**Supporting Information**

**Methods**

**Database.** We collated a trait database for 4,461 terrestrial mammal species. We based our analysis on the taxonomy provided in ‘Mammal Species of the World 3’ [1]. We classified species as

non-threatened (LC, NT), threatened (VU, EN, CR) and Data Deficient (DD) (IUCN 2008). We treated species as threatened or non-threatened, as highly imbalanced categories (2,826 LC species versus 157 CR species) are difficult to discriminate using predictive models (Webb 2002) and uncertainty around classifications with multiple categories is difficult to interpret and communicate. In contrast, machine learning predictions from our binary classification provide a simple quantification of both the likely probability of threatened status for each species and the level of uncertainty around that prediction. We did not classify species as LC vs. (NT, VU, EN, CR) as experience in Red Listing from one of the authors (B. Collen) for over 15,000 species indicates that the NT category may be one of the least consistently applied categories among taxonomic groups, especially in mammals where assessments were coordinated by more than 20 Specialist Groups. The NT category is bounded on its upper limit (species are not threatened) but not its lower limit (NT species may likely be LC) and may be less robust than the limit between the NT and VU categories for modelling risk. However, future studies could investigate the impact of classifying NT species as threatened.

We selected the following life-history and ecological variables due to their high completeness [2]: body mass (68.4% complete), litter size (48.8% complete), habitat breadth (52% complete), trophic level (42% complete), and number of IUCN listed habitats (100% complete). Because some ML methods cannot cope with missing data, we phylogenetically imputed missing life-history and ecological variables using a global mammal phylogeny [3] and the PhyloPARS method [4]. PhyloPARS estimates missing values at the nodes of a phylogeny using a limited number of observations and a specified evolutionary model. Additionally, it allows for correlated evolution of different traits and intraspecific variability of traits. Species which were not present on the phylogeny were assigned the median trait value for their genus or family. We recorded the biogeographical realm of each species from the IUCN Red List assessments [5], as well as their geographical range size [6] and latitude of range centroid. We extracted habitat suitability information from [7] and computed the proportion of each species’ range deemed ‘highly suitable’. For each species, we also derived External Threat Index (ETI) values following the method proposed by [8]. The ETI for a given species is the mean threat status of all species present in the focal species’ range, weighted by the overlap in range between the focal species and all other species. The ETI is therefore a proxy measure of the level of threat within a species’ distribution. Using species range maps, environmental and anthropogenic threat variables were derived from global grids. All data extractions were conducted in ArcGIS version 9.2. All geographical variables were 100% complete for each species. Trait distributions were similar for data-sufficient and Data Deficient species (Figure S1).

**Machine Learning tools.** We compared the ability of seven commonly used Machine Learning (ML) algorithms (classification trees, random forests, boosted trees, k-nearest neighbours, support vector machines, neural networks and decision stumps) to predict species’ threat status. We briefly introduce each ML tool.

**Classification Trees.** Classification Trees (CT) were first introduced by Breiman [9] and explain variation in a response variable by repeatedly splitting the data into more homogeneous groups, using combinations of explanatory variables. Each terminal leaf is characterized by the value of the response variable, the number of observations in the group and the corresponding threshold values of the explanatory variables that define it. Predictions are made by sorting new species down the CT until a leaf is reached. CTs make no distributional assumptions about the explanatory or response variables [10], can fit non-linear relationships and high-order interactions and can handle missing values in the explanatory variables [11]. However, they can be sensitive to small changes in the underlying data and can only approximate linear functions [10]. CTs have been widely used in ecology [11], including threatened species classification [12–14]. We optimized tree depth (number of splits) during model training.

**Random Forests.** Random Forests (RF) are an ensemble method related to classification trees: many classification trees are constructed and classes are predicted by a majority vote [15]. For each tree, only a randomly chosen subset of the explanatory variables is used at each node, which reduces correlation between trees and improves the overall classification accuracy of the RF. RFs are generally robust to overfitting, outliers and noise. RFs give direct estimates of variable importance and can model complex interactions between explanatory variables [16], but unlike CTs they do not provide a simple representation of the classifier in the form of a single tree. RFs have become increasingly popular in ecology due to their high predictive power [10,16]. We grew 500 trees at each model iteration and optimized the number of variables chosen randomly at each node.

