Received: 22 April 2022 | Accepted: 21 December 2022

DOI: 10.1111/2041-210X.14066

Methods in Ecology and Evolution
Open Access
BRITISH ECOLOGICAL SOCIETY

RESEARCH ARTICLE

ENphylo: A new method to model the distribution of extremely rare species

Alessandro Mondanaro[1] | **Mirko Di Febbraro**[2] | **Silvia Castiglione**[3] | **Marina Melchionna**[3] | **Carmela Serio**[4] | **Giorgia Girardi**[3] | **Arianna Morena Belfiore**[3] | **Pasquale Raia**[3]

[1]Department of Earth Sciences, University of Florence, Florence, Italy

[2]Department of Biosciences and Territory, University of Molise, Pesche, Italy

[3]Department of Earth Sciences, Environment and Resources, University of Naples Federico II, Naples, Italy

[4]Research Centre in Evolutionary Anthropology and Palaeoecology, School of Biological and Environmental Sciences, Liverpool John Moores University, Liverpool, UK

Correspondence
Pasquale Raia
Email: pasquale.raia@unina.it

Handling Editor: Veronica Zamora-Gutierrez

Abstract

1. Species distribution models (SDMs) are a useful mean to understand how environmental variation influences species geographical distribution. SDMs are implemented by several different algorithms. Unfortunately, these algorithms consistently lose accuracy exactly when they are needed the most, that is with rare species, originating the so-called rare-species modelling paradox. Although approaches exist to tackle this problem, most notably by performing and then averaging a number of bivariate models, they are usually computationally intensive and were never shown to apply successfully to the rarest species (i.e. with less than 20 geographical occurrences).
2. Here, we present a new algorithm, *ENphylo*, embedded in the readily-available R package *RRdtn*, which couples Environmental Niche Factor Analysis (ENFA) and phylogenetic imputation to model the distribution of rare species.
3. Using the fossil record of 31 species of large mammals that lived during the late Pleistocene as the source data to sample from, we demonstrate *ENphylo* provides good SDM evaluation scores, with area under the curve and Sørensen Index both consistently above 0.75, True Skills Statistics above 0.4 and Boyce Index above 0.5 in most cases, when just 10 fossil occurrences are randomly drawn from their respective fossil records. *ENphylo* proved significantly more accurate than ENFA and the ensemble of bivariate models using Maxent, Generalized Linear Model and Random Forest algorithms. Intriguingly, we found that randomly drawing as little as 10 occurrence data points per species allows *ENphylo* to perform equally well as Maxent run using the entire fossil record of these same species and data.
4. *ENphylo* provides a fast and accurate solution to perform species distribution modelling with rare species, which will help predicting their distribution in the light of climate change, and to delineate how rare extinct species reacted to past climatic variation.

Alessandro Mondanaro and Mirko Di Febbraro equally contributed.

KEYWORDS
climate change, *ENphylo*, overfitting, rare species, *RRdtn*, species distribution models

1 | INTRODUCTION

Species distribution models (SDMs) algorithms are powerful tools for predicting species distributions across the landscape under the hypothesis that environmental conditions influence species' geography (Elith & Leathwick, 2009). In the last decades, SDMs have been widely applied to assess the effect of climate change on species distributions and the impact of invasive species, to select suitable sites for species reintroductions, or to address conservation objectives (Barbet-Massin et al., 2018; Fois et al., 2018). Although a number of different algorithms have been successfully used to calibrate SDMs, their applicability is limited by several factors, including species intrinsic characteristics (e.g. range size, dispersal ability) and methodological issues (Fourcade et al., 2018; Tessarolo et al., 2021). Most, if not all, SDM algorithms fail to apply to scanty occurrence records, as with rare species or at modelling extinct species with sparse fossil records (Raia et al., 2020; Svenning et al., 2011; Varela et al., 2011). Several studies have demonstrated that low sample size impacts negatively on SDM accuracy (Jiménez-Valverde, 2020; Pearman et al., 2008; Santini et al., 2021). It has repeatedly been suggested that more than 20, and ideally more than 50 geographical occurrences are necessary to provide robust distribution models predictions (Santini et al., 2021; Wisz et al., 2008). Thus, although rare species are the most demanding in terms of reliable SDM estimates (since they are the most exposed to extinction risk, Blomqvist et al., 2010; Eaton et al., 2018), they are also those where the SDM algorithms perform worst (Breiner et al., 2015, 2018; Lomba et al., 2010; Sousa-Silva et al., 2014). The same applies to most extinct species, which is disappointing because the inclusion of fossil information to living species distributions allows approaching the species fundamental niche (Maiorano et al., 2013; Raia et al., 2020; Timmermann et al., 2022) and because fossil species provide the most genuine evidence of past response to climate change (Di Febbraro et al., 2017; Mondanaro et al., 2021; Tóth et al., 2019). The main problem with modelling rare species is that their few geographical occurrences are usually coupled to numerous environmental explanatory variables, causing a strong imbalance between the poor information on the species actual environmental preferences and the rich environmental information. The imbalance likely causes model overfitting, which, in turn, reduces SDM transferability (Vaughan & Ormerod, 2005). A possible solution proposed in literature is to fit several bivariate models (i.e. including two environmental variables at a time) and then averaging their predictions within a weighted ensemble model (Lomba et al., 2010). Although viable and effective (Breiner et al., 2015, 2018), this approach is computationally intensive and time-consuming, and does not address the problem of the weak starting information about the species preferences. In principle, species environmental predilections and tolerance limits are determined by the traits they inherit so that the climatic niche can be studied as it were a phenotype (Pearman et al., 2008; Rolland et al., 2018). This implies phylogenetic position might supplement the scarce environmental data related to the species occurrences that typically comes with modelling rare species. Starting from this assumption, we propose *ENphylo*, a new modelling algorithm able to provide fast and accurate distribution predictions for rare species by combining Ecological Niche Factor Analysis (ENFA, Hirzel et al., 2002) and phylogenetic imputation (Garland & Ives, 2000). To test *ENphylo* performance, we applied the algorithm to model extremely under-sampled species (i.e. with sample size as low as 10 or 20 occurrences, respectively) and compared its predictive accuracy to ENFA and to the ensembles of small models (ESM, Breiner et al., 2015, 2018) approaches. In applying ESM, we included three widely used modelling techniques: Maxent, Random Forest (RF) and Generalized Linear Models (GLM). We demonstrated *ENphylo* consistently outperforms both ESM and ENFA at 10 occurrences. At 20 occurrences ESM performs best, yet it fails for some 15% of the species for which *ENphylo* still provides decent performance. Crucially, *ENphylo* results at sampling a mere 10 occurrences from the fossil record of individual species, performs as well as Maxent using the full fossil record on the same set of data and climatic variables, as implemented in a previous study.

2 | MATERIALS AND METHODS

ENphylo's workflow includes two consecutive steps. The first, embodied in the R function ENphylo_modeling, formats the input data (i.e. species occurrence/background points and the phylogenetic tree), calibrates ENFA and phylogenetic imputation, and then evaluates their predictive accuracy. The second R function, ENphylo_prediction, relies on the output of ENphylo_modeling to predict species marginality, specialization and habitat suitability on a new dataset provided by the user (e.g. to generate spatially explicit predictions). ENphylo_modeling and ENphylo_prediction are embedded in a single R package, named *RRdtn*, that is made available as part of the current study. The modelling steps involved in both functions are described in the following paragraphs.

2.1 | ENphylo_modeling: Calculating species marginality and specialization via ENFA and phylogenetic imputation

ENphylo_modeling takes three objects as inputs: (i) presence/background points for all the species under analysis provided as a named list (argument input_data), (ii) the phylogenetic tree including the species present in input_data (argument tree), and (iii) a geographical mask defining the spatial domain encompassing the background area enclosing all the species (argument input_mask). For each species in input_data, the relevant information should be provided as a data frame or a spatial object (as handled by 'sp' or 'sf' R packages) and must include a column with occurrence data in binary format (1 for presences, 0 for background), two columns with occurrence/background coordinates (in case the data are not in the form of spatial objects), and one column for each environmental variable to be used in the modelling process. Environmental predictors must be the same for all the species on the phylogenetic tree. Optionally, the user can provide a column specifying the age of each occurrence (which is useful when modelling fossil species). The phylogenetic tree must be provided in Newick or Nexus format. The geographical mask must be a RasterLayer object. ENphylo_modeling contains three internal functions: (i) DATA_PREPARATION, (ii) ENFA_CALIBRATION and (iii) IMPUTED_CALIBRATION. DATA_PREPARATION matches the geographical mask with species occurrence/background data provided as input, rearranging them in a proper way as to be used in the subsequent modelling steps. ENFA_CALIBRATION relies on the R package 'CENFA' (Rinnan, 2021) to compute marginality and specialization factors for all the species in input_data having a number of occurrence records above a minimum threshold (e.g. >50) as set by the user (argument min_occ_enfa). Marginality and specialization are at the core of ENFA modelling. Marginality can be thought of as the distance between the centroid of the niche occupied by the species and the centroid calculated for all the available habitat (Rinnan & Lawler, 2019). Specialization is the ratio of the variance in the available habitat distribution to that in the species distribution (Rinnan & Lawler, 2019). Occupied and available habitats are defined by a given set of environmental variables, which are expected to show some degree of multicollinearity. Under ENFA, multicollinearity is accounted for by means of factor analysis, performed to extract the linear combinations of environmental variables maximizing the focal species' marginality and specialization eigenvectors. The n x m matrix of marginality/specialization coefficients having the n environmental variables as rows and m marginality/specialization eigenvectors as columns, named the CO matrix, represents the amount of marginality and specialization of the focal species on each variable (Hirzel et al., 2002). The first eigenvector represents marginality, the following, orthogonal to the first, eigenvectors represent specialization. The number of columns m of the CO matrix is reduced dropping specialization eigenvectors accounting for little variance, according to the broken-stick criterion (Jackson, 1993).

