Reductions in global biodiversity loss predicted from conservation spending

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

Affiliations:
1. Edward Grey Institute, Department of Zoology, Oxford University, Oxford OX1 3PS, UK.
2. Department of Natural Resources and Environmental Science, University of Illinois, Urbana-Champaign, IL61801, USA.
3. Department of Genetics, Evolution and Environment, University College London, London WC1E 6BT, UK.
4. Biology Department, Simon Fraser University, Burnaby, BC V5A 1S6, Canada.
5. Scimitar Scientific, Whitehorse, YT, Y1A 6V6, Canada.
6. Warnell School of Forestry & Natural Resources, University of Georgia, Athens, GA 30602, USA.
7. Institute for the Study of Environment and Society, Brown University, Providence, RI 02912, USA.
8. Department of Life Sciences, Imperial College London, Silwood Park, Buckhurst Road, Ascot, Berkshire, SL5 7PY, UK.
9. Odum School of Ecology, University of Georgia, Athens, GA, 30602, USA.

Correspondence and requests for materials should be addressed to anthony.waldron@zoo.ox.ac.uk/ anthonywaldron@hotmail.com

Keywords: Conservation, Aichi targets, Sustainable Development Goals, Convention on Biological Diversity, conservation finance, conservation impact, conservation effectiveness, economic development, IUCN Red List, extinction

Halting global biodiversity loss is central to both the Convention on Biological Diversity (CBD) and United Nations Sustainable Development Goals (SDGs)[1,2], but success to date has been very limited[3–5]. A critical determinant of overall strategic success (or failure) is the financing committed to biodiversity[6–9]; however, financing decisions are still hindered by considerable uncertainty over what any investment is likely to achieve[6–9]. For greater effectiveness, we need an evidence-based model (EBM)[10–12] showing how conservation spending quantitatively reduces the rate of loss. Here, we empirically quantify how i$14.4 billion of conservation investment reduced biodiversity loss across 109 signatory countries between 1996 and 2008, by an average 29% per country. We also show that biodiversity change in signatory countries can be predicted with high accuracy, using a dual model that combines the positive impact of conservation investment with the negative impact of economic, agricultural and population growth (i.e. human development pressures)[13–18]. Decision-makers can use this dual model to forecast the improvement that any proposed biodiversity budget would achieve under various scenarios of human development pressure, comparing those forecasts to any chosen policy target (including the CBD and SDGs). Importantly, we further find that spending impacts shrink as human development pressures grow, implying that funding may need to increase over time. The model therefore offers a flexible tool for balancing the SDGs of human development and biodiversity, by predicting the dynamic changes needed in conservation finance as human development proceeds.

The rapid loss of global biodiversity has major consequences for human wellbeing[5,19] and so governments worldwide have committed reducing those losses through multiple international agreements, including the CBD and SDG frameworks[1,2]. However, strategic outcomes to date have been poor: we missed the 2010 CBD target and now seem likely to also miss the 2020 Aichi biodiversity targets[3,4]. As outlined in Aichi target 20 and SDG17, one of the most important determinants of policy success is our ability to correctly decide (and secure) the level of financing needed to resource overall biodiversity-conservation strategies[1,2,6–8]. A second key way to improve on currently poor outcomes is to take a more evidence-based approach, in which decision making is guided by reliable evaluations of past successes and failures ("conservation impact assessments")[10–12]. In many fields, the financing of strategic goals is fundamentally evidence-based, analysing previous spending outcomes to guide current budget decisions[20,21]. 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[15,22] (as already occurs with climate change[23]). In bio-political science, predictive models typically quantify how biodiversity loss is driven by human socioeconomic pressures, such as economic or agricultural expansion[14–16,24]. To date, conservation impact assessments and predictive decline models have largely developed as separate major fields, despite their outcomes being strongly interdependent. It is rarely possible to accurately measure the impact of one factor (either spending or pressures) on biodiversity without accounting for the influence of the other factor[3,25]. To make accurate predictions for policy use, we therefore need unified models that treat biodiversity change as the simultaneous outcome of pressures and their impact, plus conservation and its impact (henceforth, "pressures-and-conservation-impact (PACI) models"). Indeed, one of the core challenges for the SDGs is to balance (or trade off) the often-conflicting goals of human development (e.g. SDGs 1, 2 & 8) and biodiversity conservation (SDG 15)[2,14–18,24]. To measure this trade-off, policymakers need models that unite these two aspects. Finally, such models need to apply to the key geopolitical decision-making scale for the CBD and SDGs – sovereign countries – demanding finer geographic resolution than common planet-scale approaches[3,7].

Here, we use empirical evidence to develop a unified PACI model at the sovereign country scale, by statistically quantifying how changing human pressures drive biodiversity decline while conservation spending reduces it. As such, the model informs policymakers not just what biodiversity losses to expect but more constructively, how changes in conservation resourcing can reduce those expected losses[3]. We also show how the impacts of spending and pressures depend predictably upon national socioeconomic contexts, and thus how they may change over time.

A standard policy measure of biodiversity change (usually, decline) is the planet-scale sum of all changes in individual species' IUCN Red List status, using well-known taxa as a proxy for biodiversity[3,26]. To calculate biodiversity change at the decision-making scale of sovereign signatory countries (hereafter each country's "biodiversity decline score" or BDS), we took Red List status changes for all global bird and mammal species for 1996–2008 (see Methods for justification and details) and portioned them out among all countries where each species is found (treating the few status improvements as negative fractions). We then summed all decline fractions for each country to calculate BDSs[8,26] (Figure 1, Supplementary Table 1). It is noteworthy that 60% of total BDS for the globe was found in only seven countries: Indonesia, Malaysia, Papua New Guinea, China, India, Australia, and the USA (principally Hawai'i). Seven countries had net biodiversity improvements (negative BDSs): Mauritius, Seychelles, Fiji, Samoa, Tonga, Poland and Ukraine. (See Extended Data Figure 1 for average BDS per species).

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 BDS, using national-level data on strict-sense conservation spending (annualised, see Methods) plus the broad socioeconomic pressures of GDP growth, 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 zero-inflated due to multiple species-poor countries with no status changes, so we used two-part models[27] 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 decline (i.e. BDS, Figure 2), whereas GDP growth and agricultural expansion tended to increase it (Figure 3). Although forest loss was often significant, the best-fitting predictive model favoured more generalized terms (Table 1, Supplementary Discussion). Interaction terms revealed several context-dependent nuances (see Supplementary Discussion). The GDP growth effect decreased as baseline GDP decreased, becoming non-significant in the poorest countries (Figure 3). Agricultural expansion had a deleterious impact 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 (Figure 3). The binomial part also suggested that the impact of agricultural expansion could be greatly reduced by improvements in the quality of national governance (Extended Data Figure 2), and that the deleterious impact of GDP became stronger as human population growth increased, i.e. the combined impact of two pressures was greater than the sum of its parts (Table 1). Finally, conservation spending was more effective in poorer countries than in higher-income ones, and spending also had a greater impact when more species were threatened in the first place (Extended Data Figure 3).

Both model parts accurately predicted historical declines ($R^2 = 0.85$ in the continuous part; accuracy = 94% in binomial part; Extended Data Figure 4) and were robust to several sensitivity tests (Supplementary Results, Extended Data Table 4). They also had high forecasting accuracy in cross-validation (82% continuous part; 85% binomial part). Our PACI models therefore have immediate application to several major policy needs. They can predict not only future biodiversity declines[15,22], 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 impact of spending an extra i$5 million in each country (such that the overall global annual budget was increased by 42%, Supplementary Table 3). Outcomes for all countries are shown in Supplementary Table 3 (see also Figure 1) but to give an example: in the mega-diverse countries of PNG and Peru, the model predicted reductions in decline (BDS) of 33% and 54% respectively. We also used the model to back-predict how much biodiversity loss was prevented by post-Earth Summit conservation financing[8,28], estimating that on average (median), losses per country were 29% less than would otherwise have occurred (Methods).

