Predicting predator-prey interactions in terrestrial endotherms using random forest

John Llewelyn[1,2], Giovanni Strona[3], Christopher R. Dickman[4], Aaron C. Greenville[5,6,7], Glenda M. Wardle[5], Michael S. Y. Lee[8,9], Seamus Doherty[1,2], Farzin Shabani[10], Frédérik Saltré[1,2], Corey J. A. Bradshaw[1,2]

[1]Global Ecology | *Partuyarta Ngadluku Wardli Kuu*, College of Science and Engineering, Flinders University, GPO Box 2100, Adelaide, South Australia 5001, Australia

[2]ARC Centre of Excellence for Australian Biodiversity and Heritage. EpicAustralia.org.au

[3]Faculty of Biological and Environmental Sciences, University of Helsinki, Viikinkaari 1, Biocentre 3, 00790, Helsinki, Finland

[4]Desert Ecology Research Group, School of Life and Environmental Sciences, University of Sydney, New South Wales 2006, Australia

[5]School of Life and Environmental Sciences, University of Sydney, Sydney, New South Wales 2006, Australia

[6]School of Life Sciences, University of Technology Sydney, Sydney, New South Wales 2007, Australia

[7]National Environmental Science Program, Threatened Species Recovery Hub, University of Sydney, Sydney, New South Wales 2006, Australia

[8]College of Science and Engineering, Flinders University, GPO Box 2100, Adelaide, SA 5001, Australia

[9]Earth Sciences Section, South Australian Museum, North Terrace, Adelaide, SA 5000, Australia

[10]Department of Biological and Environmental Sciences, College of Arts and Sciences, Qatar University, PO Box 2713, Doha, Qatar

Corresponding Author: John Llewelyn (johnsllewelyn@gmail.com)

Abstract

Species interactions play a fundamental role in ecosystem function and stability. However, few ecological communities have complete data describing such interactions, which is an obstacle to understanding how ecosystems function and respond to perturbations. Because it is often impractical to collect empirical data for all potential interactions in a community, various methods have been developed to infer interactions. Random forest—a machine learning technique—is emerging as one of the most frequently used methods for making interaction predictions, but its performance in inferring predator-prey interactions in terrestrial vertebrates and its sensitivity to variation in quality of training data remain untested. We examined predator-prey interactions within and between two diverse, primarily terrestrial vertebrate classes: birds and mammals. Combining data from a global interaction dataset and a specific community (Simpson Desert, Australia), we tested how well random forests predict predator-prey interactions for mammals and birds using species' ecomorphological and phylogenetic traits. We also tested how variation in training data quality—manipulated by removing records and switching interaction records to non-interactions—affected model performance. We found that random forests could predict predator-prey interactions for birds and mammals using ecomorphological or phylogenetic traits, correctly predicting up to 88% and 67% of interactions and non-interactions in the global and community-specific datasets, respectively. These predictions were accurate even when there were no records in the training data for the focal predator or prey species. In contrast, false non-interactions for focal predators in the training data strongly degraded model performance. Our results demonstrate that random forests can identify predator-prey interactions for birds and mammals that have few or no trophic interaction records. Furthermore, our study provides a roadmap for predicting interactions using machine learning, which might help ecologists (*i*) address knowledge gaps and explore network-

related questions in data-poor situations, and (*ii*) predict interactions for range-expanding species.

Introduction

Species interactions are essential for maintaining biodiversity (Thompson 1999) because they determine how energy and nutrients flow among organisms (Traill et al. 2010, Allan et al. 2021) and affect the distribution, survival, and abundance of species (Van der Putten et al. 2010). Species interactions also determine how ecological disturbances—such as those involving invasive species, habitat modification, and climate change—reverberate through ecological communities (Tylianakis et al. 2008, Blois et al. 2013). This is because environmental changes that directly alter one or a subset of species in a community can have flow-on effects for other species via interaction networks (Kaneryd et al. 2012, Morton et al. 2022). Given the need to predict and mitigate human-driven biodiversity declines, interest in modelling interaction networks for biological communities is increasing (Bohan et al. 2017, Strydom et al. 2021; Strona and Bradshaw 2022). However, most species interactions in most ecological communities have not been documented (Hortal et al. 2015, Jordano 2016, Caron et al. 2022). Indeed, it is usually logistically impossible to measure how all species interact in a community—even when only one type of interaction (e.g., predator-prey) is considered (Brousseau et al. 2018, Strona 2022). In addition to the challenge of identifying long-established interactions, the rapid and widespread translocation of species across the planet due to migratory responses to climate change and movement by humans means we also need to identify interactions that could occur between species that previously did not co-occur, and therefore had not previously had the opportunity to interact (Smith and Phillips 2006, Valdovinos et al. 2018).

To overcome gaps in species interaction data, various methods of inferring interactions using species traits/proxies can be applied, such as quantile-regression trophic niche-space modelling (Gravel et al. 2013, Llewelyn et al. 2021), Bayesian linear modelling (Caron et al. 2022), and several machine-learning approaches (e.g., *K* nearest-neighbour—(Desjardins-Proulx et al. 2017, Pichler et al. 2020), neural networks/deep learning methods—(Fricke et al. 2022), and random forests—(Sydenham et al. 2021)). Of these, random forest is emerging as a popular method due to its ability to learn nonlinear interactions between predictor variables, outperform other inference methods (Desjardins-Proulx et al. 2017, Laigle et al. 2018, Sydenham et al. 2021), and its availability to a wide range of users—including those without expert programming skills—through several R packages (e.g., Liaw and Wiener 2002, Wright and Ziegler 2017, Seligman 2022). Random forests have been used to predict various interaction types, including plant-pollinator (Ornai and Keasar 2020, Sydenham et al. 2021), parasite-host (Kotula et al. 2020), and predator-prey interactions (Desjardins-Proulx et al. 2017, Parravicini et al. 2020). They have also been used to identify which species' traits are most important for predicting whether species interact (Dellinger et al. 2019, Klomberg et al. 2022). Using training data consisting of interacting and non-interacting species, random forests can learn how ecomorphological traits, phylogenetic position, and/or taxonomy (as a proxy for phylogenetic position; Laigle et al. 2018) relate to which species interact. However, questions remain regarding the ability of random forests to predict different types of interactions in different taxonomic groups, which species' data/traits are best for inferring interactions, and how to fine-tune training datasets (Poisot 2022).

Identifying predator-prey interactions is central to community ecology because these interactions influence both top-down and bottom-up processes such as energy transfer and extinction cascades (Sinclair et al. 2003). Random forests have been applied to predict such interactions, but these applications have largely been limited to fish communities (Parravicini

et al. 2020, Strona et al. 2021) and soil invertebrates (Desjardins-Proulx et al. 2017, Laigle et al. 2018). Indeed, the performance of random forests for inferring predator-prey interactions among terrestrial vertebrates has not been assessed despite this group including many species of high conservation concern (Tilman et al. 2017, Cox et al. 2022) and of important ecological function (e.g., ecological engineers; Samson and Knopf 1996).

If we are to model terrestrial ecosystems realistically to predict and minimise the risk of extinction cascades, reliable inference of trophic interactions is essential. However, trophic-related traits and network structure differ among terrestrial vertebrates, fish, and invertebrates (Liem 1984, Liem 1990, Shurin and Gruner 2006, Miller-ter Kuile et al. 2022), suggesting traits and methods (such as random forest and other machine learning techniques) that effectively predict interactions in fish and invertebrates might not necessarily work well for terrestrial vertebrates. These vertebrates are diverse structurally, phylogenetically, and functionally, potentially making the 'rules' determining interactions complex and difficult for random forests and other machine-learning classifiers to identify (whereas just one trait—body size—largely determines trophic interactions in fish due to their gape limitation; Gravel et al. 2013, Morales-Castilla et al. 2015). Furthermore, interaction data for many terrestrial vertebrate species have been recorded opportunistically or not at all (although detailed diet data do exist for some species; Pringle and Hutchinson 2020), potentially leading to sparse (i.e., missing most prey for most predators) and taxonomically biased datasets for training and testing random forests. Having sparse interaction records for predators in training data is problematic because machine learning classifiers such as random forests need to be trained with data that include both interacting (e.g., predator and prey) and non-interacting species pairs, with non-interacting species pairs usually assigned by assuming that undocumented interactions imply no interaction (i.e., pseudo-non-interactions generated from presence-only data; Pichler et al. 2020). When applied to incomplete (sparse) datasets, this process can

identify undocumented but truely interacting species pairs as non-interacting pairs (but see Strona et al. 2021). Importantly, such 'false non-interactions' in training data likely reduce the reliability of interactions and non-interactions inferred from random forests and other classifiers. Similarly, random forests and other classifiers trained on datasets that are taxonomically biased might perform poorly when making predictions for species not represented in the training data (Strydom et al. 2022).

We designed analyses to test whether random forest could accurately predict predator-prey interactions within and between two diverse terrestrial vertebrate classes—birds and mammals—using both a global (*GloBI*) and a single-community (Simpson Desert, Australia) predator-prey interaction dataset. We also assessed (*i*) the importance of ecomorphological traits and phylogenetically based latent traits (i.e., as proxies for phylogenetically conserved traits associated with trophic interactions; Benandi et al. 2022) for inference accuracy, (*ii*) whether prediction accuracy improved by removing predators from the training dataset that had few interaction records (i.e., those with scarce diet records and were, therefore, prone to generating false non-interactions), and (*iii*) the sensitivity of model performance to the quality of the training data. To test sensitivity to the quality of training data, we compared two scenarios that could occur as a result of sparse training data: (a) a reduction in the number of interaction and non-interaction records used to train the models, and (b) a reduction in the number of interaction records and an increase in the number of false non-interaction records used to train the models. We also investigated how these changes affected model performance when they were restricted to different components of the training dataset (i.e., testing the effect of taxonomic bias on performance).