**Boosted Classification Trees.** Boosted Classification Trees (BT) are an ensemble method constructed with a boosting algorithm [17] which begin by constructing a single classification tree. Instances (in our case, species) are then weighted by whether the initial tree predicted their class correctly. If the model did not predict the class correctly, the instance is given extra weight in the following classification tree. The process continues until a stopping criterion is reached, such as a predefined number of trees. The final model predicts the class membership of a new example with a weighted voting scheme, where the voting power of each tree is proportional to its accuracy. BTs can therefore modify a classification tree with low predictive accuracy (‘weak classifier’) into a ‘strong classifier’ by focusing on difficult cases [17]. The final model can be understood as an additive regression model in which individual terms are classification trees. BTs often show high predictive accuracy [18] but are prone to overfitting. We optimized the number of trees grown, tree depth and learning rate.

**K-Nearest Neighbours.** The K-Nearest Neighbour (KNN) is a learning algorithm based on instances [19]. Given an instance (in our case, species) its *k* closest neighbours are found in the *n*-dimensional feature space, where *n* denotes the number of explanatory variables. The class label of the instance is determined using a majority vote of the neighbours. A number of distance metrics have been proposed, but the most commonly used is Euclidian distance. KNNs have low memory requirements, but are sensitive to irrelevant explanatory variables and can exhibit higher error rates than more advanced methods. For each dataset, we chose the best performing classifiers created with a range of *k* values.

**Support Vector Machines.** Support Vector Machines (SVM) rely on processing the data to represent the pattern in a high dimension, typically much higher than the original feature space [19]. Using a kernel function, a SVM constructs a separating hyperplane between the training instances of both classes in the new space. Training of the SVM allows the determination of the separating hyperplane with the largest margin between the two classes. The support vectors are the training samples that define the optimal separating hyperplane and are the most difficult cases to classify; they are the patterns most informative for the classification task. SVMs are highly accurate classifiers, which do not suffer from local optima and are less prone to over-fitting than other methods [20]. However, the parameters of the model are difficult to interpret. We used a Radial Basis kernel function and optimized sigma (inverse kernel width) and *c*, the cost of constraint violation.

**Neural Networks.** Neural networks are non-linear mapping structures based on Rosenblatt’s perceptron. The most popular neural network is the multi-layer feed-forward network trained by a back-propagation algorithm [21]. Neurons are arranged in successive layers, and information flows uni-directionally from the input layer (explanatory variables) to the output layer (response variable) through the hidden layer(s). Each hidden neuron is connected to each input and output neuron, and the strength of the initial connections are determined at the start training. Predicted and observed classes are compared, and learning is achieved through the updating of weights at each connection using back-propagation. Neural Networks often show high predictive performance and have been used in a wide range of ecological studies, however they suffer from slow training, and have been criticized for being a “black box” method with a tendency to overfit the data [22].

**Decision stumps.** In order to assess the role of geographical range size in determining extinction risk, we computed a decision stump (DT) for each dataset. Decision stumps are CTs derived from a single explanatory variable, in our case, geographical range size. Decision stumps effectively identify a geographical range size threshold above and below which species are attributed to a threat level.

**Training of Machine Learning tools.** We pre-processed the predictor variables as described in the package *caret* [23]. Numeric predictors were transformed, centred and scaled to a mean of zero and standard deviation of one, a common procedure in ML data pre-processing. For each taxonomic dataset separately, we then removed variables with near-zero variance, as these predictors may acquire zero-variance when the data are split into cross-validation sub-samples. We also removed highly correlated predictors (correlation coefficient > 0.9) as these can bias model fitting procedures. We set aside all DD species to form a prediction set. We partitioned non-DD species into a training set comprising 75% of species and a validation set comprising 25% of species, to independently assess the performance of different ML methods. For each ML tool in turn, we optimized tuning parameters using ten-fold cross-validation on the training set. During each iteration of the cross-validation, the algorithm was trained on nine tenth of the data and tested on the excluded tenth (test fold), creating a set of built classifiers. Classifier performance was estimated by comparing the predicted and observed threat level of the species in the test folds. For each combination of tuning parameters, we measured the area under the receiver operating characteristic curve (AUC). The ROC curve is a graphical plot of the sensitivity against the false positive rate (1- specificity) of a classifier. The sensitivity of the classifier is the proportion of threatened species correctly identified, while the specificity is the proportion of non-threatened species correctly identified. AUC provides a tool for model selection which is insensitive to class imbalance and does not require the specification of misclassification costs [24]. Values of AUC higher than 0.7 indicate a good fit of the classifier to the data, while values higher than 0.9 indicate an extremely good fit [25]. We selected the optimal tuning parameters for each ML tool using AUC rather than overall accuracy. Given the large class imbalance in some of our datasets, accuracy would provide a skewed measure of classifier performance. For example, 22.1% of mammals are threatened in our global dataset, hence any classifier that would classify all species as non-threatened (i.e. make no decision) would achieve an accuracy of 77.9%. Optimal tuning parameters for each ML tool can be found in Table S2. ML tools were compared independently on the validation sets previously set aside, and the best ML tool for each dataset was selected using AUC. As predictions of threat were probabilistic, predicting the threat category of a species required the determination of a predicted probability of threat above which a species should be classified as threatened. We used Youden’s index [26], to identify the optimal cutoff point. The Youden index Y is defined as (Y= sensitivity + specificity -1), and can be intuitively interpreted as the point on the ROC curve farthest from chance [27]. This method effectively assigns equal importance to sensitivity and specificity. Using the optimal cutoff point, we predicted the binary threat status of species in the validation sets and computed additional performance metrics, including specificity and sensitivity. We computed performance metrics for order-level predictions from the global model using both globally and ordinally optimized cutoff points (Table S3).

