroes, and the ‘binomial’ part ( $n = 109$ ) models whether BDS is zero or non-zero across all countries. We tested for context dependence by fitting several hypothesized interactions (Methods, Extended Data Table 1).

In the best-fitting regression models (Table 1), we found that conservation spending strongly reduced BDS (Fig. 2) whereas GDP growth and agricultural expansion tended to increase it (Fig. 3). Although forest loss was often an important factor, the best-fitting predictive model favoured more generalized terms (Table 1, Supplementary Discussion). Interaction terms revealed several context-dependent nuances (see Supplementary Discussion). The effect of GDP growth decreased as baseline GDP decreased, becoming non-significant in the poorest countries (Fig. 3). Agricultural expansion had a deleterious effect in countries with relatively low percentages of land devoted to agriculture (such as Malaysia and Peru), but was not statistically significant in countries with mid-to-high percentages (such as Bangladesh) (Fig. 3). The binomial part of the model also suggested that the

consequences of agricultural expansion could be considerably reduced by improvements in the quality of national governance (Extended Data Fig. 2) and that the effect size for GDP growth increased as human population growth increased (that is, the combined effect of the two pressures was greater than the sum of its parts; see Table 1). Finally, conservation spending was more effective in low-income countries than it was in higher-income ones, and also had a greater impact when there was greater threatened species richness (Extended Data Fig. 3).

Both parts of the model accurately predicted historical declines ( $R^2 = 0.85$  in the continuous part; accuracy = 94% in the binomial part; Extended Data Fig. 4) and were robust to several sensitivity tests (Extended Data Table 2, Supplementary Results). They also had high forecasting accuracy in cross-validation (82% for the continuous part; 85% for the binomial part). Our PACI models therefore have immediate application to several major policy needs. They can predict not only future biodiversity declines<sup>15,22</sup> but also how changes to a key policy instrument (the high-level financial resourcing of biodiversity conservation) will quantifiably reduce the declines expected. To illustrate this feature, we used the model to predict the effect of spending an extra five million constant international dollars (\$5) in each country (increasing the overall global annual budget by 42%, Supplementary Table 3). Outcomes for all countries are shown in Supplementary Table 3; see also Fig. 1) but to give an example: the model predicted that this would lead to BDS reductions of 33% and 54%, respectively, in the biologically mega-diverse countries of Papua New Guinea and Peru. We also used the model to estimate how much biodiversity loss has been prevented by conservation financing since the Earth Summit<sup>8,28</sup>, finding that median loss per country was 29% less than would otherwise have occurred (see Methods).

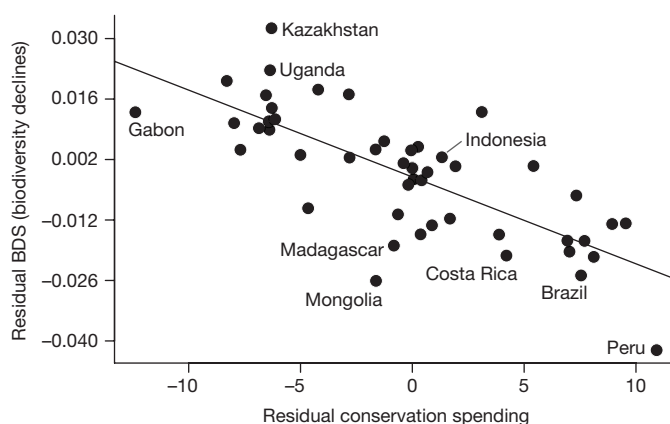
The model can also be used to predict the funding each country requires to achieve specific biodiversity policy goals, including the

**Table 1 | Best-fit models predicting biodiversity decline**

Predictor variable	Continuous part of the model (BDS)	Binomial part of the model (BDSb)
Conservation spending	−0.251	−4.800
Agricultural growth	−0.012	−3.065
GDP growth	0.035	−0.152
Population growth	NA	−2.738
Declines in period $t-1$	0.024	NA
Declines in period $t-2$	0.048	NA
Threatened species richness	0.155	5.421
Country area	NA	8.754
GDP	0.037	−5.426
Per cent agricultural land	0.049	−1.226
GDP growth $\times$ GDP	0.031	NA
Spending $\times$ GDP	NA	5.026
Spending $\times$ threatened species richness	−0.247	NA
Population growth $\times$ GDP growth	NA	1.044
Agricultural growth $\times$ per cent agricultural land	−0.045	−10.143
Spending $\times$ per cent agricultural land	0.065	NA
Agricultural growth $\times$ governance improvement	NA	−9.603

Standardized conditional coefficients are shown, reflecting effects at median values of any moderator variables; however, for all terms that interact, the interaction plots provided must be used to correctly interpret these effects (Fig. 3, Extended Data Figs 2, 3). Per cent agricultural land, mean percentage of agricultural land 1992–2003;  $t-1$ , 1994–2000;  $t-2$ , 1988–1994; GDP, GDP per capita at purchasing power parity; population, rural population density; governance improvement, change in the government effectiveness score.  $n = 50$  independent countries and index parameter is 1.01 in the continuous part of the model;  $n = 109$  countries in binomial part of the model, with a 42:67 ratio of ones to zeroes. NA, not applicable.

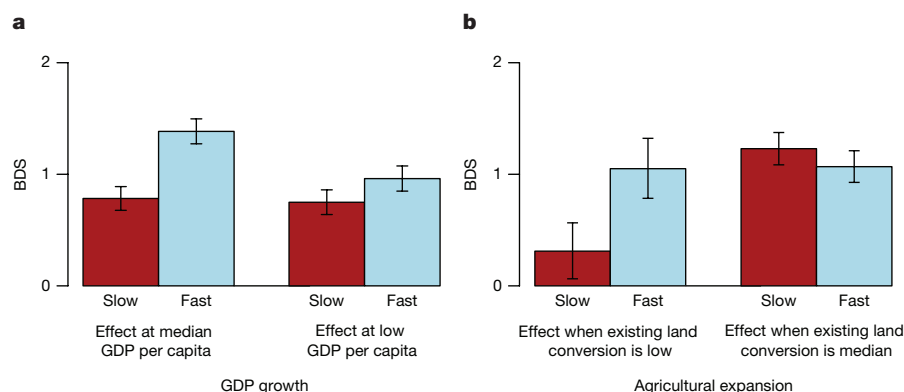
CBD and SDG targets. Saliently, our results demonstrate how the cost of meeting any target changes constantly as socioeconomic pressures change. For example, when socioeconomic pressures were at their historical 1992–2003 levels, Peru would have needed an extra \$4.6 million per year to achieve 50% less decline in biodiversity. However, if socioeconomic pressures are set to their current levels (2001–2012 mean), this figure rises to \$5.7 million. Our model explicitly accounts for such changes in socioeconomic context. Policy-makers could therefore use it to predict biodiversity outcomes at various funding levels for a range of scenarios of economic, agricultural and population change, and then compare these outcomes to any policy target, including those for the CBD and SDGs. In particular, the model can be used to help resolve problems of discordance between the various SDGs, by quantifying the ways in which any negative effects of economic and agricultural growth can be balanced by short-term increases in conservation funding (thereby allowing time for the development of more sustainable pathways to national growth<sup>18</sup>).



**Figure 2 | The country-scale rate of biodiversity decline (BDS) depends on conservation spending levels.** The continuous part of the model (which focuses on high-decline countries,  $n = 50$  independent countries) is shown and both variables are corrected for all other predictors in a residual–residual plot (Pearson's  $r = -0.69$ ). See Table 1 for spending effects in the binomial part of the model.

We caution that an unmeasured variable that is correlated with conservation spending may explain some of the effects of this spending; that the way in which national-level spending on birds and mammals might benefit other taxa remains unknown; that species declines that are too small to affect Red List status will not be accurately predicted and will require alternative approaches<sup>29</sup>; and that long-distance effects (such as Chinese demand for African ivory<sup>30</sup>) were beyond the scope of our model. However, our general approach should be flexible enough to accommodate these additions in the future.

At a time when the outlook for biodiversity often seems bleak<sup>4,5</sup>, our results present a constructive opportunity for shaping global biodiversity policy, by showing how increased conservation investment can lead to major and quantifiable improvements. Set against this note of optimism, however, our model also underlines the fact that conservation spending may need to constantly increase to counterbalance the continuing intensification of human development pressures<sup>5,18,24</sup>. By empirically demonstrating how limited levels of investment have reduced biodiversity loss part-way towards the CBD and SDG targets, our findings may ultimately encourage decision-makers to commit the greater amounts needed<sup>7</sup> to achieve a much larger reduction in global biodiversity loss, as was committed to in these international agreements<sup>1,2</sup>.



**Figure 3 | Conditional influence of human pressures on biodiversity.** **a**, The effect of GDP growth on BDS depends on the existing GDP per capita. Red, slow GDP growth (10th percentile); blue, fast growth (90th percentile). 'Low' GDP per capita, 10th percentile; 'median' GDP per capita, 50th percentile (effects are still significant above the 50th percentile). **b**, The effect of agricultural expansion on BDS depends on the

existing percentage of land converted to agriculture. Colours as in **a**; 'low' agricultural expansion, 10th percentile; 'median' agricultural expansion, 50th percentile (effects are still non-significant below the 50th percentile). Error bars show conditional 95% confidence intervals from the continuous part of the model. Centre is the median.  $n = 50$  independent countries.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 11 November 2016; accepted 21 September 2017.**

**Published online 25 October 2017.**

- Convention on Biological Diversity. Conference of the Parties Decision X/2: Strategic plan for biodiversity 2011–2020 [www.cbd.int/decision/cop?id=12268](http://www.cbd.int/decision/cop?id=12268) (2010).
- United Nations. Transforming our world: the 2030 agenda for sustainable development <https://sustainabledevelopment.un.org/post2015/transformingourworld/publication> (2015).
- Hoffmann, M. *et al.* The impact of conservation on the status of the world's vertebrates. *Science* **330**, 1503–1509 (2010).
- Tittensor, D. P. *et al.* A mid-term analysis of progress toward international biodiversity targets. *Science* **346**, 241–244 (2014).
- Pimm, S. L. *et al.* The biodiversity of species and their rates of extinction, distribution, and protection. *Science* **344**, 1246752 (2014).
- James, A. N., Gaston, K. J. & Balmford, A. Balancing the Earth's accounts. *Nature* **401**, 323–324 (1999).
- McCarthy, D. P. *et al.* Financial costs of meeting global biodiversity conservation targets: current spending and unmet needs. *Science* **338**, 946–949 (2012).
- Waldron, A. *et al.* Targeting global conservation funding to limit immediate biodiversity declines. *Proc. Natl Acad. Sci. USA* **110**, 12144–12148 (2013).
- United Nations Environment and Convention on Biological Diversity. Decisions adopted by the conference of the parties to the Convention on Biological Diversity at its 11th meeting <https://www.cbd.int/decisions/cop/?m=cop-11> (2013).
- Sutherland, W. J., Pullin, A. S., Dolman, P. M. & Knight, T. M. The need for evidence-based conservation. *Trends Ecol. Evol.* **19**, 305–308 (2004).
- McKinnon, M. C., Cheng, S. H., Garside, R., Masuda, Y. J. & Miller, D. C. Sustainability: map the evidence. *Nature* **528**, 185–187 (2015).
- Miteva, D. A., Pattanayak, S. K. & Ferraro, P. J. Evaluation of biodiversity policy instruments: what works and what doesn't? *Oxf. Rev. Econ. Policy* **28**, 69–92 (2012).
- Cardillo, M. *et al.* The predictability of extinction: biological and external correlates of decline in mammals. *Proc. R. Soc. Lond. B* **275**, 1441–1448 (2008).
- Naidoo, R. & Adamowicz, W. L. Effects of economic prosperity on numbers of threatened species. *Conserv. Biol.* **15**, 1021–1029 (2001).
- Freytag, A., Vietze, C. & Völkl, W. What drives biodiversity? An empirical assessment of the relation between biodiversity and the economy. *Int. J. Ecol. Econ. Stat.* **24**, 1–16 (2012).
- Smith, R. J., Muir, R. D., Walpole, M. J., Balmford, A. & Leader-Williams, N. Governance and the loss of biodiversity. *Nature* **426**, 67–70 (2003).
- Roe, D., Elliott, J., Sandbrook, C. & Walpole, M. *Biodiversity Conservation and Poverty Alleviation* (Wiley-Blackwell, 2013).
- Adams, W. M. *et al.* Biodiversity conservation and the eradication of poverty. *Science* **306**, 1146–1149 (2004).
- Cardinale, B. J. *et al.* Biodiversity loss and its impact on humanity. *Nature* **486**, 59–67 (2012).
- Schwartländer, B. *et al.* Towards an improved investment approach for an effective response to HIV/AIDS. *Lancet* **377**, 2031–2041 (2011).
- HM Treasury. Public sector business cases: using the five case model <https://www.gov.uk/government/publications/the-green-book-appraisal-and-evaluation-in-central-government> (2013).
- Collen, B. & Nicholson, E. Taking the measure of change. *Science* **346**, 166–167 (2014).
- Stern, N. *The Economics of Climate Change: the Stern Review* (Cambridge Univ. Press, 2007).
- Maxwell, S. L., Fuller, R. A., Brooks, T. M. & Watson, J. E. Biodiversity: the ravages of guns, nets and bulldozers. *Nature* **536**, 143–145 (2016).
- Ferraro, P. J. *et al.* Estimating the impacts of conservation on ecosystem services and poverty by integrating modeling and evaluation. *Proc. Natl Acad. Sci. USA* **112**, 7420–7425 (2015).
- Rodrigues, A. S. L. *et al.* Spatially explicit trends in the global conservation status of vertebrates. *PLoS One* **9**, e113934 (2014).
- Basu, A. & Manning, W. G. Issues for the next generation of health care cost analyses. *Med. Care* **47** (Suppl 1), S109–S114 (2009).
- Miller, D. C., Agrawal, A. & Timmons Roberts, J. Biodiversity, governance, and the allocation of international aid for conservation. *Conserv. Lett.* **6**, 12–20 (2013).
- Donald, P. F. *et al.* International conservation policy delivers benefits for birds in Europe. *Science* **317**, 810–813 (2007).
- Vandegrift, J. Elephant poaching: CITES failure to combat the growth in Chinese demand for ivory. *Virginia Environ. Law J.* **31**, 102 (2013).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** This research was supported by UKDWP (A.W.), the USDA National Institute of Food and Agriculture Hatch project 1009327 (D.C.M. and A.W.), the MacArthur Foundation through the Advancing Conservation in a Social Context research initiative (D.C.M. and J.T.R.), Natural Sciences and Engineering Research Council Canada Discovery and Accelerator Grants (A.M.), the UK Natural Environment Research Council grants NE/I028068/1 and NE/K016431/1 (J.A.T.) and the Odum School of Ecology (J.L.G.). We thank J. Drake, P. Holland and P. Stephens for comments on earlier manuscripts.