Our results confirm that: (*i*) random forests can predict interactions in birds and mammals, (*ii*) models trained on interaction datasets including either ecomorphological traits or phylogenetic position (or both) can perform well, (*iii*) removing predators from the training

data that have few recorded interactions therein can improve prediction accuracy, (*iv*) models can identify the prey of predators that have no interaction records in the training data, and (*v*) model performance is especially sensitive to false non-interactions in the training data that involve records of predatory species for which predictions are being made (i.e., focal predators). Our results highlight the unrealised potential for the application of random forests for predicting interactions among terrestrial vertebrates. Furthermore, they show to which components of the training data random forests are most sensitive, thereby identifying the most efficient way to increase accuracy via fine-tuning of training data.

Methods

We used two interaction datasets combined with ecomorphological and phylogenetic information to test the performance of random forests for predicting predator-prey interactions in birds and mammals (Appendix S1: Figure S1 shows flowchart of dataset preparation). We also tested how the performance of these models was influenced by the type of species data used, the filtering of training data by removing predators with few interaction records in that data, and the quality of the training data. The two interaction datasets were: (1) a dataset from the ***Glo**bal **B**iotic **I**nteractions database* (globalbioticinteractions.org; Poelen et al. 2014) supplemented with interaction data from diet studies done in Australia (Llewelyn 2022), and (2) an ecosystem-specific dataset from the Simpson Desert in Australia that focuses on seven predators for which detailed dietary information is available (Llewelyn 2022). We focused on birds and mammals rather than on all tetrapods due to the availability of detailed trait (Wilman et al. 2014) and phylogenetic (vertlife.org) information for birds and mammals (detailed trait databases for reptiles and amphibians are far from complete), thereby providing a wide range of traits for predicting interactions.

Datasets

Global data

We extracted all predator-prey interactions involving birds and mammals from *GloBI* (accessed 29/05/2020 and 26/06/2020 for birds and mammals as consumers, respectively). However, *GloBI* contains few records of species found in Oceania (including Australia) relative to other regions of the world: of the 3329 unique species combinations in the records we extracted from *GloBI*, only 109 involved a predator and prey species both found in Oceania. Supervised learners, including random forests, need to be trained on data that capture the variation present in the species for which predictions are being made, and so geographic and taxonomic biases in training data could reduce prediction accuracy. We therefore added interactions involving Simpson Desert species and their relatives to the *GloBI* data (Appendix S1: Figure S1). We added data for these species by merging two additional datasets with *GloBI*: (*i*) predator-prey records from the Simpson Desert that did not involve the seven focal predators (focal predator = species for which we made predictions), and (*ii*) interactions of the focal Simpson Desert predators with non-Simpson Desert prey elsewhere in Australia (i.e., from diet studies of the focal predators outside the Simpson Desert; no interactions between the focal predators and Simpson Desert species were included in the global/training data). In both cases, we built these datasets by aggregating predator-prey records from diet studies (Appendix S1: Figure S1). We then combined these additional interaction datasets with the *GloBI* data (adding 363 records) to create an enhanced global interaction dataset (Llewelyn 2022).

To predict species interactions using random forests, we required a training dataset that included interacting and non-interacting pairs of species. Like most interaction datasets, our global dataset did not include non-interactions i.e., since relationships are based on presence-only data. We therefore generated pseudo-non-interactions ('non-interactions' henceforth) by

taking all possible two-species combinations from the species in our global interaction dataset and removing those combinations documented as 'interacting' (Appendix S1: Figure S1). We also removed combinations where either the 'predator' does not consume vertebrates (e.g., strictly herbivorous species could not be potential predators) or the combination involved species not found on the same continent (using the `rgbif` package in R; Chamberlain et al. 2022). We removed these non-interactions because they would not help models predict which sympatric species known predators preyed upon (i.e., the type of interactions we aimed to predict in the Simpson Desert). Although the non-interactions we generated would have included many sympatric, truly non-interacting species combinations, they would have also included false non-interactions (species combination assigned as not interacting, but that do indeed interact) as well as allopatric species pairs that would interact if they were sympatric.

Random forests use explanatory variables such as species' traits to learn and predict associations (Desjardins-Proulx et al. 2017). We therefore added trait data to all the predator and prey species in the interaction and non-interaction datasets (Appendix S1: Figure S1). These traits included: body mass, time of activity (whether the species displays nocturnal, diurnal, and/or crepuscular activity), ability to fly, resource groups consumed (plants, vertebrates, and/or invertebrates), and percentage use of different habitat strata (8 habitat categories for birds and 5 for mammals; Wilman et al. 2014; Appendix S1: Table S1). We added phylogenetic information by calculating eigenvectors describing a species' position in the phylogeny (eigenvector mapping; Guénard et al. 2013), a potential proxy for explanatory variables which are phylogenetically conserved. We extracted phylogenies from VertLife.org, and used the `MPSEM` (Guénard and Legendre 2022) and `ape` (Paradis and Schliep 2019 p. 0) R packages to handle phylogenetic data and extract eigenvectors.

We also identified which species pairs in the non-interaction data included a potential prey species from within the predator's preferred prey-size range (Appendix S1: Figure S1). We did this so the ratio of non-interaction records from inside *versus* outside the preferred size range in training datasets could be adjusted to optimise performance of random forests. This is analogous to optimising species distribution models by adjusting the ratio of pseudo-absences sampled from among *versus* outside observed occurrences (Barbet-Massin et al. 2012). To identify a preferred size range, we log-transformed ($\log_e$) predator and prey body masses to linearise body-size relationships (Gravel et al. 2013). We then resampled the interaction dataset to create 100 training and testing dataset pairs using all 3692 interaction records in each of the 100 pairs (75:25 interaction records in training:testing datasets).

To each training set we fitted a series of upper and lower quantile regressions capturing the middle 0.99 to 0.8 quantiles at 0.01 quantile increments. We included a prey mass × predator taxonomic class interaction in the regression models to allow different body-size relationships for predatory birds and mammals (Llewelyn et al. 2021). We then combined each of the 100 testing sets with an equal number of species pairs randomly sampled from the non-interaction set (also with body mass $\log_e$-transformed), and applied the quantile regression models (i.e., upper and lower quantile regressions combinations) built in the previous step to the testing sets. Species combinations in the testing sets that included a potential prey species from within the upper and lower quantile ranges derived from the regressions were predicted as prey, and those outside these limits were excluded as prey.

We assessed model performance by comparing predictions to documented interactions (i.e., asking whether the quantile regression models correctly identified which combinations came from the interaction and non-interaction datasets), and used these results to calculate the true skill statistic to identify the optimal quantile regression model for assigning predator-prey interactions. The true skill statistic is a metric that can effectively evaluate the

performance of classifiers such as those predicting species interactions (Poisot 2022). A true skill statistic = 1 indicates a model that perfectly predicts interactions and non-interactions, whereas when it is 0, the model performs no better than assigning links randomly. The model capturing the middle 0.84 quantiles performed best (true skill statistic = 0.26), so we used that model to classify species pairs in the non-interaction dataset as those that included a potential prey from within the predator's preferred size range and those that included a potential prey from outside the predator's preferred size range.

Simpson Desert testing data

It is unclear how the accuracy of random forests trained and tested on data including species from many different communities (such as was the case for our global datasets) translates to model performance when applied to single ecological communities (Parravicini et al. 2020). We therefore tested model performance when applied to seven common predators from the Simpson Desert, a community we chose because it has been intensively studied for over 30 years and has detailed information on species composition and the diets of some predators. We used the species documented at established field sites in the Simpson Desert as the species assemblage (i.e., potential prey), rather than all species documented in the Simpson Desert bioregion, because the former includes sympatric species that have the potential to interact whereas the latter would likely include species that seldom have the opportunity to interact.

The Simpson Desert bird and mammal assemblage from these sites includes 64 vertebrate-consuming bird and mammal species (49 birds and 15 mammals). However, the diets of many of these species have not been studied in detail. When testing the performance of random forests for predicting predator-prey interactions, we therefore restricted the Simpson Desert testing dataset to seven predatory species with detailed dietary information from

studies both in the Simpson Desert and other arid regions in Australia. These seven focal predators included four birds—*Aquila audax*, *Falco berigora*, *Hamirostra melanosternon*, *Tyto alba*, and three mammals—*Canis dingo*, *Felis catus*, *Vulpes vulpes*. *Felis catus* (domestic cats) and *Vulpes vulpes* (red fox) are recent introductions to Australia (since European colonisation in 1788), *Canis dingo* (dingo) was brought to Australia by people > 3000 years ago (Fillios et al. 2012), and the remaining focal predators naturally occur in Australia.

To construct the dataset, we aggregated data from published and unpublished diet studies (excluding species consumed as carrion) of the focal predators in the Simpson Desert (Appendix S1: Figure S1; Llewelyn 2022). We also included records from diet studies done outside the Simpson Desert involving species pairs found in the Simpson Desert (i.e., we assumed that if trophic interactions occurred between the species in other communities, they would also occur in the Simpson Desert; Llewelyn 2022), and records involving the focal predators and Simpson Desert species in the global interaction dataset (we removed these records from the global dataset). None of the studies we used to build the Simpson Desert interaction dataset reported non-interactions. We therefore generated non-interactions by assuming any species pairs involving the focal predator and the species recorded from the Simpson Desert field sites (157 bird and mammal species) that have not been documented as predator and prey are not predator and prey. We added the same trait and phylogenetic information to the Simpson Desert data as described for the global dataset. However, we did not include a column indicating if non-interacting species pairs in the Simpson Desert data involved potential prey from within the predator's preferred size range because we only used the Simpson Desert dataset to test model performance (i.e., we did not need to optimise the sampling of these non-interactions because they were not used for training random forests).

