

1-16-2019

# Standards for distribution models in biodiversity assessments

Miguel B. Araújo

*Spanish National Research Council*

Robert P. Anderson

*CUNY City College*

A. Márcia Barbosa

*University of Evora*

Colin M. Beale

*University of York*

Carsten F. Dormann

*University of Freiburg*

*See next page for additional authors*

## [How does access to this work benefit you? Let us know!](#)

Follow this and additional works at: [https://academicworks.cuny.edu/cc\\_pubs](https://academicworks.cuny.edu/cc_pubs)



Part of the [Biodiversity Commons](#), and the [Population Biology Commons](#)

---

### Recommended Citation

Araújo, Miguel B.; Anderson, Robert P.; Barbosa, A. Márcia; Beale, Colin M.; Dormann, Carsten F.; Early, Regan; Garcia, Raquel A.; Guisan, Antoine; Maiorano, Luigi; Naimi, Babak; O'Hara, Robert B.; Zimmermann, Niklaus E.; and Rahbek, Carsten, "Standards for distribution models in biodiversity assessments" (2019). *CUNY Academic Works*.

[https://academicworks.cuny.edu/cc\\_pubs/707](https://academicworks.cuny.edu/cc_pubs/707)

---

**Authors**

Miguel B. Araújo, Robert P. Anderson, A. Márcia Barbosa, Colin M. Beale, Carsten F. Dormann, Regan Early, Raquel A. Garcia, Antoine Guisan, Luigi Maiorano, Babak Naimi, Robert B. O'Hara, Niklaus E. Zimmermann, and Carsten Rahbek

## ECOLOGY

## Standards for distribution models in biodiversity assessments

Miguel B. Araújo<sup>1,2,3\*</sup>, Robert P. Anderson<sup>4,5,6</sup>, A. Márcia Barbosa<sup>3</sup>, Colin M. Beale<sup>7</sup>, Carsten F. Dormann<sup>8</sup>, Regan Early<sup>9</sup>, Raquel A. Garcia<sup>2,3,10,11</sup>, Antoine Guisan<sup>12,13</sup>, Luigi Maiorano<sup>14,15</sup>, Babak Naimi<sup>2</sup>, Robert B. O'Hara<sup>16,17</sup>, Niklaus E. Zimmermann<sup>18,19</sup>, Carsten Rahbek<sup>2,20</sup>

Demand for models in biodiversity assessments is rising, but which models are adequate for the task? We propose a set of best-practice standards and detailed guidelines enabling scoring of studies based on species distribution models for use in biodiversity assessments. We reviewed and scored 400 modeling studies over the past 20 years using the proposed standards and guidelines. We detected low model adequacy overall, but with a marked tendency of improvement over time in model building and, to a lesser degree, in biological data and model evaluation. We argue that implementation of agreed-upon standards for models in biodiversity assessments would promote transparency and repeatability, eventually leading to higher quality of the models and the inferences used in assessments. We encourage broad community participation toward the expansion and ongoing development of the proposed standards and guidelines.

## INTRODUCTION

The Earth system is rapidly undergoing changes of enormous magnitude, and anthropogenic pressures on biodiversity are now of geological significance (1, 2). While the effects of biodiversity changes on human welfare and ecosystem services are increasingly recognized (3), our ability to forecast changes in biodiversity remains limited (4–8). One reason for this limitation is that these forecasts require projecting models to conditions for which we have no current or past analogs (9, 10). In addition, available biodiversity models struggle to deal with data limitations and the inherent complexities of biological systems (5, 8, 11). Consequently, different modeling studies and approaches often lead to inferences and projections that vary in both magnitude and direction (7, 12).

Despite difficulties in forecasting the responses of biodiversity to multiple drivers of change, many studies use models and apply their conclusions to conservation, management, and risk assessment (13). A search of articles in peer-reviewed journals over the past 20 years

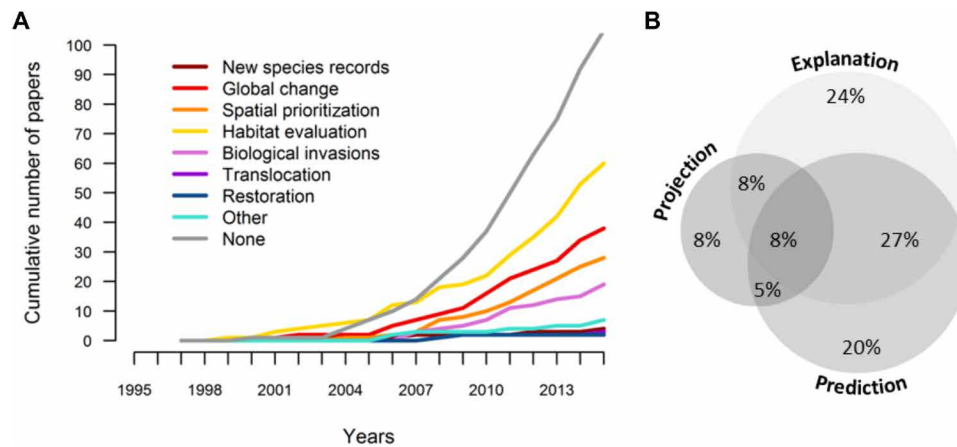
found more than 6000 studies using or mentioning one of the most common classes of biodiversity modeling: species distribution models (SDMs). Over half of the studies using SDMs sought to apply their results to at least one type of biodiversity assessment, including forecasting the effects of climate change on biodiversity and/or selecting places for protected areas, habitat restoration, and/or species translocation (Fig. 1A and text S1). Quantifying or forecasting the effects of anthropogenic factors on biodiversity was common among conservation applications. In nearly a third of the studies (30%), models were used to inform assessments of population declines or the ability of landscapes to support existing populations. Changes in suitable areas for species faced with global change now constitute a major focus of models (19% of the studies by 2015). These trends are reflected in major global assessments of the impacts of human activities on the living world, such as those by the IPBES (Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services) (6, 14), the IUCN (International Union for Conservation of Nature) (15), and the IPCC (Intergovernmental Panel on Climate Change) (16), which draw heavily on existing modeling studies to establish scientific consensus (Fig. 2). The same can be said for local, national, and regional assessments (17–19). While applied use of SDMs is extensive (56% of the papers in our review used this sort of models), a considerable fraction of the studies (44%) did not perform any direct biodiversity assessment (Fig. 1A). These figures reveal that many species distribution modeling studies are being developed in the context of basic science. That is, significant efforts are being spent to further the scientific underpinnings of these models so that they might eventually contribute to increasing the information content of biodiversity assessments.

Despite the growing body of species distribution modeling literature and the high demand for their use in biodiversity assessments, no generally agreed-upon standards for best practices yet exist for guiding the building of these models and for evaluating the adequacy of the models that feed into these assessments (6). In practice, assessors often make ad hoc judgments about which studies to include, and papers with greater visibility, such as those published in high-profile journals, are frequently favored (20). However, journal decisions depend on many factors that extend well beyond the appropriateness of the data and models, and the impact of a journal—or even the number

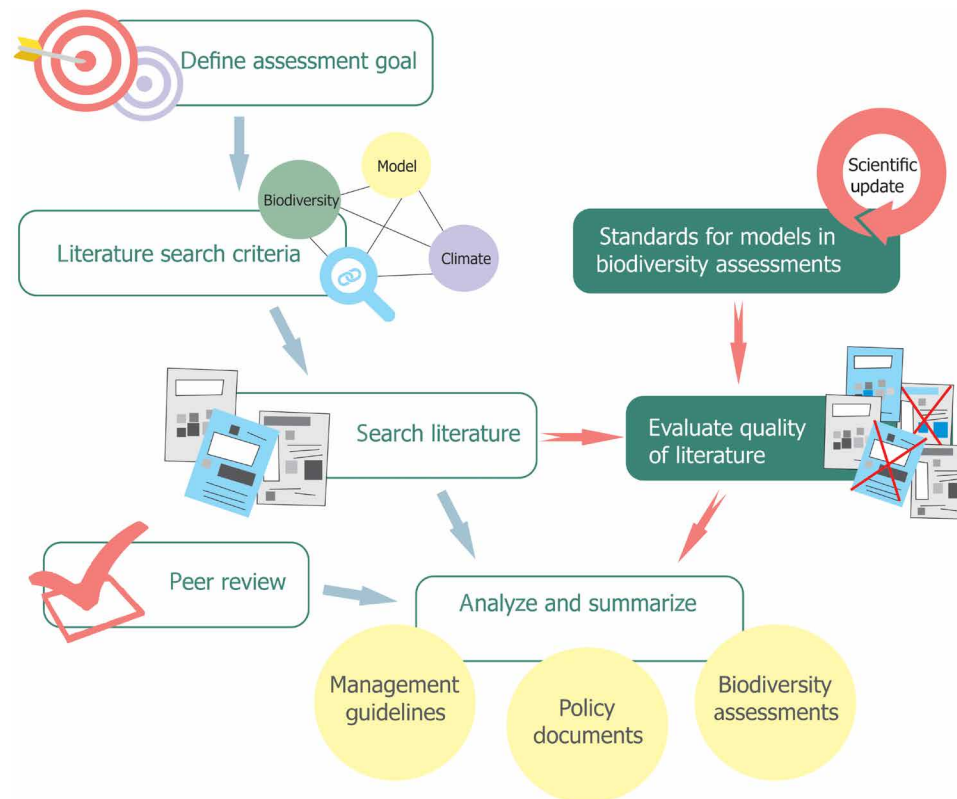
Copyright © 2019  
The Authors, some  
rights reserved;  
exclusive licensee  
American Association  
for the Advancement  
of Science. No claim to  
original U.S. Government  
Works. Distributed  
under a Creative  
Commons Attribution  
NonCommercial  
License 4.0 (CC BY-NC).