**Author Contributions** A.W. conceived the study and analysed the data, based on ideas from A.M., J.L.G. and D.C.M.; A.W., D.C.M., D.R., N.N. and J.T.R. collected the data; A.W., A.M., T.S.K., D.C.M., J.L.G. and J.A.T. wrote the paper with contributions from all authors.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to A.W. ([anthonywaldron@hotmail.com](mailto:anthonywaldron@hotmail.com)).

**Reviewer Information** Nature thanks A. Balmford, H. Possingham and the other anonymous reviewer(s) for their contribution to the peer review of this work.



## METHODS

**Country-scale biodiversity decline scores.** To quantify biodiversity decline, we used equally weighted changes in the IUCN Red List status of all global bird and mammal species up to the most recent Global Mammal Assessment in 2008 (changes in extinction risk between 1996 and 2008 for mammals, and between 2000 and 2008 for birds (there was no 1996 global assessment for birds)). We considered only 'genuine' changes in Red List status, excluding any changes that were not related to extinction risk (in particular, changes that were the result of taxonomic shifts)<sup>3,31,32</sup>. Our BDS approach is similar to the Red List indices of global biodiversity change adopted by governments to measure performance against CBD and SDG targets<sup>3,31–33</sup>, but adjusted to allow global declines to be portioned out among signatory countries while preserving the original magnitude of declines. We focused on birds and mammals because these taxa received the majority of conservation investment and supply robust and directly observed data on changes in Red List status<sup>3,34</sup>. We excluded amphibians (the only other taxon that could have been included in our model), because they received almost no conservation investment during the study period<sup>3</sup>; their declines for 1980–2004 are modelled rather than directly observed<sup>3,35</sup> (whereas robust spending data are only available from 1992 onwards<sup>8</sup>); and they are also highly data deficient and "enigmatic" in terms of their declines<sup>3,35</sup>.

To convert species-based Red List changes into country-level indices of biodiversity change, we divided up each status change into 'decline fractions', which corresponded to the percentage of the species range ( $p_{ij}$ ) that overlapped with each country<sup>8,26</sup>. These decline fractions are estimates for the underlying responsibility fraction ( $R_{ij}$ ), defined as the proportion of the status change for species  $i$  that is attributable to country  $j$  (see below, Additional method details). For greater accuracy, we corrected these range-based fractions in two ways. The losses underpinning a species decline are not homogeneously distributed in space but are instead frequently concentrated in a part of the range in which human pressures have suddenly increased<sup>36</sup>. Both empirically and at random, these concentrations of pressure-driven loss are unlikely to lie at the periphery of a range<sup>36</sup> (see Additional method details). However, a raw range-based algorithm assumes that losses are spatially homogeneous and so will often assign an erroneous and trivial responsibility fraction to any country that holds a small percentage of the species range (a 'range edge') of a species that is found almost entirely in a neighbouring country. Formally,  $R_{ij}$  for small  $p_{ij}$  is often (but not always) likely to be zero. Small  $p_{ij}$  values were also extremely numerous and generated very high noise-to-signal ratios. To address both the signal loss and the bias introduced as a result of an unknown proportion of small  $p_{ij}$  values being incorrect overestimates of zero  $R_{ij}$ , we used signal detection theory<sup>37</sup> and the numerical population changes required to trigger a change in Red List categories to estimate a range of theoretically optimal thresholds ( $T$ ), in such a way that  $R_{ij}$  is set to zero if  $p_{ij}$  is less than  $T$ . To account for uncertainty, we then carried out our analyses using three possible thresholds in this range (see Additional method details). Results in the main text are calculated for  $T = 0.17$ , the approximate optimal trade-off between noise reduction and sample size (see Additional method details). Supplementary Results and Extended Data Table 2 show sensitivity tests with alternative thresholds, including the finding that the explanatory power of the model at  $T = 0.17$  is considerably stronger than at the other thresholds.

We also analysed the Red List reports for each individual bird and mammal species and altered the decline fractions whenever a report suggested that the distribution of responsibilities across countries differed from the range-based estimate (Supplementary Table 4). We then calculated the BDS for each country by summing all decline fractions for birds and mammals, treating status improvements as negative fractions<sup>8,26</sup>. Supplementary Table 1 contains the final BDSs for each country.

**Predictors of country-scale biodiversity decline scores.** To be useful for policy formation, conservation models need to have broad general applicability and also be able to accurately forecast outcomes when presented with situations that differ from the original dataset on which they were parameterized. Using specific rather than general variables often leads to very poor forecasting and limited applicability if the model is used beyond the original data<sup>37</sup>. We therefore selected a relatively small set of simple, generalized and publicly available explanatory variables to represent socioeconomic pressures at a national level; conservation spending also captures overall conservation effort in a broad, quantifiable and publicly reported way.

For conservation spending, we took data on average annual conservation investment levels from a recently published collation<sup>8</sup>, adding new data for countries (such as Turkey<sup>38</sup>) that had been data-deficient in the original published study. Consistent with previous approaches<sup>8</sup>, finance data were collated at 2005 constant US dollar values but for analysis were converted to US\$ at local purchasing power parity (PPP) values, to account for the variation in the purchasing power of US dollars (when exchanged) in each country<sup>39</sup>. Two types of conservation

investment data were available: (i) 'strict-sense' funding that is directly linked to biodiversity conservation, and (ii) 'mixed funding' that mainly targets social and development goals but which potentially has indirect long-term (and often unclear) influences on biodiversity (for example, school building or agricultural assistance in forest communities)<sup>28,40</sup>. A priori, we hypothesized that strict-sense biodiversity funding was likely to be the better predictor of rates of decline, whereas using mixed funding (which involves much larger sums than strict-sense funding) could potentially obscure the effects of strict conservation spending. Strict-sense funding also produced lower corrected Akaike information criterion (AICc) scores in exploratory modelling and so we used it in our final analysis.

Good governance is also hypothesized to positively affect biodiversity, both directly (for example, by reducing conflict) and indirectly (for example, by making conservation investment more efficient)<sup>16,41–43</sup>. Governance has been measured using multiple indicators<sup>44</sup>; we modelled the effects of change in the six indicators published in the World Governance Indicators dataset<sup>44</sup>. These include government effectiveness, political stability and conflict, rule of law, corruption, regulatory quality (largely a measure of openness to business activity) and 'voice' (a measure of the democratic accountability of governments). All the governance indicators are very tightly correlated with one another ( $r > 0.9$  for all pairwise combinations), and to avoid collinearity we therefore tested them individually. Government effectiveness gave the best fit in exploratory analysis (as it has in previous analyses<sup>8</sup>) and is reported in the results as 'governance'.

For the country-level socioeconomic pressures aspect of our PACI model, we followed previous authors in using national rates of human population growth, economic growth and agricultural expansion<sup>13–15,17,45–53</sup>. These country-level aggregators are likely to capture the overall effects of multiple smaller-scale drivers (with agriculture being the main pressure that drives threats to biodiversity<sup>24</sup>). For example, forest clearance for food or commodity production would generally cause changes in the area of agricultural land and in economic output, and GDP levels have been associated with both hunting pressure and deforestation trends<sup>54–56</sup>. For economic growth, we used change in GDP per capita at PPP. For agricultural growth, we used change in the percentage of land converted to agriculture; for population growth, we used change in human population density (using total and rural population density as alternatives). Data on GDP, agricultural land and human populations were taken from World Bank statistical tables<sup>57</sup>. We also tested the direct impacts of forest loss, estimated per country for the period 1990–2000 using statistics from the Food and Agriculture Organization of the United Nations<sup>58,59</sup> (acknowledging the limitations of this historical dataset<sup>60</sup>).

The number of declining species in a country (and therefore its BDS) is likely to be strongly influenced by the total number of species present and/or the area of the country, and by levels of pre-existing risk and decline. We calculated total threatened species richness in the same way as we calculated total species decline (BDS); we summed all species fractions in each country, weighting them by the level of extinction risk as an index of threat<sup>61</sup>. We compiled country areas from ref. 8. In exploratory analysis, however, we found that the inclusion of area in the continuous part of any of the models consistently led to a worse fit ( $\Delta\text{AICc} > 6.5$ ), probably because species richness absorbed most of the variance explained by area in this sample ( $n = 50$  countries). By contrast, the binomial part ( $n = 109$  countries) of each of the models detected separate area and species richness effects (without collinearity; see Extended Data Tables 3, 4). We therefore included the area term in the binomial part models, but excluded it from our final set of continuous part models. We note, however, that parameter estimates with and without area were extremely similar.

**Lags between predictors and responses.** Conservation investment and/or action takes at least five years, and often over a decade, to have an effect on biodiversity<sup>29,62</sup>, especially for taxa such as birds and mammals. For mammals, the two global Red List assessments from which status changes can be calculated occurred in 1996 and 2008<sup>3</sup>. We therefore assumed that status changes detected in the 2008 assessment may have been driven by conservation finance allocations occurring as recently as five years earlier (2003), but would also have been influenced by spending occurring a decade or more earlier (in the early 1990s). Similarly, when the Rio Earth Summit<sup>28</sup> in 1992 led to some of the first major global commitments to conservation spending, the impact would have been felt in the years following the 1996 Global Mammal Assessment and have continued for more than a decade. Following this logic, we used predictor variables for 1992–2003 (annualized values) to model changes in the response value for 1996–2008, using the same lag for the four socioeconomic growth variables to avoid the analysis becoming intractable. We tested an alternative predictor period of 1992–2000 but preferred 1992–2003 on the basis of the lower AICc values associated with the latter period.

Our response variable is therefore a lagged variable<sup>63</sup> taking the form  $Y_t - Y_{t-n}$  ( $Y$  at time  $t$  is the aggregate national Red List status); our variables for socioeconomic change are similarly lagged. Although predicting change occurring in a

time block by using variables from an earlier time block is necessarily approximate, year-by-year species changes were not available. Nevertheless, country-level patterns of change in predictor variables were strongly correlated across time periods; for example, when comparing mean annual values for 1992–2000 and 1992–2003, the correlations for population growth, population size, GDP growth and GDP are 0.91, 0.999, 0.89, and 0.999, respectively. These strong correlations imply that the precise choice of year or period is unlikely to have had an important effect on the results.

The rate of decline over a fixed period is also likely to be influenced by the 'inertia' from declines in the years immediately preceding that period. To explore this effect, we calculated avian BDSs for the two IUCN assessment periods that preceded our study period (1988–1994 and 1994–2000) and added both measures to our candidate regression models. No earlier-period BDSs were available for mammals. However, as mammal and bird BDSs are highly correlated in our study period (Pearson's  $r = 0.998$ ), we assumed that earlier-period bird BDSs were a reasonable proxy for combined bird and mammal BDSs in the earlier period.