The model could also be used to predict the funding each country needs to achieve specific biodiversity policy goals, including the CBD and SDG targets. Importantly, however, our results demonstrate how the cost of meeting any target constantly changes as the levels of socioeconomic pressure change. For example, if Peru had wanted to achieve 50% less decline by 2008, then with pressures at their 1992-2003 levels, the model predicts that an extra $4.6m annually would have been needed annually. However, at current (2001-2012 mean) levels of pressure, that figure would rise to $5.7m (constant international dollars). Our model explicitly accounts for these changes in socioeconomic context, and so an appropriate policy use would be to take various scenarios of economic, agricultural and population change, and then predict biodiversity outcomes at different funding levels for each scenario, comparing them to targets. In particular, the model can be used to help resolve problems of discordance between the SDGs for biodiversity and human development, by quantifying how any negative effects of economic and agricultural growth can be balanced out by short-term increases in conservation funding (thereby creating a breathing space to develop more sustainable pathways to national growth[18].)

We caution that an unmeasured variable correlated with conservation spending could conceivably explain some of the spending impact; that the co-benefits of spending for taxa other than birds and mammals remain unknown; that species declines too small to affect Red List status will not be accurately predicted and will require different approaches[29]; and that long-distance effects such as Chinese demand for African ivory[30] were beyond the scope of our model. However, our general PACI approach should be flexible enough to accommodate such additions in the future.

At a time when the outlook for biodiversity often seems very bleak[4,5], our results present a constructive opportunity for global biodiversity policy, showing how increases in conservation investment can lead to major, quantifiable improvements. However, set against this note of optimism, our model also underlines how conservation spending may need to constantly increase (or evolve) to counterbalance the continuing intensification of human development pressures[5,18,24]. By empirically demonstrating how limited levels of investment have already led to a partial reduction in biodiversity loss, our findings may ultimately encourage decision-makers to commit the full finance needed[7] to significantly reduce or halt global losses, in line with our CBD and SDG commitments[1,2].

Acknowledgments: This research was supported by UKDWP (A.W.), the USDA National Institute of Food and Agriculture Hatch project 1009327 (A.W. and D.C.M.), 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.O.M.), the UK Natural Environment Research Council (J.A.T.) and the Odum School of Ecology (J.L.G.). We thank J. Drake, P. Holland and P. Stephens and four anonymous referees for comments on earlier manuscripts. Detailed methods, additional results and data are available in the online supplementary material.

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

Author Statement: Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to anthonywaldron@hotmail.com.

REFERENCES

1. CBD. Conference of the Parties Decision X/2: Strategic plan for biodiversity 2011-2020. (2010). Available at: www.cbd.int/decision/cop?id=12268.
2. UN. *Transforming our world: the 2030 agenda for sustainable development*. (2015).
3. Hoffmann, M. *et al.* The impact of conservation on the status of the world's vertebrates. *Science* **330,** 1503–1509 (2012).
4. Tittensor, D. P. *et al.* A mid-term analysis of progress toward international biodiversity targets. *Science* **346,** 241–244 (2014).
5. Pimm, S. L. *et al.* The biodiversity of species and their rates of extinction, distribution, and protection. *Science* **344,** 1246752 (2014).
6. James, A. N., Gaston, K. J. & Balmford, A. Balancing the Earth's accounts. *Nature* **401,** 323–324 (1999).
7. McCarthy, D. P. *et al.* Financial costs of meeting global biodiversity conservation targets: current spending and unmet needs. *Science* **338,** 946–949 (2012).
8. Waldron, A. *et al.* Targeting global conservation funding to limit immediate biodiversity declines. *PNAS* **110,** 12144–12148 (2013).
9. UNEP/CBD. *Decisions adopted by the conference of the parties to the Convention on Biological Diversity at its 11th meeting*. (2013).
10. Sutherland, W. J., Pullin, A. S., Dolman, P. M. & Knight, T. M. The need for evidence-based conservation. *Trends Ecol. Evol.* **19,** 305–8 (2004).
11. McKinnon, M., Cheng, S. H., Garside, R., Masuda, Y. J. & Miller, D. C. Sustainability: map the evidence. *Nature* **528,** 185–187 (2015).
12. Miteva, D., Pattanayak, S. & Ferraro, P. J. Analysis of biodiversity policy instruments: what works, and what doesn't. *Oxford Rev. Econ. Policy* **28,** 69–92 (2012).
13. Cardillo, M. *et al.* The predictability of extinction: biological and external correlates of decline in mammals. *Proc. Biol. Sci.* **275,** 1441–8 (2008).
14. Naidoo, R. & Adamowicz, W. L. Effects of economic prosperity on numbers of threatened species. *Conserv. Biol.* **15,** 1021–1029 (2001).
15. Freytag, A., Vietze, C. & Volkl, W. What drives biodiversity? An empirical assessment of the relation between biodiversity and the economy. *Int. J. Ecol. Econ. Stat.* **24,** 1–16 (2012).
16. Smith, R., Muir, R., Walpole, M., Balmford, A. & Leader-Williams, N. Governance and the loss of biodiversity. *Nature* **426,** 67–70 (2003).
17. Roe, D., Elliott, J., Sandbrook, C. & Walpole, M. *Biodiversity conservation and poverty alleviation*. (Wiley-Blackwell, 2013).
18. Adams, W., Aveling, R. & Brockington, D. Biodiversity conservation and the eradication of poverty. *Science* **306,** 1146–1149 (2004).
19. Cardinale, B. J. *et al.* Biodiversity loss and its impact on humanity. *Nature* **486,** 59–67 (2012).
20. Schwartländer, B. *et al.* Towards an improved investment approach for an effective response to HIV/AIDS. *Lancet* **377,** 2031–2041 (2011).
21. HM Treasury. Public sector business cases: using the five case model. Green Book Supplementary Guidance on Delivering Public Value from Spending Proposals. *HM Treasury* 1–152 (2013). doi:10.1007/s13398-014-0173-7.2
22. Collen, B. & Nicholson, E. Taking the measure of change. *Science* **346,** 166–167 (2014).

23. Stern, N. *The economics of climate change: the Stern review*. (Cambridge University Press, 2007).
24. 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).
25. Ferraro, P. J. *et al.* Estimating the impacts of conservation on ecosystem services and poverty by integrating modeling and evaluation. *Proc. Natl. Acad. Sci.* **112,** 7420–7425 (2015).
26. Rodrigues, A. S. L. *et al.* Spatially Explicit Trends in the Global Conservation Status of Vertebrates. *PLoS One* **9,** e113934 (2014).
27. Basu, A. & ManningWG. Issues for the next generation of health care cost analyses. *Med Care* **47,** S109-14 (2009).
28. Miller, D. C., Agrawal, A. & Timmons Roberts, J. Biodiversity, governance, and the allocation of international aid for conservation. *Conserv. Lett.* **6,** 12–20 (2013).
29. Donald, P. F. *et al.* International conservation delivers benefits for birds in Europe. *Science* **317,** 810–813 (2007).
30. Vandegrift, J. Elephant poaching: CITES failure to combat the growth in Chinese demand for ivory. *Virginia Environ. Law J.* **31,** 102 (2013).