301 **Modelling**

302 We used the `ranger` package (Wright and Ziegler 2017) in R to apply random forests using
303 the model-based recursive partitioning algorithm (Breiman 2001). In the first phase, we
304 optimised the random forest models and tested how performance was affected by (a) the
305 number and choice of variables included in the random forests, and (b) removing predators
306 that had few documented interactions in the training data (i.e., those whose diet is likely not
307 adequately captured in the data). In the second phase, we tested how changes to the quality of
308 the training dataset—including removing records and increasing the number of false non-
309 interactions—influenced model performance (Appendix S1: Figure S2 shows flowchart of
310 modelling steps, Table S2 shows what different models were used for). When optimising
311 models, we assessed model performance by comparing predictions to documented
312 interactions and used these results to calculate the true skill statistic (we also calculated other
313 performance metrics to confirm our results; see Appendix S1: Table S3). A true skill statistic
314 of 1 indicates a model that perfectly predicts interactions and non-interactions, a score > 0.75
315 is 'excellent', between 0.4 to 0.75 is 'fair to good', and < 0.4 indicates 'poor' performance
316 (Fleiss et al. 2003).

317

318 *1. Model optimisation and variables*

319 We used the global dataset and cross-validation to optimise training datasets and model
320 parameters, and we compared performance of models trained using: (*i*) only
321 ecomorphological traits, (*ii*) only phylogenetic eigenvectors, or (*iii*) both types of data
322 (Appendix S1: Figure S2). We also compared performance of models that incorporated many
323 *versus* few explanatory variables/traits to test whether more variables resulted in higher
324 accuracy or overfitting. There were 42 variables in the many-variable ecomorphological
325 model (21 each for predator and prey; Llewelyn 2022) because there were 21

326 ecomorphological variables in the databases we built (Appendix 1: Table S1). To be
327 consistent, we used 42 phylogenetic eigenvectors in the many-variable phylogenetic model
328 (first 21 phylogenetic eigenvectors for predator and prey). In the few-variable models, we
329 included 10 variables (a number we chose arbitrarily that is substantially smaller than the 42
330 variables in the many-variable models). These 10 variables were the most important
331 ecomorphological or phylogenetic variables (5 each for predator and prey) for predicting
332 interactions, as determined by permutation, which measures the impact of changing a
333 variable's value on model accuracy, followed by sequentially removing the least-important
334 variable and repeating the permutation (retained variables for each model listed in Appendix
335 1: Table S4).

336 The many- and few-variable models that incorporated both types of data simply combined
337 the variables from the respective ecomorphological and phylogenetic models (i.e., either 84
338 or 20 variables in total). We also tested whether performance was enhanced or eroded by
339 removing records from the global dataset involving predators with few (< 5) interaction
340 records. We posited that predators that had only a small proportion of their diet documented
341 in the dataset could bias results and lead to the generation of many false non-interactions
342 because, for example, models might be unduly influenced by predators that had not been
343 studied in detail – rather than matching of predator and prey traits. Removing records
344 involving these predators reduced the number of interactions from 3,692 in the original
345 global dataset to 3,424 in the modified global dataset, and reduced the number of non-
346 interactions from 454,924 to 89,184.

347 To optimise training datasets and model parameters, we divided the global dataset's
348 interactions (original and modified data) into training and testing datasets (ratio: 75:25
349 training:testing), and combined them with non-interaction data. The number of non-
350 interaction records added to the testing datasets equalled the number of interaction records in

351 that dataset, whereas we varied the number of non-interactions in the training datasets to
352 identify the optimal (highest true skill statistic) ratio. We replicated the process of generating
353 training and testing datasets each time we applied a random forest in the parameter-
354 optimization and cross-validation steps described below.

355 We weighted non-interactions such that their combined weight was equal to that of the
356 observed interactions because models built with weighted non-interactions performed slightly
357 better than those built using unweighted non-interactions (Appendix S1: Table S5; similar to
358 results from species distribution modelling; Barbet-Massin et al. 2012). These weightings
359 determine the probability an interaction/non-interaction is selected in the bootstrap samples
360 for the trees built from the training datasets. We optimised the training datasets and random
361 forests in terms of (*i*) number of non-interactions in the training data (see above), (*ii*) the ratio
362 of non-interactions from inside *versus* outside preferred size range in the training data, (*iii*)
363 the probability threshold for assigning interactions—because random forests can estimate
364 probabilities for binary responses such as interact *versus* not interact (Malley et al. 2012), (*iv*)
365 the number of variables randomly sampled as candidates at each split in the trees, (*v*) the
366 number of decision trees in the random forest, and (*vi*) the maximum depth of the trees.

367 We optimised these parameters by iterating through values for each parameter
368 individually, running 5 random forests at each value (on 5 resampled training and testing
369 datasets) and calculating the mean true skill statistic for models at each value. We identified
370 the parameter values that gave models with the highest true skill statistic. We separately
371 optimised the 12 models that differed in terms of included variables (i.e., the many- *versus*
372 few-variable ecomorphological, phylogenetic, and combined data models) and whether the
373 original or modified global dataset was used (Appendix S1: Table S6 shows optimised
374 parameter sets). We then compared performance of these optimised models by applying them

375 to 100 training and testing datasets generated using the global dataset cross-validation design
376 described above, and calculated the true skill statistic for each of these 100 cases.
377 After comparing performance of the twelve models with the global dataset, we tested their
378 performance when trained on the global dataset (with hyper-parameters set to optimise
379 performance, as identified in the previous step) and applied to the Simpson Desert data
380 (Appendix S1: Figure S2, Table S2). Again, we trained and applied 100 random forests for
381 each of the twelve models, and assessed their performance with the true skill statistic. We
382 resampled non-interactions for the training dataset in these 100 analyses because there were
383 more non-interactions than needed, whereas we used all available interaction records in the
384 global datasets (3,692 and 3,424 for the original and modified global datasets, respectively)
385 in each of the 100 training datasets.
386
387 *2. Quality of training dataset and model performance*
388 To test the influence of the size of the training dataset, taxonomic coverage, and prevalence
389 of false non-interactions (true interactions classified as non-interactions) on model
390 performance, we reran the analyses with the optimised model that performed best on the
391 Simpson Desert data (i.e., the many ecomorphological and phylogenetic variable model
392 trained on the modified global dataset; Figure 1f; Appendix S1: Table S3) but varied the
393 number of records, taxonomic coverage, and prevalence of false non-interactions in the
394 training datasets (Appendix S1: Figure S2, Table S2).
395 We deliberately reduced the quality of training data in eight ways to test specific
396 sensitivity hypotheses. We modified the entire training dataset by (*i*) randomly removing
397 records and (*ii*) randomly switching interaction records to non-interactions (false non-
398 interactions). We hypothesised that model performance would be sensitive to removal of
399 training data because having more records helps the models learn associations between

species traits and trophic links, and we hypothesised that models would be sensitive to false non-interactions because random forests would learn the wrong association when trained with erroneous data. We also tested what impact data removal and false non-interactions had when they were restricted to subsets of the training data that included either: (*iii*, *iv*) focal prey, (*v*, *vi*) focal predators, or (*vii*, *viii*) interactions where neither prey nor predators were focal/Simpson Desert species. We hypothesised that model performance would be especially sensitive to changes involving focal predators and prey because these records involve the species for which predictions are made. Here, the associations between these species' traits and trophic interactions should be especially important for the model to learn. Furthermore, we hypothesised that model performance would be sensitive, albeit less so, to removing records and increasing the proportion of false non-interactions in non-Simpson Desert (non-focal) interactions because these interactions teach the random forests trait-trophic associations that also apply to Simpson Desert species. Assessing the sensitivity to changes in the quality of training data involving the focal species is important because it assesses the suitability of random forests for inferring interactions of species that have few/no interaction records, a common phenomenon.

For each of the eight ways training dataset quality was reduced, we made these changes at 1% increments relative to the initial number of records in the modified group (e.g., 1% increments of the total number of interactions involving the focal predator), generating 10 training datasets at each increment. We trained the random forests on these reduced-quality datasets, applied them to the Simpson Desert data, and calculated true skill statistics 10 times at each increment. We plotted results to make visual comparison of the sensitivity of model performance to each of the eight types of artificial variation in the quality of training data, and we compared model performance when all relevant data were modified to quantify the impacts of each type of modification.

425 We also examined the mechanisms underlying changes in model performance by testing
426 how removing records or increasing false non-interactions for the focal predator influenced
427 the: (*i*) relative probability (or suitability) of interacting assigned by the random forest to each
428 potential prey for each predator (i.e., asking whether changes to the training dataset affected
429 which species were identified as the most likely prey), and (*ii*) mean probability assigned to
430 potential prey for each predator (i.e., asking whether changes to the training dataset cause an
431 overall increase or decrease in the probabilities assigned to potential prey). If lower model
432 performance reflects a systematic shift in all the probabilities (upward or downward), rather
433 than a change in which potential prey are identified as the most likely, the models could still
434 be used to predict a predator's prey because the rank order of probability of different prey
435 remains similar irrespective of absolute model performance and absolute probability.
436 To test for changes in relative prey suitability, we measured the correlation (Pearson's *r*)
437 between prey probabilities assigned by models trained on reduced-quality training datasets
438 with that assigned by models trained on the full training dataset (Appendix S1: Figure S2).
439 To test for changes in mean probability assigned to potential prey overall, we subtracted the
440 mean probability assigned to potential prey for each predator as calculated by the full
441 training-data models from that calculated by the models trained on the reduced-quality
442 training datasets (Appendix S1: Figure S2). For both analyses, we reduced the quality of the
443 training dataset at 1% increments (10 replicate training datasets/models at each increment)
444 and plotted the results to compare visually. We hypothesised that removing records and
445 including more false non-interactions in the training dataset would disrupt the assignment of
446 relative suitability and reduce assigned probabilities overall.
447

448 # Results

449 **Variables and model performance**

Random forests trained and tested on the global data had mean true skill statistics ranging from 0.53 to 0.77 (Figure 1a,c,e,g), indicating 'good' to 'excellent' performance (Fleiss et al. 2003). These models tended to perform better when we used the original global interaction dataset rather than the modified dataset that had predators with few prey records removed (mean difference in true skill statistic between models = 0.14; Figure 1a and Figure 1c *versus* 1e and 1g). Using few or many variables in the phylogenetic or phylogenetic/ecomorphological variable models had little effect on performance when applied to the global datasets (difference in mean true skill statistic < 0.02 in all cases; Figure 1), but there was a slight decrease in performance associated with using few variables in the ecomorphological-only models (difference in mean true skill statistic = 0.08 and 0.06 between the many- and few-variable ecomorphological models when the original or modified global dataset was used, respectively; Figure 1a *versus* 1c, Figure 1e *versus* 1g).