Multiple classification performance measures are commonly used among different research fields, reflecting varying attitudes towards misclassification costs [28]. To investigate the role of performance measure on our results, we repeated all analyses by maximizing the H measure, a recently developed alternative to AUC which allows the specification of misclassification costs [29, but see 30]. We selected the prior distribution of misclassification costs based on the Beta() distribution [31], where is the proportion of threatened species in the sample, and the proportion of non-threatened species in the sample. The distribution takes a balanced view of misclassification costs when faced with unbalanced datasets, setting the mode of the relative misclassification severity distribution at c= . As with models trained with AUC, we found a significant difference in performance among tools (Friedman test,**χ**2= 17.8, p=0.006, df=6). *Post hoc* symmetry tests showed that this difference was caused by the difference between highly predictive boosted trees vs. k-nearest neighbours (p=0.01, df=1), and boosted trees vs. decision stumps based on geographical range size alone (p=0.03, df=1). The best model for all mammals and rodents remained random forests, and the best model for bats remained boosted trees (Table S5). The best models in carnivores and primates were boosted trees, in contrast to neural networks and support vector machines respectively for models trained on AUC (Table S5). Model predictions between the best global model trained on AUC and the best model trained with the H measure were highly consistent (Table S7).

**Tables**

Table S1. Sources for the terrestrial mammal database.

|  |  |  |  |
| --- | --- | --- | --- |
| Variables | Unit | Source | Resolution |
| Taxonomy (Order, Family, Genus) |  | [5] |  |
| Body Mass | Grams | [2] |  |
| Litter Size |  | [2] |  |
| Habitat Breadth |  | [2] |  |
| Trophic Level |  | [2] |  |
| Number of IUCN Habitats |  | [5] |  |
| Biogeographical Realms |  |  |  |
| Range Size | km² | [6] |  |
| Latitude of Range Centroid | Degrees Latitude | [6] |  |
| High Habitat Suitability | Percent of Range Size | [6,7] |  |
| Mean Annual Temperature | Degrees (°C) | [32] | 30 arc seconds |
| Mean Temperature Seasonality | Standard deviation\*100 |  | 30 arc seconds |
| Mean Annual Precipitation | Millimetres |  | 30 arc seconds |
| Mean Precipitation Seasonality | Coefficient of variation | [32] | 30 arc seconds |
| Mean Annual Net Primary Productivity (1976-2000) | Grams per m² per year | [33] | 0.25 degrees |
| Minimum Elevation | Meters | [32] | 30 arc seconds |
| Elevation Range | Meters | [32] | 30 arc seconds |
| External Threat Index |  | [5,6] |  |
| Mean Human Population Density (2000) | People per unit area | [34] | 2.5 arc minutes |
| Minimum Human Population Density (2000) | People per unit area | [34] | 2.5 arc minutes |
| Mean Human Footprint | Human Influence Index normalized per region and biome | [35] | 30 arc seconds |
| Mean Human Appropriation of Net Primary Productivity | Percent of NPP | [33] | 0.25 degrees |
| Mean GDP (1990) | Dollars per person per year | [36] | 0.25 degrees |