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. UNSD. Sustainable Development Goals indicators. (2016). Available at: http://unstats.un.org/sdgs/. (Accessed: 15th October 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. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **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 World Bank. Global Purchasing Power Parities and Real Expenditures. 2005 International Comparison Programme. (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,** 1358–66 (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*. (2010). Available at www.govindicators.org.
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,** 17–37 (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. Majumder, 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. Effects of human activity on global extinction risk. *Conserv. Biol.* **9,** 1528–1538 (1995).
54. Meyfroidt, P. & Lambin, E. 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.* **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. World Bank. The World Bank Databank. (2013). Available at: databank.worldbank.org.
58. FAO and JRC. *Global forest land-use change 1990-2005. FAO Forestry Paper 169, Food and Agriculture Organization of the United Nations.* (2012).
59. FAO. *Global Forest Resources Assessment 2010: Main Report: FAO Forestry Paper 163, Food and Agriculture Organization of the United Nations*. (2010).
60. Hansen, M. C. *et al.* High-Resolution Global Maps of 21st-Century Forest Cover Change. *Sci.* **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 (2011).
64. Zuur, A. F., Ieno, E. N., Walker, N. J., Savaliev, 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. *TREE* **11,** 193–196 (1996).
66. Cragg, J. 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 and Hall, 1997).
69. Tweedie, M. An index which distinguishes between some important exponential families. in *Statistics: Applications and New Directions. Proceedings of the Indian Statistical Institute Golden Jubilee International Conference* 579–604 (Indian Statistical Institute, 1984).
70. Zhang, W. cplm: Compound Poisson Linear models version 0.7-2. Available at the Comprehensive R Archive Network (CRAN). (2014).
71. Burnham, K. P. & Anderson, D. R. *Model selection and multimodel inference: a practical information-theoretic approach*. (Springer-Verlag, 2002).
72. Cohen, J. *Statistical power analysis for the behavioural sciences*. (Laurence Erlbaum, 1988).
73. Ferraro, P. & Hanauer, M. Advances in measuring the environmental and social impacts of environmental programs. *Annu. Rev. Environ. Resour.* 495–517 (2014).
74. Imai, K. & Ratkovic, M. Covariate Balancing Propensity Score. *J. R. Stat. Soc. B* **76,** 243–263 (2012).
75. Fong, C., Ratkovic, M. & Imai, K. CBPS: R package for covariate balancing propensity score. available at the Comprehensive R Archive Network (CRAN). (2015).
76. Pandit, R. & Laband, D. N. Spatial autocorrelation in country-level models of species imperilment. *Ecol. Econ.* **60,** 526–532 (2006).
77. Amin, A. & Choumert, J. Development and biodiversity conservation in Sub-Saharan Africa: A spatial analysis. *Econom. Bull.* **35,** 729–744 (2015).
78. Drukker, D. M., Prucha, I. R. & Raciborski, 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. (2013).
80. Fox, J. *An R and S-Plus companion to applied regression*. (Sage Publications, 1997).
81. IUCN. *The IUCN Red List of threatened species*. (2009).
82. Schipper, J. *et al.* The status of the world's land and marine mammals: diversity, threat, and knowledge. *Science* **322,** 225–30 (2008).
83. Birdlife International. Species distribution data download. (2014).
84. Sandvik, B. World Borders Dataset. (2009). Available at: http://thematicmapping.org/downloads/world_borders.php.
85. Pebesma, E. J. & Bivand, R. S. Classes and methods for spatial data in R. *R News* **5,** (2005).
86. Bivand, R. . & Kewin-Koh, N. Maptools: Tools for reading and handling spatial objects. *R package version 0.8-27* (2013). Available at: http://cran.r-project.org/pacakge=maptools.
87. Balmford, A., Gaston, K. J., Blyth, S., James, A. & Kapos, V. Global variation in terrestrial conservation costs, conservation benefits, and unmet conservation needs. *PNAS* **100,** 1046–1050 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

FIGURE LEGENDS

Figure 1 | Global biodiversity declines and conservation spending impacts. Colours show percentage of all global declines (total BDS) associated with each country: dark red = >10% (Indonesia only, 21%); dark, mid and light orange = 5–10, 2.5–5 and 1–2.5% respectively; yellow = 0–1%; grey indicates BDS = 0; blue indicates a net improvement in national biodiversity status. Pies show the predicted reduction in decline (in black) if spending had been i$5million higher (for selected countries); pie size represents $\sqrt{}$BDS. Inset shows predicted vs. observed BDS (ln-transformed) for the continuous model (see also Extended Data Figure 4). Country outlines supplied by esri_dm https://www.arcgis.com/home/item.html?id=d86e32ea12a64727b9e94d6f820123a2#overview

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

Figure 3 | Conditional impacts of human pressures on biodiversity. (a) Impact of GDP growth on BDS depends on the existing level of GDP/capita. Red = slow GDP growth (10^{ile}), blue = fast growth (90^{ile}), "low" GDP/capita = 10^{ile}, "median" = 50^{ile} (effects are still significant at >50^{ile}). (b) Impact of agricultural expansion on BDS depends on the existing % of land converted to agriculture: colours as in (a), "low" agricultural expansion = 10^{ile}, "median" = 50^{ile} (effects are still non-significant at <50^{ile}). Error bars show conditional 95% confidence intervals from the continuous model-part. N=50 independent countries. Centre is the median.

Table 1 | Best-fit models predicting biodiversity decline. Note that for all terms that interact, the interaction plots provided must be used to interpret the reported standardised coefficients correctly (Figure 3 and Extended Data Figures 2-3). "Agric. land" = mean percentage of agricultural land; t-1 = 1994-2000, t-2 = 1988-1994; GDP = Gross domestic product per capita PPP; population = rural population density; governance improvement = change in the government effectiveness score. N=50 independent countries and index parameter = 1.01 in the continuous part, n=109 countries in binomial part with a 42:67 ratio of ones to zeroes.

Predictor variable	Continuous model part (BDS)	Binomial model part (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
% agric. land	0.049	-1.226
GDP growth x GDP	0.031	NA
Spending x GDP	NA	5.026
Spending x threatened species richness	-0.247	NA
Population growth x GDP growth	NA	1.044
Agric. growth x % agric. land	-0.045	-10.143
Spending x % agric. land	0.065	NA
Agric growth x governance improvement	NA	-9.603

Materials and Methods

Country-scale biodiversity decline scores

To quantify biodiversity decline, we used equally-weighted genuine changes in the IUCN Red List status of all global bird and mammal species up to the last Global Mammal Assessment in 2008 (i.e. changes in extinction risk between 1996-2008 for mammals and 2000-2008 for birds, there being no 1996 global bird assessment; the term "genuine" excludes any Red List changes not related to changing extinction risk, in particular those due simply to taxonomic changes)[3,31,32]. Our approach is therefore similar to planet-scale Red List Indices (RLI) of global biodiversity change adopted by governments to measure performance against CBD and SDG targets[3,31–33], 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 received the vast majority of conservation investment and supply robust, directly-observed data on changes in Red List status[3,34]; thus, we excluded the other possible taxon (amphibians) because they received almost no conservation investment during the study period[3], only have modelled (rather than directly observed) declines available for 1980-2004[3,35] (whereas robust spending data are only available from 1992 onwards[8]), and are also highly data deficient and "enigmatic" in terms of their declines[3,35].

To convert species-based Red List changes into country-level indices of biodiversity change, we divided up each species change as "decline fractions", based on the percentage of the species range p_{ij} held by each country[8,26]. However, decline fractions are estimates for the underlying responsibility fraction R_{ij} = the proportion of the status change for species *i* attributable to country *j* (see Additional Method Details at end). For greater accuracy, we therefore corrected these range-based fractions in two ways. First, the losses underlying a species decline are not homogeneously distributed in space but instead, are frequently concentrated in some part of the range where human pressures have suddenly increased[36]. Both empirically and at random, those concentrations of pressure-driven loss are unlikely to lie at the range periphery (Additional Method Details and[36]). However, a raw range-based algorithm assumes spatially homogeneous losses right up to the range periphery, and so will often assign an erroneous and trivial responsibility fraction to any country holding a small range-edge (p_{ij}) of a species found almost entirely in a neighbouring country. Formally, R_{ij} for small p_{ij} is often but not always likely to be zero. These small p_{ij} values were also extremely numerous, generating very high noise-to-signal ratios in analysis. To address these problems of extreme signal loss and bias when an unknown proportion of small p_{ij} were incorrect overestimates of zero R_{ij}, we used Signal Detection Theory[37] and the mathematics of the Red List categories to estimate a range of theoretically optimal thresholds T such that R_{ij} is set to zero if $p_{ij}<T$, and then carried out our analyses using three possible thresholds within this range, to account for uncertainty (see Additional Method Details). The main text shows results for T=0.17 (being the approximate optimal trade-off between noise reduction and sample size, Additional Method Details) and Supplementary Results and Extended Data Table 4 shows sensitivity tests with alternative thresholds (including the finding that explanatory power at T=0.17 is considerably stronger than occurs with the other thresholds).

Second, we analysed the Red List reports for each individual bird and mammal species and altered the range-based fractions wherever a report suggested a different distribution of responsibilities across countries (Supplementary Table 4). We then calculated the Biodiversity Decline Score (BDS) for each country by summing all decline fractions for birds and mammals, treating the rare status improvements as negative fractions[8,26]. Supplementary Table 1 contains the final BDS scores per country.