ENFA_CALIBRATION calculates predictive accuracy by splitting the data into 80%–20% training/testing samples (the split percentage is indicated in the boot_test_perc argument) for the calibration–evaluation of the ENFA models. Specifically, a new CO matrix (CO80) is obtained calibrating ENFA on the 80% (training) data and used to predict marginality and specialization factors on the entire dataset through the row per column product of the $n \times m$ (CO80) matrix multiplied by the n columns of environmental variable values in the entire $g \times n$ dataset, where the number of rows g is equal to the number of geographic cells in the data. The predicted marginality and specialization values are then converted into Mahalanobis distances calculated from the species position to the barycenter of the ENFA axes in the multivariate marginality/specialization space (Fonderflick et al., 2015; Hengl et al., 2009; Préau et al., 2018). This step allows to (i) accommodate for the residual collinearity among specialization axes (Calenge et al., 2008; Hirzel et al., 2002) and (ii) convert predicted marginality/specialization values into habitat suitability values in the 0–1 range. This latter operation is possible because squared Mahalanobis distances approximate a Chi-square distribution (Clark et al., 1993). The suitability values predicted on the remaining 20% (testing) occurrence data are eventually used to assess model accuracy. The entire procedure is repeated several times (as regulated by the boot_rep argument), relying by default on multi-core parallel processing. Predictive accuracy is assessed through a set of discrimination-, reliability-, and similarity-based evaluation metrics (Leroy et al., 2018), namely the area under the operating characteristic curve (AUC; Fielding & Bell, 1997), the true skill statistic (TSS; Allouche et al., 2006), the continuous Boyce index (CBI; Hirzel et al., 2006) and the Sørensen similarity index (SSI; Leroy et al., 2018; Li & Guo, 2013). Evaluation metrics are calculated using PresenceAbsence (Freeman & Moisen, 2008) and ecospat (Broennimann, 2022) R packages. Omission error rate is further supplied to evaluate the incidence of false positives, that is recommended at using AUC/TSS as discrimination metrics (Jiménez-Valverde, 2012).

The taxa present in INPUT_DATA which have either a number of occurrences lower than the min_occ_enfa, or for which ENFA reports predictive accuracy below the reference level (as set by the argument eval_metric_threshold), form a subset I of all the species in the phylogenetic tree for which CO matrices are either unavailable or little reliable, respectively. For them, the last internal function, IMPUTED_CALIBRATION, will automatically estimate the CO matrices by means of phylogenetic imputation. Imputation relies on the Brownian motion model of evolution to provide sensible estimates of CO for each species in I. Imputation is implemented using the function phylopars in the Rphylopars R package (Goolsby et al., 2017). To account for phylogenetic uncertainty, alternative trees are created by altering topology and branch lengths of the original tree using the function swapONE in the RRphylo R package (Castiglione et al., 2018). The user can set the number of alternative phylogenies to create (which is 100 by default), the proportion of tree tips whose topologic arrangement will be swapped, and the proportion of tree nodes whose age will be changed. To keep the same number of marginality/specialization axes—hence the same number of columns in CO—for all the species in I, the dimensionality of the imputed CO matrices is forced to be equal to the median number of axes retained by all the species not present in I. Phylogenetically imputed CO matrices obtained from the tree swapping iterations are used to predict marginality and specialization values and to evaluate predictive accuracy through bootstrap cross-validation, as it is done with ENFA-modelled species. No accuracy is computed for species represented by less than 10 occurrences. Once predictive accuracies are calculated over all tree swapping replicates, the function returns different outputs according to three alternative strategies specified by the user according to the "output_options" argument: (i) CO matrices and evaluation metrics for all the tree swapping iterations ("output_options" = "full"); (ii) a subset of CO matrices and evaluation metrics limited to tree swapping iterations achieving a predictive accuracy above a user-defined threshold ("output_options" = "weighted.mean"), according to a specific metric ("eval_metric_for_imputation") and the relative score ("eval_threshold"); (iii) a single CO matrix and the corresponding evaluation scores relative to the most accurate tree swapping iteration ("output_options" = "best"). Eventually, for the species forming the subset *I* the CO matrix are always imputed for those with less than min_occ_enfa occurrences, whereas for those > min_occ_enfa the validation metrics are compared for the two methods (ENFA versus phylogenetic imputation) and the one performing best is retained.

2.2 | ENphylo_prediction: Generating predictions of species marginality, specialization and habitat suitability

ENphylo_prediction takes as input the output of ENphylo_modeling and the "newdata" object including the same environmental variables used to calibrate models. The newdata object can be a different time frame, or a different geographical area or both, as compared to the calibration data. For each species present in the ENphylo_modeling output object, ENphylo_prediction automatically recognizes whether ENFA or phylogenetic imputation was performed for modelling. In the latter case, the function retrieves marginality and specialization predictions from the tree swapping iterations according to ENphylo_modeling "output_options" specification. In particular, by specifying "full" or "best" as outputs, either all or the best-performing tree are used to generate predictions. By specifying "weighted.mean" a subset of trees is used, limited to trees generating model performance above a certain value (AUC > 0.7, by default) and the weighted average performance for the selected evaluation metric is calculated.

ENFA and phylogenetic imputation predictions can be performed on either a data frame or RasterLayer objects (provided as the newdata argument). Optionally, the user can convert predictions into habitat suitability values using the argument convert_to_suitability.

2.3 | Comparing *ENphylo* predictive accuracy toESM approach

We compared the predictive performance achieved by *ENphylo* to other state-of-the-art SDM algorithms, that is ENFA and ESM built using Maxent (Phillips et al., 2006), RF (Breiman, 2001) and GLM (McCullagh & Nelder, 1983), posing the explicit hypothesis that *ENphylo* would outperform competing methods under low-sampling conditions and achieve good absolute predictive performance. To this aim, we gathered fossil occurrences for 21 extinct and 10 extant large mammals living in Eurasia during the last 200 ka from Mondanaro et al. (2021). Occurrence data included 4651 mammal records distributed over 916 fossil layers. For each species, the gathered data include occurrences along with 10,000 background points (for further details, see Mondanaro et al., 2021). Each occurrence/background datapoint is spatially and temporally associated (depending on the age of the fossil layer) to a vector of climate values retrieved from the paleoclimate emulator described in Holden et al. (2019), corresponding to six non-collinear predictors (Mondanaro et al., 2021): BIO4 (temperature seasonality), BIO8 (mean temperature of the wettest quarter), BIO 10 (Mean Temperature of Warmest Quarter), BIO13 (precipitation of wettest month), BIO14 (precipitation of driest month) and BIO18 (precipitation of warmest quarter). Along with the occurrence data, we constructed a phylogenetic tree of mammal species, using the R function tree.merger in RRphylo (Castiglione et al., 2022). The function combines different phylogenies into a single synthetic time-calibrated tree. Here, the source phylogenies were published in Carotenuto et al. (2016) and Castiglione et al. (2021). The data, species and climatic variables are the same as in Mondanaro et al. (2021), which is welcome since it allows comparing SDM accuracy in this study directly to the SDM results we got in that study, where the entire fossil record was used.

As a preliminary step, we applied *ENphylo* to all the 31 species, relying on their entire fossil record. Since all the species included more than 50 occurrence records, they were initially modelled with ENFA. We evaluated model predictive performances of these preliminary ENFA models splitting randomly species occurrence records into an 80% training and 20% testing data, and calculating AUC, TSS, CBI and Sørensen Index. We repeated this splitting procedure 20 times, then averaging the evaluation scores. From the ENFA models, we retrieved the marginality/specialization CO matrix for each species.