Table S2. Optimal tuning parameters for models trained with AUC among datasets. CT: Classification Tree, RF: Random Forests, BT: Boosted Trees, KNN: K-Nearest Neighbours, SVM: Support Vector Machines, NNET: Neural Networks. AUC: area under the receiver operator characteristic curve.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | CT | RF | BT | |  | KNN | SVM | | NNET | |
| Tree depth | Number of variables randomly sampled at each split | Number of trees | Tree depth | Learning rate | Number of neighbours | Sigma inverse kernel width | Cost of constraint violation | Number of units in the hidden layer | Weight decay |
| Global | 0 | 6 | 212 | 11 | 0.1 | 29 | 0.0244 | 1 | 1 | 0.1 |
| Bats | 0.00931 | 9 | 96 | 13 | 0.1 | 27 | 0.0261 | 0.5 | 15 | 0.0422 |
| Carnivores | 0.0329 | 30 | 127 | 7 | 0.1 | 15 | 0.0227 | 1 | 1 | 0.00133 |
| Primates | 0.0106 | 8 | 206 | 6 | 0.1 | 9 | 0.0296 | 2 | 13 | 0 |
| Rodents | 0.0325 | 2 | 53 | 7 | 0.1 | 29 | 0.03 | 1 | 17 | 0.1 |

Table S3. Measures of model performance among validation sets for models trained on AUC. AUC: area under the receiver operator characteristic curve.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Cutoff | Sensitivity | Specificity | Accuracy | H | AUC | Youden |
| Dataset predictions | | | | | | | | |  |
| Global | 0.282 | 0.935 | 0.887 | 0.898 | 0.785 | 0.944 | 0.754 |
| Bats | 0.067 | 0.914 | 0.842 | 0.854 | 0.597 | 0.897 | 0.756 |
| Carnivores | 0.808 | 0.900 | 0.917 | 0.913 | 0.759 | 0.961 | 0.817 |
| Primates | 0.547 | 0.861 | 0.727 | 0.803 | 0.499 | 0.873 | 0.588 |
| Rodents | 0.24 | 0.843 | 0.933 | 0.918 | 0.728 | 0.951 | 0.790 |
| Order-level predictions from the global model (globally optimized cutoff point) | | | | | | | | |  |
| Bats | 0.282 | 0.821 | 0.937 | 0.916 | 0.74 | 0.956 | 0.758 |
| Carnivores | 0.282 | 0.778 | 0.939 | 0.905 | 0.773 | 0.969 | 0.717 |
| Primates | 0.282 | 1 | 0.743 | 0.888 | 0.732 | 0.955 | 0.743 |
| Rodents | 0.282 | 0.908 | 0.898 | 0.899 | 0.795 | 0.969 | 0.806 |
| Order-level predictions from the global model (ordinally optimized cutoff point) | | | | | | | | |  |
| Bats | 0.192 | 0.897 | 0.856 | 0.864 | 0.74 | 0.956 | 0.779 |
| Carnivores | 0.162 | 0.888 | 0.849 | 0.857 | 0.773 | 0.969 | 0.85 |
| Primates | 0.472 | 0.844 | 0.886 | 0.863 | 0.732 | 0.955 | 0.752 |
| Rodents | 0.566 | 0.846 | 0.973 | 0.853 | 0.795 | 0.969 | 0.834 |

Table S4. Optimal tuning parameters for models trained with the H measure among datasets. CT: Classification Tree, RF: Random Forests, BT: Boosted Trees, KNN: K-Nearest Neighbours, SVM: Support Vector Machines, NNET: Neural Networks.

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | CT | RF | BT | | |  | KNN | | SVM | | NNET | |
| Tree depth | Number of variables randomly sampled at each split | Number of trees | Tree depth | Learning rate | | | Number of neighbours | Sigma inverse kernel width | Cost of constraint violation | Number of units in the hidden layer | Weight decay |
| Global | 0.0025 | 3 | 50 | 12 | | 0.1 | 29 | | 0.0245 | 2 | 1 | 0.1 |
| Bats | 0.0124 | 10 | 175 | 9 | | 0.05 | 53 | | 0.0246 | 0.25 | 15 | 0.0422 |
| Carnivores | 0.011 | 25 | 239 | 7 | | 0.01 | 25 | | 0.0224 | 0.25 | 1 | 0.00133 |
| Primates | 0.0106 | 23 | 89 | 11 | | 0.05 | 9 | | 0.0258 | 2 | 13 | 0 |
| Rodents | 0.0135 | 5 | 118 | 5 | | 0.05 | 35 | | 0.0295 | 1 | 17 | 0.1 |

Table S5. H measure for each combination of tool and dataset on the validation sets, for models trained on H measure. CT: Classification Tree, RF: Random Forests, BT: Boosted Trees, KNN: K-Nearest Neighbours, SVM: Support Vector Machine, NNET: Neural Networks, DS: Decision Stump.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | CT | RF | BT | KNN | SVM | NNET | DS |
| Global | 0.518 | 0.642 | 0.611 | 0.519 | 0.619 | 0.557 | 0.389 |
| Bats | 0.582 | 0.555 | 0.61 | 0.479 | 0.523 | 0.509 | 0.382 |
| Carnivores | 0.505 | 0.726 | 0.789 | 0.404 | 0.663 | 0.759 | 0.372 |
| Primates | 0.356 | 0.497 | 0.560 | 0.334 | 0.525 | 0.331 | 0.272 |
| Rodents | 0.62 | 0.727 | 0.721 | 0.591 | 0.71 | 0.702 | 0.623 |