<sup>1</sup>National Museum of Natural Sciences, Spanish National Research Council (CSIC), 28006 Madrid, Spain. <sup>2</sup>Center for Macroecology, Evolution and Climate, Natural History Museum of Denmark, University of Copenhagen, DK-2100 Copenhagen, Denmark. <sup>3</sup>Rui Nabeiro Biodiversity Chair, University of Evora, 7000 Évora, Portugal. <sup>4</sup>Department of Biology, City College of New York, New York, NY 10031, USA. <sup>5</sup>Program in Biology, Graduate Center, City University of New York, New York, NY 10016, USA. <sup>6</sup>Division of Vertebrate Zoology, American Museum of Natural History, New York, NY 10024, USA. <sup>7</sup>Department of Biology, University of York, York YO19 5PR, UK. <sup>8</sup>Biometry and Environmental System Analysis, University of Freiburg, D-79106 Freiburg, Germany. <sup>9</sup>Centre for Ecology and Conservation, University of Exeter, Penryn, Cornwall TR10 9FE, UK. <sup>10</sup>Centre for Invasion Biology, Department of Botany and Zoology, Stellenbosch University, Matieland 7602, South Africa. <sup>11</sup>Centre for Statistics in Ecology, Environment and Conservation (SEEC), University of Cape Town, Private Bag, Rondebosch, 7701 Cape Town, South Africa. <sup>12</sup>Department of Ecology and Evolution, University of Lausanne, Lausanne, Switzerland. <sup>13</sup>Institute of Earth Surface Dynamics, University of Lausanne, Lausanne, Switzerland. <sup>14</sup>Department of Biology and Biotechnologies Charles Darwin, University of Rome La Sapienza, Rome, Italy. <sup>15</sup>Department of Integrative Marine Ecology, Stazione Zoologica Anton Dohrn, Naples, Italy. <sup>16</sup>Senckenberg Biodiversity and Climate Research Centre, Senckenberganlage 25, 60325 Frankfurt, Germany. <sup>17</sup>Department of Mathematical Sciences and Centre for Biodiversity Dynamics, NTNU, 7491 Trondheim, Norway. <sup>18</sup>Landscape Dynamics, Swiss Federal Research Institute WSL, Birmensdorf, Switzerland. <sup>19</sup>Environmental Systems Science, Swiss Federal Institute of Technology ETH, Zürich, Switzerland. <sup>20</sup>Department of Life Sciences, Imperial College London, Silwood Park, Ascot SL5 7PY, UK.

\*Corresponding author. Email: maraujo@mncn.csic.es



**Fig. 1. Uses of SDMs.** Classification of published species distribution modeling studies by (A) type of biodiversity assessment accomplished with the trend in the numbers of studies shown over time and (B) purpose of the model (see glossary in text S4). In (A), the trend for translocation is very similar to that of restoration, and hence is hardly visible. The classification is based on a random sample of 400 papers (of 6483 identified articles mentioning statistical models of species distributions); 238 of the randomly selected papers used SDMs and were included in this analysis. Details on the literature search and analyses appear in text S1, figs. S1.1, S1.2, S1.3, and S1.4, and tables S1.1, S1.2, and S1.3.



**Fig. 2. Steps in biodiversity assessments.** Assessment process flow as typically implemented by international and national initiatives on biodiversity and/or climate change (e.g., IPBES, IPCC, IUCN, and national governments) and the suggested addition of agreed-upon (and updated) standards to ensure the adequacy of studies feeding into the assessments. Blue arrows and hollow boxes represent the current procedure, and red arrows and green-filled boxes represent the suggested additional steps.

of citations of a given paper—is a poor indicator for a study’s appropriateness for inclusion in biodiversity assessments (21, 22). Therefore, specific best-practice standards and guidelines must be established and then agreed upon by the scientific community to support the evaluation of the appropriateness of data and models used in the assessments supporting policy recommendations and decisions. The need for standards in biodiversity assessments was recently acknowledged

by IPBES’s Methodological Assessment Report on Scenarios and Models (6) in its guidance point 4: “The scientific community may want to consider developing practical and effective approaches to evaluating and communicating levels of uncertainty associated with scenarios and models, as well as tools for applying those approaches to assessments and decision making. This would include setting standards for best practices.”

When developing best-practice standards for models in biodiversity assessments, it is important to recognize that the criteria for judging the data and the models will differ according to the particular objectives of the assessments. For example, specific data and models might be well suited for understanding the assembly of butterfly communities in a particular region while being inappropriate for generalizing management decisions across taxonomic groups in several regions. Similarly, the appropriateness of a given dataset and model will change with the type of applied question being addressed. For example, management personnel might readily use particular models to predict areas suitable for reintroduction of a locally extinct species within its historical range (23) or to target additional specific areas for sampling within the distribution of a poorly known species (13, 24). However, projections of SDMs into regions that a species might invade (25, 26), or inhabit under future climate scenarios (27–29), are fundamentally more challenging and seldom can be directly used in practical conservation and/or management applications. Therefore, guidelines for scoring modeling studies should account for the specific assumptions and uncertainties associated with different kinds of model uses (Fig. 1B) and for the weight of evidence required for the various types of assessments (Fig. 1A).

### Best-practice standards for models in biodiversity assessments

The aims of establishing best-practice standards for models in biodiversity assessments are to provide a hierarchy of reliability (30–32), ensure transparency and consistency in the translation of scientific results into policy, and encourage improvements in the underlying science (33, 34). These standards do not aim to govern or guide publishing of research on species distribution modeling in general, but rather focus on the applicability of these models for biodiversity assessments. Best-practice standards should be general—applicable to a variety of available data and modeling approaches—and reflect the evidence required for the particular type of question addressed or decision being taken. A set of best-practice standards can often be first proposed by an expert group, but embedded within an ongoing dynamic process that is transparent and open to improvements over time by interested and experienced scientists to reflect the changing community consensus (34).

Here, we present such a framework for implementing best-practice standards together with detailed guidelines for scoring key aspects of SDMs used in biodiversity assessments. To start the process, we propose four levels of standards. The first level is aspirational (gold standard). It usually requires ideal data (seldom available) and next-generation modeling approaches that remain under development, as well as results obtained through multiple sources of evidence including manipulative experiments. Therefore, this level typically provides targets for excellence and directions for future research. The second level (silver standard) corresponds to current cutting-edge approaches, typically involving imperfect (but best available) data combined with analyses that allow uncertainty and bias to be reduced, accounted for, or at least estimated. The third level (bronze standard) encompasses data and procedures that represent the minimum currently acceptable practices for models to be included in biodiversity assessments. It includes approaches to characterize and address limitations of data and models, and to interpret their implications on the results. The final category (deficient) involves the use of data and/or modeling practices that are considered unacceptable for models used in driving policy and practice. Although models

characterized as “deficient” may be useful for addressing specific research questions, the outcomes of these studies should be interpreted with extreme caution when considering their inclusion in biodiversity assessments (Fig. 1).

### Applying best-practice standards to species distributions and environmental correlates

Real-world implementation of the best-practice standards requires that detailed yet flexible criteria be identified and agreed upon (35). We propose an extensive set of criteria that enables the scoring of modeling studies to be used for biodiversity assessments (tables S2.1, S2.2, S2.3, and S2.4). As mentioned above, we focus on the particular case of empirical models of species distributions (36, 37). These models constitute, by far, the most widely used tools for making inferences about biodiversity dynamics in space and time. However, many of the principles and issues raised in our scoring scheme can be co-opted for use with other classes of biodiversity models. Notably, the broad criteria that we outline are sufficiently general to apply to any approach in statistical species distribution modeling.