To test the ability of *ENphylo* to predict the observed occurrence records of a given species starting from an extremely sub-sampled fossil occurrence pool, we randomly selected 10 occurrences datapoints per species to calibrate the models, and kept the remaining datapoints to evaluate predictive accuracy (i.e. external evaluation). This entire procedure was iterated 20 times, changing the 10 randomly selected occurrences at each iteration. The same set of 20, 10-occurrence-wide subsets were used, in turn, to calibrate ENFA, *ENphylo*, and ESM (see below), which underwent a second evaluation round under an 80%–20% bootstrap cross-validation scheme (i.e. internal evaluation) performed on each fossil occurrences set. To model a sub-sampled species with *ENphylo*, CO matrices calculated in the preliminary phase for the remaining 30 species were used as phenotypic traits on the tree to perform imputation. To account for phylogenetic uncertainty, we created 20 alternative phylogenies, altering 50% of the topology and 50% of the branch lengths of the original tree according to swapONE specifications. For each species, *ENphylo* outputs were generated according to all the three strategies implemented, that is "full", "best" and "weighted.mean", where the last two were, respectively, applied selecting the single most accurate tree (in terms of AUC; 'best.tree', hereafter) and calculating a weighted average of the trees with an AUC > 0.7 ('selected.tree', hereafter), relying on AUC values from the internal evaluation for selection.

To optimally tune ESM, we followed the procedure described in Breiner et al. (2018). Specifically, for each bivariate model (i.e. 15 bivariate predictor combinations of the six climatic variables), we varied parameters and complexity as generated by alternative settings and then chose the configuration that yielded the highest AUC from an 80%–20% bootstrap procedure. We implemented the ensemble according to GLM and Maxent algorithms, which were shown to provide a good compromise between computational time and accuracy (Breiner et al., 2018) and RF which was shown to provide good performance when there is limited environmental overlap between presence and background data (Valavi et al., 2021) which is more easily met by our historical data. For GLMs, we tested the shape of the relationship (i.e. linear, quadratic or cubic) and the interaction level (absent or present). For Maxent, we used the ENMEVAL R package to test regularization values between 0.5 and 4, with 0.5 steps and the following alternative combinations of feature classes in turn: linear, linear + quadratic, hinge, linear + quadratic + hinge, linear + quadratic + hinge + product and linear + quadratic + hinge + product + threshold (Muscarella et al., 2014). Among the 48 resulting combinations, we chose the model reporting the lowest Akaike information criterion corrected for a small sample size (AICc; Warren & Seifert, 2011). As for RF, we adopted the default options as reported in biomod2 package (as in Breiner et al., 2018). As with ENFA and *ENphylo*, we evaluated each bivariate model by performing a bootstrap cross-validation scheme, where calibration data are randomly split into 80%–20% training/testing samples and repeated this step 10 times (i.e. internal evaluation procedure). Among these models, we dropped those poorly calibrated (i.e. AUC < 0.7) from the subsequent analyses. Ensemble models were obtained by averaging individual GLM, RF and Maxent models projections weighted by their respective AUC scores calculated under the internal validation (Marmion et al., 2009). Both internal and external evaluation procedures were carried out calculating AUC, TSS, CBI and SSI values. The entire subsampling and testing procedure was iterated over the 31 species in the tree and repeated on the 20-occurrence randomly generated datasets.

Significant differences in predictive performance among the three modelling approaches were assessed by fitting a random-slope linear mixed effect model (LMM), where the evaluation metrics were used, in turn, as the response variable, the three modelling approaches as the explanatory variable, and the species as random effect.

2.4 | Testing *ENphylo* sensitivity to the number of species requiring phylogenetic imputation

We set a second experiment to assess whether *ENphylo* is sensitive to the proportion of species to be imputed. To this aim, we randomly imputed 3 out 31 (9.7%), using the CO matrices obtained by ENFA on the full dataset for the 28 remaining. For these three species, we imputed the new CO matrices after modelling 50 alternative swapped trees. The accuracy of *ENphylo* predictions for these three species was evaluated through the 80%–20% cross-validation scheme, repeating the data splitting 10 times, and calculating the average AUC value. Lastly, we selected the tree achieving the highest AUC values among the 50 swapped phylogenies. The entire procedure was repeated selecting, in turn 6 (19.4%) and then 9 (29%) species to be modelled through phylogenetic imputation. To test for significant differences in *ENphylo* predictive performance among the three imputation intensity scenarios, we fitted a random-slope LMM, where AUC values were used as the response variable, the three imputation scenarios as the explanatory variable and the species as the random effect.

3 | RESULTS

Overall, we generated >446.000 models (i.e. 31 species × 20 subsampling replicates × 10 bootstrap replicates × two levels of sampling intensity, all multiplied to 20 swapped trees for *ENphylo* and 15 bivariate models for ESM). Using the 10-occurrence datasets, all the three *ENphylo* strategies reported >50% species to achieve an acceptable level of evaluation accuracy (i.e. with AUC > 0.75; Elith, 2000), with the “full” strategy showing the highest absolute percentage of well-performing models (58%). The corresponding percentages for the other algorithms tested are 29% for ENFA and 3% for ESM. *ENphylo* “full” proved the best algorithm according to three out of four evaluation metrics, reporting a mean AUC value averaged among the species equal to 0.75 (0.58–0.88), mean TSS = 0.39 (0.23–0.57), and mean CBI of 0.56 (0.31–0.74), whereas ENFA achieved the highest SSI values (mean = 0.68; 0.56–0.78).

ESM resulted systematically the least performing algorithm according to all evaluation metrics (Table 1). Under the 10-occurrence subsampling scheme, LMMs indicated AUC values associated to all the three *ENphylo* strategies are significantly higher than those related to any other algorithm. This finding remains true using TSS, though in this case the differences with ENFA are statistically significant only for the “full” strategy (Figure 1). As for CBI, LMMs indicated all the three *ENphylo* strategies to significantly outperform ESM, while the differences with ENFA are not statistically significant. ENFA was the best algorithm according to SSI values (Figure 1).

When using 20-occurrence datasets, more than 70% species achieved AUC > 0.75 under ESM and *ENphylo* “best.” Under *ENphylo* “best” and ENFA more than 80% species reached a TSS > 0.4. ESM emerged as the best algorithm according to AUC (mean = 0.79; 0.50–0.92) together with *ENphylo* “best” (mean = 0.79; 0.60–0.90), and CBI (mean = 0.77; 0.00–0.99), while *ENphylo* “best” was the most accurate algorithm according to TSS (mean = 0.50; 0.30–0.68) and SSI (mean = 0.75; 0.63–0.84; Table 1, Table S1). LMMs indicated ESM to outperform other algorithms in terms of AUC and CBI values, though this difference is not significant against *ENphylo* “full” and “best” in terms of AUC, and against *ENphylo* “best” in terms of CBI (Figure 1). As for TSS and SSI, the only significant difference pertains to *ENphylo* “full” and “best” outperforming ENFA, and the opposite at comparing *ENphylo* “weighted.mean” to ENFA under SSI (Figure 1). Although using 20-occurrence species records ESM performs well overall, some 15% of the species report low evaluation scores with his algorithm, a problem that does not apply to ENFA and *ENphylo* (Figure 2, Figures S1 and S2).

TABLE 1 Evaluation metric values (mean and range) for each algorithm and for two different levels of sampling intensity (either extracting 10 or 20 occurrences from the full fossil record of each species), averaged among the 31 species. The number of training occurrences represents the number of occurrences randomly extracted from the fossil record of each species to calibrate models

Evaluation metric	# training occurrences	*ENphylo* best.tree	*ENphylo* full.tree	*ENphylo* selected. tree	ENFA	ESM
AUC	10	0.74 (0.57–0.86)	0.75 (0.58–0.88)	0.73 (0.53–0.87)	0.72 (0.57–0.81)	0.62 (0.50–0.83)
	20	0.79 (0.60–0.90)	0.78 (0.61–0.90)	0.76 (0.52–0.90)	0.76 (0.59–0.88)	0.79 (0.50–0.92)
TSS	10	0.37 (0.21–0.52)	0.39 (0.23–0.57)	0.37 (0.14–0.57)	0.37 (0.13–0.53)	0.20 (0.01–0.54)
	20	0.50 (0.30–0.68)	0.49 (0.29–0.68)	0.44 (0.06–0.68)	0.46 (0.19–0.69)	0.49 (0.00–0.71)
CBI	10	0.54 (0.28–0.71)	0.56 (0.31–0.74)	0.53 (0.27–0.76)	0.56 (−0.04–0.79)	0.36 (−0.05–0.79)
	20	0.70 (0.31–0.89)	0.70 (0.30–0.91)	0.64 (0.12–0.91)	0.69 (0.22–0.89)	0.77 (0.00–0.99)
SSI	10	0.62 (0.50–0.71)	0.65 (0.55–0.76)	0.62 (0.36–0.76)	0.69 (0.56–0.78)	0.33 (0.03–0.70)
	20	0.75 (0.63–0.84)	0.75 (0.65–0.84)	0.67 (0.13–0.84)	0.74 (0.63–0.85)	0.68 (−0.01–0.86)