Predictors of country-scale biodiversity decline scores (BDS)

Conservation models with policy relevance need to have general applicability, including being able to accurately forecast outcomes when presented with situations that are different from the original dataset on which they were parameterised. To achieve this, it is highly advisable to use broad, general variables because more specific ones often have very poor forecasting performance when used beyond the original data[37]. We therefore selected a relatively small set of simple, generalised and publicly available explanatory variables to represent national-level socioeconomic pressures, noting that 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[8], adding new data for countries that had been data-deficient in the original published study e.g. Turkey[38]. Finance data were collated at 2005 constant U.S. dollar values (consistent with[8]) but for analysis, were converted to "international dollars" (abbreviated as i$ in the main text) at local purchasing power parity values, where purchasing power parity accounts for differences in the purchasing power of U.S. dollars (when exchanged) in each country[39]. Two types of conservation investment data were available: (a) "strict-sense" funding with direct links to biodiversity conservation, and (b) "mixed funding" mainly targeted at social and development goals but with potential indirect, long-term, and often unclear impacts on biodiversity (e.g. school-building or agricultural assistance in forest communities)[28,40]. *A priori*, we hypothesized that strict-sense biodiversity funding was likely to be the better predictor of rates of decline, whereas "mixed" development funding (which involves much larger sums than strict-sense funding) was likely to obscure any effect. "Strict-sense" funding also produced lower 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 (e.g. through reducing conflict) and indirectly (e.g. through making conservation investment more efficient)[16,41–43]. Governance has been measured using multiple indicators[44], so we modelled the impact of change in the six indicators published in the World Governance Indicators dataset[44]: 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 each other ($r>0.9$ for all pairwise combinations) and so to avoid collinearity, we tested each one individually. Government effectiveness gave the best fit in exploratory analysis (as in[8]) and is reported in the results as "governance".

For the country-level pressures aspect of our PACI model, we followed previous authors in using national rates of human population growth, economic growth and agricultural expansion[13–15,17,45–53]. Such country-level aggregators likely capture the overall impact of multiple smaller-scale drivers (with agriculture being the main pressure driving threat[24]). For example, forest clearance for food production or commodities would generally cause changes in both area of agricultural land and economic output, and GDP levels have been associated with both hunting pressure and deforestation trends[54–56]. For economic growth, we used change in GDP/capita PPP (purchasing power parity). For agricultural growth, we used change in the percentage of land converted to agriculture; and 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 data from World Bank statistical tables[57]. We also tested the direct impact of forest loss, estimated per country for 1990-2000 using FAO statistics[58,59] (although we acknowledge the limitations of this historical dataset[60]).

The number of declining species in a country (and hence its BDS) is likely to be strongly influenced by the total number of species present and/or country area, plus the starting-condition levels of risk and decline. Following previous studies (e.g.[61]), we calculated total threatened species richness in the same way as we calculated total species decline (BDS), i.e. we summed all species fractions in each country, weighting them by the level of extinction risk as an index of threat. We compiled country area from[8]. However, in exploratory analysis, we found that the inclusion of area in any continuous-part model consistently led to a worse fit (delta AICc >6.5), likely because species richness absorbed most of the variance explained by area in this (n=50 countries) sample. In contrast, binomial-part models (n=109 countries) detected separate area and species richness effects (without collinearity; Extended Data Tables 2–3). Thus, we included the area term in binomial 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/action takes at least 5 years, and often over a decade, to have an impact on biodiversity[29,62], especially for taxa such as birds and mammals. For mammals, the two global Red List assessments from which status changes can be calculated were in 1996 and 2008[3]. We therefore assumed that changes detected in the 2008 assessment may have been driven by conservation finance allocations occurring as recently as five years earlier (i.e. 2003) but in all likelihood, could also be influenced by spending from a decade or more earlier (the early 1990s). Similarly, changes occurring after 1996 (i.e. starting in 1997) could have been influenced by spending allocations as early as 1992 (also the year in which global conservation spending began in earnest with the Rio Earth Summit[28]) but also by allocations up to the early 2000s. Following this logic, we used predictor variables for 1992-2003 (annualised values) to model changes in the response value for 1996-2008, using the same lag for the four different socioeconomic growth variables to avoid the analysis becoming intractable. We tested an alternative predictor period of 1992-2000 but preferred 1992-2003 based on lower AICc values.

Technically, therefore, our response variable is a lagged variable[63] taking the form $Y_t – Y_{t-n}$ and our socioeconomic change variables are similarly lagged. We acknowledge that predicting change occurring in a time block using variables from an earlier time block is necessarily approximate, but year-by-year species changes were not available. Nevertheless, country-level patterns of change in predictor variables were strongly correlated across different time periods (e.g. when comparing mean annual values for 1992-2000 and 1992-2003, the correlations for population growth, population size, GDP growth and GDP respectively are 0.91, 0.999, 0.89, and 0.999). These strong correlations imply that the precise choice of year/period seems unlikely to have 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, we calculated avian BDSs for the two IUCN assessment periods preceding our study period (1988-1994 and 1994-2000) and added both measures to our candidate regression models. No earlier-period BDS was available for mammals; however, mammal and bird BDS are highly correlated in the study period (Pearson's r = 0.998), so we assumed earlier-period bird BDSs to be reasonable proxy of combined (bird + mammal) earlier-period BDSs.

Statistical analysis

All predictor variables were z-standardized to put effect sizes on a common scale[64]. We excluded any countries for which complete, robust data were lacking (see[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 conservation spending was not disaggregated across those territories, despite strongly different values for the socioeconomic predictors and rates of decline across the territories. In particular, the USA, France and the UK were excluded from regression models under this rubric (and we therefore recommend greater geo-referenced finance reporting). See Supplementary Table 1 for all exclusions. The Solomon Islands and New Zealand represented potentially influential leverage points, so we tested models both with and without these countries. We found that inclusion of the Solomon Islands had a large impact on binomial outcomes (causing governance growth to be dropped from the best-fit binomial-part model, likely due to the extreme value of governance growth for the Solomons), so we excluded this country from all binomial models. The impact of including the Solomon Islands was smaller in the continuous part (an identical best-fit model with similar coefficients was selected whether the country was included or excluded) but for completeness, we consistently tested all continuous model variations both with and without the Solomons. Inclusion of New Zealand had a major impact on binomial-part outcomes, altering most coefficients by~20% and some by >100%, and also greatly worsened fit in the continuous part, so it was excluded overall. The leverage associated with including New Zealand may be due to this country having a negative value for agricultural growth.

We then built candidate PACI models to predict BDS, each testing hypotheses about how conservation investment and various human pressures might impact 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 may either have a reduced marginal effect on biodiversity due to an extinction filter[65], or a greater impact as the last vestiges of habitat disappear (Supplementary Discussion). Thus, we further calculated mean annual values of GDP, population, governance and % 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 covering both positive and negative values, but had a relatively dense cloud of values at zero. Although a more limited number of zeroes does not violate regression assumptions, such a long tail of zeroes can generate extreme bias[64]. We therefore used the recommended approach of a two-part model[27,66] that creates (a) a "continuous" part (n=50 countries) comprising all countries with a non-zero BDS plus informative zeroes; (b) a "binomial" part (n=109) that included all countries with data (and so all zeroes), but converted BDS to the binary response BDSb (where BDSb = 1 if BDS>0 and 0 otherwise). For the continuous part specifically, we sought to optimise the trade-off between information content and bias by including as many zeroes as possible, in order of their likely informativeness, without causing clear patterns in regression diagnostic plots (thus extending the principle of the hurdle models developed for non-negative integer data[64] to two-part analyses). A country that has many species but has experienced no declines, such as Costa Rica, suggests an important underlying process captured by zero BDS (= higher informativeness of zero decline). Conversely, when a country is species-poor, there is a strong random expectation that over a 13-year period, no species will be observed changing its Red List status (= lower informativeness of the zero). We therefore defined Ψ as country-level species richness (derived from our prior geographic analysis) and then, for various possible values of this parameter, heuristically tested the degree of regression bias arising when we excluded all cases of {BDS=0 and species richness < Ψ}. We found a tradeoff whereby setting Ψ at 40 or more left minimal patterns in residual plots but reduced sample size and statistical power, whereas Ψ values below 20 started to generate strong patterns in plots of residuals against fitted values. We therefore chose a value of $\Psi = 25$ (see Supplementary Results and "Sensitivity Testing" (below) for sensitivity testing on this parameter).