For these reasons, we define three major categories of model use (see text S4 and table S1.2). “Explanation” investigates the statistical relationships between species distributions and the environment, providing hypotheses (and sometimes testing them) regarding the environmental factors that account for species distributions. “Prediction” uses modeled species-environment relationships to map potential distributions in the same time period and geographical region. “Projection” extends these models to estimate suitable areas in the past or future and/or in different regions (see glossary in text S4 for an extended discussion of these concepts). Despite the rise in studies forecasting species distributions under climate change (Fig. 1A), explanation and prediction remain the most common uses of models (Fig. 1B).

Within each of these three broad “purpose” categories, we developed criteria for scoring the four critical aspects of modeling that affect the quality of model outputs (Table 1): (i) quality of the “response variable” (usually species occurrence data), (ii) quality of the “predictor variables” (usually environmental data), (iii) “model building,” and (iv) “model evaluation.” To help assessors characterize species distribution modeling studies, we identified 15 specific issues that should be considered within these four modeling aspects (Table 1 and text S2). Rather than being a compendium of disparate recommendations and best practices in the field, the proposed criteria were defined in a highly structured way, including four levels of quality (aspirational, cutting-edge, acceptable, and deficient) for each of the three categories of model use (explanation, prediction, and projection). For each combination of quality and model use, 15 particular issues are identified spanning four steps in modeling (choice of response variable, predictor variables, model building, and evaluation). This structure provides a framework for understanding the pros and cons of a given study. It also facilitates future refinement of the best-practice standards by the scientific community and by end-users conducting biodiversity assessments for use in the policy arena. In addition, this framework guides the production of standards for other classes of biodiversity models. Below, we provide an overview of each of the four aspects.

#### Response variable

Biases and inaccuracies in taxonomic and distributional data place heavy constraints on biodiversity assessments that use SDMs (38).



Critical questions include: whether the distributional data available for the species have been comprehensively sampled (39) while being representative of the regions and the environments in which the species lives (40) or can live (41), whether the spatial accuracy of the distributional occurrence data matches the spatial resolution of the environmental predictor variables (42), and whether species taxonomic identities are well understood, stable, and consistently applied (43). The gold-standard level often requires improvement in sampling efforts rather than conceptual or methodological challenges. Given that sampling tends to be contingent on human effort and availability of funding, gold standards in this aspect of modeling are arguably easier to achieve than gold standards in other aspects (text S2.1 and table S2.2).

### **Predictor variables**

SDMs use predictor variables to describe species–environment relationships. These modeled relationships are then transferred back to geographic space to examine biodiversity patterns. A wealth of relevant data to derive environmental predictor variables now exists. They may be collected on the ground, estimated through remote sensing, interpolated across landscapes by statistical or physically based means, or generated by combining different approaches; the quality of these estimates constitutes a major field of research on its own (44). It is important for biodiversity assessments that the processes of identification, acquisition, preparation, and selection of predictor variables be tailored to the specific goal of the models and taking the species' biology into account (see text S2.2 and table S2.2) (45, 46). For example, does solid evidence exist to demonstrate that the selected variables have causal relationships with the distributions of the species to be modeled, or do variables at least represent likely correlated surrogates (47, 48)? Is the spatial and temporal resolution of the variables consistent with the biological response being modeled (49)? Are the uncertainties associated with the choice of predictor variables quantified (50, 51)? The gold-level standard requires that predictor variables capture the conditions on which the response variable actually depends, at the relevant spatial and temporal resolutions and with uncertainty that can be quantified in the final model. The reason why the gold standard is often not currently achievable is because of the lack of sufficient biological knowledge regarding the species involved, or the lack of data on the relevant environmental variables for the species at the appropriate spatial extent and resolution. Broad extent/coarse grain studies tend to focus on climatic predictors, and local extent/fine grain studies tend to use habitat characteristics (52). Despite these tendencies, increasing evidence shows that both types of variables are important across scales (46).

### **Model building**

Building SDMs typically includes fitting a statistical relationship between species occurrence data and environmental data. Many techniques and implementations exist to fit these mathematical functions (53), but this rapidly developing field still lacks consensus regarding which algorithms are the best for which purposes. Key issues include consideration of model complexity (54, 55) and procedures to take into account imperfections in the response and predictor variables (56). The latter commonly include dealing with the effects of unrepresentative survey data (e.g., due to biased biological sampling) and, more generally, with the characterization of uncertainties in model outputs, including both those resulting from the data and those inherent to the model building process (see text S2.3 and table S2.3) (57). The gold-level standard includes a full exploration of the con-

sequences of all choices in model building, thereby quantifying the degree to which variability in the final predictions is introduced by the modeling methods themselves. Silver- and bronze-level standards correspond to different degrees of consideration but not to full quantification of the consequences of modeling choices (58, 59). The most common type of study design, namely, using only one set of response and predictor variables and one type of model, is generally acceptable today for biodiversity assessment purposes but is insufficient for inclusion in climate change assessments when model projections are involved (27, 29).

### **Model evaluation**

When evaluating biodiversity models, three critical questions need to be addressed: How robust is the model to departures from the assumptions? How meaningful are the evaluation metrics used? How predictive is the model when tested against independent data? In addressing these questions, at least three properties of the models can be evaluated (36, 60): realism, accuracy, and generality. Realism is the ability of a model to identify the critical predictors directly affecting the system and to characterize their effects and interactions appropriately. Accuracy is the ability of the model to predict events correctly within the system being modeled (e.g., species distributions in the same space and time as the input data). Generality is the ability of the model to predict events outside the modeled system via projection or transfer to a different resolution, geographic location, or time period. Depending on the question at hand, one of these properties might be more important than the other (24). For example, when forecasting the effects of climate change on biodiversity—arguably one of the most difficult questions that models can be asked to address—evaluating model generality is critical, yet realism might be key to achieving that goal. Accuracy is paramount if the goal is to predict current species distributions, such as when planning the placement of protected areas, but it is of limited value if the goal is to identify suitable conditions outside the current distribution (for example, for reintroduction purposes) (37).

We provide specific criteria for assessing realism (text S2.4 and table S2.4, Evaluation of model assumptions), accuracy, and generality (table S2.4B, Evaluation of model outputs; table S2.4C, Measures of model performance) based largely on the three categories of model use (explanation, prediction, and projection; see above). In general, gold-level standards are reached when multiple lines of evidence support model predictions and projections, or when no assumption of the models is violated. Silver- and bronze-level standards represent lesser, yet feasible, ways of evaluating model quality based on currently available data and techniques.

### **Systematic review of published studies**

We scored a sample of 400 species distribution modeling studies (see text S3 and table S3.1) using the proposed criteria. For every year between 1995 and 2015, we randomly selected peer-reviewed studies representing a variety of modeling studies with applied biodiversity focus (see text S3). We then examined studies for each of the four aspects of data and models (response variable, predictor variables, model building, and model evaluation). To quantify the effect of observer biases in scoring, we compared the obtained scores with those for 80 randomly selected papers independently re-evaluated by a different person (see text S3). We found that biases and variation among assessors were negligible (see figs. S1.6 and S1.7).

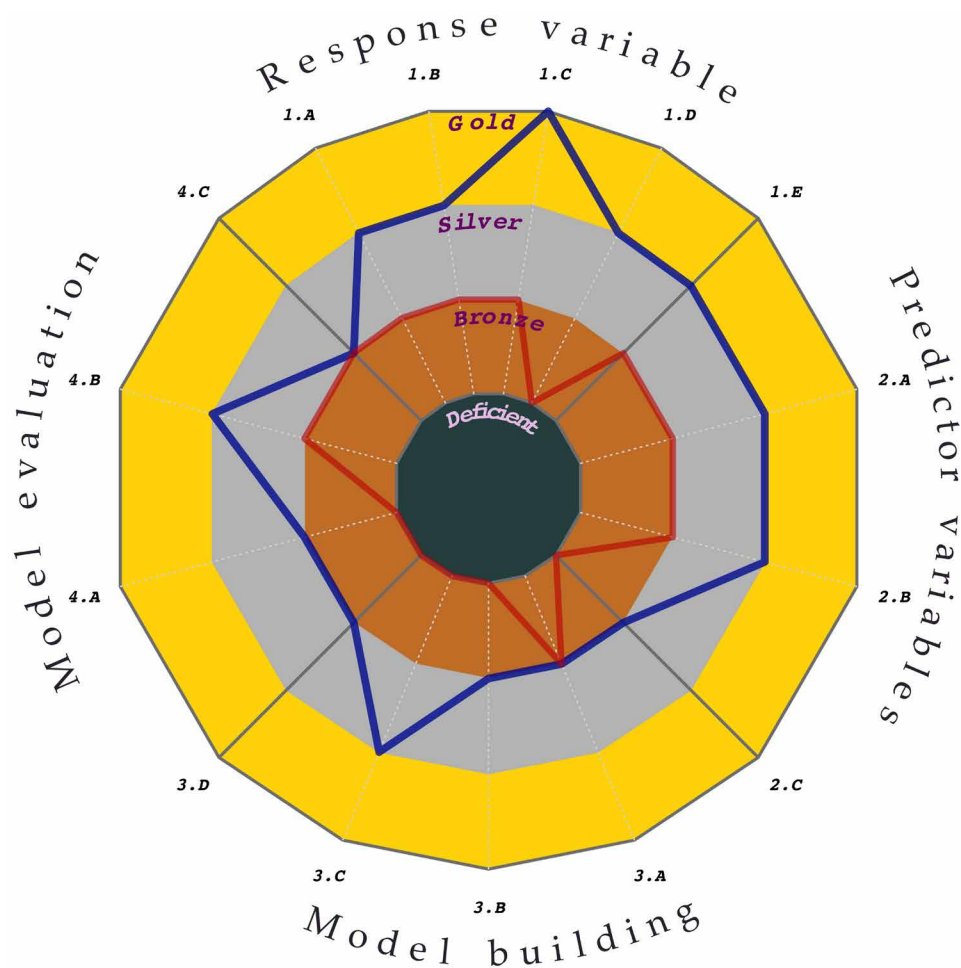
The central tendency of the scores of studies was generally low between 1995 and 2005 but varied substantially among the four key