Table S6. Measures of model performance among validation sets for models trained on H measure. AUC: area under the receiver operator characteristic curve.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Cutoff | Sensitivity | Specificity | Accuracy | H | AUC | Youden |
| Global | 0.221 | 0.867 | 0.834 | 0.906 | 0.642 | 0.93 | 0.702 |
| Bats | 0.17 | 0.914 | 0.83 | 0.889 | 0.61 | 0.889 | 0.744 |
| Carnivores | 0.232 | 0.9 | 0.86 | 0.935 | 0.789 | 0.961 | 0.761 |
| Primates | 0.566 | 0.767 | 0.818 | 0.829 | 0.56 | 0.886 | 0.586 |
| Rodents | 0.17 | 0.86 | 0.887 | 0.935 | 0.727 | 0.948 | 0.773 |

Table S7. Differences in model predictions on the validation set between the best global model trained on AUC, and the best global model trained on H measure. Predictions for the model trained on AUC are taken as ‘true’ classes in the confusion matrix and following classification performance measures. Accuracy: 0.868. Sensitivity: 0.923. Specificity: 0.852. The validation set contains 991 species. AUC: area under the receiver operator characteristic curve.

|  |  |  |  |
| --- | --- | --- | --- |
|  | AUC predictions | | |
| H measure predictions |  | Non-threatened | Threatened |
| Non-threatened | 656 | 17 |
| Threatened | 114 | 204 |

Table S8. Proportion of species in the validation set correctly identified as threatened or non-threatened by the best machine learning model trained with AUC (area under the receiver operator characteristic curve).

|  |  |  |
| --- | --- | --- |
| Criterion | Proportion correctly identified | Number of species |
| IUCN Red List categories |  |  |
| CR spell these out | 0.96 | 28 |
| EN | 0.98 | 97 |
| VU | 0.87 | 94 |
| NT | 0.66 | 67 |
| LC | 0.91 | 705 |
| Threat type |  |  |
| habitat loss | 0.94 | 186 |
| invasive species | 1 | 36 |
| utilisation | 0.94 | 88 |
| Threatened species by range size |  |  |
| 0-20,000 (km²) | 0.97 | 75 |
| 20,000-100,000 (km²) | 0.90 | 39 |
| 100,000-1,000,000 (km²) | 0.95 | 60 |
| >1,000,000 (km²) | 0.87 | 45 |
| Non-threatened species by range size |  |  |
| 0-20,000 (km²) | 0.74 | 97 |
| 20,000-100,000 (km²) | 0.87 | 108 |
| 100,000-1,000,000 (km²) | 0.93 | 287 |
| >1,000,000 (km²) | 0.9 | 280 |

Table S9. Justification for listing as Data Deficient. Multiple justifications can apply to any one species.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Type records | Few records | Old records (prior to 1970) | Unknown provenance | Taxonomic uncertainty | Unknown population status and range | Unknown threats | New species |
| Number of species | 126 | 134 | 83 | 9 | 127 | 226 | 83 | 46 |