For the continuous part, BDS retained a right skew even after log-transformation (Extended Data Figure 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[67–69] (cplm R package[70]). 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 with future scenarios where more species recoveries may occur, and where BDS may therefore become more negative). Tweedie model selection often uses the Gini index for model selection[70]. 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 re-ran model selection using AICc, a technique which penalizes overfitting and is asymptotically similar to leave-one-out cross validation[71], and 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, below), we preferred AICc approaches. We also carried out a power analysis[72], which revealed that our best-fitting models had a power of >0.99, and thus that our sample size was adequate to detect effects among the relatively large number of parameters.

In the binomial part, exploratory GAMs again suggested that linear modelling was appropriate, and so we used GLMs with binomial errors, fitting an additional dispersion parameter to account for strong underdispersion[64]. Models containing this extra parameter do not generate AIC values, so we carried out non-automated binomial model selection, using stepwise backward and forward regression with likelihood ratio tests[64]. 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(BDSb=1) <50% as a predicted 0, and p(BDSb=1) >50% as a predicted 1).

Cross validation to test for forecasting accuracy on unseen data

To test the model's forecasting accuracy, as would be needed for policy usefulness, we carried out ten-fold cross-validation, a procedure that repeatedly sets aside part of the data (as a "fold" of BDS values the model has never seen), parameterises the model on the remaining subset of data, then tests how well it forecasts the unseen BDS values[37]. 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 also be close to 1.0 and 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 (note also that percentage deviations, rather than absolute deviations, are not appropriate metrics for low-volume data containing several zeroes such as BDS[37]). For the binomial model part, we tested mean forecasting accuracy against unknown data using % correct predictions, as we had done in testing binomial explanatory power.

Covariate balancing and spatial considerations

An important issue with impact studies is "selection bias", where the likelihood of receiving the intervention of interest is non-random[25,73]. The amount of conservation investment a country receives is indeed known to be influenced by non-random factors including Red List status itself[8], potentially creating endogeneity problems[25,73] and in particular, a potential problem of reverse causality whereby decline drives changes in conservation spending rather than vice versa.

Our use of a time lag between predictors and responses was designed to reduce the issue of reverse causality in the analysis. We also note that since greater decline has been shown to cause greater investment[8,28], 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 less subsequent decline). To correct for selection bias and associated endogeneity problems more generally, we used covariate balancing propensity scores[25] for continuous treatment variables[74] (in the R package CBPS[75]), which minimises the association (the Pearson correlation) between covariates and the treatment[74,75]. 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, GDPPPP, the percentage of land that is a protected area, and the sums of threatened bird and mammal species weighted by their level of extinction risk[8,28]. We carried out covariate balancing using data on these variables (taken from[8]) plus data on forest loss between 1990 and 2005 (taken from the FAO data[58,59]) and data on 1992-2003 growth in GDP per capita PPP (taken from World Bank data[57]). Extended Data Figure 6 shows the Pearson 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, violating regression assumptions[50,64,76,77]. We tested for this effect by fitting four possible structures to the most complete GLM model using REML (restricted maximum likelihood estimate) and comparing their predictive power using AICc. The structures tested were: (i) a fixed effect for Region (see[8] and Supplementary Tables 1-2 for regions and regional intercept differences); (ii) a GLMM with a SAC (Generalized Additive Mixed model with spatial autocorrelative structure), where five possible structural models describing the spatial autocorrelative structure between country centroid coordinates were tested – linear, spherical, Gaussian, ratio and exponential[64]; (iii) a GLMM with an SAC as in (ii) plus a fixed effect for Region; (iv) a GLMM with an SAC plus a random intercept for Region. The best-fitting structure was (i) and we used this in subsequent modelling. Using Region as a fixed effect also follows logically from theory, since regional differences are a potentially important component of decline[46]. 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 have impacts on biodiversity in neighbouring countries and statistical "spatial lags" have been used to model such possible effects[50,77]. 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 OLS regression[78] and to our knowledge, no robust methodology exists for non-linear generalized 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 likely removed much of the artefactual spatial lag that arises 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[79]. We checked for violations of model assumptions using diagnostic plots of residuals against fitted values and against all candidate predictors variables[64]. 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. Collinearity was checked for using VIF scores (Extended Data Table 3).

<u>Predicting the impact of spending and pressure changes</u>

To predict the impact that an extra i$1m or i$5m dollars annually of conservation spending would have had in each country, we added those 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 the 1992–2003 levels of socioeconomic growth (i.e. change in pressures) with 2001–2012 levels. To estimate the decline that we may have avoided as a result of 1992–2003 spending, we used the fact that prior to the 1992 Earth Summit, biodiversity spending for which we have data was flat and often zero (noting that data becomes sparser prior to the 1990s, and sparser still as one goes back further in time). We therefore estimated mean annual spending for 1985–1990, then re-predicted outcomes as if post-1990 annual budgets had only increased in line with inflation (i.e. no real increase). Although reduced data quality and imputation for the 1985–1990 spending makes these estimates approximate, the median change in BDS was robust to several different spending estimates, and so the global figure for avoided decline (29%) is 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 (which was set at 0.17 in the main text, see Additional Method Details, below), we examined the model outcomes using T = 0.10. and T = 0.25. To test for sensitivity to the Ψ parameter, we repeated the analysis with multiple variations around the parameter value used in the main analysis, finding no qualitative differences in the results. To test for the effect of the influential outliers (Solomon Islands and New Zealand), we ran model selection both with and without the 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 impact 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 high BDS, since the BDS distribution is skewed (Extended Data Figure 5). Our tenfold cross-validation test already showed that the omission of various groups of countries had no substantive impact 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, i.e. they are different for each country and indeed affect each other. An appropriate measure of parameter change is therefore the average percentage change in the values of the conditional expectations across all countries. For example, if a country C (such as Indonesia) was strongly biasing the model results, then when we re-run the model without C, we should see a substantial change in the average conditional expectation of BDS across the remaining countries, indicating a strong shift across the conditional parameter estimates for the interaction model. With heteroscedastic errors, the median percentage may also be more informative than the mean, so we considered both.

Even with these tests, there remained the possibility of "joint influence" in the continuous model part[80] where the highest-value BDS countries were driving the model as a group (for example, the BDS values for the top three countries of Indonesia, Australia and China are very large, being 272%, 69% and 24% larger than the fourth-highest BDS value, and so may 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 BDS countries were omitted.

In impact assessments addressing the impact of a single variable, a further concern is "missing variable bias", where there may be a confounding variable closely correlated with both the studied impact variable and the outcome variable[87]. In other words, the observed impact of conservation spending may simply be an artefact of spending being collinear with an unknown variable that is actually driving the outcome. When only one explanatory variable is being studied for its impact, hidden variable bias can be investigated by testing whether the main variable impact is still observed after an artificially created, collinear dummy variable has been added to the analysis[25]. 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. Nevertheless, we attempted to take the spirit of the missing variable test by looking for an empirical variable that was closely correlated with our spending variable (and therefore had a natural co-correlation with all other variables in the regression formula), then adding it into the regression and testing whether the spending impact 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 GDPPPP (i.e. raw GDP rather than the GDP per capita used in the main analysis) had a correlation of 0.54 with spending. We therefore tested the impact 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 a strong correlation between the two).

Finally, we tested the possible impact of inaccuracy in national conservation spending data, following the sensitivity tests used in[8]: 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 all these sensitivity tests are shown in the Supplementary Results, but none affected our conclusions substantively.

Additional Method Details: Mathematical calculation of BDS

Although change in Red List status is a standard measure of biodiversity change used in the CBD and SDG frameworks[3,31,32], 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 biodiversity policy and targets. We therefore created an algorithm to convert species-level change to country-level change. Mathematically, we define R_{ij} = 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 such errors 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 having some level of threat in global Red List assessments)[14,15,45,46,77]. Implicitly, such an algorithm assumes that if two countries share a species, they have equal responsibility for that species' 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 (e.g. >80% of the range) and neighbouring countries hold very small fractions of the remaining range edge (e.g. <5% each)

(Extended Data Figure 5). In such cases, it would be highly inaccurate (and politically unfair) to allocate equal shares of responsibility for a species decline across all these countries. A fairer, more accurate system may be to divide up responsibility according to the fraction of each species' range found in each country[8,26]. 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 (formally, the error is defined as the difference between the p_{ij}-based estimate and R_{ij}).

For any observed p_{ij}, there is therefore a theoretical probability density function (PDF) of all possible R_{ij} that it could represent. For example, if a species is split 60:40 between two countries, then for the $p_{ij} = 0.60$ country, the underlying assumption is that there is an approximately Gaussian PDF 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.