The models trained on the global dataset and applied to the Simpson Desert data yielded mean true skill statistics between 0.25 and 0.45, indicating 'poor' to 'fair' performance (Fleiss et al. 2003; Figure 1). Those that used both phylogenetic and ecomorphological variables performed well when trained on the modified global dataset, yielding mean true skill statistics of 0.43 ± 0.02 (mean ± standard deviation) and 45 ± 0.02 when we used few or many variables, respectively (Figure 1f,h). Using few *versus* many variables did not strongly affect the performance of the phylogenetic-only or phylogenetic- and ecomorphological-variable models (difference in mean true skill statistic < 0.07 in all four cases; Figure 1). In contrast, the few-variable ecomorphological models did not perform as well as their many-variable counterparts (difference in mean true skill statistic > 0.13 between the many- and few-variable ecomorphological models when using either the original or modified global datasets Figure 1b *versus* 1d, Figure 1f *versus* 1h).

474 Using the model that performed best on the Simpson Desert data (i.e., the many-variable
475 model with phylogenetic and ecomorphological traits trained on the modified global dataset;
476 Figure 1f), performance predicting interaction for recently introduced predators—cats and
477 foxes—was better than that for native predators (true skill statistic = 0.56 ± 0.03 and 0.41 ±
478 0.02 for recently introduced and native predators, respectively; Figure 2).
479
480 **Effects of training dataset size, false non-interactions, and taxonomic coverage**
481 We altered the global (modified) training dataset by randomly removing records and
482 changing interaction records to false non-interactions, applying these changes both to the
483 entire training dataset as well as to three subsets: (*i*) records involving the focal prey (i.e.,
484 species found in the Simpson Desert), (*ii*) records involving the focal predators (i.e., the 7
485 Simpson Desert predators for which we predicted interactions), and (*iii*) records involving
486 species not found in the Simpson Desert. Removing records reduced performance when
487 modifying either the entire training dataset (Figure 3a; true skill statistic reduced to 0 when
488 all records removed because random forests cannot run without training data) or the focal-
489 predator component only (Figure 3e; true skill statistic reduced to 0.24 ± 0.01 [mean ±
490 standard deviation] when all focal-predator records were removed). However, these effects
491 became pronounced only when > 80% of the interaction and non-interaction records were
492 removed (Figure 3a,e), which equates to removing > 15,750 or 1,240 records from the entire
493 or focal-predator components of the training datasets, respectively. Removing records
494 involving the focal prey or non-Simpson Desert species had only weak or no effect on model
495 performance (Figure 3c,g; true skill static = 0.36 ± 0.02 and 0.46 ± 0.01 when removing all
496 focal prey or all non-Simpson Desert species, respectively).
497 The effect of false non-interactions on model performance was stronger than that of
498 removing records. False non-interactions assigned at random to interaction records in either

the entire training dataset or the component of the training dataset involving focal predators degraded model performance more rapidly (Figure 3b,f; true skill static reduced to 0 and 0.04 ± 0.01 for models when changing all records or all records involving focal predators to non-interactions, respectively). These declines were triggered when changing > 40% of the interaction records in the relevant group to false non-interactions (Figure 3b,f), which equates to 1,370 records in the entire training-dataset analysis and 235 in the focal-predator analysis. False non-interactions in the training data involving the focal prey (Figure 3d) or non-Simpson Desert species (Figure 3h) caused slight declines in model performance (true skill static reduced to 0.36 ± 0.03 and 0.34 ± 0.02 when changing all interactions records to false non-interactions for focal prey species and non-Simpson Desert species, respectively).

Focal-predator training data and the mechanisms underlying changes in model performance

The correlation between probabilities assigned to potential prey by models trained on the full *versus* record-removed global datasets showed a gradual decrease as more interactions involving the focal predator were removed from the training data, and this decline became more rapid when removing > 80% of records (Figure 4a). The correlation between these probabilities reduced to 0.59 ± 0.02 when all records involving focal predators were removed from the training dataset, indicating that models identified many of the same species as prey even when there were no records involving the focal predator in the training data. The mean probability assigned to potential prey for each predator was reduced by removing records of the focal predator from the training data and, like the correlation results, this change was small until 80% of records were removed (Figure 4c). After removing all records, the probabilities assigned to potential prey were on average 0.17 ± 0.01 lower compared to those assigned by models trained on the full dataset.

Both the correlation and change in mean probability assigned to potential prey were more sensitive to false non-interactions involving the focal predators than to removing their records from the training dataset, showing more rapid declines (Figure 4a *versus* 4b, Figure 4c *versus* 4d). When changing all interactions involving the focal predators in the training data to non-interactions, the correlation between probabilities assigned to prey decreased to 0.27 ± 0.03 (Figure 4b) and the mean probabilities assigned to potential prey dropped by 0.31 ± 0.01 (Figure 4d).

Discussion

Our results demonstrate that random forests can predict the prey of predatory birds and mammals even when few or no interaction records involving the focal predator are included in training data (Figure 4a). Applying random forests to predict predator-prey interactions for these predators (and terrestrial vertebrates more generally) to facilitate network modelling of terrestrial vertebrate systems is therefore supported (Brousseau et al. 2018). However, our analyses highlight that caution is needed in applying the random forest and other machine-learning approaches depending on the quality, quantity, and type of training data available, and in the face of these limitations we suggest that filtering training data could improve model performance (Figure 5).

We found that random forests could predict bird and mammal interactions in both our global and ecosystem-specific datasets (including for native and recently introduced predators). Models based on ecomorphological, phylogenetic, or both types of traits, and those using few or many traits, could perform similarly well (Figure 1)—consistent with previous studies showing that either data type can be used for predicting interactions in other taxa (Faisal et al. 2010, Morales-Castilla et al. 2015, Brousseau et al. 2018, Pomeranz et al. 2019, Elmasri et al. 2020) and that few traits/variables are required to make accurate

549 predictions (Desjardins-Proulx et al. 2017). Because phylogenetic information is available for
550 most terrestrial vertebrates (e.g., VertLife.org), ecological/morphological trait databases are
551 increasingly common and comprehensive (Jones et al. 2009, Oliveira et al. 2017, e.g.,
552 Herberstein et al. 2022), and interaction data for many species are available from open-source
553 databases (e.g., globalbioticinteractions.org; web-of-life.es) and publications, random forests
554 and other machine learning methods can now easily be applied to such data to predict
555 predator-prey interactions for most vertebrate communities, and aid in the ecological network
556 and invasive species modelling required to guide conservation decisions.

557 We found that model performance was robust to the random removal of records from the
558 training dataset. In fact, we had to remove > 80% of records (>15,750 individual records)
559 before an obvious change in performance occurred (Figure 3a). Thus, models that use large
560 training datasets are unlikely to be sensitive to random additions or removal of interaction
561 and non-interaction records.

562 The removal of records involving focal predators only caused a prominent decline in
563 model performance when we removed > 80% (Figure 3e) of records in that group, but this
564 translates to fewer (1,240) records in total needing to be removed (as compared to the random
565 removal of records). Conversely, the complete removal of records involving focal-prey
566 species or non-focal species had little impact on model performance (Figure 3c,g). Thus,
567 declines in model performance in the random-removal analyses were largely driven by the
568 loss of focal predator records. This reduction in model performance following removal of
569 focal-predator records was due to every prey species being assigned consistently lower
570 probabilities and a slight shuffling of prey preference (Figure 4a,c). Because *relative* prey
571 preferences were largely maintained even when all training records involving the focal
572 predator were removed from training data (mean Pearson $r = 0.59$; Figure 4a), our results
573 suggest prey species could be predicted for predators that do not have any interaction data,

which is likely the case for many understudied, rare, or extinct species. The ability to predict trophic interactions for extinct species accurately has important applications for understanding past extinction events (Ripple and Van Valkenburgh 2010, Pires et al. 2015, Llewelyn et al. 2021), while the capacity to predict interactions for rare species could help identify cryptic factors that threaten their persistence (Smith and Phillips 2006, Doherty et al. 2016). However, additional work to assess the requirements of training data for predicting interactions for species without interaction records is still required. For example, it is unclear how phylogenetically and/or ecologically similar species in the training data need to be to the focal species for models to be able to predict trophic interactions accurately.

In contrast to removal, false non-interactions in the training data reduced model performance more rapidly, triggering lower performance when > 40% of the interaction records were changed to non-interactions. This sensitivity highlights an important issue: most studies generate non-interactions by assuming that undocumented interactions imply true non-interaction; however, this assumption could generate many false non-interactions, especially if interaction records are sparse (i.e., if many of the predators in the training data have only a small proportion of their interactions documented). To avoid generating false non-interactions, one can restrict non-interactions in training datasets to species combinations where the predator is too small to eat the potential prey and/or the potential predator comes from a low trophic level (Strona et al. 2021). However, applying this approach could reduce the capacity of random forests to identify unsuitable prey from within a predator's preferred prey-size range or for high trophic-level species. Indeed, the optimal ratio of non-interactions sampled from within *versus* outside preferred size range in the model that performed best on our Simpson Desert data (i.e., the many-variable, both data-types model) was 5:2 (Appendix S1: Table S6), suggesting that non-interactions of suitably-sized potential prey should be included in training data.