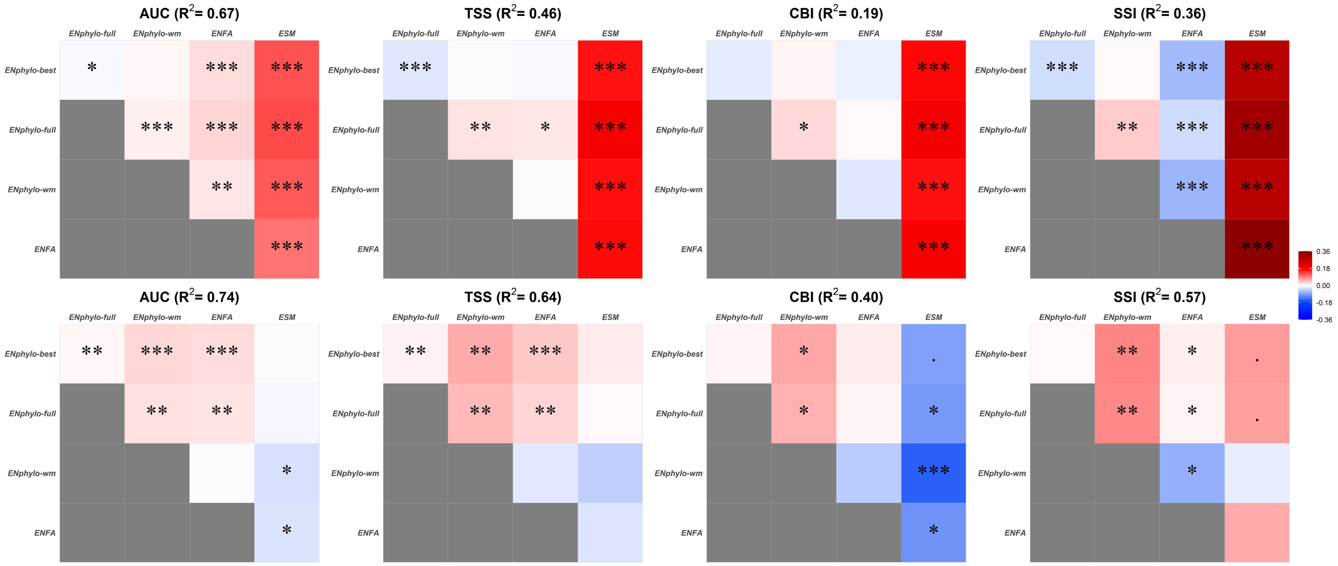

FIGURE 1 Results of linear mixed models' marginal means comparison of the different SDM algorithms applied in this study, performed on individual species by sampling from their respective fossil records either 10 (upper row) or 20 (lower row) fossil occurrences. The procedure is iterated over 31 different species. Asterisks indicate significance at 0.05 (one asterisk), 0.01 (two asterisks), 0.001 levels (three asterisks). The colour shade indicates whether the algorithm in each row is either better performing (red shades) or worse performing (blue shades) than the competing algorithm (columns), according to four different evaluation metrics.

ENphylo performance is not sensitive to the number of species imputed. LMMs suggested that there are no significant differences between the three imputation intensity scenarios in term of AUC values (Table 2) indicating imputation is still robust when some 30% CO matrices are imputed.

4 | DISCUSSION

The trade-off between the need to accurately shape the distribution of rare species and the inherent difficulty in modelling them is known as the 'rare-species modelling paradox' (Lomba et al., 2010). Lomba and colleagues proposed to overcome the problem using all possible combinations of environmental predictors fitting bivariate models and averaging their predictions with a weighted ensemble based on model performances. Breiner et al. (2015) compared this approach, named 'ensembles of small models' (ESM), to standard ensemble models calibrated on large covariate sets by using a pool of 107 rare species with number of occurrences ranging in between 10 and 140 (Breiner et al., 2015). They found ESM outperforms standard SDMs in terms of model performance and transferability, especially when the number of training data was low (Breiner et al., 2015). In a subsequent paper, Breiner and colleagues tested ESM with a range of different modelling algorithms using rare and under-sampled species with 10–25 occurrences (Breiner et al., 2018). However, how much ESM is appropriate, in absolute terms, to pursue the goal of modelling the rarest species was never directly tested, and whereas ESM effectively addresses overfitting, it remains fraught with the problem that the poor available starting information on species

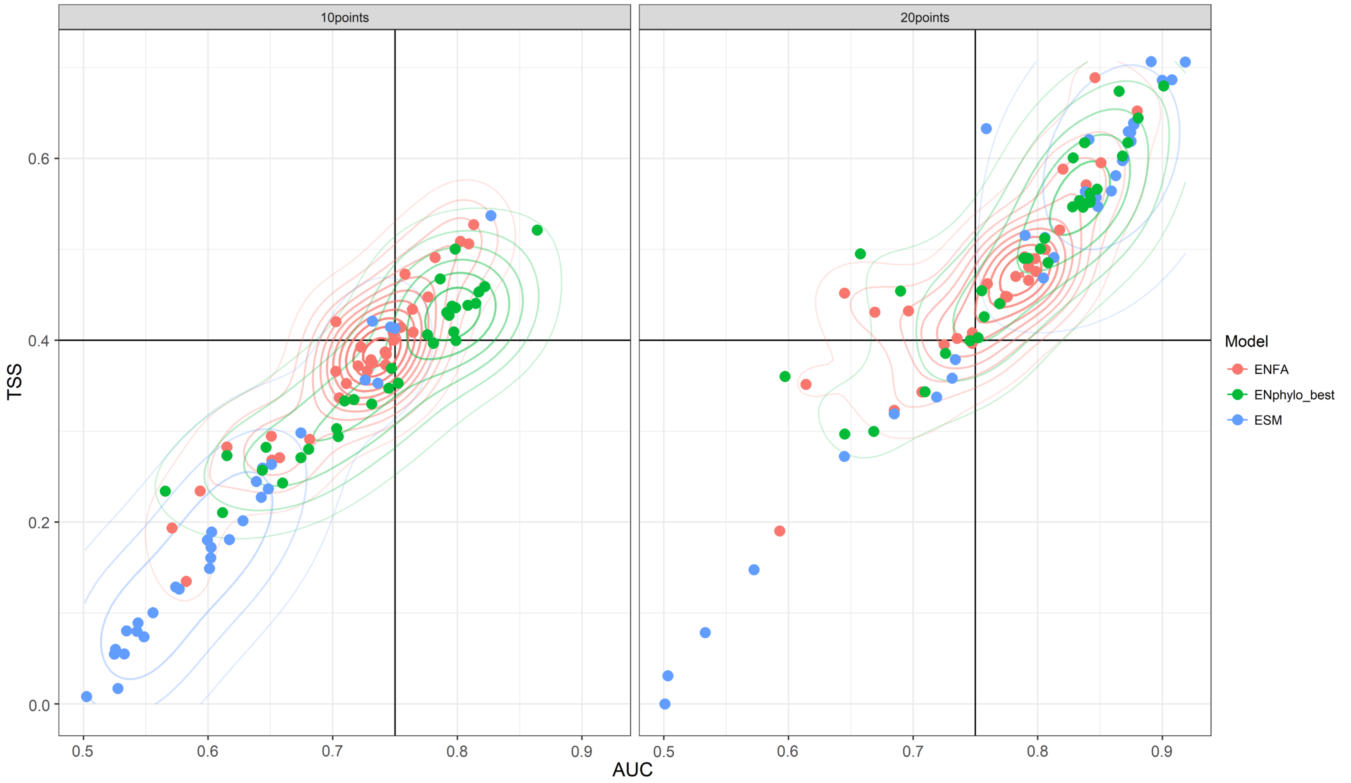

FIGURE 2 Contour plots showing model performance (according to AUC and TSS metrics) for individual species, under both 10-occurrence (10 points) and 20-occurrence (20 points) sampling strategies. The vertical (for AUC) and horizontal (for TSS) solid lines represent commonly held reference values to judge model accuracy.