**Statistical analysis.** All predictor variables were  $z$  standardized to place effect sizes on a common scale<sup>64</sup>. We excluded any countries for which complete and robust data were lacking<sup>8</sup>, including those for which reported finance commitments cannot be securely regarded as strict-sense biodiversity spending. For countries with multiple overseas territories, we excluded those countries in which conservation spending was not disaggregated by territory, as values for the socioeconomic predictors and rates of decline varied considerably across territories. In particular, the USA, France and the UK were excluded from regression models under this rubric; we therefore recommend greater geographical specificity in finance reporting. Supplementary Table 1 lists all of the exclusions. The Solomon Islands and New Zealand represented potentially influential leverage points, so we tested our models with and without these countries. We found that the Solomon Islands had a large influence on binomial outcomes (causing governance growth to be dropped from the best-fit binomial part of the model, probably as a result of the extreme value of governance growth for this country). We therefore excluded this country from all binomial models. The inclusion of the Solomon Islands had a less pronounced effect on the continuous part of the model; an identical best-fit model with similar coefficients was selected whether or not the country was included. For completeness, however, we tested all continuous model variations both with and without the Solomon Islands. The presence of New Zealand had a major effect on binomial-part outcomes, altering most coefficients by roughly 20% and some by over 100%; it also led to a greatly worsened fit in the continuous part. New Zealand was therefore excluded from both parts of the model. The leverage associated with New Zealand may be a consequence of this country having a negative value for agricultural growth.

We then built candidate PACI models to predict BDSs, each of which tested hypotheses regarding the ways in which conservation investment and various human pressures might have an effect on biodiversity (see Supplementary Table 2 for full list). We included several interactions to test whether socioeconomic context altered the impact of socioeconomic change. For example, we hypothesized that in countries that have already converted much of their land base to agriculture, additional expansion of farmland might have either a reduced marginal effect on biodiversity as a result of an extinction filter<sup>65</sup> or, alternatively, a greater effect as the last vestiges of habitat disappear (Supplementary Discussion). Thus, we calculated mean annual values of GDP, population, governance and percentage of agricultural land for 1992–2003 and added these to our interaction model specifications. Extended Data Table 1 and Supplementary Table 2 show all interactions tested.

The BDS data were non-integer values, covering both positive and negative values, with a relatively dense cloud of values at zero. Although a limited number of zeroes does not violate regression assumptions, a long tail of zeroes can generate extreme bias<sup>64</sup>. We therefore used the recommended approach of a two-part model<sup>27,66</sup> that creates (i) a continuous part comprising all countries ( $n = 50$ ) with a non-zero BDS plus those countries with informative BDSs of zero and (ii) a binomial part that included all countries ( $n = 109$ ) for which data were available, including all those with BDSs of zero. For the binomial part, we converted the BDS to a binary response (BDSb); BDSb was set at 1 if  $\text{BDS} > 0$  and otherwise was set at zero. For the continuous part specifically, we sought to optimize the trade-off between information content and bias by including as many zero-value BDSs as possible, in order of their probable informativeness, without causing clear patterns in regression diagnostic plots (thus extending the principle of hurdle models developed for non-negative integer data<sup>64</sup> to two-part analyses). For a country that has many species but has nevertheless experienced no declines (such as Costa Rica), a BDS of zero probably represents an important underlying process and, consequently, should be regarded as highly informative. Conversely, if a country is species-poor there is a strong random expectation that over a 13-year period no species will be observed to change its Red List status, leading to a BDS of zero that is much less informative. We therefore defined  $\Psi$  as country-level species richness

(summed species fractions when range-edge cases are excluded) and then, for various values of this parameter, heuristically tested the degree of regression bias arising when we excluded all cases in which  $\text{BDS} = 0$  and species richness  $< \Psi$ . We found a trade-off whereby setting  $\Psi$  at 40 or more left minimal patterns in residual plots but reduced sample size and statistical power, and setting  $\Psi$  values at 20 or lower started to generate strong patterns in plots of residuals against fitted values. We therefore chose a value of  $\Psi = 25$  (for sensitivity testing on this parameter, see below and Supplementary Results).

In both model parts, exploratory generalized additive models (GAMs) suggested that linear modelling was appropriate. For the continuous part, BDS retained a right skew even after log-transformation (Extended Data Fig. 5) and there was also heteroscedasticity in the errors, so we tested generalized linear models (GLMs) with the gamma-like Tweedie error distribution, which uses maximum likelihood to simultaneously model heteroscedastic variance as a function of the mean<sup>67–69</sup> (using the *cplm* R package<sup>70</sup>). We carried out an  $(x + 10)$  transformation on BDS to avoid violating gamma assumptions (where the value of 10 was chosen to give flexibility for modelling future scenarios in which more species recoveries may occur, and where BDS may therefore become more negative). Tweedie model selection often uses the Gini index for model selection<sup>70</sup>. However, the ratio of sample size to the number of parameters is relatively small in the Tweedie analyses, potentially indicating low power to distinguish among models and a risk of overfitting. Thus, we initially compared model fit using the Gini index but then ran model selection using AICc, a technique that penalizes overfitting and is asymptotically similar to leave-one-out cross validation<sup>71</sup>. We regarded Gini-selected models as overfitted if they contained terms that both were excluded in AICc selection and had  $P > 0.1$ . Gini and AICc approaches gave identical model selection results in the main text; in the sensitivity tests for  $T = 0.10$  and  $T = 0.25$ , however (see Sensitivity testing), we preferred AICc approaches. We also carried out a power analysis<sup>72</sup>, which revealed that our best-fitting models had a power of  $> 0.99$  and that our sample size was therefore adequate to detect effects among the relatively large number of parameters.

In the binomial part, we used GLMs with binomial errors and fitted an additional dispersion parameter to account for strong underdispersion<sup>64</sup>. Models containing this extra parameter do not generate AIC or AICc values, so we carried out non-automated binomial model selection using stepwise backward and forward regression with likelihood ratio tests<sup>64</sup>. Explanatory power was measured in the continuous part using McFadden's  $R^2$  (known to be conservative), and in the binomial part using the percentage of times that the model correctly predicted BDSb, taking  $p(\text{BDSb} = 1) < 50\%$  as a predicted 0 and  $p(\text{BDSb} = 1) > 50\%$  as a predicted 1.

**Cross validation to test for forecasting accuracy on unseen data.** To test the model's forecasting accuracy, we carried out tenfold cross-validation. This procedure repeatedly sets aside part of the data (as a 'fold' of BDS values the model has never seen), parameterizes the model on the remaining subset of data and then tests how well it forecasts the unseen BDS values<sup>37</sup>. For the continuous model part, we measured forecasting accuracy by calculating McFadden's  $R^2$  for the model fit to the unknown (hold-out) BDS in each of the ten folds. Ideally, the slope of forecast versus known values should be close to 1.0; to test for this, we regressed the complete set of forecast values (across the ten folds) against the complete set of known values in the cross-validation, using a generalized least squares regression model with a constant power function fitted to describe the heteroscedasticity in the residuals. We also calculated the median absolute deviation, although this is less informative in data with a large spread of values. Percentage deviations (rather than absolute deviations) are not appropriate metrics for low-volume data containing several zeroes, such as BDSs<sup>37</sup>. For the binomial model part, we tested mean forecasting accuracy against unknown data using the percentage of correct predictions, as we had done in testing binomial explanatory power.

**Covariate balancing and spatial considerations.** An important issue in impact studies is selection bias, whereby the likelihood of receiving the intervention of interest is non-random<sup>25,73</sup>. The amount of conservation investment a country receives is known to be influenced by non-random factors, including Red List status itself<sup>8</sup>. This has the potential to create endogeneity problems<sup>25,73</sup> and, in particular, the problem of reverse causality, in which biodiversity declines drive changes in conservation spending (rather than vice versa).

We used a time lag between predictors and responses to reduce the issue of reverse causality in the analysis. We also note that, as greater declines have been shown to lead to greater investment<sup>8,28</sup>, a simple reverse causality hypothesis would imply a positive correlation between spending and decline, whereas we observed a negative correlation (greater investment was associated with a decrease in subsequent decline). To correct for selection bias and associated endogeneity problems more generally, we used covariate balancing propensity scores<sup>25</sup> for continuous treatment variables<sup>74</sup> (in the R package CBPS<sup>75</sup>); this approach minimizes the association (Pearson's correlation) between covariates and the treatment<sup>74,75</sup>. Previous studies have explained a high proportion of the variance in conservation



finance allocation using country area, cost (the national price level), government effectiveness, political stability, GDP at PPP, percentage of land that is protected and the sums of threatened bird and mammal species weighted by their level of extinction risk<sup>8,28</sup>. We carried out covariate balancing, using data on these variables taken from previous studies<sup>8</sup> in combination with data on forest loss between 1990 and 2005<sup>58,59</sup> and data on 1992–2003 growth in GDP per capita at PPP<sup>57</sup>. Extended Data Fig. 6 shows the Pearson's correlations between the treatment and the covariates before and after the covariate balancing propensity score correction.

Analysing species declines at the country level could potentially generate spatial structure in model residuals and thereby violate regression assumptions<sup>50,64,76,77</sup>. We tested for this effect by fitting four possible structures to the most complete GLM model using restricted maximum likelihood estimates, and comparing their predictive power using AICc. The structures tested were: (i) a fixed effect for region (following previous studies<sup>8</sup>; see Supplementary Tables 1 and 2 for regions and regional intercept differences); (ii) a generalized mixed model (GLMM) with a spatial autocorrelative structure (SAC), which tested five possible structural models (linear, spherical, Gaussian, ratio and exponential) of the spatial autocorrelative structure that may exist between the centroid coordinates of different countries<sup>64</sup>; (iii) a GLMM with a SAC as in (ii), plus a fixed effect for region; (iv) a GLMM with a SAC as in (ii), plus a random intercept for region. Structure (i) provided the best fit, and we used this in subsequent modelling. Using region as a fixed effect also follows logically from theory, as regional differences are a potentially important component of decline<sup>46</sup>. Binomial models including spatial autocorrelative structures did not converge and regional effects were non-significant, so we tested for possible spatial effects by plotting residuals from the best-fit binomial model against both latitude and longitude, and also by exploring the effect of including the latitude and longitude coordinates of the country centroids in the model specification. There was no support for models including latitude and longitude and no visual relationship in the plots against residuals.

Decline drivers in one country may influence biodiversity in neighbouring countries and statistical 'spatial lags' have been used to model these possible effects<sup>50,77</sup>. However, statistical techniques to model a mixture of spatial error and spatial lag in the dependent and independent variables have only recently been developed for ordinary least squares regression<sup>78</sup> and, to our knowledge, no robust methodology exists for generalized linear models with heteroscedastic Tweedie error structures. We therefore restricted ourselves to testing and correcting for spatial error structures. However, by dividing responsibility for declines proportionally among countries, we have probably removed much of the artefactual spatial associations that will arise when neighbouring countries are given equal responsibility for any declining species that they share.

All statistical analysis was carried out in the R statistical software environment<sup>79</sup>. We checked for violations of model assumptions using diagnostic plots of residuals against fitted values and against all candidate predictor variables<sup>64</sup>. When removing a variable in model selection, we also plotted the residuals of each reduced model against the newly removed variable, checking for any pattern that the statistical tests may have missed. We checked for collinearity using variance inflation factor scores (Extended Data Table 4).

**Predicting the effects of spending and pressure changes.** To predict the effect that an extra \$1 million or \$5 million dollars of annual conservation spending would have had in each country, we added these amounts to known financing levels for each country and used the model to re-predict the outcomes. To predict the effect of changing human pressures on those outcomes, we followed the same protocol but also replaced 1992–2003 levels of socioeconomic growth with 2001–2012 levels. To estimate the decline that may have been avoided as a result of conservation spending between 1992 and 2003, we used the fact that, before the 1992 Earth Summit, biodiversity spending for which we have data was flat and often zero (noting, however, that data become sparser as one goes back further in time). We estimated mean annual spending for 1985–1990 and re-predicted outcomes as if post-1990 annual conservation budgets had increased only in line with inflation (no real increase). Although reduced data quality and imputation for 1985–1990 spending make these estimates approximate, the median change in BDS was robust across several different spending estimates, and the global figure for avoided decline (29%) is therefore likely to be a reasonable approximation (although we acknowledge that the true figure may be higher or lower).