599 Alternatively, filtering training data by removing predators that have few interaction
600 records could help avoid false non-interactions because many of the interactions/prey of these
601 predators are likely missing (i.e., they have sparse interaction records, although these species
602 would also include some predators with narrow diet breadths). When applied to the Simpson
603 Desert data, our models trained on the global dataset modified by removing predators with <
604 5 interaction records performed better than those trained on the original/full global dataset
605 (Figure 3). This result might reflect a reduction of false non-interactions in the modified
606 training dataset. Thus, removing predators/consumers that are missing most of their
607 interactions from training datasets could be an effective way to filter training data and
608 improve model performance.
609 Model performance was especially sensitive to false non-interactions in the training data
610 involving focal predators. This type of change in data quality disrupted relative prey
611 preferences and lowered probabilities assigned to interactions more extensively than did
612 reducing the number of records involving focal predators (Figure 4a *versus* 4b, Figure 4c
613 *versus* 4d), reinforcing the importance of avoiding false non-interactions—especially for
614 focal predators—when generating pseudo-non-interactions for training data; otherwise, such
615 data can change which species are identified as most-likely prey. Although removing
616 predators/consumers with few interaction records from training data and/or restricting the
617 generation of non-interactions based on relative size and trophic level are two simple
618 methods for avoiding false non-interactions as we discussed, there are other strategies that
619 could be applied to the entire- or focal-predator-component of training datasets to minimise
620 the number of false non-interactions. For example, false non-interactions could be minimised
621 by generating non-interactions on a case-by-case basis (similar to the 'target-group absences'
622 approach used in species distribution modelling; Barbet-Massin et al. 2012), an approach
623 shown to improve the accuracy of predator-prey interactions predicted by deep-learning

models (Fricke et al. 2022). We generated non-interactions for the global dataset using interaction data from different regions, studies, and sources, but this approach could generate false non-interactions, such as when pairs of species are allopatric in studies that contributed to the training data and so are designated as non-interacting, but the component species interact when/where they are sympatric. This approach could also generate misleading non-interactions, such as pairs of currently allopatric species that might interact should future conditions force them to become sympatric. Instead, training data non-interactions could be generated based on individual studies or locations of recorded interactions describing what a predator does and does not eat in a particular area. Such a nuanced approach would identify more reliable and informative non-interactions than would be achieved by assuming that any undocumented pairs from an incomplete global database are truly not predator and prey.

Our models performed better with the global dataset than with the Simpson Desert data. This difference might indicate that the random forests learnt patterns in the global dataset that are not as effective when re-applied to specific communities. For example, the random forests might have learnt from the global dataset that predators with particular traits do not eat prey with particular traits, but predators or prey with those traits might not be present in the Simpson Desert. In other words, learning the relationship did not help the random forest predict interactions in the Simpson Desert. Alternatively, random forests might learn which predators had few or many prey records within the global dataset—another pattern that would be less effective at identifying interactions and non-interactions when applied to the Simpson Desert predators due to the similarity among focal predators in terms of number of interactions in the training data (60% of predators in the original/full global dataset had < 5 interaction records whereas all Simpson Desert focal predators had > 10 interaction records in the global dataset). This might explain why our models did not perform as well on the global data when we used the modified dataset rather than the original/full global dataset for training

649 and testing, because we had removed predators with few interaction records from the
650 modified dataset. Geographic biases in the global training dataset could also lead to weaker
651 performance on our ecosystem-specific dataset (Tuia et al. 2022). Irrespective of what caused
652 the difference in model performance, our results indicate that performance based on global
653 datasets might not be a strong indicator of performance when applied to single communities.
654 Previous studies applying random forests in aquatic and invertebrate communities have
655 tended to report higher performance than our models achieved (mean true skill statistic
656 consistently > 0.6; Desjardins-Proulx et al. 2017, Laigle et al. 2018, Brousseau et al. 2018,
657 Parravicini et al. 2020, Strona et al. 2021). There are several possible explanations for this
658 discrepancy, including previous studies (*i*) pooling prey species into broad resource
659 categories (Parravicini et al. 2020), (*ii*) using higher-quality training data (Laigle et al. 2018,
660 Parravicini et al. 2020), (*iii*) including resource predictions for a wider range of trophic
661 groups, including easily-excluded non-interactions such as herbivores and/or detritivores
662 eating other animals (Desjardins-Proulx et al. 2017), and (*iv*) focusing on taxa whose
663 interactions might be easier to predict than for terrestrial vertebrates (although further testing
664 is required to evaluate this possibility). Another potential means of improving the
665 performance of random forests that neither we nor others have yet addressed is to include
666 detail on intraspecific variation in traits and interactions, including ontogenetic shifts
667 (González-Varo and Traveset 2016).
668 Random forest and other machine learning approaches are useful and flexible tools that
669 have great potential in terms of predicting species interactions. While global interaction
670 databases are a valuable resource for training and testing such models, they currently include
671 only a small fraction of real-world species interactions. Furthermore, for the consumers that
672 are included in such databases, in most cases only a small subset of their trophic interactions
673 is documented, thereby increasing the likelihood of generating false-non-interactions.

Databases therefore need to be expanded, and a framework for generating non-interactions that avoids or minimises false non-interactions and retains informative non-interactions is required (Figure 5 shows a proposed framework). By making such changes, the application of random forests and other machine-learning methods (including ensemble approaches) can be further honed to infer species interactions more accurately.

References

Allan, J. D. et al. 2021. Energy Flow and Nutrient Cycling in Aquatic Communities. - In: Allan, J. D. et al. (eds), Stream Ecology : Structure and Function of Running Waters. Springer International Publishing, pp. 357–381.

Barbet-Massin, M. et al. 2012. Selecting pseudo-absences for species distribution models: how, where and how many? - Methods Ecol. Evol. 3: 327–338.

Benandi, G. et al. 2012. Quantitative predictions of interactions in bipartite networks based on traits, abundance, and phylogeny? – Am. Nat. 3: 327–338.

Blois, J. L. et al. 2013. Climate Change and the Past, Present, and Future of Biotic Interactions. - Science 341: 499–504.

Bohan, D. A. et al. 2017. Next-Generation Global Biomonitoring: Large-scale, Automated Reconstruction of Ecological Networks. - Trends Ecol. Evol. 32: 477–487.

Breiman, L. 2001. Random forests. – Mach. Learn 45:5-32.

Brousseau, P.-M. et al. 2018. Trait matching and phylogeny as predictors of predator-prey interactions involving ground beetles (M Ferrari, Ed.). - Funct. Ecol. 32: 192–202.

Caron, D. et al. 2022. Addressing the Eltonian shortfall with trait-based interaction models. - Ecol. Lett. 25: 889–899.

Chamberlain, S. et al. 2022. rgbif: interface to the global biodiversity information facility.

Cox, N. et al. 2022. A global reptile assessment highlights shared conservation needs of tetrapods. - Nature: 1–6.

Dellinger, A. S. et al. 2019. Beyond buzz-pollination – departures from an adaptive plateau lead to new pollination syndromes. - New Phytol. 221: 1136–1149.

Desjardins-Proulx, P. et al. 2017. Ecological interactions and the Netflix problem. - PeerJ 5: e3644.

Doherty, T. S. et al. 2016. Invasive predators and global biodiversity loss. - Proceedings of the National Academy of Sciences 113: 11261–11265.

Elmasri, M. et al. 2020. A hierarchical Bayesian model for predicting ecological interactions using scaled evolutionary relationships. - Ann. Appl. Stat. 14: 221–240.

Faisal, A. et al. 2010. Inferring species interaction networks from species abundance data: A comparative evaluation of various statistical and machine learning methods. - Ecol. Inform. 5: 451–464.

Fillios, M. et al. 2012. The impact of the dingo on the thylacine in Holocene Australia. - World Archaeology 44: 118–134.

Fleiss, J.L. et al. 2003. Statistical methods for rates and proportions. – John Wiley & Sons, Ltd.

Fricke, E. C. et al. 2022. Collapse of terrestrial mammal food webs since the Late Pleistocene. - Science 377: 1008–1011.

González-Varo, J. P. and Traveset, A. 2016. The labile limits of forbidden interactions. - Trends Ecol. Evol. 31: 700–710.

Gravel, D. et al. 2013. Inferring food web structure from predator-prey body size relationships (R Freckleton, Ed.). - Methods Ecol. Evol. 4: 1083–1090.

Guénard, G. and Legendre, P. 2022. MPSEM: Modeling phylogenetic signals using eigenvector maps.

Guénard, G. et al. 2013. Phylogenetic eigenvector maps: a framework to model and predict species traits. - Methods in Ecology and Evolution 4: 1120–1131.

Herberstein, M. E. et al. 2022. AnimalTraits - a curated animal trait database for body mass, metabolic rate and brain size. - Sci Data 9: 265.

Hortal, J. et al. 2015. Seven Shortfalls that Beset Large-Scale Knowledge of Biodiversity. - Annu. Rev. Ecol. Evol. Syst. 46: 523–549.

Iturbide, M. et al. 2015. A framework for species distribution modelling with improved pseudo-absence generation. - Ecological Modelling 312: 166–174.

Jones, K. E. et al. 2009. PanTHERIA: a species-level database of life history, ecology, and geography of extant and recently extinct mammals: *Ecological Archives* E090-184 (WK Michener, Ed.). - Ecology 90: 2648–2648.

Jordano, P. 2016. Chasing ecological interactions. - PLoS Biol. 14:e1002559.

Kaneryd, L. et al. 2012. Species-rich ecosystems are vulnerable to cascading extinctions in an increasingly variable world: Extinction Cascades in a Variable World. - Ecol. Evol. 2: 858–874.