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

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

For a predictive regression model, the critical question is whether these errors $eBDS_j$ are likely to strongly affect modelling of BDS, for example by creating artefactual patterns or biased, non-random error distributions. If all range splits that make up BDS_j are relatively symmetric (i.e. similar to 60:40), then it is a reasonable expectation that the errors, being drawn from an approximately Gaussian distribution, will overestimate and underestimate with relatively equal frequency, and so the sum of errors will not depart strongly from zero. Thus, the errors are expected to be relatively random in their distribution, permitting robust modelling. It is also particularly unlikely that the errors would create artefactual impacts, since this would require a consistent, non-random association between large negative errors and higher-spending countries (sufficiently large, indeed, to strongly depress BDS_j), plus equally large and consistently positive errors for lower-spending countries.

However, when p_{ij} is closer to its limits of 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 where human pressures have most strongly increased and in general, it is very rare for such hotspots of decline to lie around the range periphery[36]. 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 onto the range will frequently fall entirely within the 97% country. Formally, therefore, when $p_{ij} = 0.03$, the associated probability density for R_{ij} will be high at 0 and decline rapidly towards a very low density at $R_{ij} = 0.03$, giving a PDF with a strong right skew and a likely 99th percentile at around p_{ij} itself.

In the example where $p_{ij} = 0.03$, therefore, nearly all errors will be overestimates, and the most common likely scenario is an overestimate of exactly 0.03. Generalising, whenever p_{ij} is small and the PDF is right skewed, a raw or "unadjusted" range-based algorithm will overestimate responsibility in almost all cases, generating highly biased errors eR_{ij} that will commonly have magnitude $+p_{ij}$. 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 great majority of cases.

The critical question is how severely this consistent bias will affect the regression analysis. We examined the data and found that empirically, a large number of countries had a BDS composed entirely of a trivially small (e.g. <5%) range edge fractions (Extended Data Figure 5). Their BDS_j estimates were therefore likely to be made up of multiple small p_{ij} that were consistently overestimating responsibility R_{ij}. In analysing BDS_j, the error metric of interest is $eBDS_j$ = the sum of eR_{ij}, Since the set of errors eR_{ij} was likely to be highly biased and the most common likely scenario was that $eR_{ij} = +p_{ij}$, then $eBDS_j$ (as the sum of eR_{ij}) would also be highly biased, with a substantial probability that $eBDS_j$ would equal sum (p_{ij}). Since all the individual p_{ij} values comprising these BDS scores were both trivially small and likely overestimates of zero, the associated BDS scores were also likely to be trivially small (and biased) overestimates of zero. We refer to these cases as range-edge BDS or "reBDS scores".

We further explored the empirical impact of this suspected bias on the information signal by making exploratory plots of BDS against its possible predictors. These plots showed that reBDS scores indeed generated a dense cloud of very small values, close to the x axis, that was visually distinct from patterns across larger (and likely more accurate) BDS. In Signal Detection Theory terms[37], therefore, reBDS cases were highly likely to represent strong signal noise that also lay 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 BDSb = 1 purely because those countries contained trivial range edges of status-changing species 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}. Formally, we explored setting a threshold value T, such that responsibility was set to zero for any country with a range fraction < T, such that

$$R_{ij} = \quad p^*_{ij} = \quad \begin{cases} p_{ij} & \text{if } p_{ij} >= T \\ 0 & \text{if } p_{ij} < T \end{cases}$$

(but see below for $p_{ij} >= (1-T)$).

To decide on appropriate values for the threshold T, we used Signal Detection Theory in combination with the mathematics of the Red List criteria. The most important aspect of this approach that when p_{ij} is small (e.g. 0.03), true R_{ij} may often but not always be zero, but 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, cases where a true zero is wrongly given a non-zero value represent "false positives". However, any threshold could also cause the algorithm to wrongly exclude (set to zero) some cases where the reBDS score represented a genuine (if small) fractional responsibility, and such incorrect exclusions are classed as “false negatives”. The higher the threshold T, the more false positives will be correctly excluded but the more false negatives will be wrongly excluded. Theoretical optimisation will therefore seek values of T large enough to avoid too many false positives (i.e. guarding against picking up too much noise) yet small enough to avoid too many false negatives (i.e. 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 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 so the ratio of sensitivity to specificity cannot be calculated. Appropriate thresholds therefore need to be estimated by theoretically estimating the optimal sensitivity/specificity trade off. Furthermore, in this analysis, sensitivity and specificity were likely to have different impacts on analytical bias and outcomes (making approaches that give equal weight to sensitivity and specificity, or that require accurate knowledge of the ratio between them e.g. area under the curve[37], less appropriate). The main deleterious effect of excessive sensitivity was to generate large amounts of biased noise, as already shown. The main impacts of excessive specificity, on the other hand, were likely to be (a) to slightly underestimate BDS (because a few small responsibility fractions had been wrongly discarded); (b) to reduce sample size for the continuous model part (because of removing reBDS countries); and (c) 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). Since high levels of noise and bias associated with lack of specificity are likely to have a much stronger impact than the small underestimates and sample size/binomial ratio effects associated with lack of sensitivity, avoiding false positives should take priority.

To allocate this priority (i.e. to avoid repeatedly replacing true zeroes with trivially small values), the algorithm needs to set T such that for all probability frequency distributions associated with all p^*_{ij}, there is a low probability density at $R_{ij} = 0.0$. Formally, we set a target that for all p^*_{ij}, prob ($R_{ij} = 0.0$) should be <0.5 and ideally $<<0.5$. However, a second consideration is that in range-edge countries, the likely probability density at zero is affected by the size of 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)[3]). When population loss occurs, the Red List assessment for any particular period is based on a rate of change over time, and so a change in Red List status expresses a second-derivative change in the rate of change i.e. additional net mortality/disappearance over and above what had occurred in the previous assessment period. Clearly, if a status change formally represented a 99% increase in mortality/disappearance for the entire species, there would be a strong probability that at least some of those additional deaths or disappearances had occurred in the 10%-holding country. However, genuine status changes generally imply an increase in loss of a few tens of percentage points. For example, a common status change is LC to VU, where LC can imply anywhere between zero decline and 29.9% loss over a period of ten years or three generations, and VU is defined as anywhere between 30% and 49.9% loss (depending on the use of the near-threatened category by assessors)[81]. If we take the midpoints of these ranges (15% and 40% respectively), then an LC-to-VU change would indicate an average 25 percentage point increase in loss (the difference between 40% and 15%), while other changes not at the exact midpoints would indicate a difference in decline rates above or below 25.

Since the additional deaths underlying a status change are generally non-randomly clustered in geographic space as wave fronts expanding from points of increased human pressure[36], this 25-point change can be imagined as a small number of clusters of additional net loss placed onto a gridded range, where the 10%-holding country occupies the leftmost 10% of the grid and another country or countries, the rightmost 90%. Often, such spatially-clustered mortality increases might be expected to fall entirely within the rightmost 90%, implying that a 10%-holding country will frequently have no responsibility. To explore this intuition this more quantitatively, we simulated a 25-point population loss as a varying (2-5) number of rectangular blocks that covered a total of 25% of a 10x10 gridded range. The first column of the grid was then treated as the 10%-holding country and the remaining 9 columns to another country or countries: (it is moot whether it is one or several countries in the 9 columns because the simulation focuses only on the likelihood that the 10% country will not have any part of any decline cluster overlapping its territory). The blocks were then placed independently of each other, for a limited number of trials (n=100) to introduce stochasticity, onto the gridded range and 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 occurs as two independently-placed blocks, the simulated probability of overlap was 0.19, giving a 0.81 probability that the range-edge country has $R_{ij} = 0$ (i.e. an 81% chance of a false positive). When the 25-point decline was modelled as five independently dropped blocks, the overlap probability rose to 0.41, indicating a 59% chance of a false positive – still appreciably greater than our target false-positive rate of <<0.50. These values are also conservative because clusters of loss are often not spatially independent of each other but rather, may be grouped due to larger-scale spatial contagion in threats and associated losses[36]. Such grouping further reduces the random probability of an overlap with the range edge and thus, would increase the false positive rate further. Similar outcomes occur for other percentage point increases in mortality, as implied by other IUCN status changes.