**Sensitivity testing.** We further tested the sensitivity of our original PACI model to various assumptions. To test for sensitivity to the threshold  $T$  (set at 0.17 in the main text; see Additional method details), we examined the model outcomes using  $T = 0.10$  and  $T = 0.25$ . To test for sensitivity to the  $\psi$  parameter, we repeated the analysis with multiple variations around the parameter value used in the main analysis. To test for the effect of the influential outliers (Solomon Islands and New Zealand), we ran model selection both with and without these outliers. To examine whether our results were sensitive to the variables used to calculate the propensity scores (the correction for non-random assignment of spending amounts across

countries; see Covariate balancing and spatial considerations, above), we tested the effect of removing various individual variables or combinations of variables from the list used to calculate the propensity weights for the regression model.

A further concern was that our model fits might be driven (biased) by a country (or countries) with a high BDS, as the distribution of BDSs is skewed (Extended Data Fig. 5). Our tenfold cross-validation test had already shown that the omission of various groups of countries had no substantive effect on results, but as a further check we carried out a jack-knife leave-one-out test to see how the omission of each individual country affected parameter estimates. When interactions between continuous terms are present, parameter estimates are conditional; they are different for each country and affect one another. An appropriate measure of parameter change is, therefore, the mean or median percentage change in the values of the conditional expectations across all countries. With heteroscedastic errors, the median percentage may be more informative than the mean, so we considered both. If a country was strongly biasing the model results, then running the model without it should produce a substantial change in the median and mean conditional expectation of BDS across the remaining countries, indicating a strong shift across the conditional parameter estimates for the interaction model.

Even with these tests, there remained the possibility of joint influence in the continuous model part<sup>80</sup>, in which the group of highest-value BDS countries might collectively be driving the model. For example, the BDS values for the three highest-value countries (Indonesia, Australia and China) are very large (272%, 69% and 24% larger, respectively, than the fourth-highest BDS value) and may therefore combine to exert joint leverage on the model parameters. To test for this, we plotted fitted against observed values for both the full dataset and the top-three-removed dataset. For completeness, we also examined changes in the individual conditional coefficients when the top three highest-value BDS countries were omitted.

In impact assessments that address the effect of a single variable, a further concern is missing variable bias; this occurs if there is a confounding variable that is closely correlated with both the studied variable and the outcome variable<sup>81</sup>. In this example, the concern would be that the observed effects of conservation spending may simply be an artefact of the fact that spending is collinear with an unknown variable that is actually driving the outcome. If only one explanatory variable is being studied for its effect, hidden variable bias can be investigated by testing whether the influence of the main variable is still observed after an artificially created, collinear dummy variable has been added to the analysis<sup>25</sup>. In multiple regression analyses, this is largely infeasible because it would also be necessary to artificially generate correlations between the dummy and all the other (interacting) variables in the regression formula. We attempted to approximate the missing variable test by looking for an empirical variable that was closely correlated with our spending variable and that, therefore, had a natural co-correlation with all other variables in the regression formula. We then added it into the regression and tested whether the spending effect disappeared. Using the same scaling standardization as in the main analysis, we found that mean total population size had a correlation (Pearson's  $r$ ) of 0.45 with spending and mean GDP PPP (raw GDP instead of the GDP per capita used in the main analysis) had a correlation of 0.54 with spending. We therefore tested the effect of adding both variables, in turn, to our regression formulae (in the second instance, removing GDP per capita and replacing it with raw GDP, on account of the strong correlation between them).

Finally, we tested the possible consequences of inaccuracy in national conservation spending data, following previously used sensitivity tests<sup>8</sup>: in summary, we varied the spending data for each country by iteratively drawing new spending values for each country from a normal distribution centred on the original value and with a standard deviation set to 25% of the original value, and then repeating the regression analysis. Owing to extremely slow processing times for our complex models, we carried out 100 such permutations.

Detailed results of these sensitivity tests are shown in the Supplementary Results, but none affected our conclusions substantively.

**Additional method details: mathematical calculation of BDS.** Change in Red List status is a standard measure of biodiversity change used in the CBD and SDG frameworks<sup>3,31,32</sup>. However, it applies to species, whereas we wished to measure change at the level of the sovereign countries that, as signatories to these agreements, have the principal political responsibility for establishing biodiversity policies and for meeting the targets. We therefore created an algorithm to convert species-level change to country-level change. Mathematically, we define  $R_{ij}$  as the proportional responsibility that country  $j$  has for a status change in species  $i$ , where for each species  $i$ :

$$\sum_j R_{ij} = 1.0$$

For brevity, we use the phrase 'proportional responsibility' (or simply 'responsibility') to refer to the relative influence that factors in each country had

on the changing conservation status of each species. Proportional responsibilities cannot be known exactly, and so the algorithm will generate estimates of responsibility with some error. For predictive modelling, an equally important condition of algorithm design is that this error should not bias regression outcomes.

The most commonly used responsibility algorithm simply counts the number of declining species in each country (usually, the number of species classified as being under some level of threat in global Red List assessments)<sup>14,15,45,46,77</sup>. This algorithm implicitly assumes that if two countries share a species, they have equal responsibility for its decline. This is reasonable if both countries have roughly equal shares of the species range. However, species are frequently distributed so that one country holds the bulk of the range and neighbouring countries hold very small fractions of the remaining range edge (Extended Data Fig. 5). In such cases, it would be highly inaccurate (and politically unfair) to allocate equal shares of responsibility for a species' decline across all the countries into which its range extends. A more accurate system may be to divide up responsibility according to the fraction of each species' range found in each country<sup>8,26</sup>. Formally, if  $p_{ij}$  is the proportion of the range of species  $i$  in country  $j$ , then the value of  $p_{ij}$  is an estimate of the true responsibility ( $R_{ij}$ ), with some error implied in that estimate; this error is formally defined as the difference between the  $p_{ij}$  estimate and  $R_{ij}$ .

For any observed  $p_{ij}$ , there is therefore a theoretical probability density function of all the possible  $R_{ij}$  that it could represent. For example, if a species range is split between two countries at a ratio of 60:40, for the country for which  $p_{ij} = 0.60$  there is an underlying assumption that there is an approximately Gaussian probability density function for  $R_{ij}$  with a central mode at 0.6 (such that the most probable value of  $R_{ij}$  is 0.60 or close to it), whereas extreme values such as 0.0 or 1.0 have a very low theoretical probability.

If we imagine that for any given country  $j$ , all  $p_{ij} = 0.60$ , then all  $R_{ij}$  will follow a Gaussian distribution around 0.6. The range-based algorithm will generate a series of positive and negative errors ( $eR_{ij}$ ), which represent overestimates and underestimates of  $R_{ij}$ . The same is true of the corresponding neighbour with  $p_{ij} = 0.40$ . However, the true quantity of interest we wish to estimate is  $BDS_j$  (the sum of  $R_{ij}$  rather than each individual  $R_{ij}$ ). There is therefore an associated set of errors:

$$eBDS_j = \left( \sum_i eR_{ij} \right)$$

For a predictive regression model, the critical question is whether these errors ( $eBDS_j$ ) are likely to strongly affect the modelling of  $BDS_j$ ; for example, by creating artefactual patterns or biased non-random error distributions. If all range splits that make up  $BDS_j$  are relatively symmetric (similar to 60:40 ratios), then it is a reasonable expectation that the errors (being drawn from an approximately Gaussian distribution) will be overestimations and underestimations in roughly equal proportion; consequently, the sum of errors will not depart strongly from zero. Thus, the errors are expected to be relatively random in their distribution, which permits robust modelling. It is also improbable that the errors would create artefactual effects, as this would require a consistent, non-random association between large negative errors and higher-spending countries (sufficiently large to strongly depress  $BDS_j$ ), as well as equally large and consistently positive errors for lower-spending countries.

However, when  $p_{ij}$  is closer to its limits (0.0 and 1.0), biased errors become highly likely. Human-induced population losses (leading to species declines and Red List status changes) are generally focused spatially in the particular part or parts of the species range in which human pressures have most strongly increased; in general, it is very rare for such hotspots of decline to be located at the periphery of any given range<sup>36</sup>. Therefore, a country that holds 3% of the species range will often have zero responsibility (rather than 3% responsibility), and the neighbour with 97% of the range will often be entirely responsible for a status change. Even in a random process (with limited trials and, therefore, stochastic outcomes), spatial clusters of increased mortality dropped at random within the range will frequently fall entirely within the country that contains 97% of the range. In formal terms, in situations in which  $p_{ij} = 0.03$ , the associated probability density for  $R_{ij}$  will be high at 0 and decline rapidly toward a very low density at  $R_{ij} = 0.03$ , giving a probability density function with a strong right skew and a 99th percentile that is likely to be located at around  $p_{ij}$  itself.

If  $p_{ij} = 0.03$ , therefore, nearly all errors will be overestimates; the most common and likely scenario is an overestimate of exactly 0.03. In general terms, if  $p_{ij}$  is small and the probability density function is right skewed, a raw or 'unadjusted' range-based algorithm will overestimate responsibility in almost all cases and will generate highly biased errors ( $eR_{ij}$ ) that will commonly be overestimations of magnitude  $p_{ij}$  ( $+p_{ij}$ ). By the same process, using  $p_{ij}$  to estimate  $R_{ij}$  at high  $p_{ij}$  (such as 0.9) will tend to underestimate true responsibility in the majority of cases.

The critical question is how severely this consistent bias affects the regression analysis. In our data, a large number of countries had a  $BDS$  that was composed

entirely of trivially small ( $<5\%$ ) range-edge fractions (Extended Data Fig. 5). Their  $BDS_j$  estimates, which were composed entirely of very small  $p_{ij}$  values, were therefore likely to have aggregated multiple, consistent overestimates of responsibility ( $R_{ij}$ ). In analysing  $BDS_j$ , the error metric of interest is  $eBDS_j$  (the sum of  $eR_{ij}$ ). As the set of errors  $eR_{ij}$  was likely to be highly biased and the most common likely scenario was that  $eR_{ij} = +p_{ij}$ ,  $eBDS_j$  (as the sum of  $eR_{ij}$ ) would also be highly biased. In particular, there was a substantial probability that  $eBDS_j$  would equal the sum of  $p_{ij}$ . As all the individual  $p_{ij}$  values that comprised these  $BDS$  scores were both trivially small and likely overestimates of a true zero value, the associated  $BDS$  scores, which aggregated these values, were also likely to be trivially small and biased overestimates of a zero value. We refer to these cases as range-edge  $BDS$  or 'reBDS'.

We further explored the empirical effect of this suspected bias on the information signal by making exploratory plots of a given  $BDS$  against its possible predictors. These plots showed that reBDSs generated dense clouds of very small values close to the  $x$  axis that were visually distinct from patterns across larger (and presumably more accurate)  $BDS$ s. In signal detection theory terms<sup>37</sup>, reBDS cases were highly likely to represent strong signal noise lying non-randomly to one side of the main information pattern, in a cloud of such density that the signal-to-noise ratio was extremely low, the ability of regression models to detect predictive relationships was compromised and any calculated model parameters were likely to be strongly biased by the non-random error. Similarly, in the binomial analysis, the same reBDS issue caused many species-poor countries to have  $BDS_b = 1$  simply because they contained the range edges of species that had changed Red List status, but which were found almost entirely elsewhere.

To reduce these issues of signal noise and bias at small  $p_{ij}$ , we explored setting  $R_{ij}$  to zero for small  $p_{ij}$ . In formal terms, this involved setting a threshold value ( $T$ ) that set responsibility to zero for any country with a range fraction that was less than  $T$ , such that

$$R_{ij} = P_{ij}^* = \begin{cases} p_{ij} & \text{if } p_{ij} \geq T \\ 0 & \text{if } p_{ij} < T \end{cases}$$

(But see below for  $p_{ij} \geq (1 - T)$ ).

To decide on appropriate values for the threshold  $T$ , we used signal detection theory in combination with the numerical population changes required to trigger a change in Red List status. The most important aspect of this approach is that when  $p_{ij}$  is small, true  $R_{ij}$  may often (but not always) be zero. However, it is impossible to know which range-edge countries genuinely had a very small responsibility and which had a true-zero responsibility. Therefore, reBDS values will often (but not always) be non-zero overestimates of a true zero. In signal detection theory, the cases in which a true zero is wrongly assigned a non-zero value represent 'false positives'. However, any threshold could also cause the algorithm to wrongly exclude cases in which the reBDS represented a genuine (if small) fractional responsibility, and such incorrect exclusions are classed as 'false negatives'. The higher the threshold, the more false positives will be correctly excluded and the more false negatives will be wrongly excluded. Theoretical optimization will therefore seek values of  $T$  large enough to avoid too many false positives (guarding against picking up too much noise) yet small enough to avoid too many false negatives (guarding against throwing away too much information). A threshold that produces too many false positives is classed as overly sensitive and one that produces too many false negatives is classed as overly specific.