aspects of modeling. Unfortunately, 46% of the studies were classified as deficient for one or more of the 15 specific issues. Nevertheless, a closer inspection of the 50 and 90% quantiles of the scores revealed that the quality of the data used by models had generally higher scores than aspects related with the implementation of models. Specifically, we quantified overall performance by an “area inside the line” measure, a relative metric that reaches 100% if all issues reach gold standard for all studies examined (at or above the given quantile) (Fig. 3). For this measure, response variables had the greatest relative value, with 7% (for studies at or above the median) and 48% (for those at or above the 90% quantile) performance, followed by predictor variables with 7 and 36% for those respective quantiles of studies (Fig. 3). In contrast, model building was scored only at 0 and 17%, and model evaluation achieved 3 and 18% for studies in the same respective quantiles (Fig. 3).

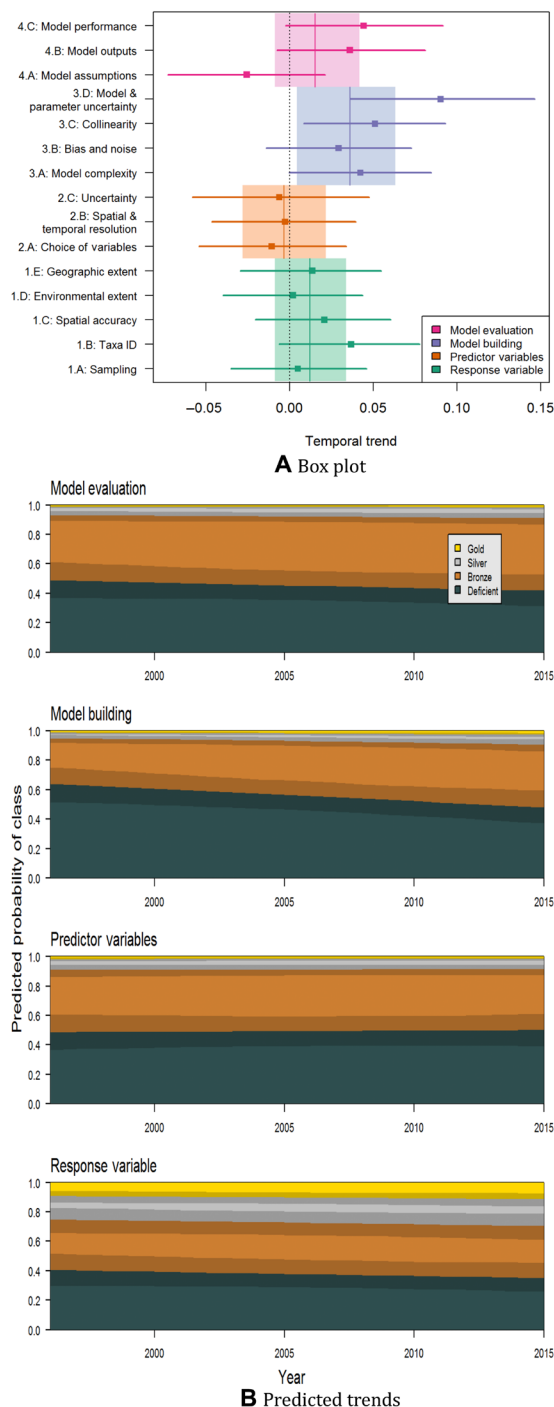
When examining whether changes existed in adequacy of the modeling studies over the 20-year period reviewed, we found positive trends in the performance scores for three of the four aspects of modeling (via ordinal regressions characterizing the change in probability that studies reached different standards across time periods;

see text S3). Despite the enormous increase in the number of studies published over this period, a general tendency existed for improvement in quality (except for predictor variables; Fig. 4 and fig. S1.7). Notably, procedures for model building were estimated to have a 3.8% yearly increase in the probability of improvement. Model evaluation and response data also showed tendencies for improvement (1.8 and 1.2% yearly, respectively; fig. S1.7), although the directions of change were less certain. Each of these three aspects saw decreases in the proportion of deficient studies, with model building and model evaluation showing a concomitant increase in those reaching the bronze level (Fig. 4B).

These general trends for each of the four aspects of modeling were consistent across the particular issues associated with response data and predictor variables, but differed across issues for model building and model evaluation. Temporal trends for each issue were slightly to moderately positive for response data and slightly negative for predictor variables (yet still within the 95% credible intervals); in contrast, temporal trends for issues of model building showed moderate to high improvement (almost always significant), while those related to model evaluation exhibited divergent trends in quality,



**Fig. 3. Best-practice standards achieved by 400 species distribution modeling studies (1995–2015).** The lines show quantiles of scores for each issue across all studies (blue = top 90% of studies; red = top 50%). See tables S2.1, S2.2, S2.3, and S2.4 for definition of standards. The area inside the respective polygon (defined by the blue and red lines) is used as a metric of overall quality of the models. The greater the area inside the polygon, the higher the overall scores for the standards. Details on the selection and scoring of articles are provided in text S3 and table S3.1.



**Fig. 4. Changes in best-practice standards of species distribution modeling studies over time.** The diagrams show the results of ordinal regression using “Year” as a continuous variable and the four key aspects of modeling as effects (including an interaction). The analysis was implemented for a sample of 400 modeling studies used for various biodiversity assessments between 1995 and 2015. **(A)** Values near zero on the x axis represent no change in standards over time, positive values indicate improvement, and bars are 95% credible intervals. **(B)** Shading represents the 95% credible intervals. Note clear increase in the number of acceptable studies regarding model building as well as lesser increases in the quality of studies with regard to model evaluation and response data. Details on the selection and scoring of articles are provided in text S3 and table S3.1. Figure S1.5 shows the raw scores used.

with wide credible intervals overlapping zero (Fig. 4A). Notably, scores for the ways of dealing with modeling and parameter uncertainty (text S2.3D) increased dramatically in quality over time. For model evaluation, issues regarding the evaluation of model outputs and measures of model performance (texts S2.4B and S2.4C) improved moderately, while evaluation of model assumptions (text S2.4A) decreased in quality. It appears that whereas researchers (including many using distribution models in associated fields) have taken up technical advances, appropriate consideration of conceptual foundations and associated assumptions has fallen behind.

In contrast to the overall temporal patterns for the full 400 papers scored, the very best studies (at or above the 90th percentile scores for each 5-year time block) demonstrated rather consistent quality over time, with two interesting exceptions. We identified the studies above the 90th percentile scores for each 5-year time block and repeated the regressions described above. This subset of high performing studies exhibited remarkably consistent scores over time for each of the four aspects of modeling (although with wider confidence intervals), as they did to a large degree for most of the 15 constituent issues (fig. S1.7). Nevertheless, these top-performing studies showed especially strong (and significant) increases regarding the ways of dealing with modeling and parameter uncertainty (text S2.3D) and measures of model performance (text S2.4C). These issues also corresponded to substantial increases for the totality of examined studies (Fig. 4A), but the even greater increases for the highest-scoring studies seems natural, as these issues remain at the forefront of development in the field.