Klomberg, Y. et al. 2022. Spatiotemporal variation in the role of floral traits in shaping tropical plant-pollinator interactions. - Ecol. Lett. 25: 839–850.

Kotula, H. J. et al. 2020. Predicting direct and indirect non-target impacts of biocontrol agents using machine-learning approaches.

Laigle, I. et al. 2018. Species traits as drivers of food web structure. - Oikos 127: 316–326.

Liaw, A. and Wiener, M. 2002. Classification and Regression by randomForest. - R News 2: 18-22.

Liem, K.F. 1984. Functional versatility, speciation, and niche overlap: are fishes different? – Trophic interactions within aquatic ecosystems: 269-305. Routledge.

Liem, K.F. 1990. Aquatic versus terrestrial feeding modes: possible impacts on the trophic ecology of vertebrates. – Am. Zool. 30: 209-221.

Llewelyn, J. 2022. Random forests for predicting predator-prey interactions in terrestrial vertebrates. - GitHub and Zenodo: DOI: 10.5281/zenodo.7037189.

Llewelyn, J. et al. 2022. Sahul's megafauna were vulnerable to plant-community changes due to their position in the trophic network. - Ecography 2022: DOI: 10.1111/ecog.06089.

Lobo, J. M. et al. 2010. The uncertain nature of absences and their importance in species distribution modelling. - Ecography 33: 103–114.

Malley, J.D. et al. 2012. Probability machines. – Methods Inf. Med. 51: 74-81.

Mateo, R. G. et al. 2010. Profile or group discriminative techniques? Generating reliable species distribution models using pseudo-absences and target-group absences from natural history collections: Reliability of models using pseudo-absences and target-group absences. - Divers. Distrib. 16: 84–94.

Miller-ter Kuile et al. 2022. Predator-prey interactions of terrestrial invertebrates are determined by predator body size and species identity. – Ecology: e3634

Morales-Castilla, I. et al. 2015. Inferring biotic interactions from proxies. - Trends Ecol. Evol. 30: 347–356.

Morton, D. N. et al. 2022. Merging theory and experiments to predict and understand coextinctions. - Trends Ecol. Evol. in press.

Oliveira, B. F. et al. 2017. AmphiBIO, a global database for amphibian ecological traits. - Sci. Data 4: 170123.

Ornai, A. and Keasar, T. 2020. Floral complexity traits as predictors of plant-bee interactions in a Mediterranean pollination web. - Plants 9: 1432.

Paradis, E. and Schliep, K. 2019. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. - Bioinformatics 35: 526–528.

Parravicini, V. et al. 2020. Delineating reef fish trophic guilds with global gut content data synthesis and phylogeny (P Jordano, Ed.). - PLoS Biol. 18: e3000702.

Pichler, M. et al. 2020. Machine learning algorithms to infer trait-matching and predict species interactions in ecological networks. - Methods Ecol. Evol. 11: 281–293.

Pires, M. M. et al. 2015. Pleistocene megafaunal interaction networks became more vulnerable after human arrival. - Proc. R. Soc. B 282: 20151367.

Poisot, T. 2022. Guidelines for the prediction of species interactions through binary classification. - EcoEvoRxiv: 16.

Poelen, J.H. et al (2014) Global biotic interactions: an open infrastructure to share and analyze species-interaction datasets. – Ecol. Inform. 24: 14-8-159.

Pomeranz, J. P. F. et al. 2019. Inferring predator–prey interactions in food webs. - Methods Ecol. Evol. 10: 356–367.

Pringle, R. M. and Hutchinson, M. C. 2020. Resolving Food-Web Structure. - Annu. Rev. Ecol. Evol. Syst. 51: annurev-ecolsys-110218-024908.

Ripple, W. J. and Van Valkenburgh, B. 2010. Linking top-down forces to the Pleistocene megafaunal extinctions. - BioScience 60: 516–526.

Samson, F. B. and Knopf, F. L. 1996. Ecosystem Management: Selected Readings. - Springer New York.

Seligman 2022. Rborist: Extensible, Parallelizable Implementation of the Random Forest Algorithm. - https://CRAN.R-project.org/package=Rborist

Shurin, J. B. and Gruner, D. S. 2006. All wet or dried up? Real differences between aquatic and terrestrial food webs. - Proc. R. Soc. B 273: 1-9.

Sinclair, A. R. E. et al. 2003. Patterns of predation in a diverse predator–prey system. - Nature 425: 288–290.

Smith, J. G. and Phillips, B. L. 2006. Toxic tucker: the potential impact of Cane Toads on Australian reptiles. - Pac. Conserv. Biol. 12: 40–49.

Strona, G. 2022. Hidden Pathways to Extinction. - Springer International Publishing.

Strona, G. and Bradshaw, C. J. A. 2022. Coextinctions dominate future vertebrate losses from climate and land use change. – Sci Adv 8: eabn4345.

Strona, G. and Bradshaw, C. J. A. 2018. Co-extinctions annihilate planetary life during extreme environmental change. - Sci Rep 8: 16724.

Strona, G. et al. 2021. Ecological dependencies make remote reef fish communities most vulnerable to coral loss. - Nat. Commun. 12: 7282.

Strydom, T. et al. 2021. A roadmap towards predicting species interaction networks (across space and time). - Proc. R. Soc. B 376: 20210063.

Strydom, T. et al. 2022. Food web reconstruction through phylogenetic transfer of low-rank network representation. – Methods Ecol. Evol. 2022: DOI: 10.1111/2041-210X.13835.

Sydenham, M. A. K. et al. 2021. MetaComNet: A random forest-based framework for making spatial predictions of plant-pollinator interactions. - Methods Ecol. Evol. in press.

Thompson, J. N. 1999. The evolution of species interactions. - Science 284: 2116–2118.

Tilman, D. et al. 2017. Future threats to biodiversity and pathways to their prevention. - Nature 546: 73–81.

Traill, L. W. et al. 2010. Mechanisms driving change: altered species interactions and ecosystem function through global warming. - J. Anim. Ecol. 79: 937–947.

Tuia, D. et al. 2022. Perspectives in machine learning for wildlife conservation. – Nat. Commun. 13: 792.

Tylianakis, J. M. et al. 2008. Global change and species interactions in terrestrial ecosystems. - Ecol. Lett. 11: 1351–1363.

Valdovinos, F. S. et al. 2018. Species traits and network structure predict the success and impacts of pollinator invasions. - Nat. Commun. 9: 2153.

Van der Putten, W. H. et al. 2010. Predicting species distribution and abundance responses to climate change: why it is essential to include biotic interactions across trophic levels. - Phil. Trans. R. Soc. B 365: 2025–2034.

Wilman, H. et al. 2014. EltonTraits 1.0: Species-level foraging attributes of the world's birds and mammals: *Ecological Archives* E095-178. - Ecology 95: 2027–2027.

Wright, M. N. and Ziegler, A. 2017. ranger: A fast implementation of Random Forests for high dimensional data in C++ and R. - J. Stat. Soft. in press.

Figure captions

Figure 1. Density plots comparing the performance (true skill statistic) of random forest models predicting predator-prey interactions in birds and mammals. The random forests used either ecomorphological traits, phylogenetic eigenvectors, or both types of data, and were applied to two interaction datasets: a global interaction dataset (panels in the left column) and a specific ecosystem dataset (7 predators and their potential prey/sympatric species in the Simpson Desert; panels in the right column). We also assessed the effects of using many variables (42, 42, and 84 variables, for ecomorphological, phylogenetic eigenvector, and both data-type models) or few variables (10, 10, and 20 variables, for ecomorphological, phylogenetic eigenvector, and both data-type models), and restricting the training dataset to predators with ≥ 5 prey records (modified training dataset). Panels (a) and (b) show results from models that use the complete training dataset and many variables; (c) and (d) show results from models that use the complete training dataset and few variables; (e) and (f) show results from models that use the modified training dataset and many variables; and (g) and (h) show results from models that use the modified training dataset and few variables.

Figure 2. The five most- and least-likely mammalian and bird prey for the red fox (*Vulpes vulpes*) in the Simpson Desert predicted by random forest. Solid blue arrows indicate prey that were predicted and observed, dashed blue arrows show prey that were predicted but not observed (potentially indicating interactions that occur but have not yet been recorded), and red lines indicate species that were not predicted or observed as prey. The five most-likely prey according to the many-variable phylogenetic and ecomorphological model (run 100 times and trained on the modified training data) were the common wallaroo (*Osphranter robustus*), long-haired rat (*Rattus villosissimus*), red kangaroo (*Osphranter rufus*), Eurasian coot (*Fulica atra*), and sandy inland mouse (*Pseudomys hermannsburgensis*). The five least-likely prey were the little eagle (*Hieraaetus morphnoides*), brown goshawk (*Accipiter fasciatus*), whistling kite (*Haliastur sphenurus*), wedge-tailed eagle (*Aquila audax*), and spotted harrier (*Circus assimilis*).

Figure 3. Effect of training-dataset quality on the performance of random forest for predicting predator-prey interactions in birds and mammals. Each panel shows performance when modifying a percentage of either the whole or specific taxonomic groups in the training dataset. Panels (a) and (b) show performance when changing the whole dataset, panels (c) and (d) show performance when restricting changes to focal prey, panels (e) and (f) show performance when restricting changes to focal predators, and panels (g) and (h) show performance when restricting changes to non-focal (non-Simpson Desert) species. Panels in the left column show the effects of removing records from the training data, whereas panels in the right column show the effects of switching interaction records to non-interactions (false non-interactions). We trained random forests on the modified global predator-prey interaction dataset (i.e., restricted to predators with ≥ 5 prey records) and applied them to seven predators and their potential prey from the Simpson Desert. Plot points indicate performance (according to the true skill statistic) when applied to the Simpson Desert predators.