TABLE 2 Results of random-slope mixed effect models at three different imputation intensity scenarios, ran under *ENphylo*. The table reports the estimate, standard error, *p*.values and significance associated with each scenario. We adopted the nomenclature 'scenarios 1-3' to refer to 3-6-9 number of imputed species (out of 31), respectively. Scenario 1 was selected as reference level in the mixed model

Imputation intensity scenario	Estimate	Std.Error	*p*.value	
(Intercept)	0.862	0.011	<0.001	***
Scenario 2	0.000	0.002	0.935	
Scenario 3	−0.001	0.002	0.670	

***Means "highly significant" in reference to the *p* value which is below 1e-3.

environmental preferences might be unrepresentative of the true species preferences. To tackle these issues, we propose *ENphylo*, a new modelling algorithm which proved to accurately predict the geographical distribution of extremely rare species by combining ENFA (Hirzel et al., 2002) and phylogenetic imputation. The method rationale is to calculate niche marginality and specialization factors, as routinely implemented under ENFA, for all the well-sampled species and then to rely on phylogenetic relatedness to derive marginality and specialization for the poorly sampled species. Although more sophisticated algorithms progressively replaced ENFA in the mainstream of modelling studies (but see e.g. Andersen et al., 2021; Cartledge et al., 2021; Mugo et al., 2020; Sutton et al., 2021), this method has been used in >200 papers since its publication in 2002 (according to Scopus database in January 2022), and is still widely adopted to describe species habitat preferences, niche characteristics and vulnerability to global change (Cordier et al., 2021; Melchionna et al., 2018; Raia et al., 2020; Rinnan & Lawler, 2019). ENFA marginality and specialization can intuitively be translated into biologically meaningful concepts, as they represent the position and width of a species niche in the environmental space relative to the habitat conditions available to the species (Hirzel et al., 2002). Since niche position and width depend on biological traits such as thermal tolerance limits, body size and fat metabolism, they can be studied as they were phenotypes (Pearman et al., 2008; Rolland et al., 2018). This makes phylogenetic effects a potential predictive factor in terms niche modelling.

Standing on this assumption, we combined ENFA and phylogenetic imputation to produce SDMs for rare species. By means of randomization experiments based on extreme subsampling of fossil species with otherwise dense fossil records, we demonstrated that *ENphylo* consistently outperforms competing methods in terms of predictive accuracy, with the difference becoming shallower as the sample size increases. Standard SDM algorithms 'learn' the species niche from occurrence data; hence, they are limited by the sampled climatic variability generating low

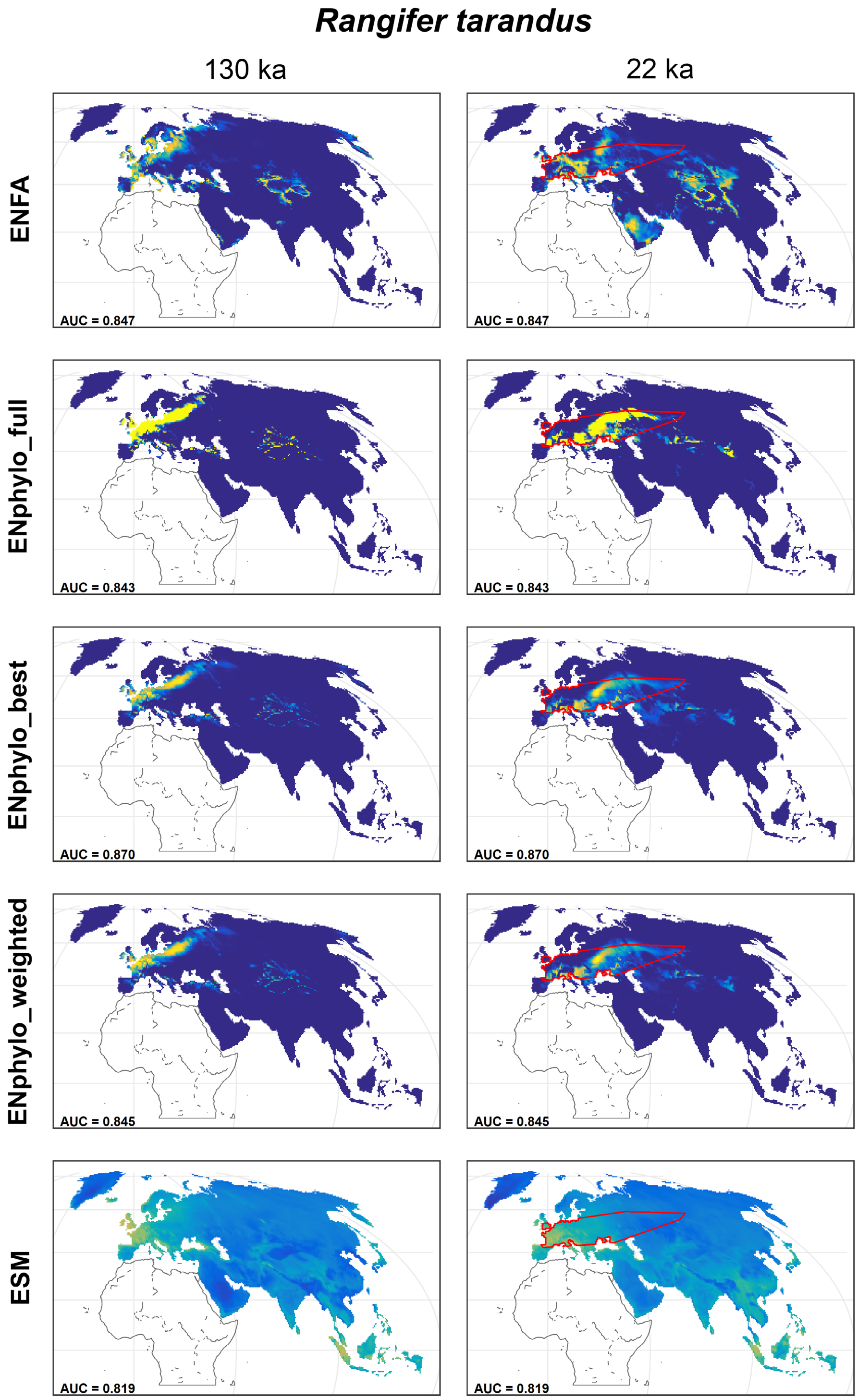

FIGURE 3 Comparison between the habitat suitability maps predicted with different algorithms for the reindeer *Rangifer tarandus*, at 130 (last interglacial, left column) and 22 ka (last glacial maximum, right column) using 10 fossil occurrences only to produce the SDM. Probability of occurrence spans from 0 (blue) to 1 (yellow). For ENphylo_full, the maps represent the consensus among the binary maps produced from habitat suitability values related to 20 alternative phylogenies, applying the committee averaging method. The red polygon indicates the minimum convex polygon (MCP) encompassing *R. tarandus* occurrence data at 22 ka (real fossil data at 130 ka are too scanty to produce a reliable MCP estimate).

transferability when projected into new climatic conditions (Liu et al., 2022; Qiao et al., 2019), a limit that *ENphylo* circumvents providing biologically meaningful suitability maps under widely differing climatic conditions (Figures 3 and 4). By using *ENphylo*, we found 21–23 species achieved large (>0.7) AUC values, depending on the approach selected. The corresponding figure in Mondanaro et al., 2021, using the full fossil record for the same set of species was 22, indicating *ENphylo* is at least as good as Maxent using only one-fifth to one-tenth of the original number of fossil occurrences. The better accuracy of *ENphylo* over competing methods progressively fades away as the number of occurrences rises above 20 (Figures 2 and 3), that is when Maxent is expected to perform better as the fundamental niche is approached (Fonderflick et al., 2015; Qiao et al., 2019). At 20 occurrences, we found that ESM performs slightly better than *ENphylo* (although the difference is not statistically significant). However, ESM still fails to produce acceptable predictions for nearly one sixth of the species (Figure 2), whereas *ENphylo*, especially under