Indeed, even if the 25-point population loss is unrealistically (and highly conservatively) modelled as spatially homogeneous, then define q = 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 10%-holding country can have for 25% change is ~40%, or 10/25. More formally, we define the 99th percentile of theoretically probable R_{ij} for the 10%-holding country as $p_{ij}/q = 0.1/0.25 = 0.4$. 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 $R_{ij} = 0$, since 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 <<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 further examined exploratory biplots of BDS against its predictors when T is varied between 0.05 and 0.25. We found that as T was reduced, and as expected from our theoretical treatment, increasingly large numbers of likely false positives became included in the BDS dataset, with noise increasing rapidly at T<0.1 (i.e. 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 themselves, 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, then 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 must remain uncertain. To account for this uncertainty and its possible impact on model outcomes, we performed our final analysis three times for three different values of T: 0.10, 0.17 and an extreme value of 0.25. The main text of the paper shows results for T = 0.17, being the parameter value where false positives could be reduced as far as possible, and yet without the trade-off of sample size reduction becoming severe; results for T = 0.10 and T = 0.25 are described in Extended Data Table 4 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 d can be positive and indicate a worsening extinction risk (d>0), it can be negative and indicate a reduction in extinction risk ("negative decline" i.e. an improvement, d<0) or it can be constant (d=0). Each country's baseline Biodiversity Decline Score (BDS_i) is therefore the net sum of all its decline fractions and improvements (negative decline fractions):

$$BDS_i = \sum_j d_j R_{ij} \qquad (1)$$

where

$$R_{ij} = p^*_{ij} / \sum_i p^*_{ij} \qquad (2)$$

where 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, or formally:

$$p^* = \begin{cases} \text{p\_obs where p\_obs} >= T \\ 0 \text{ where p\_obs} < T \end{cases}$$

If a species is split 95:5 between two countries and the responsibility R has been set to zero for the 5%-holding country, then for consistency, R for the 95%-holding country should be increased from 0.95 to 1.0, and equation (2) indeed 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 (e.g. 17.1%) then under equation (2), that country would receive a clearly exaggerated 100% of responsibility for the change in risk status (whereby p_obs = 0.171 but p* = 1.0). Such cases as fairly rare (widespread species rarely move out of the Least Concern category) but to avoid any such error, we reset the denominator of equation (2) to unity whenever 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 global bird and mammal species contained within the national borders of each country (the range overlap)[26]. Range overlap for mammals was extracted using ArcGIS utilities on the range maps provided by the IUCN Global

Mammal Assessment[82] (see[8] for details). This procedure gave very exact areas of overlap for the taxon Mammalia, but the calculation required us to run twenty processors in parallel for nearly a month. For the much larger taxon Aves, therefore, we used a slightly different procedure. Bird ranges were obtained as polygons in ESRI shapefiles provided by Birdlife International[83]. Species range areas that were designated as non-native or dubious presence were excluded *a priori*. For each species X, we then combined wintering and breeding ranges (because threats to bird species can occur in both their breeding and wintering ranges), and gridded all range polygons onto 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 center point lying inside a range polygon for X as 'presence cells' for X, overlaid each presence cell onto a vector dataset of the world's countries[84] using the over and wrld_simpl functions in R packages 'sp'[85] and 'maptools'[86], allocated the cell to the country found at the cell centre point, and then calculated the fraction of all presence cells for X found in each country. Prior to this calculation all countries with coastlines were enlarged by a 0.05 degree buffer into the sea to account for responsibility of sea bird ranges in coastal waters; coastal marine mammals were treated in a similar way, see[8].

As an additional accuracy check, we examined individual Red List reports for every declining species to see where the range-based approximation of responsibilities was clearly inappropriate, and revised those cases accordingly. Our revisions are listed in Supplementary Table 4 and include cases where (i) a decline had majorly affected how the geographic range was distributed across countries, including cases where a species had once been found in other countries but was now missing from them; (ii) the species population distribution across countries was poorly correlated with the range distribution; and (iii) specified actions e.g. along migratory routes had an impact clearly disproportionate to the percentage of the global range found in 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 neighboring countries. The 100%-holding country would then represent a final “oasis” at the species’ former range edge, and it would be wholly unjust to assign 100% responsibility for the decline to it. However, our assumption is that in the mere eight-to-twelve years between our IUCN assessments, there will rarely be a case where 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 such anomaly. For example, *Addax nomasculatus* (the rare screwhorn antelope) has recently disappeared from Chad and Mali and so we incremented the BDS of those two countries to reflect this (Supplementary Table 4).

Data availability

The authors declare that the data supporting the findings of this study are available within the Supplementary Information; original socioeconomic data (except governance values) can also be sourced from the World Bank http://databank.worldbank.org; original governance values can also be sourced from the Worldwide Governance Indicators dataset www.govindicators.org (governance data).

Code availability

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

EXTENDED DATA LEGENDS

Extended Data Figure 1 | The average per-species BDS for each country (i.e. BDS/total fractional species richness, expressed as a percentage). Dark red = > 5%, dark orange/red = 2.5 – 5%, mid 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 e.g. much of Europe and the Arab geographic crescent, zeroes are expected at random (supplementary methods). See Supplementary Table 1 for precise values per country. Country outlines supplied by esri_dm
https://www.arcgis.com/home/item.html?id=d86e32ea12a64727b9e94d6f820123a2#overview

Extended Data Figure 2 | The effect of agricultural expansion on decline (binomial part, n=109 independent countries) depends on both governance improvement and the existing percentage of land converted.
The effect (coefficient) of agricultural expansion on the probability of a decline occurring 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, v.v. for <0. However, the coefficient further depends on a second moderator, the % of land converted to agriculture: red =50^{ile} of % land conversion, grey =25^{ile}; lines show mean and coloured bands show conditional 95% confidence intervals. Note how effects are most strongly deleterious on less heavily converted landbases. Rug plot at bottom shows empirical distribution of x-axis values (but note that countries with more % agric. land generally have slow governance improvement). All variables are z-standardised.

Extended Data Figure 3 | The impact of conservation spending on decline depends on threatened species richness and on GDP. (a) Spending effect size and threatened species richness (continuous part, n=50 independent countries); (b) spending effect size and GDP (binomial part, 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. The more negative the coefficient is on the y axis, the more strongly spending reduces declines (continuous) or the probability of a decline occurring (binomial). Conditional confidence bands are shown; rug plots at bottom show empirical distribution of x-axis values. All variables are z-standardised.

Extended Data Figure 4 | Observed declines versus model-predicted declines. (a) BDS versus predicted BDS in the continuous part (n=50 independent countries). Both axes are ln-transformed for clarity; (b) As (a), but zooming in to the lower-BDS countries only (note axes values in (a) and (b)); (c) Observed decline events (BDSb) versus the predicted probabilities of a decline event, from the binomial part (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 3 BDS values are excluded: black line = full dataset prediction, dashed red line = prediction with exclusions.

Extended Data Figure 5 | Distributions of BDS and species range fractions across countries. (a) Index plot of BDS scores. For clarity, BDS has been ln(x+10) transformed, and so the straight line at 2.3 shows the long tail of zeroes. (b) Distribution of all range fractions in all countries, showing the very large number of small, range-edge fractions (<10% of a species is found in a country). (c) Distribution of the maximum range fraction for all species, showing how a large number of species have >90% of their range in one country. (d) Distribution of the minimum range fraction for all species, showing how very many species have a small range edge (<10% of their range) in a second country.