The proposed best-practice standards and first systematic analysis of the species distribution modeling literature hint at possible factors that may have caused the temporal trends in quality, and point toward future directions for progress. Increases in response (occurrence) data quality likely derive from efforts in biodiversity informatics, which have led to huge online data sources that sometimes also include estimates of spatial error (61, 62). Unfortunately, in contrast, the explosion in availability of environmental data (e.g., via remote sensing) (63) has not produced an increase in the appropriate use of predictor variables in modeling studies; rather, researchers typically rely on climatic data interpolated from weather stations without characterization of spatial uncertainty (64). The observed increases in the quality of model building correspond to conceptual and methodological advances in the field (29, 65). While studies with gold- or silver-level quality for those issues remain only a narrow slice of the overall sampled literature (fig. S1.5), many key relevant methodological advancements have been published in the past ca. 5 years, potentially leading to improvements in the quality of studies in the coming decade. Last, regarding model evaluation, technical aspects of studies are improving, but still only a small proportion of studies achieve the highest levels of quality (fig. S1.5). In contrast, the decreasing scores for evaluation of the assumptions of modeling call for renewed attention to links to ecological theory (24, 37, 48, 66).

### Strengths and limitations of best-practice standards

Despite occasional controversies (67), the use of best-practice standards has shown to improve outcomes in several applied fields. For example, the use of guidelines for application of best practice standards in health care has been shown to save lives in surgical procedures (68). They now proliferate in a variety of medical applications, including anesthesia, mechanical ventilation, childbirth, and swine flu, and—although there are ongoing discussions about how guidelines



should be developed, tested, and applied (69)—they are unlikely to be dropped from medical practice. Guidelines for best-practice standards have also existed for quite some time in aviation to determine whether every step in complex machinery operations has been taken. One might ask why similar guidelines have not previously been established in applications of models for biodiversity science? We argue that there are at least three reasons. First, there is still relatively little pressure for findings in biodiversity research to percolate through biodiversity management decisions. It was only recently that calls for the adoption of best-practice standards for data and models used in biodiversity assessments were made by the international organization charged with examining the status of the world's biodiversity and ecosystems (6). In practice, many decisions are still based on opportunistic considerations, expert judgment, or intuition. For example, over the past two decades, spatial conservation prioritization science has developed a range of concepts and tools for optimally identifying critical areas for biodiversity conservation (70). Yet, protected areas are still identified mainly based on “ad hoc” valuations involving considerations of opportunity and cost, without taking advantage of the other quantitative tools (71).

Second, while human survival, or passenger safety on aircraft, can be easily measured, the myriad facets of biodiversity are considerably harder to define, let alone measure. What is to be maximized? Species richness? Functional diversity? Persistence? These are complex questions but, as with human health, the problem becomes tractable with pragmatism. Human health is hard to measure, but survival is simple. All aspects of biodiversity cannot be unequivocally measured with a single metric, but useful descriptors, such as probability of occurrence or persistence, can and indeed are (72, 73).

Third, perhaps a more proximate reason for the absence of guidelines of best-practice standards for data and models in biodiversity applications is the lack of agreement among the modelers themselves about what constitutes a best practice (24). Even among the authors of this article, there have been disagreements on fundamental conceptual and methodological issues (74–76). The consensus reached here, embedded in an ongoing dynamic process for updating the guidelines (tables S2.1, S2.2, S2.3, and S2.4), also published as an editable wiki (77), represents a landmark that we hope will help biodiversity assessors navigate through the plethora of published papers as well as contribute to increasing the quality of the models used in biodiversity assessments.

Another positive outcome of the proposed best-practice standards framework is that it raises awareness about model quality among authors, reviewers, and editors. Although our standards are aimed at applied uses, they will lead to a larger proportion of studies that acknowledge and, to the best of current ability, address the 15 fundamental issues identified in the current proposed standards. Clearly, even if considered in the realm of peer review, failing to meet high standards for every possible issue should not be considered as automatic grounds for accepting or rejecting an article submitted to a scientific journal. For example, some studies might be deficient in one or several issues, while presenting a significant advance for another, thus meriting publication. For peer review, the critical issue is adherence to basic methodological and reporting quality guidelines that ensure reproducibility (78) and several systems have been proposed to this end, e.g. ROSES (RepOrting standards for Systematic Evidence Syntheses) (79). More fundamentally, while the best-practice standards and guidelines developed here represent a consensus among several experienced authors in the field of species

distribution modeling, the very nature of the scientific endeavor is that consensus can and should be challenged. Nevertheless, progress is faster when it builds on comprehensive synthesis of existing knowledge. We hope that by providing a synthetic understanding of the different strengths and weaknesses of the data and models used for species distribution modeling, we will stimulate future developments in modeling and incremental improvements of the proposed framework. It is important to acknowledge, though, that while scientific progress requires unhindered questioning and challenging of established “truths,” the application of scientific knowledge to issues of societal relevance (such as human health, security, or environmental protection) is ideally based on existing scientific consensus. That is, biodiversity assessment and policy decision making need to take into account known possible uncertainties and deficiencies in all aspects of the models deemed relevant, and use them to weight the results of each study.

### Feasibility of application of best-practice standards

The wealth of published studies using SDMs to make inferences about spatiotemporal dynamics of biodiversity is huge and increasing daily, but the application of best-practice standards, such as those proposed here, is nevertheless realistic and valuable. Scoring of studies that used SDMs required between 15 to 30 min per study based on the proposed standards. Therefore, scaling the assessments to, for instance, 5000 studies would take between 1250 and 2500 person-hours. Because this effort is always likely to be spread among several individuals working in parallel, the magnitude of the task is feasible. The cost of not using agreed best-practice standards to evaluate the quality of the modeling studies feeding into biodiversity assessments is potentially far greater. The risk of not acting in the face of imminent threats, or of prioritizing action toward incorrectly inflated or underestimated threats, increases if inadequate data and models are used (80). Naturally, biases and differences of interpretation of the standards might arise among assessors, especially when many assessors with different backgrounds are involved. In our test case, differences among assessors were small, but it is generally important to quantify uncertainties by repeating the scoring with different assessors in a subsample of the studies (as here). Such an approach is possible in scientific consensus-driven assessments, which often involve many scientists, like the four regional assessments currently being undertaken by IPBES (6, 14).

Critically, some taxa and regions will not be amenable to the highest level of modeling quality, but the proposed standards still identify reasonable approaches likely to be applicable to large swaths of Earth's biodiversity. For example, it is not currently feasible to survey most species distributions comprehensively and systematically across environmental gradients (silver-level standard for sampling the response variable, table S2.1) in less-studied, difficult-to-access regions, such as most of the tropics (81). Nevertheless, the field already has developed procedures for avoiding, or at least flagging, extrapolation to conditions outside the extent of each predictor variable used to train the models (82, 83), a sufficient practice for this issue (bronze level) that is achievable for huge numbers of tropical plants, vertebrates, and even many groups of invertebrates (which are limited more by taxonomic knowledge). As another example, whereas gold- and silver-level standards require uncertainty to be substantially reduced or quantitatively included in models, bronze-level standards call for the likely components of uncertainty to be qualitatively characterized or carefully minimized and interpreted. Often due to data

limitations (such as those pervasive for some taxa and regions), bronze-standard models cannot do this quantitatively, but they still can be appropriate for inclusion in biodiversity assessments because they include explicit consideration of uncertainty.

## CONCLUSIONS

Global and regional biodiversity assessments based on scientific consensus are instrumental in providing the scientific foundation for society-wide, local-to-international discussions and policy-making. The vision we present here for structured, community agreed-upon standards supports this goal. Consensus building on environmental issues with policy relevance is far from trivial and, arguably, best informed by objective analysis and synthesis of the existing scientific literature, aimed at identifying and using only studies of sufficient quality. The best-practice standards we propose constitute a tangible step toward improving the scientific foundation of future biodiversity assessments while providing a cornerstone of increased transparency and accountability. Our detailed first implementation of the best-practice standards for widely used SDMs demonstrates the viability of the overall framework for this particular class of modeling approach. We encourage the scientific community to improve the proposed best-practice standards via a process that leads to periodic updates. We provide a wiki version of the detailed guidelines (77) that enables two-way communication between the stewards of the standards and both the scientific community and end-users. As noted above, the overall structure of the standards likely also will prove useful for other classes of models. Our proposed standards, therefore, could be emulated across the vast array of biodiversity models via the general principles of consistency, repeatability, transparency, and framework provided here.

## SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at <http://advances.sciencemag.org/cgi/content/full/5/1/eaat4858/DC1>

Supplementary Text

Text S1. Uses of models in biodiversity assessments.

Text S2. Guidelines for scoring models in biodiversity assessments.

Text S2.1. Guidelines for the response variable.

Text S2.1A. Sampling of response variables.

Text S2.1B. Identification of taxa.