Figure 4. Effects of the quality of the training dataset on (*i*) the predicted relative suitability of prey and (*ii*) the mean probabilities assigned to potential prey by random forests. Panels in the left column show the effects of reducing the number of interactions records involving the focal predator in the training dataset, whereas panels in the right column show the effect of switching interaction records involving the focal predator to non-interactions (false non-interactions). Panels (a) and (b) show the correlation between prey probabilities assigned by

the models trained on the changed datasets with that assigned by models trained on the complete (modified) training dataset. Panels (c) and (d) show the mean probability (suitability) assigned to potential prey by models trained on the changed datasets minus those values assigned by the models trained on the complete (modified) training datasets. We trained these random forests on the modified global predator-prey interaction dataset (restricted to predators with ≥ 5 prey records) and applied them to seven predators and their potential prey from the Simpson Desert.

Figure 5. Framework for preparing and optimising random forest models for predicting species interactions. Most of these steps and methods are also relevant to the application of other machine-learning approaches for inferring species interactions.

Fig. 1

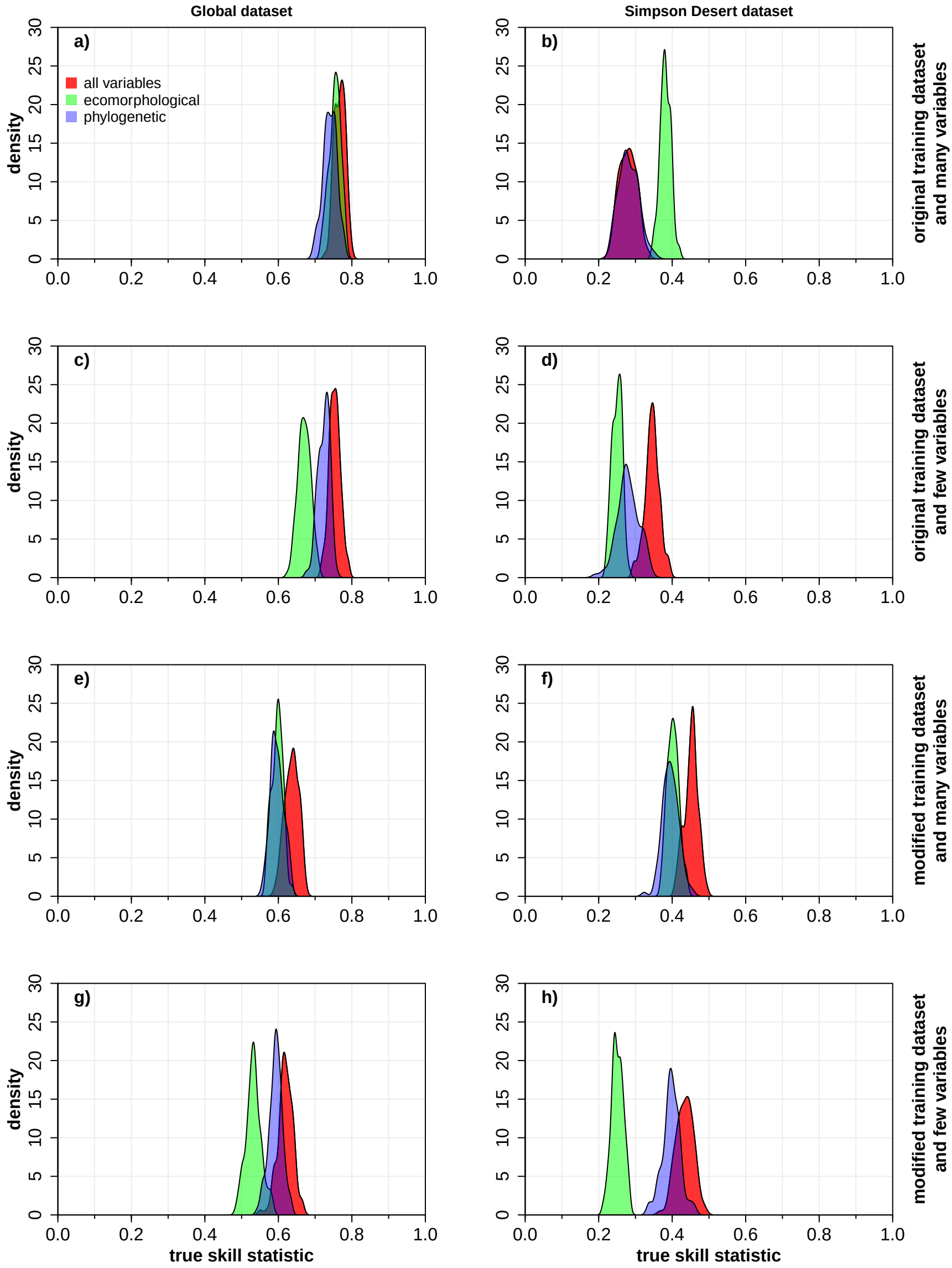
Global dataset
Simpson Desert dataset
a)
b)
c)
d)
e)
f)
g)
h)
all variables
ecomorphological
phylogenetic
density
true skill statistic
original training dataset and many variables
original training dataset and few variables
modified training dataset and many variables
modified training dataset and few variables
0.0
0.2
0.4
0.6
0.8
1.0
0
5
10
15
20
25
30

Fig. 2

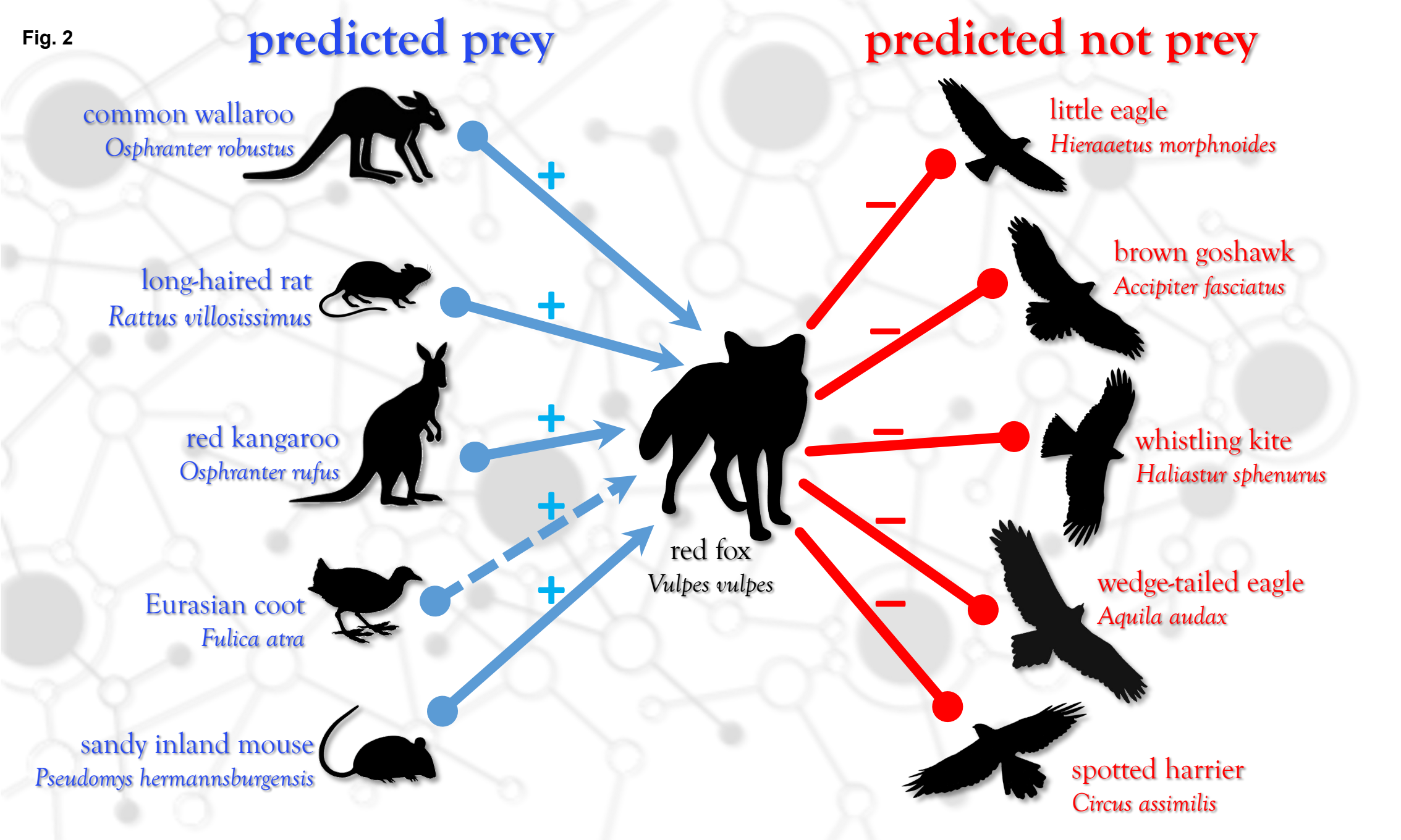
predicted prey
predicted not prey
common wallaroo
Osphranter robustus
long-haired rat
Rattus villosissimus
red kangaroo
Osphranter rufus
Eurasian coot
Fulica atra
sandy inland mouse
Pseudomys hermannsburgensis
+
+
+
+
+
red fox
Vulpes vulpes
−
−
−
−
−
little eagle
Hieraaetus morphnoides
brown goshawk
Accipiter fasciatus
whistling kite
Haliastur sphenurus
wedge-tailed eagle
Aquila audax
spotted harrier
Circus assimilis