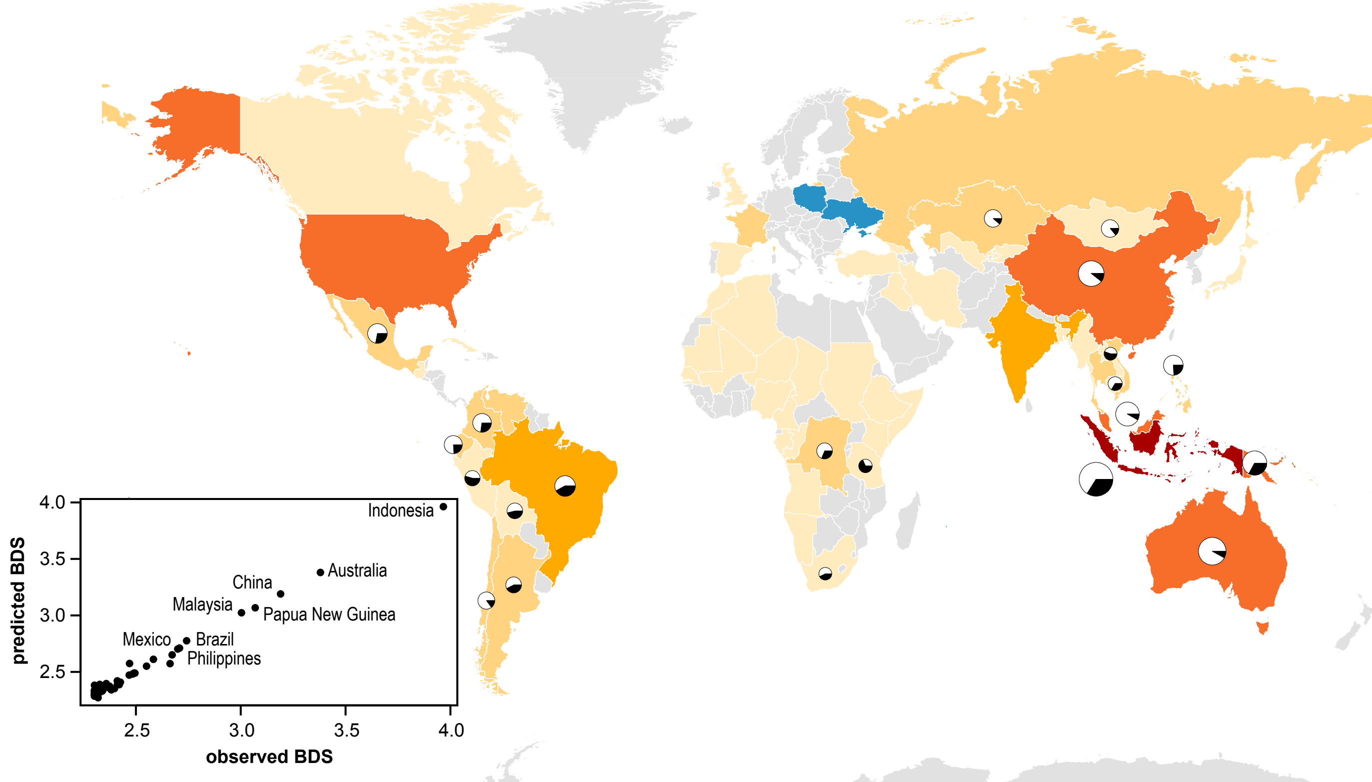
Extended Data Figure 6 | Differences in absolute Pearson's correlations between conservation spending and each of its covariates before and after carrying out covariate balancing propensity score weighting (CBPS). (a) continuous analysis; (b) binomial analysis. Upper bars show absolute Pearson correlations prior to CBPS, lower bars after CBPS. Box shows interquartile range with the median (bold central line). Whiskers show 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. Also shown are the best-fitting four models from continuous analysis with their AICc values, Akaike weights and variables (see Supplementary Table 2 for full continuous-model results). Spending = conservation spending PPP; Agric. = agricultural; governance = government effectiveness indicator. In main body of table, 1 = term included, 0 = term not included.

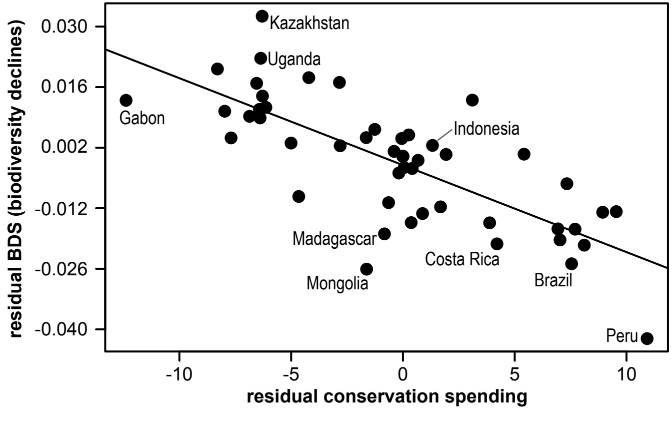
Extended Data Table 2 | Cross correlations between variables. $$ = conservation spending PPP; Agric. = agricultural; Pop = population; Gov = governance; Decl = declines; Spp. Rich = threatened species richness; For. Loss = % forest loss; Area = country area.

Extended Data Table 3 | Variance inflation factors (VIFs) for the continuous and binomial model parts. Spending = conservation spending PPP; Agric. = agricultural; Pop = population; Gov = governance; Spp. Rich = threatened species richness; Area = country area.

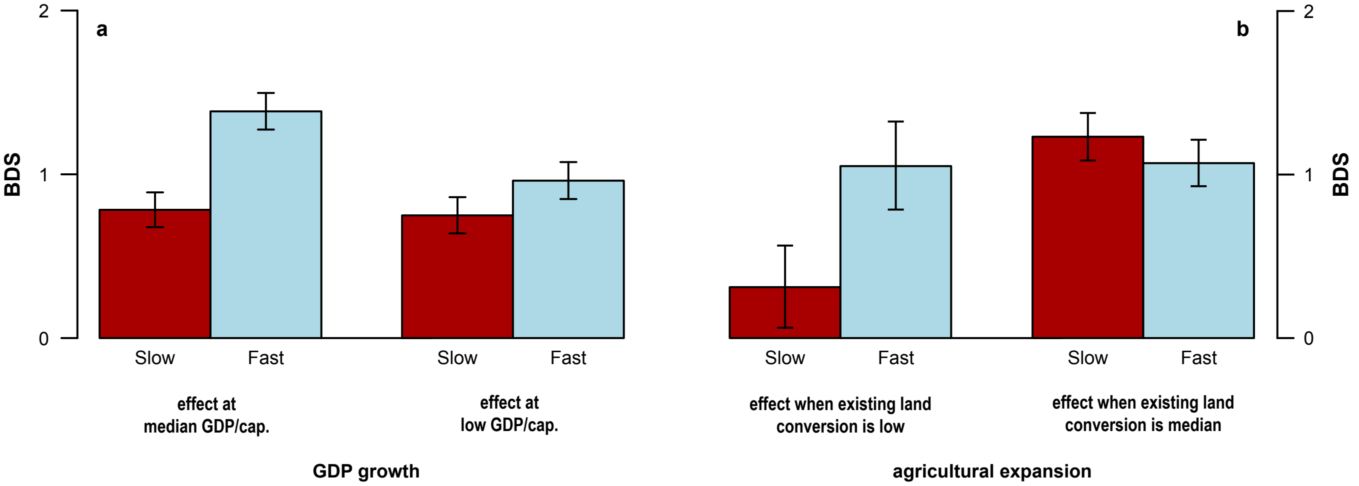
Extended Data Table 4 | Standardized coefficients for best-fitting models under alternative assumptions. Best-fit models that used alternative values of the threshold T are shown. We very strongly caution that for interacting variables (marked *), the coefficients shown cannot be interpreted by simply reading the table (refer to the Supplementary Results for their complex interpretation). "Agric. land" = mean percentage of agricultural land; t-1 = 1994-2000, t-2 = 1988-1994; GDP = Gross domestic product per capita 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 (index parameter = 1.99), and the ratio of ones to zeroes was 44:65 in the binomial part. Equivalent values for T=0.25 are n=43 independent countries (i.e. a large sample size decrease) and a ratio of 37:74.



predicted BDS
observed BDS
4.0
3.5
3.0
2.5
2.5
3.0
3.5
4.0
Indonesia
Australia
China
Malaysia
Papua New Guinea
Mexico
Brazil
Philippines



residual BDS (biodiversity declines)
0.030
0.016
0.002
-0.012
-0.026
-0.040
residual conservation spending
-10
-5
0
5
10
Kazakhstan
Uganda
Gabon
Indonesia
Madagascar
Mongolia
Costa Rica
Brazil
Peru



2
1
0
BDS
a
b
Slow
Fast
Slow
Fast
Slow
Fast
Slow
Fast
effect at
median GDP/cap.
effect at
low GDP/cap.
effect when existing land
conversion is low
effect when existing land
conversion is median
GDP growth
agricultural expansion