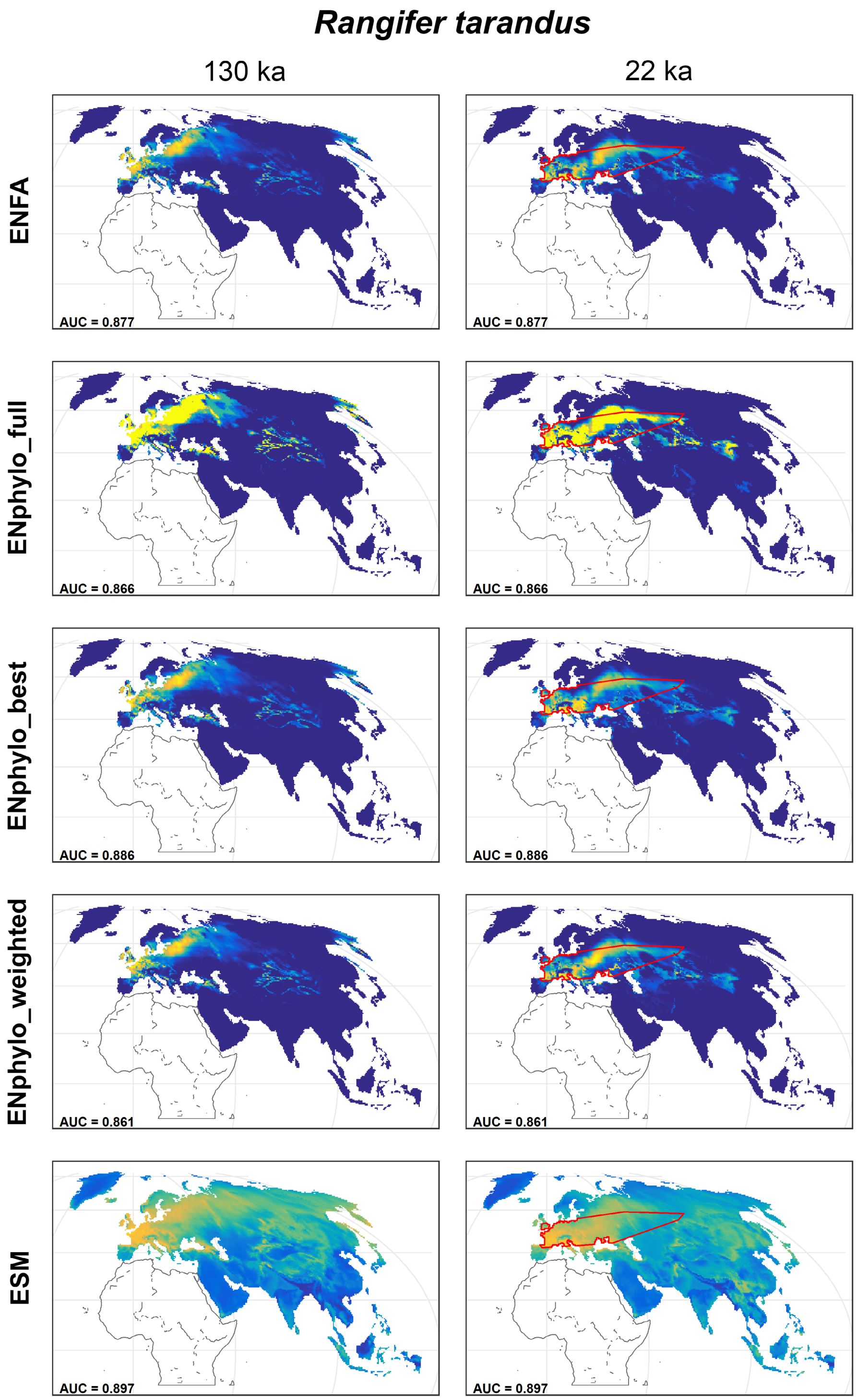

FIGURE 4 Comparison between the habitat suitability maps predicted with different algorithms for the reindeer *Rangifer tarandus*, at 130 (last interglacial, left column) and 22 ka (last glacial maximum, right column) using 20 fossil occurrences only to produce the SDM. Probability of occurrence spans from 0 (blue) to 1 (yellow). For ENphylo_full, the maps represent the consensus among the binary maps produced from habitat suitability values related to 20 alternative phylogenies, applying the committee averaging method. The red polygon indicates the minimum convex polygon (MCP) encompassing *R. tarandus* occurrence data at 22 ka (real fossil data at 130 ka are too scanty to produce a reliable MCP estimate).

the 'best' strategy, remains stable (Figure 2). A key point of concern is whether phylogenetic imputation, on which *ENphylo* stands, is appropriate under all conditions. It has been shown that phylogenetic effects might be negligible or even misleading in terms of inferring climatic preferences (Münkemüller et al., 2015). However, under *ENphylo* phylogenetic effects are not derived from the tree topology and branch lengths as they are. In contrast, the species positions on the tree are swapped, and the branching times altered (100 times by default), as to maximize the model predictive performance while accounting for phylogenetic uncertainty. This effectively relaxes the assumption (intrinsic to phylogenetic imputation) that the tree is 'correct' and by altering branch lengths relieves the assumption that a single rate of evolution in the climatic niche applies to all branches in the tree.

We meant *ENphylo* to extend species distribution modelling to species for which limited observational information is available. We demonstrate the method is especially appropriate when dealing with rare species, and with extinct species whose fossil record is necessarily scarce because of the nature of the preservation process.

AUTHOR CONTRIBUTIONS

Alessandro Mondanaro, Mirko Di Febbraro and Pasquale Raia conceived the study. Alessandro Mondanaro, Mirko Di Febbraro and Silvia Castiglione wrote the R codes and ran the analyses. All authors contributed equally to writing the text, preparing data and figures.

ACKNOWLEDGEMENT

We are grateful to the Associate Editor, Dr. Tim Lucas, and one anonymous reviewer for the precious advice we were kindly provided on an earlier version of the manuscript.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

PEER REVIEW

The peer review history for this article is available at https://www.webofscience.com/api/gateway/wos/peer-review/10.1111/2041-210X.14066.

DATA AVAILABILITY STATEMENT

The source codes and raw data to perform the analyses, and the R package including ENphylo functions is available via a permanent GitHub link at: https://github.com/pasraia/RRdtn (the R package) and via Zenodo at: https://zenodo.org/badge/latestdoi/588226643 (Raia, 2023; the data and code to reproduce the experiments).

ORCID

Alessandro Mondanaro https://orcid.org/0000-0003-0325-7066
Mirko Di Febbraro https://orcid.org/0000-0001-8898-7046
Silvia Castiglione https://orcid.org/0000-0002-6140-1495
Marina Melchionna https://orcid.org/0000-0001-7919-4431
Carmela Serio https://orcid.org/0000-0001-5256-5511
Pasquale Raia https://orcid.org/0000-0002-4593-8006

REFERENCES

Allouche, O., Tsoar, A., & Kadmon, R. (2006). Assessing the accuracy of species distribution models: Prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology*, *43*(6), 1223–1232. https://doi.org/10.1111/j.1365-2664.2006.01214.x

Andersen, D., Borzée, A., & Jang, Y. (2021). Predicting global climatic suitability for the four most invasive anuran species using ecological niche factor analysis. *Global Ecology and Conservation*, *25*, e01433. https://doi.org/10.1016/j.gecco.2020.e01433

Barbet-Massin, M., Rome, Q., Villemant, C., & Courchamp, F. (2018). Can species distribution models really predict the expansion of invasive species? *PLoS ONE*, *13*(3), e0193085. https://doi.org/10.1371/journal.pone.0193085

Blomqvist, D., Pauliny, A., Larsson, M., & Flodin, L. Å. (2010). Trapped in the extinction vortex? Strong genetic effects in a declining vertebrate population. *BMC Evolutionary Biology*, *10*(1), 33. https://doi.org/10.1186/1471-2148-10-33

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Breiner, F. T., Guisan, A., Bergamini, A., & Nobis, M. P. (2015). Overcoming limitations of modelling rare species by using ensembles of small models. *Methods in Ecology and Evolution*, *6*(10), 1210–1218. https://doi.org/10.1111/2041-210X.12403

Breiner, F. T., Nobis, M. P., Bergamini, A., & Guisan, A. (2018). Optimizing ensembles of small models for predicting the distribution of species with few occurrences. *Methods in Ecology and Evolution*, *9*(4), 802–808. https://doi.org/10.1111/2041-210X.12957

Broennimann, O. (2022). Package 'ecospat'. Retrieved from https://cran.r-project.org/web/packages/ecospat/ecospat.pdf

Calenge, C., Darmon, G., Basille, M., Loison, A., & Jullien, J. M. (2008). The factorial decomposition of the Mahalanobis distances in habitat selection studies. *Ecology*, *89*(2), 555–566. https://doi.org/10.1890/06-1750.1

Carotenuto, F., Di Febbraro, M., Melchionna, M., Castiglione, S., Saggese, F., Serio, C., Mondanaro, A., Passaro, F., Loy, A., & Raia, P. (2016). The influence of climate on species distribution over time and space during the late quaternary . *Quaternary Science Reviews* , *149* , 188 – 199. https://doi.org/10.1016/j.quascirev.2016.07.036

Cartledge, E. L., Baker, M., White, I., Powell, A., Gregory, B., Varley, M., Hurst, J. L., & Stockley, P. (2021). Applying remotely sensed habitat descriptors to assist reintroduction programs: A case study in the hazel dormouse . *Conservation Science and Practice*, *3* (12), e544. https://doi.org/10.1111/csp2.544

Castiglione, S., Serio, C., Mondanaro, A., Melchionna, M., & Raia, P. (2022). Fast production of large, time-calibrated, informal supertrees with tree.merger. *Palaeontology*, *65*(1), e12588. https://doi.org/10.1111/pala.12588