Fig. 3

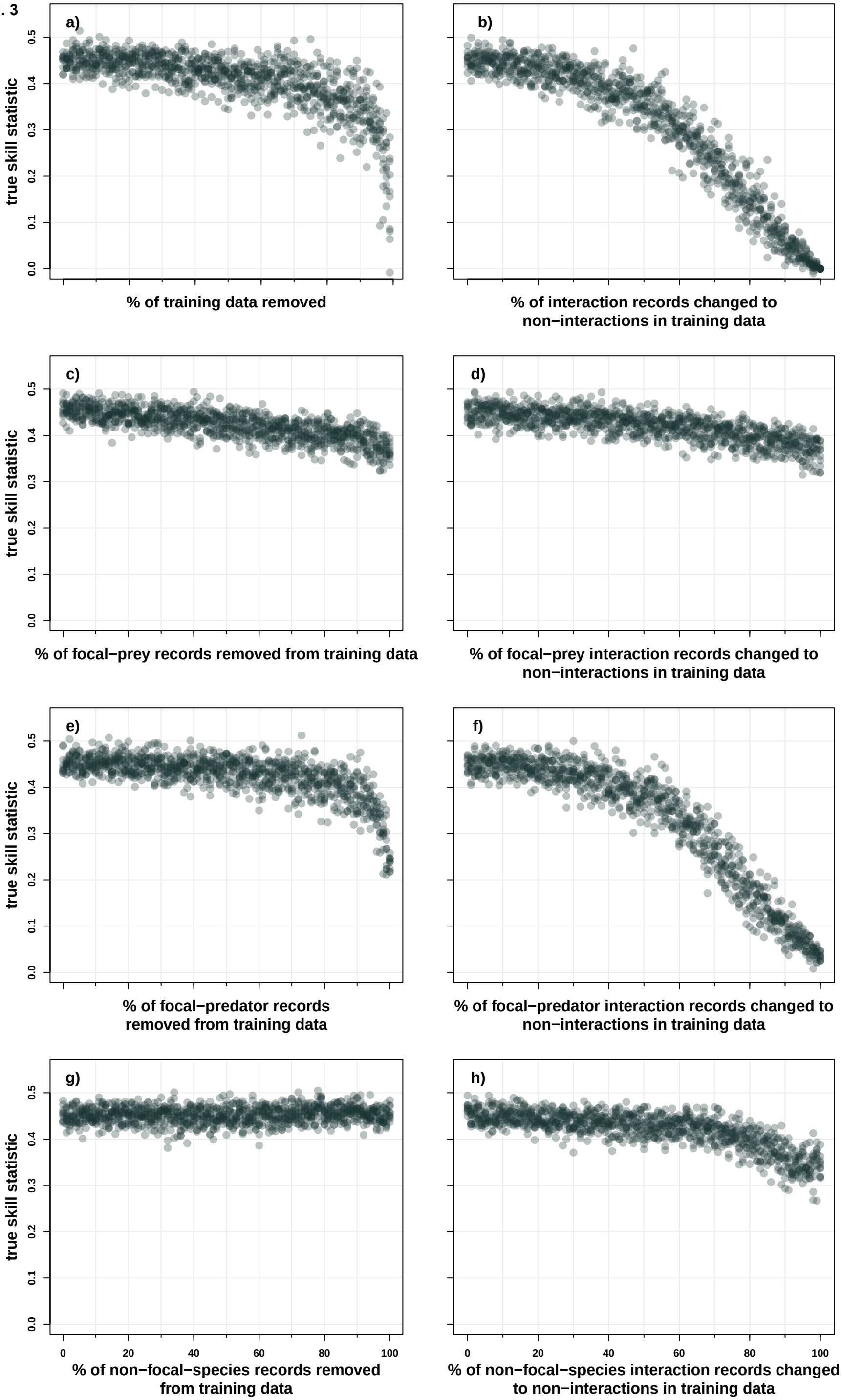
a)
b)
c)
d)
e)
f)
g)
h)
true skill statistic
0.0
0.1
0.2
0.3
0.4
0.5
% of training data removed
% of interaction records changed to non−interactions in training data
% of focal−prey records removed from training data
% of focal−prey interaction records changed to non−interactions in training data
% of focal−predator records removed from training data
% of focal−predator interaction records changed to non−interactions in training data
% of non−focal−species records removed from training data
% of non−focal−species interaction records changed to non−interactions in training data
0
20
40
60
80
100

Fig. 4

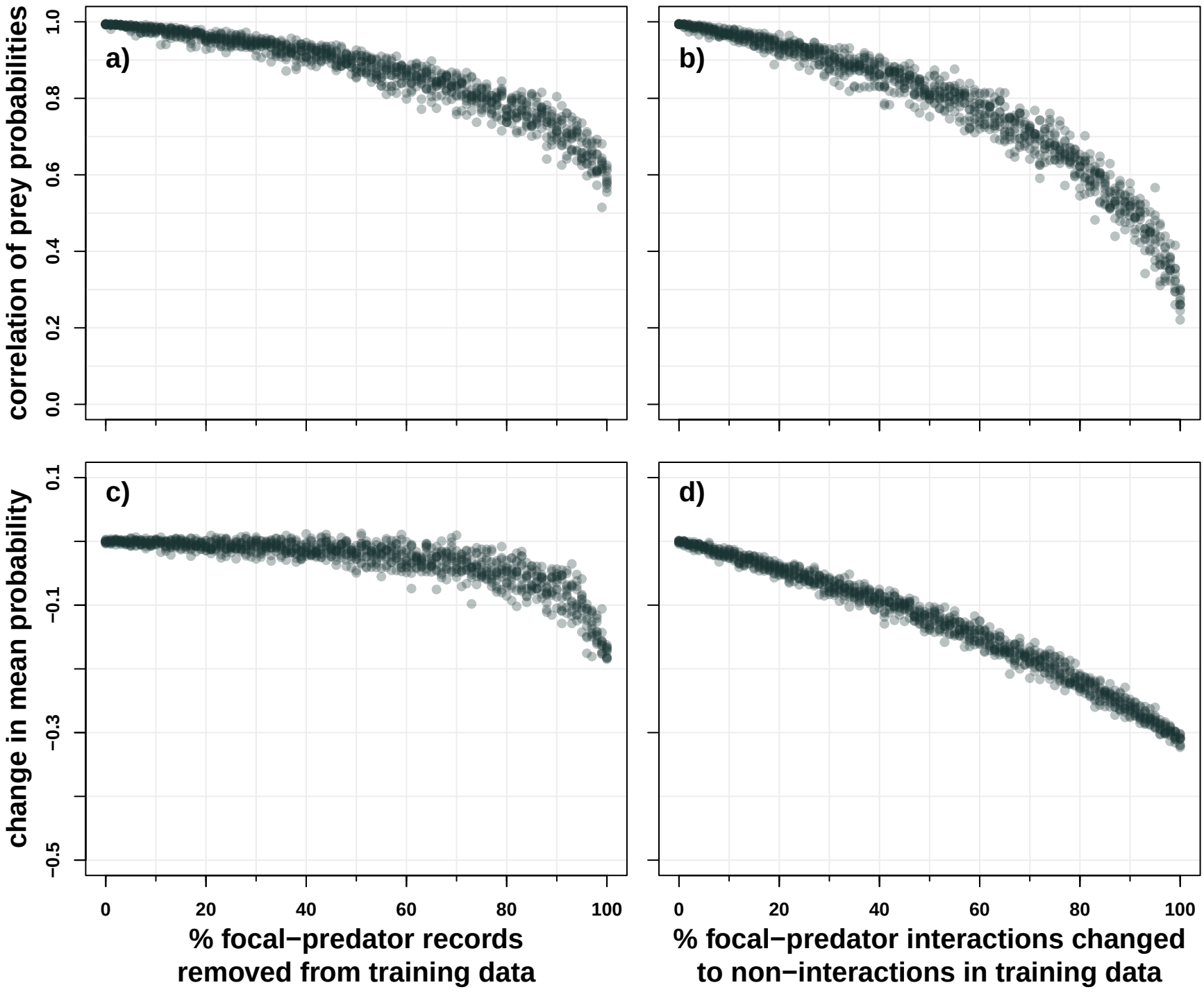
a)
b)
c)
d)
correlation of prey probabilities
change in mean probability
1.0
0.8
0.6
0.4
0.2
0.0
0.1
−0.1
−0.3
−0.5
0
20
40
60
80
100
% focal−predator records removed from training data
% focal−predator interactions changed to non−interactions in training data

Fig. 5

random forest for predicting community interactions

steps	methods
1 obtain interaction & non-interaction data for training	**extract documented interaction and non-interaction records** from online databases, publications, unpublished data, and/or targeted collection of new data via experiments or field observations (see below for non-interactions) **prioritise collecting interaction data for focal predators & ecologically similar species**
	generate pseudo-non-interactions Few studies report non-interactions; pseudo-non-interactions usually assigned instead by assuming undocumented interactions imply no interaction. Strategies to assign reliable (truly non-interacting) & informative pseudo-non-interactions required. For example: 1. generate non-interactions from individual studies rather than from pooled data; 2. include only co-occurring species if predictions needed for co-occurring species; 3. remove predators with few interaction records; 4. restrict non-interactions based on trait-matching rules (e.g., body size, trophic level) **important to avoid incorrect non-interactions involving focal predators**
2 add trait data to training dataset	**collate ecomorphological/phylogenetic data** sources: online databases, publications, unpublished data, targeted collection of new trait data
3 optimise training dataset	**optimise interactions:non-interactions ratio (dataset balance)** out-of-bag design; vary interactions:non-interactions ratio; compare performance of random forests trained & tested on these datasets
	optimise ratio of suitable to unsuitable predator/prey species combinations in <u>non-interaction</u> data based on trait matching & out-of-bag approach (see above)
4 optimise model hyper-parameters	**identify & tune important hyper-parameters** e.g., # variables at each tree split, # trees, probability threshold for assigning interactions, max tree depth, max/min leaf node size
5 build dataset for making predictions	**collate data for focal species/ecological community (focal species = those for which predictions needed)** add same trait and/or phylogenetic variables as in training data & arrange into species pairs for making predictions
6	**train model on optimised training dataset & apply to focal species pairs to make predictions**