Text S2.1C. Spatial accuracy of response variable.

Text S2.1D. Environmental extent across which response variable is sampled.

Text S2.1E. Geographic extent across which response variable is sampled (includes occurrence data and absence, pseudo-absence, or background data).

Text S2.2. Guidelines for the predictor variables.

Text S2.2A. Selection of candidate variables.

Text S2.2B. Spatial and temporal resolution of predictor variables.

Text S2.2C. Uncertainty in predictor variables.

Text S2.3. Guidelines for model building.

Text S2.3A. Model complexity.

Text S2.3B. Treatment of bias and noise in response variables.

Text S2.3C. Treatment of collinearity.

Text S2.3D. Dealing with modeling and parameter uncertainty.

Text S2.4. Guidelines for model evaluation.

Text S2.4A. Evaluation of model assumptions.

Text S2.4B. Evaluation of model outputs.

Text S2.4C. Measures of model performance.

Text S3. Scoring a representative sample of the literature according to the guidelines.

Text S4. Glossary.

Fig. S1.1. Classification of 400 randomly sampled papers applying SDMs to biodiversity assessments according to the number and taxonomic group of species modeled.

Fig. S1.2. Classification of 400 randomly sampled papers applying SDMs to biodiversity assessments according to the continent and ecological realm of focus.

Fig. S1.3. Accumulated percentage of papers reviewed falling in different classes as the size of the random sample is increased.

Fig. S1.4. The continents used to classify papers applying SDMs to biodiversity assessments.

Fig. S1.5. Frequencies of scores of different categories of issues assessed.

Fig. S1.6. Differences between scores obtained in the first assessment of the studies and the second independent reevaluation by a different assessor.

Fig. S1.7. Changes in species distribution modeling standards over time (1995–2015).

Fig. S1.8. Magnitude of standard deviations (0, no error; 1, maximum error) between first and second independent scoring of the studies over annual steps for each aspect and issue judged.

Table S1.1. Search terms used to select papers for the literature characterization.

Table S1.2. Classification of the purpose for which SDMs are used.

Table S1.3. Classification of the conservation applications of SDMs.

Table S2.1. Guidelines—Response variable.

Table S2.2. Guidelines—Predictor variables.

Table S2.3. Guidelines—Model building.

Table S2.4. Guidelines—Model evaluation.

Table S3.1. Search terms used to select papers using SDMs for biodiversity assessments, for the purpose of scoring according to the guidelines.

## REFERENCES AND NOTES

1. S. L. Lewis, M. A. Maslin, Defining the Anthropocene. *Nature* **519**, 171–180 (2015).
2. A. D. Barnosky, N. Matzke, S. Tomiya, G. O. U. Wogan, B. Swartz, T. B. Quental, C. Marshall, J. L. McGuire, E. L. Lindsey, K. C. Maguire, B. Mersey, E. A. Ferrer, Has the Earth's sixth mass extinction already arrived? *Nature* **471**, 51–57 (2011).
3. D. U. Hooper, E. C. Adair, B. J. Cardinale, J. E. K. Byrnes, B. A. Hungate, K. L. Matulich, A. Gonzalez, J. E. Duffy, L. Gamfeldt, M. I. O'Connor, A global synthesis reveals biodiversity loss as a major driver of ecosystem change. *Nature* **486**, 105–108 (2012).
4. H. M. Pereira, P. W. Leadley, V. Proença, R. Alkemade, J. P. W. Scharlemann, J. F. Fernandez-Manjarrés, M. B. Araújo, P. Balvanera, R. Biggs, W. W. L. Cheung, L. Chini, H. D. Cooper, E. L. Gilman, S. Guénette, G. C. Hurtt, H. P. Huntington, G. M. Mace, T. Oberdorff, C. Revenga, P. Rodrigues, R. J. Scholes, U. R. Sumaila, M. Walpole, Scenarios for global biodiversity in the 21st century. *Science* **330**, 1496–1501 (2010).
5. M. C. Urban, G. Bocedi, A. P. Hendry, J.-B. Mihoub, G. Pe'er, A. Singer, J. R. Bridle, L. G. Crozier, L. De Meester, W. Godsoe, A. Gonzalez, J. J. Hellmann, R. D. Holt, A. Huth, K. Johst, C. B. Krug, P. W. Leadley, S. C. F. Palmer, J. H. Pantel, A. Schmitz, P. A. Zollner, J. M. J. Travis, Improving the forecast for biodiversity under climate change. *Science* **353**, eaad8466 (2016).
6. S. Ferrier, K. N. Ninan, P. Leadley, R. Alkemade, *The Methodological Assessment Report on Scenarios and Models of Biodiversity and Ecosystem Services* (IPBES, 2016), 348 p.
7. M. B. Araújo, C. Rahbek, How does climate change affect biodiversity? *Science* **313**, 1396–1397 (2006).
8. T. P. Dawson, S. T. Jackson, J. I. House, I. C. Prentice, G. M. Mace, Beyond predictions: Biodiversity conservation in a changing climate. *Science* **332**, 53–58 (2011).
9. R. A. Garcia, M. Cabeza, C. Rahbek, M. B. Araújo, Multiple dimensions of climate change and their implications for biodiversity. *Science* **344**, 1247579 (2014).
10. W. Steffen, K. Richardson, J. Rockström, S. E. Cornell, I. Fetzer, E. M. Bennett, R. Biggs, S. R. Carpenter, W. de Vries, C. A. de Wit, C. Folke, D. Gerten, J. Heinke, G. M. Mace, L. M. Persson, V. Ramanathan, B. Reyers, S. Sörlin, Planetary boundaries: Guiding human development on a changing planet. *Science* **347**, 1259855 (2015).
11. C. Bellard, C. Bertelsmeier, P. Leadley, W. Thuiller, F. Courchamp, Impacts of climate change on the future of biodiversity. *Ecol. Lett.* **15**, 365–377 (2012).
12. W. Thuiller, M. B. Araújo, R. G. Pearson, R. J. Whittaker, L. Brotons, S. Lavorel, Uncertainty in predictions of extinction risk. *Nature* **430**, 34 (2004).
13. A. Guisan, R. Tingley, J. B. Baumgartner, I. Naujokaitis-Lewis, P. R. Sutcliffe, A. I. T. Tulloch, T. J. Regan, L. Brotons, E. McDonald-Madden, C. Mantyka-Pringle, T. G. Martin, J. R. Rhodes, R. Maggini, S. A. Setterfield, J. Elith, M. W. Schwartz, B. A. Wintle, O. Broennimann, M. Austin, S. Ferrier, M. R. Kearney, H. P. Possingham, Y. M. Buckley, Predicting species distributions for conservation decisions. *Ecol. Lett.* **16**, 1424–1435 (2013).
14. C. Perrings, A. Duraipapp, A. Larigauderie, H. Mooney, The biodiversity and ecosystem services science-policy interface. *Science* **331**, 1139–1140 (2011).
15. W. B. Foden, B. E. Young, *IUCN SSC Guidelines for Assessing Species Vulnerability to Climate Change* (IUCN Species Survival Commission, Cambridge, UK, and Gland, Switzerland, 2016), vol. 59, 114 p.
16. IPCC, *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (Cambridge Univ. Press, 2014), 1132 p.
17. M. B. Araújo, D. Alagador, M. Cabeza, D. Nogués-Bravo, W. Thuiller, Climate change threatens European conservation areas. *Ecol. Lett.* **14**, 484–492 (2011).