Castiglione , S., Serio , C., Piccolo , M., Mondanaro , A., Melchionna , M., Di Febbraro , M. , Sansalone, G., Wroe, S., & Raia , P. (2021). The influence of domestication, insularity and sociality on the tempo and mode of brain size evolution in mammals . *Biological Journal of the Linnean Society* , *132*(1), 221–231 . https://doi.org/10.1093/biolinnean/blaa186

Castiglione, S., Tesone, G., Piccolo, M., Melchionna, M., Mondanaro, A., Serio, C., Di Febbraro, M., & Raia, P. (2018). A new method for testing evolutionary rate variation and shifts in phenotypic evolution. *Methods in Ecology and Evolution*, *9*(4), 974– 983. https://doi.org/10.1111/2041-210X.12954

Clark, J. D., Dunn, J. E., & Smith, K. G. (1993). A multivariate model of female black bear habitat use for a geographic information system. *The Journal of Wildlife Management*, *57*(3), 519. https://doi.org/10.2307/3809276

Cordier, J. M., Rojas-Soto, O., Semhan, R., Abdala, C. S., & Nori, J. (2021). Out of sight, out of mind: Phylogenetic and taxonomic gaps imply great underestimations of the species' vulnerability to global climate change. *Perspectives in Ecology and Conservation*, *19*(2), 225–231. https://doi.org/10.1016/J.PECON.2021.02.002

Di Febbraro, M., Carotenuto, F., Castiglione, S., Russo, D., Loy, A., Maiorano, L., & Raia, P. (2017). Does the jack of all trades fare best? Survival and niche width in late Pleistocene megafauna. *Journal of Biogeography*, *44*(12), 2828–2838. https://doi.org/10.1111/jbi.13078

Eaton, S., Ellis, C., Genney, D., Thompson, R., Yahr, R., & Haydon, D. T. (2018). Adding small species to the big picture: Species distribution modelling in an age of landscape scale conservation. *Biological Conservation*, *217*, 251–258. https://doi.org/10.1016/J.BIOCON.2017.11.012

Elith, J. (2000). Quantitative methods for modeling species habitat: Comparative performance and an application to Australian plants. In S. Ferson & M. Burgman (Eds.), *Quantitative methods for conservation biology* (pp. 39–58). Springer-Verlag. https://doi.org/10.1007/0-387-22648-6_4

Elith, J., & Leathwick, J. R. (2009). Species distribution models: Ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics*, *40*(1), 677–697. https://doi.org/10.1146/annurev.ecolsys.110308.120159

Fielding, A. H., & Bell, J. F. (1997). A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation*, *24*(1), 38–49. https://doi.org/10.1017/S0376892997000088

Fois, M., Cuena-Lombraña, A., Fenu, G., & Bacchetta, G. (2018). Using species distribution models at local scale to guide the search of poorly known species: Review, methodological issues and future directions. *Ecological Modelling*, *385*, 124–132. https://doi.org/10.1016/j.ecolmodel.2018.07.018

Fonderflick, J., Azam, C., Brochier, C., Cosson, E., & Quékenborn, D. (2015). Testing the relevance of using spatial modeling to predict foraging habitat suitability around bat maternity: A case study in Mediterranean landscape. *Biological Conservation*, *192*, 120–129. https://doi.org/10.1016/j.biocon.2015.09.012

Fourcade, Y., Besnard, A. G., & Secondi, J. (2018). Paintings predict the distribution of species, or the challenge of selecting environmental predictors and evaluation statistics. *Global Ecology and Biogeography*, *27*(2), 245–256. https://doi.org/10.1111/GEB.12684

Freeman, E. A., & Moisen, G. (2008). PresenceAbsence: An R package for presence absence analysis. *Journal of Statistical Software*, *23*(11), 1–31. https://doi.org/10.18637/jss.v023.i11

Garland, T., & Ives, A. R. (2000). Using the past to predict the present: Confidence intervals for regression equations in phylogenetic comparative methods. *The American Naturalist*, *155*(3), 346–364. https://doi.org/10.1086/303327

Goolsby, E. W., Bruggeman, J., & Ané, C. (2017). Rphylopars: Fast multivariate phylogenetic comparative methods for missing data and within-species variation. *Methods in Ecology and Evolution*, *8*(1), 22–27. https://doi.org/10.1111/2041-210X.12612

Hengl, T., Sierdsema, H., Radović, A., & Dilo, A. (2009). Spatial prediction of species' distributions from occurrence-only records: Combining point pattern analysis, ENFA and regression-kriging. *Ecological Modelling*, *220*(24), 3499–3511. https://doi.org/10.1016/J.ECOLMODEL.2009.06.038

Hirzel, A. H., Hausser, J., Chessel, D., & Perrin, N. (2002). Ecological-niche factor analysis: How to compute habitat-suitability maps without absence data? *Ecology*, *83*(7), 2027. https://doi.org/10.2307/3071784

Hirzel, A. H., Le Lay, G., Helfer, V., Randin, C., & Guisan, A. (2006). Evaluating the ability of habitat suitability models to predict species presences. *Ecological Modelling*, *199*(2), 142–152. https://doi.org/10.1016/j.ecolmodel.2006.05.017

Holden, P. B., Edwards, N. R., Rangel, T. F., Pereira, E. B., Tran, G. T., & Wilkinson, R. D. (2019). PALEO-PGEM v1.0: A statistical emulator of Pliocene-Pleistocene climate. *Geoscientific Model Development*, *12*(12), 5137–5155. https://doi.org/10.5194/gmd-12-5137-2019

Jackson, D. A. (1993). Stopping rules in principal components analysis: A comparison of heuristical and statistical approaches. *Ecology*, *74*(8), 2204–2214. https://doi.org/10.2307/1939574

Jiménez-Valverde, A. (2012). Insights into the area under the receiver operating characteristic curve (AUC) as a discrimination measure in species distribution modelling. *Global Ecology and Biogeography*, *21*(4), 498–507. https://doi.org/10.1111/j.1466-8238.2011.00683.x

Jiménez-Valverde, A. (2020). Sample size for the evaluation of presence-absence models. *Ecological Indicators*, *114*, 106289. https://doi.org/10.1016/j.ecolind.2020.106289

Leroy, B., Delsol, R., Hugueny, B., Meynard, C. N., Barhoumi, C., Barbet-Massin, M., & Bellard, C. (2018). Without quality presence-absence data, discrimination metrics such as TSS can be misleading measures of model performance. *Journal of Biogeography*, *45*(9), 1994–2002. https://doi.org/10.1111/jbi.13402

Li, W., & Guo, Q. (2013). How to assess the prediction accuracy of species presence-absence models without absence data? *Ecography*, *36*(7), 788–799. https://doi.org/10.1111/j.1600-0587.2013.07585.x

Liu, C., Wolter, C., Courchamp, F., Roura-Pascual, N., & Jeschke, J. M. (2022). Biological invasions reveal how niche change affects the transferability of species distribution models. *Ecology*, *103*, e3719. https://doi.org/10.1002/ECY.3719

Lomba, A., Pellissier, L., Randin, C., Vicente, J., Moreira, F., Honrado, J., & Guisan, A. (2010). Overcoming the rare species modelling paradox: A novel hierarchical framework applied to an Iberian endemic plant. *Biological Conservation*, *143*(11), 2647–2657. https://doi.org/10.1016/j.biocon.2010.07.007

Maiorano, L., Cheddadi, R., Zimmermann, N. E., Pellissier, L., Petitpierre, B., Pottier, J., Laborde, H., Hurdu, B. J., Pearman, P. B., Psomas, A., Singarayer, J. S., Broennimann, O., Vittoz, P., Dubuis, A., Edwards, M. E., Binney, H. A., & Guisan , A. (2013). Building the niche through time: Using 13,000 years of data to predict the effects of climate change on three tree species in Europe. *Global Ecology and Biogeography*, *22*(3), 302–317. https://doi.org/10.1111/j.1466-8238.2012.00767.x

Marmion, M., Parviainen, M., Luoto, M., Heikkinen, R. K., & Thuiller, W. (2009). Evaluation of consensus methods in predictive species distribution modelling. *Diversity and Distributions*, *15*(1), 59–69. https://doi.org/10.1111/j.1472-4642.2008.00491.x

McCullagh, P., & Nelder, J. A. (1983). *Generalized linear models. Generalized linear models*. Routledge. https://doi.org/10.1201/9780203753736

Melchionna, M., Di Febbraro, M., Carotenuto, F., Rook, L., Mondanaro, A., Castiglione, S., Serio, C., Vero, V. A., Tesone, G., Piccolo, M., Diniz-Filho, J. A. F., & Raia, P. (2018). Fragmentation of Neanderthals' pre-extinction distribution by climate change . *Palaeogeography, Palaeoclimatology, Palaeoecology*, *496*, 146–154. https://doi.org/10.1016/j.palaeo.2018.01.031