For the  $BDS$ , the optimal signal detection threshold cannot be precisely estimated because the proportions of false positives and false negatives at any value of  $T$  are not empirically known, and the ratio of sensitivity to specificity cannot be calculated. Appropriate thresholds therefore need to be assigned by theoretically estimating the optimal sensitivity–specificity trade off. Furthermore, in this analysis, sensitivity and specificity were likely to have had distinct effects on analytical bias and outcomes; this meant that approaches that gave equal weight to sensitivity and specificity, or that required accurate knowledge of the sensitivity–specificity ratio (for example, area under the curve<sup>37</sup>), were less appropriate than they would otherwise have been. The main deleterious effect of excessive sensitivity was likely to have been the generation of large amounts of biased noise. The main effects of excessive specificity were likely to have been (i) to slightly underestimate  $BDS$  (because a few small responsibility fractions had been wrongly discarded); (ii) to reduce sample size for the continuous model part (by removing reBDS countries); and (iii) to change the ratio of ones to zeroes in the binomial analysis (because reBDS countries have  $BDS > 0$  before adjustment and  $BDS = 0$  after adjustment). As the high levels of noise and bias associated with a lack of specificity are likely to have a much greater effect than the small underestimates and sample size and/or binomial ratio effects associated with a lack of sensitivity, avoiding false positives should take priority.



To allocate this priority (to avoid repeatedly replacing true zeroes with trivially small values), the algorithm needs to set  $T$  in such a way that there is a low probability density at  $R_{ij} = 0$ , for all probability frequency distributions associated with all  $p_{ij}^*$ . Formally, we set a target that for all  $p_{ij}^*$ , the probability ( $R_{ij} = 0$ ) should be less than 0.5 (and ideally, much less than 0.5). However, a second consideration is that, in range-edge countries, the theoretically expected probability density at zero is affected by the size of the decline implied by a status change. To illustrate this, we take the example of a country that holds 10% of a species' range and the most frequent criterion justifying a status change, population loss (Red List category A(2–4)<sup>3</sup>). When population loss occurs, the Red List assessment for any particular period is based on the rate of change over time. Therefore, a change in Red List status expresses a second-derivative change in the rate of change (net mortality above that which had occurred in the previous assessment period). If a status change represented a 99% increase in mortality for the entire species, there would be a strong probability that at least some of those additional deaths had occurred in the country that incorporated 10% of the species' range. However, genuine status changes generally imply an increase in loss of a few tens of percentage points. For example, a common status change is from Least Concern (LC) to Vulnerable (VU); LC implies anything between no decline and a 29.9% loss over a period of ten years (or three generations) and VU is defined as anything between 30% and 49.9% loss (depending on whether and how assessors use the near-threatened category; [www.iucnredlist.org](http://www.iucnredlist.org)<sup>82</sup>). If we take the midpoints of these ranges (15% and 40%, respectively), an LC–VU change would typically indicate a 25 per cent increase in loss (the difference between 40% and 15%) and LC–VU changes that did not occur at the exact midpoints would indicate differences in decline rates above or below 25 per cent.

As the additional deaths that underlie a status change are generally non-randomly clustered in geographic space (as wave fronts expanding from points of increased human pressure<sup>36</sup>), this change of 25 percentage points can be imagined as a small number of clusters of additional net loss placed on a gridded range, in which the country holding 10% of the species range occupies the leftmost 10% of the grid and another country (or countries) occupies the rightmost 90%. Spatially clustered mortality increases such as this might be expected to often fall entirely within the rightmost 90% (implying that the country holding 10% of the range will frequently bear no responsibility for the decline). To explore this intuition quantitatively, we simulated a 25-percentage-point population loss as a varying number (between two and five) of rectangular blocks that covered a total of 25% of a  $10 \times 10$  gridded range. The first column of the grid was then treated as the country holding 10% of the range and the remaining nine columns as another country or countries. The simulation focuses only on the likelihood that the 10% country will not have any part of any decline cluster overlapping its territory, and it is therefore moot whether one or several countries occupy the remaining nine columns. The blocks were then placed independently of one another on the gridded range, for a limited number ( $n = 100$ ) of trials to introduce stochasticity. For each placement, we tested whether any part of the leftmost column had been overlapped. Overall, we found that the probability of any overlap between a block and the leftmost 10% of the grid was generally  $< 0.5$ , varying with the number of blocks. For example, if the decline occurred as two independently placed blocks, the simulated probability of overlap was 0.19, giving a 0.81 probability that the range-edge country has  $R_{ij} = 0$  (an 81% chance of a false positive). When the 25-percentage-point decline was modelled as five independently dropped blocks, the overlap probability rose to 0.41, indicating a 59% chance of a false positive, which was still appreciably greater than our target false-positive rate of much less than 0.50. These values are conservative because clusters of loss are often not spatially independent of one another but instead may be grouped as a result of larger-scale spatial contagion in threats and associated losses<sup>36</sup>. Such grouping further reduces the random probability of an overlap with the range edge and thus would increase the false positive rate. Similar outcomes occur for other percentage point increases in mortality, as implied by other IUCN status changes.

The 25-percentage-point population loss can also be unrealistically and highly conservatively modelled as spatially homogeneous. Define  $q$  as the change in rate of species decline required to trigger a change in Red List status (such that in the example,  $q = 0.25$ ). Under an assumption of homogeneity, the theoretical maximum responsibility that a country holding 10% of the species range can have for a 25% change is roughly 40% (10 divided by 25). More formally, we define the 99th percentile of theoretically probable  $R_{ij}$  for the country holding 10% of the range as ( $p_{ij}$  divided by  $q$ ) = (0.1 divided by 0.25) = 0.4. A distribution with a 99th percentile at 0.4 is likely to have a relatively strong skew and consequently a relatively high probability density at  $R_{ij} = 0$ , because skewness in the theoretical probability distribution for  $R_{ij}$  increases at an accelerating rate as the entire distribution moves to the left.

There is therefore a strong likelihood that even for non-trivial  $p_{ij}$  (such as 10% or more) the probability that  $R_{ij} = 0$  will be greater than the algorithm's target of

'much less than 0.5'. Therefore, the theoretical expectation is that to avoid false positives to a sufficient degree, the threshold  $T$  may need to be set at greater than 0.1 and potentially as high as 0.2 (or more). To further explore this expectation empirically, we examined exploratory biplots of BDS against its predictors in which  $T$  was varied between 0.05 and 0.25. As expected from our theoretical treatment, we found that as  $T$  was reduced increasingly large numbers of likely false positives were included in the BDS dataset, with noise increasing rapidly at  $T < 0.1$  (an increasingly dense cloud of points with trivially small BDS values developed). On the other hand, increasing  $T$  from 0.14 to 0.25 caused little variation in  $R_{ij}$  values but progressively reduced sample size (and so power) in the continuous analysis, with the drop off in sample size being small between  $T = 0.1$  and  $T = 0.17$  and larger between  $T = 0.17$  and  $T = 0.25$  (see Supplementary Results).

Simulation and probability theory can therefore suggest the approximate range for appropriate values of  $T$ , but the exact optimal value remains uncertain. To account for this uncertainty and its possible effect on model outcomes, we performed our final analysis three times for three different values of  $T$  (0.10, 0.17 and 0.25). The main text of the paper shows results for  $T = 0.17$ , which represents the parameter value at which false positives were substantially reduced without sample size reduction becoming severe. Results for  $T = 0.10$  and  $T = 0.25$  are described in Extended Data Table 2 and Supplementary Results.

In formal summary, for each species  $j$ , each country  $i$  holds  $R$  proportional responsibility for the total decline ( $d$ ) of  $j$ . Decline can be positive and indicate an increasing risk of extinction ( $d > 0$ ); it can be negative and indicate a reduction in extinction risk ('negative decline' or improvement,  $d < 0$ ); or it can be constant ( $d = 0$ ). The baseline BDS for each country (BDS<sub>*i*</sub>) is the net sum of all its decline fractions and improvements (negative decline fractions):

$$\text{BDS}_i = \sum_j d_j R_{ij}$$

where

$$R_{ij} = \frac{p_{ij}^*}{\sum_i p_{ij}^*} \quad (1)$$

and  $p^*$  indicates the range proportion of each species  $j$  in country  $i$ , after range fractions below the minimum percentage  $T$  have been set to zero:

$$p^* = \begin{cases} p_{\text{obs}} & \text{where } p_{\text{obs}} \geq T \\ 0 & \text{where } p_{\text{obs}} < T \end{cases}$$

If a species' range is split at a ratio of 95:5 between two countries and the responsibility  $R$  has been set to zero for the country that holds 5% of the range, then for consistency the  $R$  of the other country should be increased from 0.95 to 1.0; equation (1) performs this function. However, a widespread species can be spread in small fractions across multiple countries without any one country having a major proportion of the range. In such cases, if only one country has a range fraction exceeding the threshold (for example, 17.1%) equation (1) would assign it 100% of responsibility for the change in risk status (whereby  $p_{\text{obs}} = 0.171$  but  $p^* = 1.0$ ), which clearly exaggerates its responsibility. Such cases are uncommon (widespread species rarely move out of the Least Concern category) but to avoid errors of this type, we reset the denominator of equation (1) to unity for cases in which a widespread species was scattered in small fractions across multiple countries.

To calculate the  $p_{ij}$  fractions themselves, we extracted the percentage of the geographic range of all bird and mammal species contained within the national borders of each country (the range overlap)<sup>26</sup>. Range overlap for mammals was extracted using ArcGIS utilities on the range maps provided by the IUCN Global Mammal Assessment<sup>83</sup>, as previously published<sup>8</sup>. This procedure gave very exact areas of overlap for Mammalia, but required us to run twenty processors in parallel for nearly a month. For Aves (a much larger taxon) we therefore used a slightly different procedure. Bird ranges were obtained as polygons in ESRI shapefiles from Birdlife International (<http://datazone.birdlife.org/species/requestdis>). Species range areas that were designated as non-native or dubious presence were excluded a priori. For each species  $X$ , we combined wintering and breeding ranges (threats to bird species can occur in either range) and gridded all range polygons on a 0.1 degree raster grid, using a cylindrical equal-area projection to match the projection of the original vector data. We designated all grid cells that had a centre point lying inside a range polygon for  $X$  as 'presence cells' for  $X$ , overlaid each presence cell onto a vector dataset of country borders ([http://thematicmapping.org/downloads/world\\_borders.php](http://thematicmapping.org/downloads/world_borders.php)) using the `over` and `wrld_simpl` functions in the R packages 'sp'<sup>84</sup> and 'maptools'<sup>85</sup>, allocated the cell to the country found at the cell centre point and calculated the fraction of all presence cells for  $X$  found in each country. Before performing this calculation, all countries with coastlines were enlarged by a 0.05 degree buffer into the sea to account for responsibility for sea bird ranges

in coastal waters; coastal marine mammals were treated in a similar way, as in previously published approaches<sup>8</sup>.