18. F. V. Faleiro, R. B. Machado, R. D. Loyola, Defining spatial conservation priorities in the face of land-use and climate change. *Biol. Conserv.* **158**, 248–257 (2013).
19. C. Kremen, A. Cameron, A. Moilanen, S. J. Phillips, C. D. Thomas, H. Beentje, J. Dransfield, B. L. Fisher, F. Glaw, T. C. Good, G. J. Harper, R. J. Hijmans, D. C. Lees, E. Louis Jr., R. A. Nussbaum, C. J. Raxworthy, A. Razafimphanana, G. E. Schatz, M. Vences, D. R. Vieites, P. C. Wright, M. L. Zjhra, Aligning conservation priorities across taxa in Madagascar with high-resolution planning tools. *Science* **320**, 222–226 (2008).
20. C. D. Thomas, A. Cameron, R. E. Green, M. Bakkenes, L. J. Beaumont, Y. C. Collingham, B. F. N. Erasmus, M. F. de Siqueira, A. Grainger, L. Hannah, L. Hughes, B. Huntley, A. S. van Jaarsveld, G. F. Midgley, L. Miles, M. A. Ortega-Huerta, A. Townsend Peterson, O. L. Phillips, S. E. Williams, Extinction risk from climate change. *Nature* **427**, 145–148 (2004).
21. H. R. Akçakaya, S. H. M. Butchart, G. M. Mace, S. N. Stuart, C. Hilton-Taylor, Use and misuse of the IUCN Red List Criteria in projecting climate change impacts on biodiversity. *Glob. Chang. Biol.* **12**, 2037–2043 (2006).
22. D. B. Botkin, H. Saxe, M. B. Araújo, R. Betts, R. H. W. Bradshaw, T. Cedhagen, P. Chesson, T. P. Dawson, J. R. Etterson, D. P. Faith, S. Ferrier, A. Guisan, A. S. Hansen, D. W. Hilbert, C. Loehele, C. Margules, M. New, M. J. Sobel, D. R. B. Stockwell, Forecasting the effects of global warming on biodiversity. *BioScience* **57**, 227–236 (2007).
23. D. A. Fordham, H. R. Akçakaya, B. W. Brook, A. Rodríguez, P. C. Alves, E. Civantos, M. Triviño, M. J. Watts, M. B. Araújo, Adapted conservation measures are required to save the Iberian lynx in a changing climate. *Nat. Clim. Chang.* **3**, 899–903 (2013).
24. M. B. Araújo, A. T. Peterson, Uses and misuses of bioclimatic envelope modeling. *Ecology* **93**, 1527–1539 (2012).
25. O. Broennimann, U. A. Treier, H. Müller-Schärer, W. Thuiller, A. T. Peterson, A. Guisan, Evidence of climatic niche shift during biological invasion. *Ecol. Lett.* **10**, 701–709 (2007).
26. R. Tingley, M. Vallinoto, F. Sequeira, M. R. Kearney, Realized niche shift during a global biological invasion. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 10233–10238 (2014).
27. M. B. Araújo, R. J. Whittaker, R. J. Ladle, M. Erhard, Reducing uncertainty in projections of extinction risk from climate change. *Glob. Ecol. Biogeogr.* **14**, 529–538 (2005).
28. R. G. Pearson, W. Thuiller, M. B. Araújo, E. Martínez-Meyer, L. Brotons, C. McClean, L. Miles, P. Segurado, T. P. Dawson, D. C. Lees, Model-based uncertainty in species' range prediction. *J. Biogeogr.* **33**, 1704–1711 (2006).
29. R. A. Garcia, N. D. Burgess, M. Cabeza, C. Rahbek, M. B. Araújo, Exploring consensus in 21st century projections of climatically suitable areas for African vertebrates. *Glob. Chang. Biol.* **18**, 1253–1269 (2012).
30. J. C. Walsh, L. V. Dicks, W. J. Sutherland, The effect of scientific evidence on conservation practitioners' management decisions. *Conserv. Biol.* **29**, 88–98 (2014).
31. M. C. McKinnon, S. H. Cheng, R. Garside, Y. J. Masuda, D. C. Miller, Sustainability: Map the evidence. *Nature* **528**, 185–187 (2015).
32. A.-C. Mupepele, J. C. Walsh, W. J. Sutherland, C. F. Dormann, An evidence assessment tool for ecosystem services and conservation studies. *Ecol. Appl.* **26**, 1295–1301 (2016).
33. A. Schmolke, P. Thorbek, D. L. DeAngelis, V. Grimm, Ecological models supporting environmental decision making: A strategy for the future. *Trends Ecol. Evol.* **25**, 479–486 (2010).
34. M. W. Schwartz, K. Deiner, T. Forrester, P. Grof-Tisza, M. J. Muir, M. J. Santos, L. E. Souza, M. L. Wilkerson, M. Zylberberg, Perspectives on the open standards for the practice of conservation. *Biol. Conserv.* **155**, 169–177 (2012).
35. A. S. L. Rodrigues, J. D. Pilgrim, J. F. Lamoreux, M. Hoffmann, T. M. Brooks, The value of the IUCN Red List for conservation. *Trends Ecol. Evol.* **21**, 71–76 (2006).
36. A. Guisan, N. E. Zimmermann, Predictive habitat distribution models in ecology. *Ecol. Model.* **135**, 147–186 (2000).
37. A. T. Peterson, J. Soberón, R. G. Pearson, R. P. Anderson, E. Martínez-Meyer, M. Nakamura, M. B. Araújo, *Ecological Niches and Geographical Distributions. Monographs in Population Biology* (Princeton Univ. Press, 2011).
38. R. P. Anderson, Harnessing the world's biodiversity data: Promise and peril in ecological niche modeling of species distributions. *Ann. N. Y. Acad. Sci.* **1260**, 66–80 (2012).
39. G. Tassarolo, T. F. Rangel, M. B. Araújo, J. Hortal, Uncertainty associated with survey design in species distribution models. *Divers. Distrib.* **20**, 1258–1269 (2014).
40. M. Barbet-Massin, W. Thuiller, F. Jiguet, How much do we overestimate future local extinction rates when restricting the range of occurrence data in climate suitability models? *Ecography* **33**, 878–886 (2010).
41. S. Varela, J. Rodríguez, J. M. Lobo, Is current climatic equilibrium a guarantee for the transferability of distribution model predictions? A case study of the spotted hyena. *J. Biogeogr.* **36**, 1645–1655 (2009).
42. A. Guisan, C. H. Graham, J. Elith, F. Huettmann; NCEAS Species Distribution Modelling Group, Sensitivity of predictive species distribution models to change in grain size. *Divers. Distrib.* **13**, 332–340 (2007).
43. C. Mora, D. P. Tittensor, S. Adl, A. G. B. Simpson, B. Worm, How many species are there on Earth and in the ocean? *PLOS Biol.* **9**, e1001127 (2011).
44. E. Waltari, R. Schroeder, K. McDonald, R. P. Anderson, A. Carnaval, Bioclimatic variables derived from remote sensing: Assessment and application for species distribution modelling. *Methods Ecol. Evol.* **5**, 1033–1042 (2014).
45. J. Franklin, F. W. Davis, M. Ikegami, A. D. Syphard, L. E. Flint, A. L. Flint, L. Hannah, Modeling plant species distributions under future climates: How fine scale do climate projections need to be? *Glob. Chang. Biol.* **19**, 473–483 (2013).
46. M. P. Austin, K. P. Van Niel, Improving species distribution models for climate change studies: Variable selection and scale. *J. Biogeogr.* **38**, 1–8 (2011).
47. M. P. Austin, T. M. Smith, K. P. Van Niel, A. B. Wellington, Physiological responses and statistical models of the environmental niche: A comparative study of two co-occurring *Eucahyptus* species. *J. Ecol.* **97**, 496–507 (2009).
48. R. P. Anderson, A framework for using niche models to estimate impacts of climate change on species distributions. *Ann. N. Y. Acad. Sci.* **1297**, 8–28 (2013).
49. J. A. Wiens, D. Stralberg, D. Jongsomjit, C. A. Howell, M. A. Snyder, Niches, models, and climate change: Assessing the assumptions and uncertainties. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 19729–19736 (2009).
50. N. W. Synes, P. E. Osborne, Choice of predictor variables as a source of uncertainty in continental-scale species distribution modelling under climate change. *Glob. Ecol. Biogeogr.* **20**, 904–914 (2011).
51. V. Braunisch, J. Coppes, R. Arlettaz, R. Suchant, H. Schmid, K. Bollmann, Selecting from correlated climate variables: A major source of uncertainty for predicting species distributions under climate change. *Ecography* **36**, 971–983 (2013).
52. R. G. Pearson, T. P. Dawson, Predicting the impacts of climate change on the distribution of species: Are bioclimate envelope models useful? *Glob. Ecol. Biogeogr.* **12**, 361–371 (2003).
53. J. Elith, C. H. Graham, R. P. Anderson, M. Dudík, S. Ferrier, A. Guisan, R. J. Hijmans, F. Huettmann, J. R. Leathwick, A. Lehmann, J. Li, L. G. Lohmann, B. A. Loiselle, G. Manion, C. Moritz, M. Nakamura, Y. Nakazawa, J. M. M. Overton, A. T. Peterson, S. J. Phillips, K. Richardson, R. Scachetti-Pereira, R. E. Schapire, J. Soberón, S. Williams, M. S. Wisz, N. E. Zimmermann, Novel methods improve prediction of species' distributions from occurrence data. *Ecography* **29**, 129–151 (2006).
54. C. Merow, M. J. Smith, T. C. Edwards Jr., A. Guisan, S. M. McMahon, S. Normand, W. Thuiller, R. O. Wüest, N. E. Zimmermann, J. Elith, What do we gain from simplicity versus complexity in species distribution models? *Ecography* **37**, 1267–1281 (2014).
55. D. García-Callejas, M. B. Araújo, The effects of model and data complexity on predictions from species distributions models. *Ecol. Model.* **326**, 4–12 (2016).
56. S. Varela, R. P. Anderson, R. García-Valdés, F. Fernández-González, Environmental filters reduce the effects of sampling bias and improve predictions of ecological niche models. *Ecography* **37**, 1084–1091 (2014).
57. M. B. Araújo, M. New, Ensemble forecasting of species distributions. *Trends Ecol. Evol.* **22**, 42–47 (2007).
58. J. A. F. Diniz-Filho, L. M. Bini, T. F. Rangel, R. D. Loyola, C. Hof, D. Nogués-Bravo, M. B. Araújo, Partitioning and mapping uncertainties in ensembles of forecasts of species turnover under climate change. *Ecography* **32**, 897–906 (2009).
59. L. Buisson, W. Thuiller, N. Casajus, S. Lek, G. Grenouillet, Uncertainty in ensemble forecasting of species distribution. *Glob. Chang. Biol.* **16**, 1145–1157 (2010).
60. R. Levins, The strategy of model building in population biology. *Amer. Sci.* **54**, 421–431 (1966).
61. J. Soberón, A. T. Peterson, Biodiversity informatics: Managing and applying primary biodiversity data. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **359**, 689–698 (2004).
62. B. R. Stein, J. R. Wiecek, Mammals of the World: MaNIS as an example of data integration in a distributed network environment. *Biodiv. Inform.* **1**, 14–22 (2004).
63. B. A. Bradley, E. Fleishman, Can remote sensing of land cover improve species distribution modelling? *J. Biogeogr.* **35**, 1158–1159 (2008).
64. R. J. Hijmans, S. E. Cameron, J. L. Parra, P. G. Jones, A. Jarvis, Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* **25**, 1965–1978 (2005).
65. R. P. Anderson, I. Gonzalez Jr., Species-specific tuning increases robustness to sampling bias in models of species distributions: An implementation with Maxent. *Ecol. Model.* **222**, 2796–2811 (2011).
66. J. Franklin, *Mapping Species Distributions-Spatial Inference and Prediction. Ecology, Biodiversity, and Conservation* (Cambridge Univ. Press, 2009).
67. A. A. Gawande, *The Checklist Manifesto: How to Get Things Right* (Metropolitan Books, 2011).
68. A. B. Haynes, T. G. Weiser, W. R. Berry, S. R. Lipsitz, A.-H. Breizat, E. P. Dellinger, T. Herbosa, S. Joseph, P. L. Kibatala, M. C. M. Lapitan, A. F. Merry, K. Moorthy, R. K. Reznick, B. Taylor, A. A. Gawande; Safe Surgery Saves Lives Study Group, A surgical safety checklist to reduce morbidity and mortality in a global population. *N. Engl. J. Med.* **360**, 491–499 (2009).
69. E. Anthes, Hospital checklists are meant to save lives—So why do they often fail? *Nature* **523**, 516–518 (2015).
70. A. Moilanen, K. Wilson, H. Possingham, *Spatial Conservation Prioritization: Quantitative Methods and Computational Tools* (Oxford Univ. Press, 2009).