Mondanaro, A., Di Febbraro, M., Melchionna, M., Maiorano, L., Di Marco, M., Edwards, N. R., Holden, P. B., Castiglione, S., Rook, L., & Raia , P. (2021). The role of habitat fragmentation in Pleistocene megafauna extinction in Eurasia. *Ecography*, *44*(11), 1619–1630. https://doi.org/10.1111/ecog.05939

Mugo, R., Saitoh, S. I., Igarashi, H., Toyoda, T., Masuda, S., Awaji, T., & Ishikawa, Y. (2020). Identification of skipjack tuna (*Katsuwonus pelamis*) pelagic hotspots applying a satellite remote sensing-driven analysis of ecological niche factors: A short-term run. *PLoS ONE*, *15*(8 August), e0237742. https://doi.org/10.1371/journal.pone.0237742

Münkemüller, T., Boucher, F. C., Thuiller, W., & Lavergne, S. (2015). Phylogenetic niche conservatism – Common pitfalls and ways forward. *Functional Ecology*, *29*(5), 627–639. https://doi.org/10.1111/1365-2435.12388

Muscarella, R., Galante, P. J., Soley-Guardia, M., Boria, R. A., Kass, J. M., Uriarte, M., & Anderson, R. P. (2014). ENMeval: An R package for conducting spatially independent evaluations and estimating optimal model complexity for Maxent ecological niche models. *Methods in Ecology and Evolution*, *5*(11), 1198–1205. https://doi.org/10.1111/2041-210x.12261

Pearman, P. B., Guisan, A., Broennimann, O., & Randin, C. F. (2008). Niche dynamics in space and time. *Trends in Ecology & Evolution*, *23*(3), 149–158. https://doi.org/10.1016/j.tree.2007.11.005

Phillips, S. B., Aneja, V. P., Kang, D., & Arya, S. P. (2006). Modelling and analysis of the atmospheric nitrogen deposition in North Carolina. *International Journal of Global Environmental Issues*, *6*(2–3), 231–252. https://doi.org/10.1016/j.ecolmodel.2005.03.026

Préau, C., Trochet, A., Bertrand, R., & Isselin-Nondedeu, F. (2018). Modeling potential distributions of three european amphibian species comparing enfa and maxent. *Herpetological Conservation and Biology*, *13*(1), 91–104.

Qiao, H., Feng, X., Escobar, L. E., Peterson, A. T., Soberón, J., Zhu, G., & Papeş, M. (2019). An evaluation of transferability of ecological niche models. *Ecography*, *42*(3), 521–534. https://doi.org/10.1111/ecog.03986

Raia, P. (January 2023). pasraia/ENphylo-test-files: Publication files and package. https://zenodo.org/badge/latestdoi/588226643

Raia, P., Mondanaro, A., Melchionna, M., di Febbraro, M., Diniz-Filho, J. A. F., Rangel, T. F., Holden, P. B., Carotenuto, F., Edwards, N. R., Lima-Ribeiro, M. S., Profico, A., Maiorano, L., Castiglione, S., Serio, C., & Rook, L. (2020). Past extinctions of homo species coincided with increased vulnerability to climatic change. *One Earth*, *3*(4), 480–490. https://doi.org/10.1016/j.oneear.2020.09.007

Rinnan, D. S. (2021). Package 'CENFA' type package title climate and ecological niche factor analysis. https://doi.org/10.1111/ecog.03937

Rinnan, D. S., & Lawler, J. (2019). Climate-niche factor analysis: A spatial approach to quantifying species vulnerability to climate change. *Ecography*, *42*(9), 1494–1503. https://doi.org/10.1111/ecog.03937

Rolland, J., Silvestro, D., Schluter, D., Guisan, A., Broennimann, O., & Salamin, N. (2018). The impact of endothermy on the climatic niche evolution and the distribution of vertebrate diversity. *Nature Ecology and Evolution*, *2*(3), 459–464. https://doi.org/10.1038/s41559-017-0451-9

Santini, L., Benítez-López, A., Maiorano, L., Čengić, M., & Huijbregts, M. A. J. (2021). Assessing the reliability of species distribution projections in climate change research. *Diversity and Distributions*, *27*(6), 1035–1050. https://doi.org/10.1111/ddi.13252

Sousa-Silva, R., Alves, P., Honrado, J., & Lomba, A. (2014). Improving the assessment and reporting on rare and endangered species through species distribution models. *Global Ecology and Conservation*, *2*, 226–237. https://doi.org/10.1016/J.GECCO.2014.09.011

Sutton, L. J., Anderson, D. L., Franco, M., McClure, C. J. W., Miranda, E. B. P., Vargas, F. H., Vargas Gonzalez, J. J., & Puschendorf, R. (2021). Geographic range estimates and environmental requirements for the harpy eagle derived from spatial models of current and past distribution. *Ecology and Evolution*, *11*(1), 481–497. https://doi.org/10.1002/ece3.7068

Svenning, J. C., Fløjgaard, C., Marske, K. A., Nógues-Bravo, D., & Normand, S. (2011). Applications of species distribution modeling to paleobiology. *Quaternary Science Reviews*, *30*(21–22), 2930–2947. https://doi.org/10.1016/j.quascirev.2011.06.012

Tessarolo, G., Ladle, R. J., Lobo, J. M., Rangel, T. F., & Hortal, J. (2021). Using maps of biogeographical ignorance to reveal the uncertainty in distributional data hidden in species distribution models. *Ecography*, *44*(12), 1743–1755. https://doi.org/10.1111/ecog.05793

Timmermann, A., Yun, K.-S., Raia, P., Ruan, J., Mondanaro, A., Zeller, E., Zollikofer, C., Ponce de León, M., Lemmon, D., Willeit, M., & Ganopolski, A. (2022). Climate effects on archaic human habitats and species successions. *Nature*, *604*(7906), 495–501. https://doi.org/10.1038/s41586-022-04600-9

Tóth, A. B., Lyons, S. K., Barr, W. A., Behrensmeyer, A. K., Blois, J. L., Bobe, R., … Alroy, J. (2019). Reorganization of surviving mammal communities after the end-Pleistocene megafaunal extinction. *Science*, *365*(6459), 1305–1308. https://doi.org/10.1126/science.aaw1605

Valavi, R., Elith, J., Lahoz-Monfort, J. J., & Guillera-Arroita, G. (2021). Modelling species presence-only data with random forests. *Ecography*, *44*(12), 1731–1742. https://doi.org/10.1111/ECOG.05615

Varela, S., Lobo, J. M., & Hortal, J. (2011). Using species distribution models in paleobiogeography: A matter of data, predictors and concepts. *Palaeogeography, Palaeoclimatology, Palaeoecology*, *310*(3–4), 451–463. https://doi.org/10.1016/j.palaeo.2011.07.021

Vaughan, I. P., & Ormerod, S. J. (2005). The continuing challenges of testing species distribution models. *Journal of Applied Ecology*, *42*, 720–730 . https://doi.org/10.1111/j.1365-2664.2005.01052.x

Warren, D. L., & Seifert, S. N. (2011). Ecological niche modeling in Maxent: The importance of model complexity and the performance of model selection criteria. *Ecological Applications*, *21*(2), 335–342. https://doi.org/10.1890/10-1171.1

Wisz, M. S., Hijmans, R. J., Li, J., Peterson , A. T., Graham , C. H., Guisan, A., & NCEAS. (2008). Effects of sample size on the performance of species distribution models . *Diversity and Distributions*, *14*(5), 763–773 . https://doi.org/10.1111/j.1472-4642.2008.00482.x

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

Table S1. AUC statistics for individual species calculated according to each SDM algorithm at 10 (up) and 20 (down) occurrence data points.

Figure S1. Contour plots showing model performance (according to AUC and Boyce Index (CBI) metrics) for individual species, under both 10-occurrences (10points) and 20-occurrences (20points) sampling strategies. The vertical (for AUC) and horizontal (for TSS) solid lines represent commonly held reference values to judge model accuracy.

Figure S2. Contour plots showing model performance (according to AUC and Soerensen Index (SSI) metrics) for individual species, under both 10-occurrences (10points) and 20-occurrences (20points) sampling strategies. The vertical (for AUC) and horizontal (for TSS) solid lines represent commonly held reference values to judge model accuracy.

How to cite this article: Mondanaro, A., Di Febbraro, M., Castiglione, S., Melchionna, M., Serio, C., Girardi, G., Belfiore, A. M., & Raia, P. (2023). *ENphylo*: A new method to model the distribution of extremely rare species. *Methods in Ecology and Evolution*, *14*, 911–922. https://doi.org/10.1111/2041-210X.14066