**Figures**

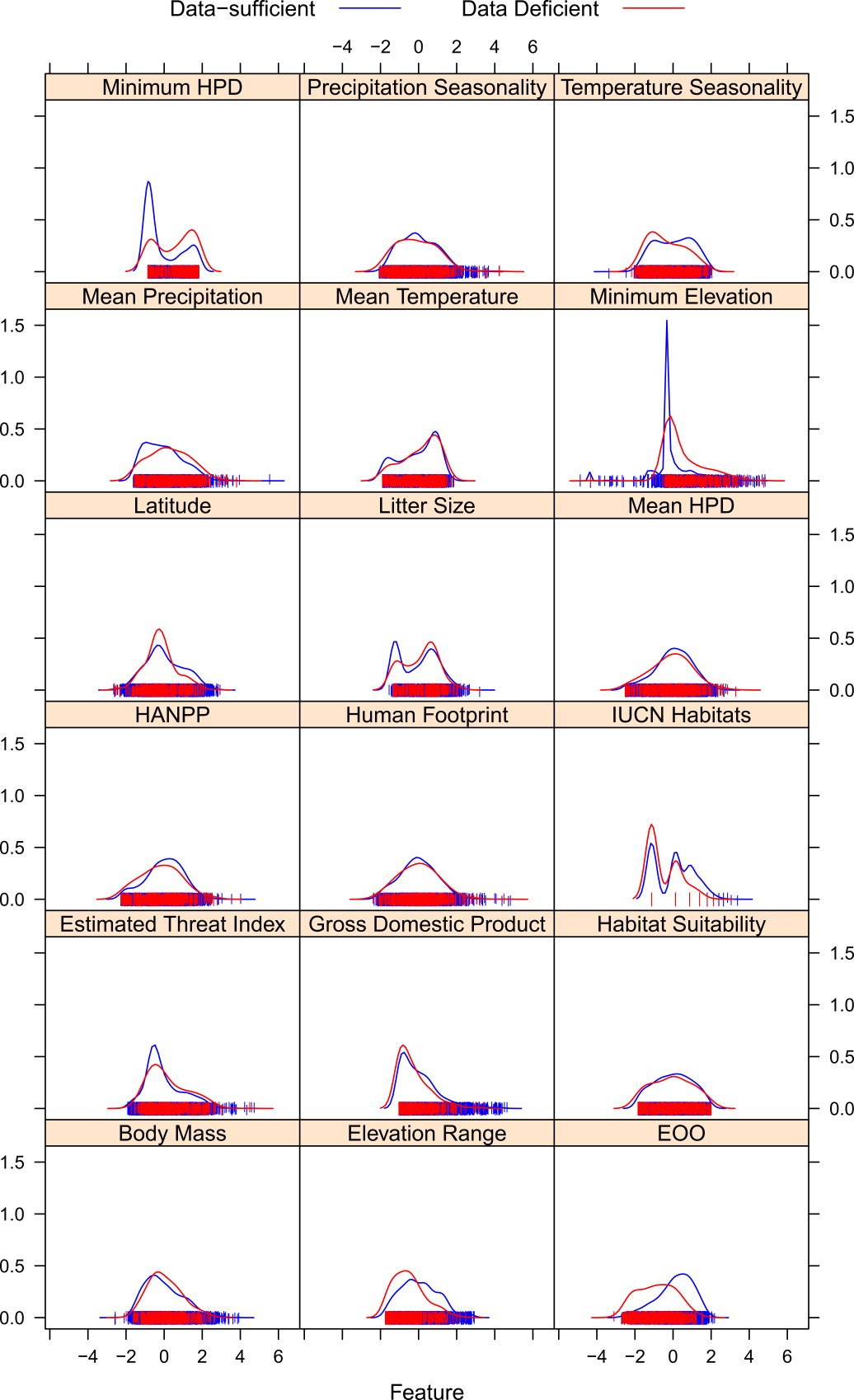


Figure S1. Overlay density plot of numeric explanatory variables in Data Deficient (n=493) and data-sufficient species (n=3,967). HPD: Human population density. HANPP: Human appropriation of net primary productivity.



Figure S2. Correlogram of residual extinction risk in assemblages in the validation set. OLS: Ordinary Least Squares model. SAR: Spatial Autoregressive model.



Figure S3. Observed and predicted assemblage threat across assemblage sizes in the global validation set (n=4,505).



Figure S4. Predicted probability of threat from the global model against Red List category in the global validation set (n=991).

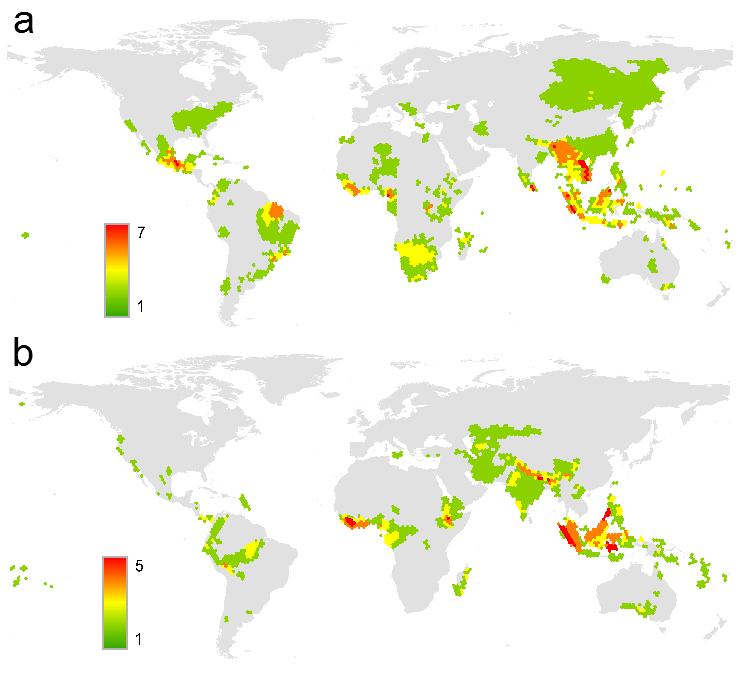


Figure S5. Number of false negative (a) and false positive (b) species classification in the global validation set (n=991). False negatives are threatened species incorrectly predicted as non-threatened. False positives are non-threatened species incorrectly predicted as threatened.