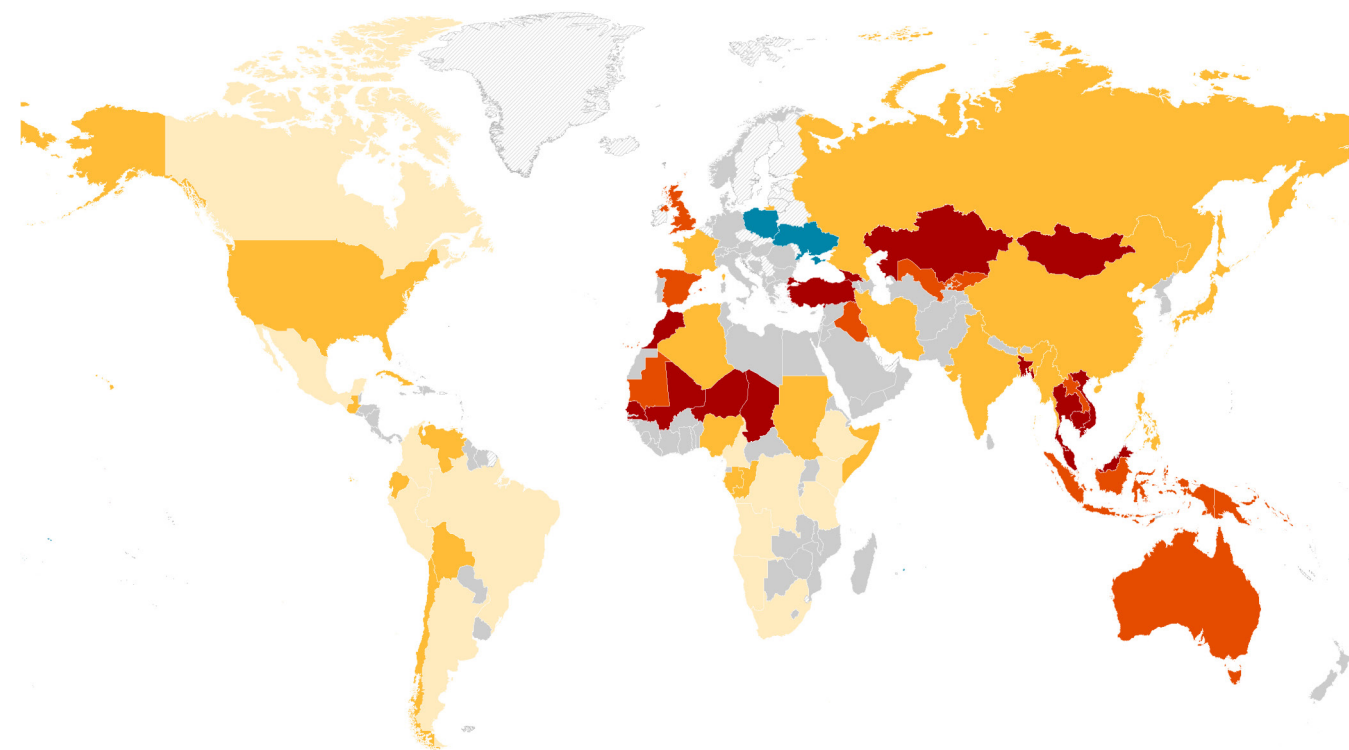
As an additional accuracy check, we examined individual Red List reports for every declining species to see whether our range-based approximations of responsibilities were appropriate; if they were not, we revised the case accordingly. Our revisions are listed in Supplementary Table 4 and include cases in which (i) a decline had had a major effect on the distribution of the geographic range across countries (including cases in which a species was missing from countries that had once been part of its range); (ii) the species population distribution across countries was poorly correlated with the range distribution; and (iii) specified actions (for example, along migratory routes) had an influence that was clearly disproportionate to the percentage of the global range that was located within the country carrying out those actions.

At a theoretical extreme, a 100% range fraction for a declining species could indicate that one country contains the last extant individuals of a species that used to be widespread in neighbouring countries. The country holding the remaining species population would then represent a final 'oasis' at the species' former range edge and it would be unjust to assign 100% responsibility for the decline to this country. However, our assumption is that in the eight-to-twelve years between the IUCN assessments, there will rarely be a case in which a species has been extirpated from its main homeland countries without some record of this event existing. We applied the BDS adjustments based on Red List reports after the adjustments for range edges (reBDS), and so our method corrected for any anomaly of this type. For example, *Addax nomasculatus* (the screw-horn antelope) has recently disappeared from Chad and Mali, and we therefore incremented the BDS of these two countries to reflect this (Supplementary Table 4).

**Code availability.** R scripts used in analysis are available upon request from the corresponding author.

**Data availability.** The authors declare that all the data that support the findings of this study are available within the Supplementary Information. Original socio-economic data are available from the World Bank (<http://databank.worldbank.org>) and original governance values from the Worldwide Governance Indicators dataset ([www.govindicators.org](http://www.govindicators.org)).

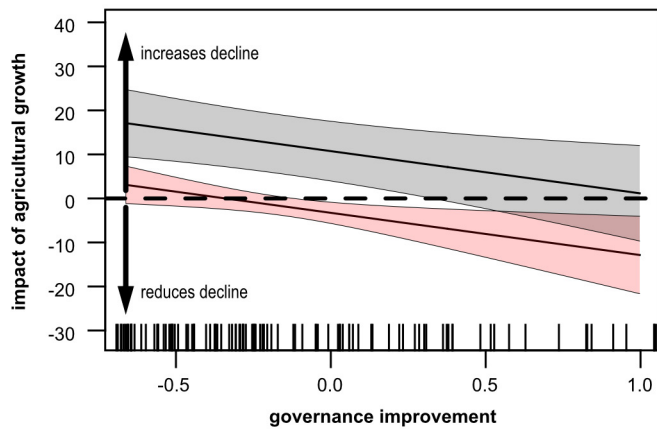
31. Butchart, S. H. M. *et al.* Measuring global trends in the status of biodiversity: Red List indices for birds. *PLoS Biol.* **2**, e383 (2004).
32. Butchart, S. H. M., Akcakaya, H. R., Kennedy, E. & Hilton-Taylor, C. Biodiversity indicators based on trends in conservation status: strengths of the IUCN Red List Index. *Conserv. Biol.* **20**, 579–581 (2006).
33. United Nations Statistics Division. Sustainable Development Goals indicators <http://unstats.un.org/sdgs/report/2016/> (2016).
34. Metrick, A. & Weitzmann, M. L. Conflicts and choices in biodiversity preservation. *J. Econ. Perspect.* **12**, 21–34 (1998).
35. Butchart, S. H. M. *et al.* Using Red List Indices to measure progress towards the 2010 target and beyond. *Phil. Trans. R. Soc. Lond. B* **360**, 255–268 (2005).
36. Channell, R. & Lomolino, M. V. Dynamic biogeography and conservation of endangered species. *Nature* **403**, 84–86 (2000).
37. Hastie, T., Tibshirani, R. & Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (Springer, 2009).
38. Waldron, A. *et al.* Turkey's biodiversity funding on the rise. *Science* **341**, 1173 (2013).
39. International Bank for Reconstruction and Development and The World Bank. Global purchasing power parities and real expenditures: 2005 international comparison programme <http://documents.worldbank.org/curated/en/842221468166145357/Global-purchasing-power-parities-and-real-expenditures-2005-international-comparison-program> (2008).
40. Miller, D. C. Explaining global patterns of international aid for linked biodiversity conservation and development. *World Dev.* **59**, 341–359 (2014).
41. Barrett, C. B., Gibson, C. C., Hoffman, B. & McCubbins, M. D. The complex links between governance and biodiversity. *Conserv. Biol.* **20**, 1358–1366 (2006).
42. McNeely, J. A. Conserving forest biodiversity in times of violent conflict. *Oryx* **37**, 142–152 (2003).
43. Hanson, T. *et al.* Warfare in biodiversity hotspots. *Conserv. Biol.* **23**, 578–587 (2009).
44. Kaufmann, D., Kraay, A. & Mastruzzi, M. The Worldwide Governance Indicators: methodology and analytical issues [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=1682130](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1682130) (2010).
45. Mills, J. H. & Waite, T. A. Economic prosperity, biodiversity conservation, and the environmental Kuznets curve. *Ecol. Econ.* **68**, 2087–2095 (2009).
46. Dietz, S. & Adger, W. N. Economic growth, biodiversity loss and conservation effort. *J. Environ. Manage.* **68**, 23–35 (2003).
47. Asafu-Adjaye, J. Biodiversity loss and economic growth: a cross-country analysis. *Contemp. Econ. Policy* **21**, 173–185 (2003).
48. Donald, P. F. Biodiversity impacts of some agricultural commodity production systems. *Conserv. Biol.* **18**, 17–38 (2004).
49. Cardillo, M. *et al.* Human population density and extinction risk in the world's carnivores. *PLoS Biol.* **2**, e197 (2004).
50. McPherson, M. A. & Nieswiadomy, M. L. Environmental Kuznets curve: threatened species and spatial effects. *Ecol. Econ.* **55**, 395–407 (2005).
51. Mozumder, P., Berrens, R. P. & Bohara, A. K. Is there an environmental Kuznets curve for the risk of biodiversity loss? *J. Dev. Areas* **39**, 175–190 (2006).
52. Pandit, R. & Laband, D. N. Economic well-being, the distribution of income and species imperilment. *Biodivers. Conserv.* **18**, 3219–3233 (2009).
53. Kerr, J. T. & Currie, D. J. Effects of human activity on global extinction risk. *Conserv. Biol.* **9**, 1528–1538 (1995).
54. Meyfroidt, P. & Lambin, E. F. Global forest transition: prospects for an end to deforestation. *Annu. Rev. Environ. Resour.* **36**, 343–371 (2011).
55. Brashares, J. S., Golden, C. D., Weinbaum, K. Z., Barrett, C. B. & Okello, G. V. Economic and geographic drivers of wildlife consumption in rural Africa. *Proc. Natl Acad. Sci. USA* **108**, 13931–13936 (2011).
56. Allen, J. C. & Barnes, D. F. The causes of deforestation in developing countries. *Ann. Assoc. Am. Geogr.* **75**, 163–184 (1985).
57. The World Bank. The World Bank databank <http://databank.worldbank.org/data/home.aspx> (2013).
58. Food and Agriculture Organization of the United Nations and European Commission Joint Research Centre. *Global Forest Land-use Change 1990–2005* (FAO, 2012).
59. Food and Agriculture Organization of the United Nations. *Global Forest Resources Assessment 2010: Main Report* (FAO, 2010).
60. Hansen, M. C. *et al.* High-resolution global maps of 21st-century forest cover change. *Science* **342**, 850–853 (2013).
61. Redding, D. W. & Mooers, A. Ø. Incorporating evolutionary measures into conservation prioritization. *Conserv. Biol.* **20**, 1670–1678 (2006).
62. Male, T. D. & Bean, M. J. Measuring progress in US endangered species conservation. *Ecol. Lett.* **8**, 986–992 (2005).
63. Clemens, M. A., Radelet, S., Bhavnani, R. R. & Bazzi, S. Counting chickens when they hatch: timing and the effects of aid on growth. *Econ. J.* **122**, 590–617 (2012).
64. Zuur, A. F., Ieno, E. N., Walker, N. J., Saveliev, A. A. & Smith, G. M. *Mixed Effects Modelling and Extensions in Ecology with R* (Springer, 2009).
65. Balmford, A. Extinction filters and current resilience: the significance of past selection pressures for conservation biology. *Trends Ecol. Evol.* **11**, 193–196 (1996).
66. Cragg, J. G. Some statistical models for limited dependent variable with application to the demand for durable goods. *Econometrica* **39**, 829–844 (1971).
67. Joergensen, B. Exponential dispersion models. *J. R. Stat. Soc. B* **49**, 127–162 (1987).
68. Joergensen, B. *Theory of Dispersion Models* (Chapman & Hall, 1997).
69. Tweedie, M. in *Statistics: Applications and New Directions. Proceedings of the Indian Statistical Institute Golden Jubilee International Conference* (ed. Ghosh, J. K.) 579–604 (Indian Statistical Institute, 1984).
70. Zhang, W. cplm: Compound Poisson Linear models version 0.7-2 <https://CRAN.R-project.org/package=cplm> (2014).
71. Burnham, K. P. & Anderson, D. R. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach* (Springer, 2002).
72. Cohen, J. *Statistical Power Analysis for the Behavioural Sciences* (Lawrence Erlbaum, 1988).
73. Ferraro, P. J. & Hanauer, M. M. Advances in measuring the environmental and social impacts of environmental programs. *Annu. Rev. Environ. Resour.* **39**, 495–517 (2014).
74. Imai, K. & Ratkovic, M. Covariate balancing propensity score. *J. R. Stat. Soc. B* **76**, 243–263 (2014).
75. Fong, C., Ratkovic, M., Hazlett, C., Yang, X. & Imai, K. CBPS: R package for covariate balancing propensity score <https://CRAN.R-project.org/package=CBPS> (2015).
76. Pandit, R. & Laband, D. N. Spatial autocorrelation in country-level models of species imperilment. *Ecol. Econ.* **60**, 526–532 (2007).
77. Amin, A. & Choumert, J. Development and biodiversity conservation in Sub-Saharan Africa: A spatial analysis. *Econ. Bull.* **35**, 729–744 (2015).
78. Drukker, D. M., Egger, P. & Prucha, I. R. On two-step estimation of a spatial autoregressive model with autoregressive disturbances and endogenous regressors. *Econom. Rev.* **32**, 686–733 (2013).
79. R Core Team. R: A Language and Environment for Statistical Computing <https://www.r-project.org/> (2013).
80. Fox, J. *An R and S-Plus Companion to Applied Regression* (Sage, 1997).
81. Rosenbaum, P. *Observational Studies* (Springer, 2002).
82. Vié, J.-C., Hilton-Taylor, C. & Stuart, S. N. (eds) *Wildlife in a Changing World: An Analysis of the 2008 IUCN Red List of Threatened Species* (IUCN, 2009).
83. Schipper, J. *et al.* The status of the world's land and marine mammals: diversity, threat, and knowledge. *Science* **322**, 225–230 (2008).
84. Pebesma, E. J. & Bivand, R. S. Classes and methods for spatial data in R. *R News* **5**, 9–13 (2005).
85. Bivand, R. & Kewin-Koh, N. Maptools: Tools for reading and handling spatial objects <https://CRAN.R-project.org/package=maptools> (2013).