71. F. L. Gibson, A. A. Rogers, A. D. M. Smith, A. Roberts, H. Possingham, M. McCarthy, D. J. Pannell, Factors influencing the use of decision support tools in the development and design of conservation policy. *Environ. Sci. Policy* **70**, 1–8 (2017).
72. S. H. M. Butchart, M. Walpole, B. Collen, A. van Strien, J. P. W. Scharlemann, R. E. A. Almond, J. E. M. Baillie, B. Bomhard, C. Brown, J. Bruno, K. E. Carpenter, G. M. Carr, J. Chanson, A. M. Chenery, J. Csirke, N. C. Davidson, F. Dentener, M. Foster, A. Galli, J. N. Galloway, P. Genovesi, R. D. Gregory, M. Hockings, V. Kapos, J.-F. Lamarque, F. Leverington, J. Loh, M. A. McGeoch, L. McRae, A. Minasyan, M. H. Morcillo, T. E. E. Oldfield, D. Pauly, S. Quader, C. Revenga, J. R. Sauer, B. Skolnik, D. Spear, D. Stanwell-Smith, S. N. Stuart, A. Symes, M. Tierney, T. D. Tyrrell, J.-C. Vie, R. Watson, Global biodiversity: Indicators of recent declines. *Science* **328**, 1164–1168 (2010).
73. H. M. Pereira, S. Ferrier, M. Walters, G. N. Geller, R. H. G. Jongman, R. J. Scholes, M. W. Bruford, N. Brummitt, S. H. M. Butchart, A. C. Cardoso, N. C. Coops, E. Dulloo, D. P. Faith, J. Freyhof, R. D. Gregory, C. Heip, R. Höft, G. Hurtt, W. Jetz, D. S. Karp, M. A. McGeoch, D. Obura, Y. Onoda, N. Pettorelli, B. Reyers, R. Sayre, J. P. W. Scharlemann, S. N. Stuart, E. Turak, M. Walpole, M. Wegmann, Essential Biodiversity Variables. *Science* **339**, 277–278 (2013).
74. C. F. Dormann, Promising the future? Global change projections of species distributions. *Basic Appl. Ecol.* **8**, 387–397 (2007).
75. C. M. Beale, J. J. Lennon, A. Gimona, Opening the climate envelope reveals no macroscale associations with climate in European birds. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 14908–14912 (2008).
76. M. B. Araújo, W. Thuiller, N. G. Yoccoz, Re-opening the climate envelope reveals macroscale associations with climate in European birds. *Proc. Natl. Acad. Sci. U.S.A.* **106**, E45–E46 (2009).
77. Standards for biodiversity models; <http://macroecology.ku.dk/resources/stdBiodivModels>.
78. M. R. Munafò, B. A. Nosek, D. V. M. Bishop, K. S. Button, C. D. Chambers, N. Percie du Sert, U. Simonsohn, E.-J. Wagenmakers, J. J. Ware, J. P. A. Ioannidis, A manifesto for reproducible science. *Nat. Hum. Behav.* **1**, 0021 (2017).
79. K. Pussegoda, L. Turner, C. Garrity, A. Mayhew, B. Skidmore, A. Stevens, I. Boutrou, R. Sarkis-Onofre, L. M. Bjerre, A. Hróbjartsson, D. G. Altman, D. Moher, Systematic review adherence to methodological or reporting quality. *Syst. Rev.* **6**, 131 (2017).
80. K. S. Shrader-Frechette, E. D. McCoy, *Method in Ecology: Strategies for Conservation* (Cambridge Univ. Press, 1993).
81. K. J. Feeley, J. T. Stroud, T. M. Perez, Most 'global' reviews of species' responses to climate change are not truly global. *Divers. Distrib.* **23**, 231–234 (2017).
82. J. Elith, M. Kearney, S. Phillips, The art of modelling range-shifting species. *Methods Ecol. Evol.* **1**, 330–342 (2010).
83. M. B. Mesgaran, R. D. Cousins, B. L. Webber, Here be dragons: A tool for quantifying novelty due to covariate range and correlation change when projecting species distribution models. *Divers. Distrib.* **20**, 1147–1159 (2014).

**Acknowledgments:** We thank K. Boehning-Gaese, F. Guilhaumon, and C. H. Graham for discussion at early stage of this study and N. Melo for assistance with Fig. 2. **Funding:** This study is the output of Working Group 3 (Agreed Standards for Biodiversity Models) of EC COST Action ES1101: Harmonizing Global Biodiversity Modelling (HarmBio). Additional sponsorship was obtained through the Center for Macroecology, Evolution and Climate funded by the Danish National Research Foundation (no. DNRF96). We also acknowledge funding support by Portuguese Foundation for Science and Technology (no. PTDC/AAG-GLO/0463/2014) to M.B.A., National Science Foundation (no. DEB-1119915) to R.P.A., Swiss National Science Foundation (nos. CR23I2\_162754 and 31003A-152866) to A.G., and Swiss National Science Foundation (no. 310030L-170059) to N.Z. **Author contributions:** All authors contributed to the working-group discussions, writing, and analysis of the review article. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials, or the references cited here within. Additional data related to this paper may be requested from the authors.

Submitted 3 March 2018  
Accepted 11 December 2018  
Published 16 January 2019  
10.1126/sciadv.aat4858

**Citation:** M. B. Araújo, R. P. Anderson, A. M. Barbosa, C. M. Beale, C. F. Dormann, R. Early, R. A. Garcia, A. Guisan, L. Maiorano, B. Naimi, R. B. O'Hara, N. E. Zimmermann, C. Rahbek, Standards for distribution models in biodiversity assessments. *Sci. Adv.* **5**, eaat4858 (2019).