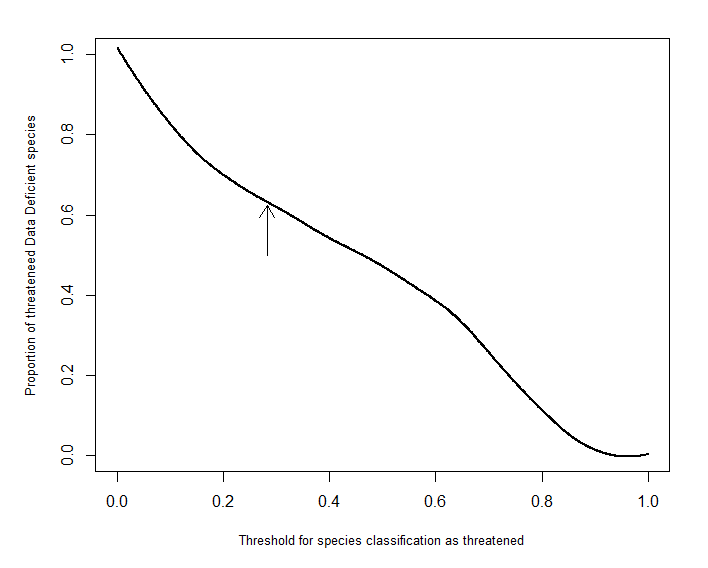


Figure S6. Relationship between the threshold for Data Deficient species classification as threatened and the proportion of threatened Data Deficient species (n=493). The arrow indicates the proportion of threatened Data Deficient species for the cutoff obtained from the best global model trained on AUC.

**References**

1. Wilson, D. E. & Reeder, D. M. 2005 *Mammal species of the world. A taxonomic and geographic reference*. Baltimore, MD: Johns Hopkins University Press.

2. Jones, K. E. et al. 2009 PanTHERIA: a species-level database of life history, ecology, and geography of extant and recently extinct mammals. *Ecology* **90**, 2648–2648.

3. Fritz, S. A., Bininda-Emonds, O. R. P. & Purvis, A. 2009 Geographical variation in predictors of mammalian extinction risk: big is bad, but only in the tropics. *Ecol. Lett.* **12**, 538–49. (doi:10.1111/j.1461-0248.2009.01307.x)

4. Bruggeman, J., Heringa, J. & Brandt, B. W. 2009 PhyloPars: estimation of missing parameter values using phylogeny. *Nucleic Acids Res.* **37**, 179–184. (doi:10.1093/nar/gkp370)

5. IUCN 2008 2008 IUCN Red List of Threatened Species.

6. IUCN 2010 2010 IUCN Red List of threatened species. Version 2010.3.

7. Rondinini, C. et al. 2011 Global habitat suitability models of terrestrial mammals. *Philos. Trans. R. Soc. B Biol. Sci.* **366**, 2633–41. (doi:10.1098/rstb.2011.0113)

8. Cardillo, M., Purvis, A., Sechrest, W., Gittleman, J. L., Biebly, J. & Mace, G. M. 2004 Human population density and extinction risk in the world’s carnivores. *PLoS Biol.* **2**, 909–914. (doi:10.1371/journal.pbio.0020197)

9. Breiman, L., Friedman, J., Olshen, R. & Stone, C. J. 1984 *Classification and regression trees*. Belmont, CA: Wadsworth International Group.