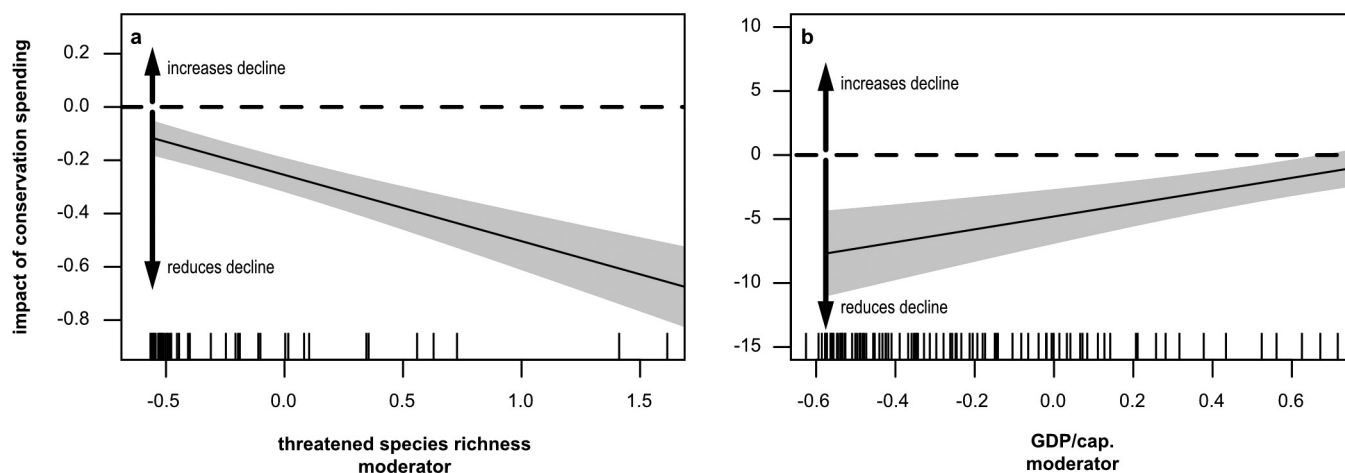
**Extended Data Figure 1 | The mean BDS per species for each country (BDS per total fractional species richness, expressed as a percentage).** Dark red,  $\geq 5\%$ ; dark orange-red, 2.5–5%; orange, 1–2.5%; pale yellow, 0–1%; grey, 0%; blue, improving (negative percentage); light grey hatching, cannot be calculated (zeroes in the denominator). Note that in more

species-poor countries (for example, much of Europe, North Africa and the Middle East), zeroes are expected at random (see Supplementary methods). See Supplementary Table 1 for precise values per country. Country outlines from esri\_dm (<https://www.arcgis.com/home/item.html?id=d86e32ea12a64727b9e94d6f820123a2#overview>).



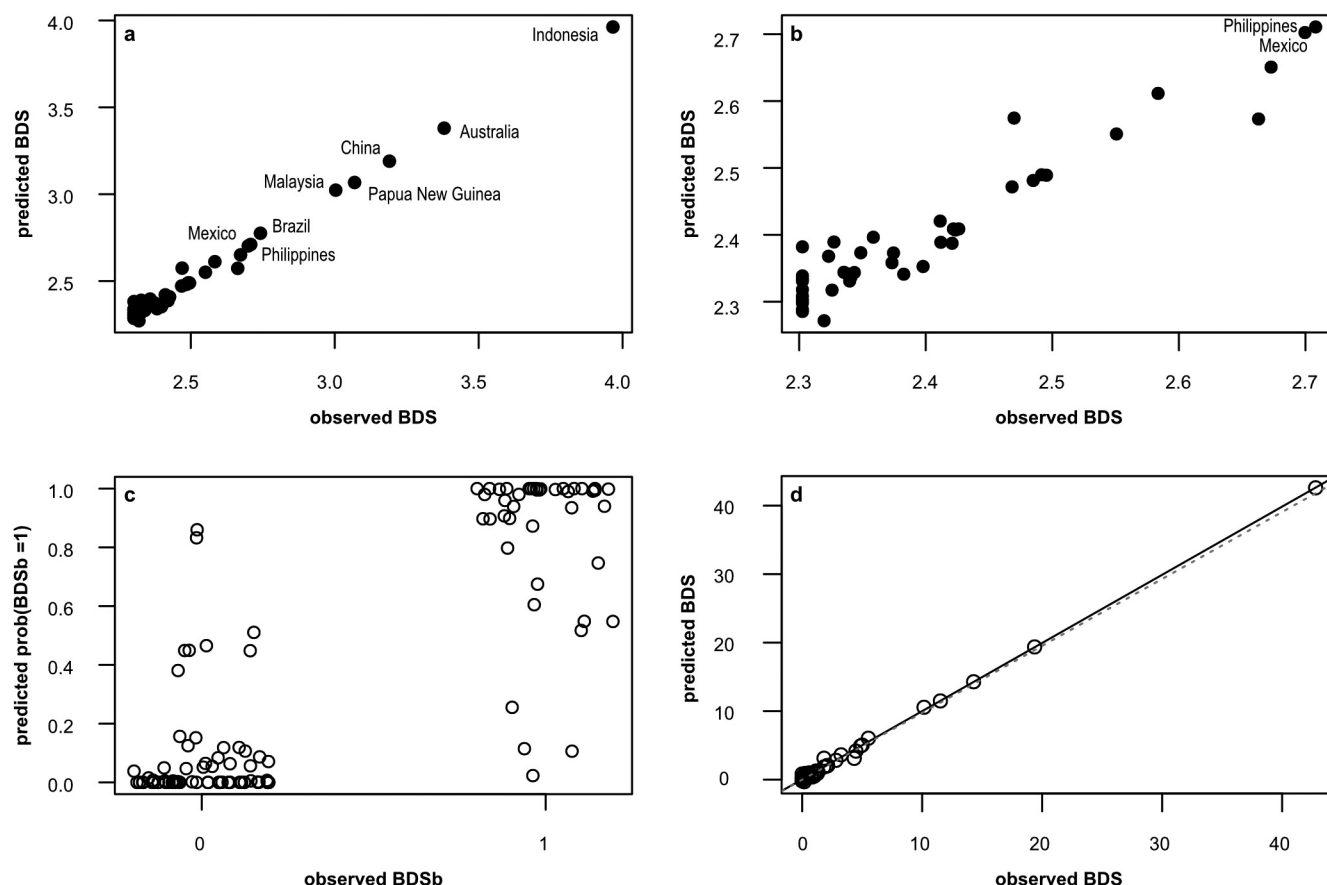


**Extended Data Figure 2 | The context-specific effect of agricultural expansion on decline.** In the binomial part of the model ( $n = 109$  independent countries), the effect of agricultural expansion on decline depends on governance improvement and on the pre-existing percentage of land converted to agricultural use. The effect (coefficient) of agricultural expansion is shown on the  $y$  axis and varies with the rate of governance improvement on the  $x$  axis. Coefficients  $>0$  (above the dashed line) indicate that agricultural growth increases the probability of a decline occurring; coefficients  $<0$  indicate that agricultural growth decreases the probability of a decline occurring. The coefficient also depends on a second moderator, the percentage of land already converted to agriculture: red, 50th percentile of percentage of land converted; grey, 25th percentile; lines show mean, and coloured bands show conditional 95% confidence intervals. The effects of agricultural expansion are most strongly deleterious on land bases that are less heavily converted to agriculture overall. Rug plot along the bottom shows empirical distribution of  $x$ -axis values (but note that countries with higher percentages of agricultural land generally have slower rates of governance improvement). All variables are  $z$  standardized.



**Extended Data Figure 3 | The effect of conservation spending on decline depends on threatened species richness and on GDP. a,** Spending effect size and threatened species richness, in the continuous part of the model ( $n = 50$  independent countries). **b,** Spending effect size and GDP, in the binomial part of the model ( $n = 109$  independent countries). The effect size (coefficient) for spending is shown on the y axis and varies with

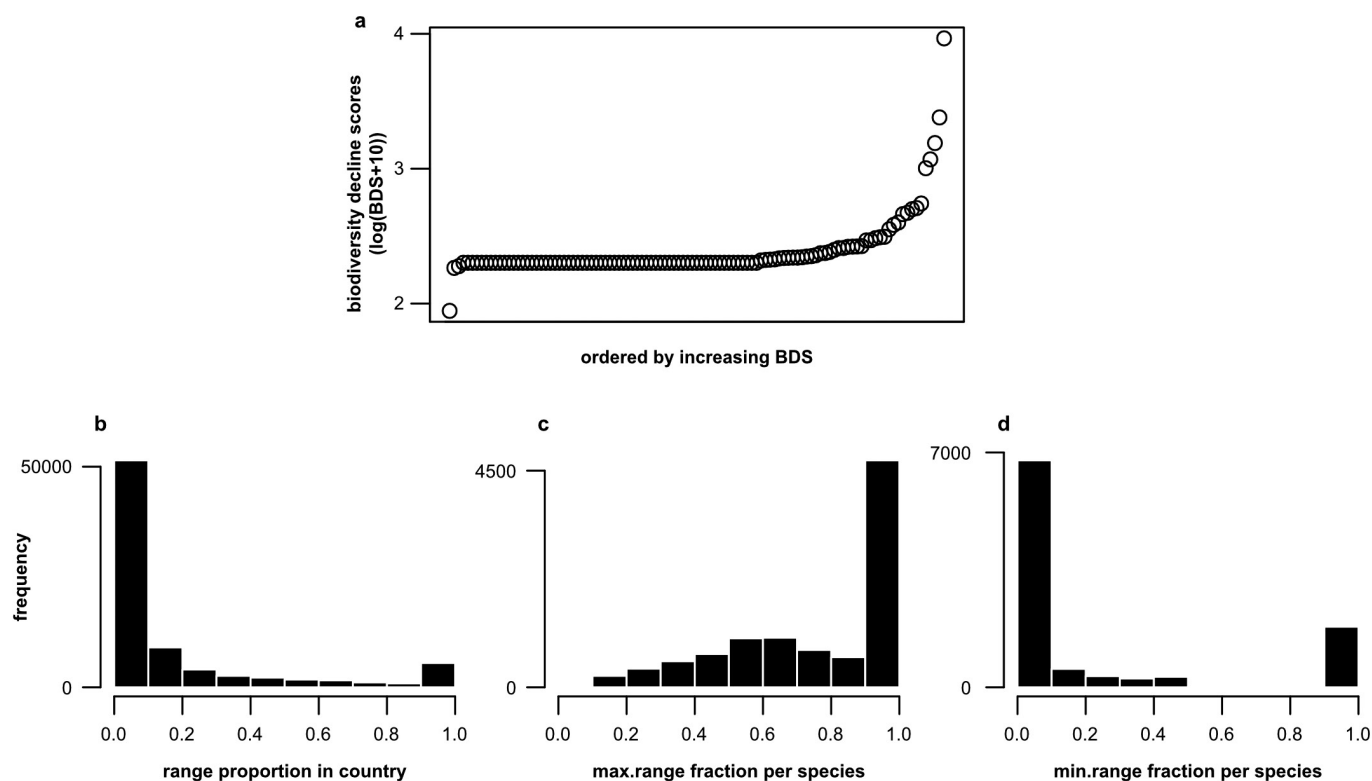
the value of species richness on the x axis. As coefficients on the y axis become increasingly negative, spending produces more marked reductions in biodiversity decline (continuous) or the probability of such a decline occurring (binomial). Conditional 95% confidence bands are shown; rug plots along the bottom show empirical distribution of x-axis values. All variables are z standardized.