10. Prasad, A. M., Iverson, L. R. & Liaw, A. 2006 Newer classification and regression tree techniques: bagging and random forests for ecological prediction. *Ecosystems* **9**, 181–199. (doi:10.1007/s10021-005-0054-1)

11. De’ath, G. & Fabricius, K. E. 2000 Classification and regression trees: a powerful yet simple technique for ecological data analysis. *Ecology* **81**, 3178–3192. (doi:10.1890/0012-9658(2000)081[3178:CARTAP]2.0.CO;2)

12. Boyer, A. G. 2008 Extinction patterns in the avifauna of the Hawaiian islands. *Divers. Distrib.* **14**, 509–517. (doi:10.1111/j.1472-4642.2007.00459.x)

13. Larson, E. R. & Olden, J. D. 2010 Latent extinction and invasion risk of crayfishes in the southeastern United States. *Conserv. Biol.* **24**, 1099–1110. (doi:10.1111/j.1523-1739.2010.01462.x)

14. Bielby, J., Cardillo, M., Cooper, N. & Purvis, A. 2010 Modelling extinction risk in multispecies data sets: phylogenetically independent contrasts versus decision trees. *Biodivers. Conserv.* **19**, 113–127. (doi:10.1007/s10531-009-9709-0)

15. Breiman, L. 2001 Random Forests. *Statistics (Ber).* , 1–33.

16. Cutler, R. D., Edwards, T. C., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J. & Lawler, J. J. 2007 Random forests for classification in ecology. *Ecology* **88**, 2783–92. (doi:10.1890/07-0539.1)

17. Freund, Y. & Schapire, R. E. 1996 Experiments with a New Boosting Algorithm. *Proc. Thirteen. Int. Conf. Mach. Learn.* , 148–156. (doi:10.1.1.133.1040)

18. Elith, J., Leathwick, J. R. & Hastie, T. 2008 A working guide to boosted regression trees. *J. Anim. Ecol.* **77**, 802–13. (doi:10.1111/j.1365-2656.2008.01390.x)

19. Hastie, T., Tibshirani, R. & Friedman, J. 2009 *The Elements of Statistical Learning*. NY, USA: Springer.

20. Duda, R. O., Hart, P. E. & Stork, D. G. 2001 *Pattern Classification*. USA: Wiley.

21. Recknagel, F. 2001 Applications of machine learning to ecological modelling. *Ecol. Modell.* **146**, 303– 310. (doi:10.1016/S0304-3800(01)00316-7)

22. Ozesmi, S., Tan, C. & Ozesmi, U. 2006 Methodological issues in building, training, and testing artificial neural networks in ecological applications. *Ecol. Modell.* **195**, 83–93. (doi:10.1016/j.ecolmodel.2005.11.012)

23. Kuhn, M. 2008 Building predictive models in R using the caret package. *J. Stat. Softw.* **28**, 1–26.

24. Fawcett, T. 2006 An introduction to ROC analysis. *Pattern Recognit. Lett.* **27**, 861–874. (doi:10.1016/j.patrec.2005.10.010)

25. Pearce, J. & Ferrier, S. 2000 Evaluating the predictive performance of habitat models developed using logistic regression. *Ecol. Modell.* **133**, 225–245. (doi:10.1016/S0304-3800(00)00322-7)

26. Youden, W. J. 1950 An index for rating diagnostic tests. *Cancer* **3**, 32–35. (doi:10.1002/1097-0142(1950)3:1<32::AID-CNCR2820030106>3.0.CO;2-3)

27. Perkins, N. J. & Schisterman, E. F. 2006 The inconsistency of “optimal” cutpoints obtained using two criteria based on the receiver operating characteristic curve. *Am. J. Epidemiol.* **163**, 670–5. (doi:10.1093/aje/kwj063)

28. Hand, D. J. 2012 Assessing the Performance of Classification Methods. *Int. Stat. Rev.* **80**, 400–414. (doi:10.1111/j.1751-5823.2012.00183.x)

29. Hand, D. J. 2009 Measuring classifier performance: a coherent alternative to the area under the ROC curve. *Mach. Learn.* **77**, 103–123. (doi:10.1007/s10994-009-5119-5)

30. Flach, P., Hernandez-Orallo, J. & Ferri, C. 2011 A Coherent Interpretation of AUC as a Measure of Aggregated Classification Performance. *Proc. 28th Int. Conf. Mach. Learn.*

31. Hand, D. J. & Anagnostopoulos, C. 2012 A better Beta for the H measure of classification performance.

32. Hijmans, S. E., Cameron, J. L., Parra, P. G., Jones, A. & Jarvis, R. J. 2005 Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* **25**, 1965–1978. (doi:10.1002/joc.1276)

33. Imhoff, M. L., Bounoua, L., Ricketts, T., Loucks, C., Harriss, R. & Lawrence, W. T. 2004 Global patterns in human consumption of net primary production. *Nature* **429**, 870–873. (doi:10.1038/nature02619)

34. CIESIN 2005 Gridded Population of the World (2000), Version 3 (GPWv3).

35. CIESIN 2005 Last of the Wild Data Version 2 (LWP-2): Global Human Footprint dataset (HF).

36. CIESIN 2002 Country-level Population and Downscaled Projections based on the B2 Scenario (1990).