**Extended Data Figure 4 | Observed declines versus model-predicted declines.** **A**, BDS versus predicted BDS in the continuous part of the model ( $n = 50$  independent countries). Both axes are log-transformed for clarity. **B**, As **a**, but focusing on countries with lower BDSs (note difference between values for the axes in **a** and **b**). **C**, Observed decline events (BDSb) versus the predicted probabilities of a decline event, from the binomial

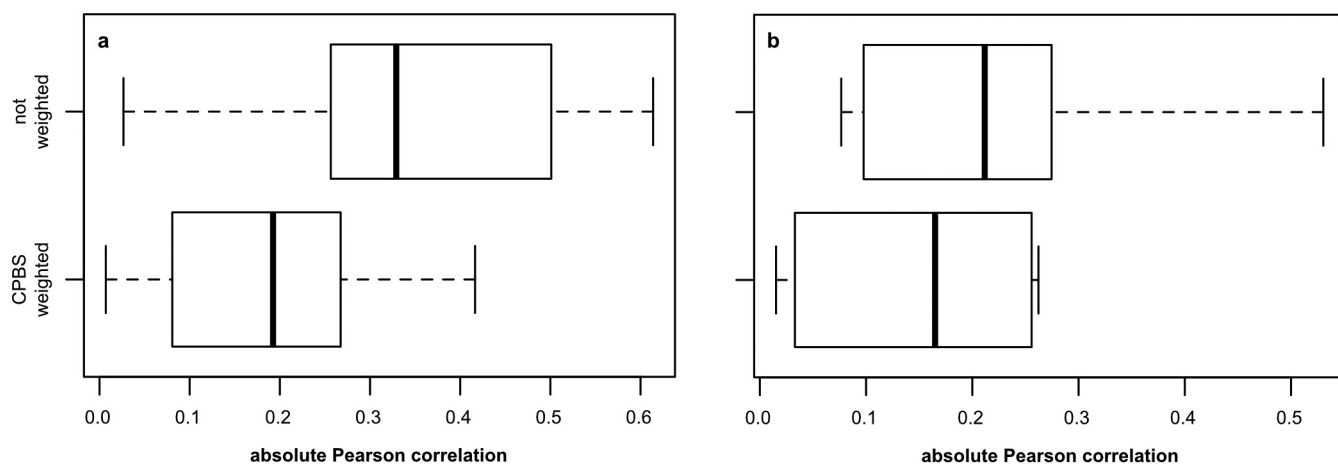
part of the model ( $n = 109$  independent countries). Observed decline events on the  $x$  axis (0 = no decline occurred, 1 = decline occurred) have been jittered for visibility. **D**, Change in model prediction when top three BDS values are excluded. Black line, full dataset prediction; dashed grey line, prediction with exclusions.





**Extended Data Figure 5 | Distributions of BDS and species range fractions across countries.** **a**, Index plot of BDSs. For clarity, BDS has been  $\log(x+10)$ -transformed; consequently, the straight line at 2.3 shows the long tail of zeroes. **b**, Distribution of all range fractions in all countries, showing the large number of small, range-edge fractions (in which <10%

of a species range is found in a country). **c**, Distribution of the maximum range fraction for all species; note that a large number of species have >90% of their range in a single country. **d**, Distribution of the minimum range fraction for all species; note that many species have a small range edge (<10% of their range) in a second country.



**Extended Data Figure 6 | Differences in absolute Pearson's correlations between conservation spending and each of its covariates before and after carrying out covariate balancing propensity score weighting.** **a**, Continuous analysis. **b**, Binomial analysis; absolute Pearson's

correlations before (upper bars) and after (lower bars) covariate balancing propensity score weighting. Box shows interquartile range; central line, median. Whiskers show the most extreme data point, no more than  $1.5 \times$  the interquartile range.  $n = 50$  independent countries.

**Extended Data Table 1 | List of regression terms tested, and the best-fitting four models from continuous analysis with their AICc values, Akaike weights and variables (see Supplementary Table 2 for full results from the continuous part of the model)**

Aicc	124.11	130.58	132.56	133.69
Delta aicc	0.00	6.47	8.45	9.58
Akaike weight	0.93	0.04	0.01	0.01
Conservation spending	1	1	1	1
Threatened species richness	1	1	1	1
GDP growth	1	1	1	1
GDP	1	1	1	1
Agric. Growth	1	1	1	1
% Agric. Land	1	1	1	1
Population growth	0	0	0	0
Population	0	0	0	0
Governance growth	0	0	1	0
Governance	0	0	0	1
Forest loss (%)	0	1	0	0
Declines (t-1)	1	1	1	1
Declines (t-2)	1	1	1	1
Country area	0	0	0	0
Region	1	1	1	1
GDP x GDP growth	1	1	1	1
GDP x conservation spending	0	0	0	1
GDP growth x conservation spending	0	0	0	1
Governance x conservation spending	0	0	0	0
Threatened species richness x conservation spending	1	1	1	1
GDP x GDP growth x conservation spending	0	0	0	1
Population growth x conservation spending	0	0	0	0
Population growth x GDP growth	0	0	0	0
GDP ^2	0	0	0	0
Agricultural growth x % agricultural land	1	1	1	0
Agricultural growth x conservation spending	0	0	0	1
% Agricultural land x conservation spending	1	1	1	0
Agricultural growth x % agricultural land x conservation spending	0	0	0	0
Population x population growth	0	0	0	0
Population x conservation spending	0	0	0	0
Population x population growth x conservation spending	0	0	0	0
Population^2	0	0	0	0
Governance growth x governance	0	0	0	0
Governance growth x agric. growth	0	0	1	0
Forest loss (%) x agric. growth	0	0	0	0

Spending, conservation spending at PPP; Agric., agricultural; governance, government effectiveness indicator. In the main body of the table: 1, term included; 0, term not included.



Extended Data Table 2 | Standardized coefficients for best-fitting models under alternative assumptions

Predictor variable	T = 0.10		T = 0.25	
	Continuous model part (BDS)	Binomial model part (BDSb)	Continuous model part (BDS)	Binomial model part (BDSb)
Conservation spending*	-0.221	-15.430	-0.232	-25.10
Agricultural growth*	-0.007	-9.852	-0.049	NA
GDP growth*	NA	-9.231	NA	0.069
Population growth*	NA	-6.322	NA	-1.421
Declines in period t-1	0.029	NA	0.024	NA
Declines in period t-2	0.022	NA	0.023	NA
Threatened species richness*	0.201	11.595	0.201	2.537
Country area	NA	18.060	NA	4.270
GDP*	NA	-6.912	NA	NA
% agric. land*	0.016	0.184	0.048	NA
GDP growth x GDP	NA	-21.065	NA	NA
Governance	0.020	NA	0.020	NA
Spending x GDP	NA	15.869	NA	2.775
Spending x threatened species richness	-0.397	4.719	-0.167	NA
Population growth x GDP growth	NA	NA	NA	0.743
Agric. growth x % agric. land	-0.049	-9.767	-0.049	NA
Spending x % agric. land	0.065	NA	0.075	NA
Population growth x population density	NA	2.325	NA	NA
Agric growth x governance improvement	NA	-9.603	NA	NA

Best-fit models that used alternative values of the threshold  $T$  are shown. For interacting variables (marked \*), the coefficients shown cannot be interpreted by reading the table (see Supplementary Results for their interpretation). Agric., agricultural; t-1, 1994–2000; t-2, 1988–1994; GDP, gross domestic product per capita at PPP; Population, rural population density; governance improvement, change in the government effectiveness score. For  $T=0.10$ , sample size increased to  $n=53$  independent countries in the continuous part of the model (index parameter = 1.99) and the ratio of ones to zeroes was 44:65 in the binomial part of the model. For  $T=0.25$ ,  $n=43$  independent countries and the ratio was 37:74.

Extended Data Table 3 | Cross correlations between variables

	\$\$	GDP grow	Agric. grow	Pop grow	Gov grow	GDP	% Agric land	Pop	Gov	Decl (t-1)	Decl (t-2)	Spp. rich	For. loss	Area
\$\$	1.00	0.10	-0.11	-0.14	-0.11	0.54	0.17	-0.09	0.36	0.14	0.46	0.35	-0.13	0.39
GDP growth	0.10	1.00	0.04	0.34	-0.12	0.14	-0.23	0.08	0.22	0.07	-0.05	0.04	-0.03	0.01
Agric. growth	-0.11	0.04	1.00	0.14	-0.10	-0.02	0.19	-0.06	0.08	-0.01	-0.05	-0.05	0.10	-0.06
Pop. growth	-0.14	0.34	0.14	1.00	0.16	-0.16	-0.18	-0.04	-0.25	0.02	0.07	0.04	0.15	0.14
Gov. growth	-0.11	-0.12	-0.10	0.16	1.00	-0.09	-0.26	-0.01	-0.16	0.08	0.16	0.08	-0.08	-0.08
GDP	0.54	0.14	-0.02	-0.16	-0.09	1.00	0.10	-0.02	0.20	0.11	0.19	0.33	-0.15	0.36
% Agric. land	0.17	-0.23	0.19	-0.18	-0.26	0.10	1.00	0.20	-0.02	-0.13	-0.01	-0.15	0.07	0.14
Pop.	-0.09	0.08	-0.06	-0.04	-0.01	-0.02	0.20	1.00	0.04	-0.03	-0.20	-0.09	-0.12	-0.40
Gov.	0.36	0.22	0.08	-0.25	-0.16	0.20	-0.02	0.04	1.00	-0.02	0.00	-0.02	-0.29	-0.28
Decl. (t-1)	0.14	0.07	-0.01	0.02	0.08	0.11	-0.13	-0.03	-0.02	1.00	0.30	0.69	0.10	0.17
Decl. (t-2)	0.46	-0.05	-0.05	0.07	0.16	0.19	-0.01	-0.20	0.00	0.30	1.00	0.53	0.02	0.38
Spp. rich	0.35	0.04	-0.05	0.04	0.08	0.33	-0.15	-0.09	-0.02	0.69	0.53	1.00	0.03	0.38
For. loss	-0.13	-0.03	0.10	0.15	-0.08	-0.05	0.07	-0.12	-0.29	0.10	0.02	0.03	1.00	0.16
Area	0.39	0.01	-0.06	0.14	-0.08	0.36	0.14	-0.40	-0.28	0.17	0.38	0.38	0.15	1.00

\$\$, conservation spending at PPP; Agric., agricultural; Pop, population density; Gov, governance effectiveness indicator; Decl, declines; Spp. Rich, threatened species richness; For. Loss, percentage of forest loss; Area, country area.

**Extended Data Table 4 | Variance inflation factors for the continuous and binomial parts of the model**

	<b>Continuous model part</b>	<b>Binomial model part</b>
Spending	1.61	1.44
GDP growth	1.63	1.00
Agric. growth	1.19	1.10
Pop. Growth	NA	1.33
Gov. growth	NA	1.07
GDP	1.15	1.71
% Agric. land	1.84	1.33
Decline (t-1)	2.81	NA
Decline (t-2)	1.77	NA
Spp. Rich	2.89	1.72
Area	NA	1.52

Spending, conservation spending at PPP; Agric., agricultural; Pop, population density; Gov, governance effectiveness indicator; Spp. Rich, threatened species richness; Area, country area.

## Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see [Reporting Life Sciences Research](#). For further information on Nature Research policies, including our [data availability policy](#), see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

### ► Experimental design

#### 1. Sample size

Describe how sample size was determined.

Since the study analyses as many countries as data are available for, we did not use statistical tests to pre-determine sample size. Nevertheless, we did use power analysis to confirm that the number of countries analysed was sufficiently large for the regression model structures tested

#### 2. Data exclusions

Describe any data exclusions.

We excluded any countries for which complete, robust data were lacking (see reference 8), including where reported finance commitments cannot be safely regarded as strict-sense biodiversity spending. We also excluded countries that had multiple overseas territories but where conservation spending was not disaggregated across those territories, despite strongly varying values for the socioeconomic predictors and rates of decline.

#### 3. Replication

Describe whether the experimental findings were reliably reproduced.

Since we studied countries and the biodiversity losses they have experienced, no replication was performed.

#### 4. Randomization

Describe how samples/organisms/participants were allocated into experimental groups.

Not applicable to our study, which predicts biodiversity change and its relationship to conservation across world countries.

#### 5. Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

Not applicable to our study

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.



## 6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

n/a Confirmed

- ☐ ☒ The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
- ☒ ☐ A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- ☒ ☐ A statement indicating how many times each experiment was replicated
- ☐ ☒ The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
- ☐ ☒ A description of any assumptions or corrections, such as an adjustment for multiple comparisons
- ☐ ☒ The test results (e.g.  $P$  values) given as exact values whenever possible and with confidence intervals noted
- ☐ ☒ A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
- ☐ ☒ Clearly defined error bars

See the web collection on [statistics for biologists](#) for further resources and guidance.

## ► Software

Policy information about [availability of computer code](#)

## 7. Software

Describe the software used to analyze the data in this study.

We used the R statistical program to analyse the data (version 3.2.2), with CRAN packages mgcv, MuMIn, cplm, Tweedie, CPLS.

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* [guidance for providing algorithms and software for publication](#) provides further information on this topic.

## ► Materials and reagents

Policy information about [availability of materials](#)

## 8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.

No physical materials. Our study analyses countries and their rates of biodiversity decline.

## 9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).

No antibodies were used

## 10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.

No eukaryotic cell lines were used

b. Describe the method of cell line authentication used.

No eukaryotic cell lines were used

c. Report whether the cell lines were tested for mycoplasma contamination.

No eukaryotic cell lines were used

d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by [ICLAC](#), provide a scientific rationale for their use.

No commonly misidentified cell lines were used

## ► Animals and human research participants

Policy information about [studies involving animals](#); when reporting animal research, follow the [ARRIVE guidelines](#)

## 11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.

No animals were used

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.

The study did not involve human research participants