

Research Article

Models based on chronological data correctly predict the spread of freshwater aliens, and reveal a strong influence of river access, anthropogenic activities and climate regimes

Marta Rodríguez-Rey¹, Sofia Consuegra², Carlos Garcia de Leaniz²

¹ Universidad de Alcalá, Facultad de Ciencias, Departamento de Ciencias de la Vida, Grupo de Ecología y Restauración Forestal, Crta. A-2 Km. 33.6, 28805, Alcalá de Henares, Madrid, Spain

² Department of Biosciences, Swansea University, Singleton Park, Swansea SA2 8PP, UK

Corresponding author: Marta Rodríguez-Rey (marta.rodriguez.rey@gmail.com)



Academic editor: F. J. Oficialdegui

Received: 23 January 2023

Accepted: 1 August 2023

Published: 8 November 2023

Citation: Rodríguez-Rey M, Consuegra S, Garcia de Leaniz C (2023) Models based on chronological data correctly predict the spread of freshwater aliens, and reveal a strong influence of river access, anthropogenic activities and climate regimes. *Aquatic Invasions* 18(4): 455–472. <https://doi.org/10.3391/ai.2023.18.4.111481>

Copyright: © Marta Rodríguez-Rey et al. This is an open access article distributed under terms of the Creative Commons Attribution License ([Attribution 4.0 International – CC BY 4.0](https://creativecommons.org/licenses/by/4.0/)).

Abstract

Alien species constitute one of the main threats to freshwater ecosystems, negatively impacting biodiversity, economy, biosecurity and ecosystem services. Predicting the arrival and spread of alien species is of paramount importance to prevent new introductions and control the expansion and establishment of already introduced species. We modelled the distribution of four freshwater invaders in Great Britain, using environmental and anthropogenic predictors, to help focus management actions. The species grouped different taxa including signal crayfish (*Pacifastacus leniusculus*), the marsh frog (*Pelophylax ridibundus*), the red-eared slider (*Trachemys scripta*) and the pike-perch (*Sander lucioperca*). The modelling approach accounted for methodological limitations and implemented two evaluations, a temporal evaluation using data corresponding to 70% of the oldest records to calibrate models and the remaining 30% for evaluation using various performance metrics (the common AUC, TSS and also null models) and an independent evaluation using the most recent range expansion of the species in the last six years. The distribution of the species was facilitated by multiple environmental and anthropogenic predictors. Road density was the second most important predictor of the occurrence of signal crayfish and red-eared slider preceded by the distance to ports and isothermality for each species respectively. Human population density was the most important predictor of marsh frog presence whereas pike-perch was mostly related to the proximity of boat ramps and precipitation regimes. Our distribution models were accurate and predicted the most recent range expansion of all of the species, highlighting their usefulness for preventing alien species spread and the value of using historical projections, usually available for non-native species, to calibrate and evaluate Species Distribution Models.

Key words: aquatic non-native species, human ecology, introduction pathways, management, species distribution models, forecasting

Introduction

Freshwater ecosystems are among the most endangered ecosystems on the planet (Sala et al. 2000; Tickner et al. 2020), yet they tend to be neglected by conservation efforts compared to terrestrial or marine ecosystems (Darwall et al. 2011). Biological invasions are of special concern in freshwater ecosystems, threatening the native biodiversity by disrupting trophic links (Gallardo et al. 2016) or competing with species of commercial interest (Cardona et al. 2008), and also impacting the economy (Cuthbert et al. 2021) and human health (Tricarico et al. 2017; Mazza and Tricarico 2018). Information about the dispersal patterns and vectors of invasion is crucial to predict the invasion process of different species, for which appropriate models are essential (Barbet-Massin et al. 2018). Once an alien species has arrived in a new area or country, secondary dispersal may occur naturally or via human transport (Johnson et al. 2001; Vander Zanden et al. 2008), but for many organisms assessing the relative contribution of each of these factors is difficult (Huang et al. 2017). Therefore, understanding this post-introduction range expansion is critical to control the further spread of introduced species (Mallez and McCartney 2018; Robertson et al. 2020).

In addition to the importance of climate on the distribution of alien freshwater species (Rahel and Olden 2008), recent research has shown that anthropogenic factors can affect the range expansion more than environmental variables alone (Faurby and Araújo 2018; Chapman et al. 2020). For example, the number of invasive species tends to be higher outside protected areas than inside them, possibly reflecting the role of anthropogenic pressures (Gallardo et al. 2015; Moustakas et al. 2018). Multiple intentional introductions in distant localities are also common and can increase invasion success (Signorile et al. 2016; Huang et al. 2017), for example, most alien fishes in Great Britain have increased their numbers in the last decade (Copp et al. 2006) due to both accidental and deliberate releases (Zięba et al. 2010). A similar pattern exists for alien molluscs (Sousa et al. 2014) and pet turtles such as the red-eared slider, whose high establishment success has been aided by repeated introductions around the globe (Bomford et al. 2008). Therefore, combining anthropogenic and environmental types of variables in ecological models will help to obtain accurate predictions that support cost-effective management solutions which are usually challenging at large spatial scales (Glen et al. 2013).

Species Distribution Modelling (SDM) is a valuable tool that uses spatially explicit variables to explain and predict the range expansion of alien species and provide options for management actions (Guisan et al. 2013). However, SDM reliability to correctly predict the spread of the species has been questioned based on limitations such as lack of equilibrium of the species in the invaded ecosystems (Václavík and Meentemeyer 2012; Mainali et al. 2015), data restrictions and errors (Václavík and Meentemeyer 2009; Barbet-Massin et al. 2012), shortcomings of the performance metrics (Jiménez-Valverde 2012) or difficulties in the evaluation process (Warren et al. 2020). In fact, the evaluation step in SDM is key and requires independent distribution data which is, for most species rarely available (West et al. 2016), thus discrimination metrics are obtained through a data splitting approach [usually cross-validation (Helmstetter et al. 2021)] creating a training and testing dataset with randomly selected records from the total occurrence set. This type of evaluation has some limitations and has been criticised due to overestimation (Lobo et al. 2008). Thus, the desirable way to reliably test a model is by gathering independent data to evaluate predictions. Monitoring programs of alien species provide a great opportunity for a more independent evaluation by boosting data availability over space and time, helping to improve the evaluation approach (Barbet-Massin et al. 2018).

To overcome some of these limitations and investigate the range dynamics of alien species, we applied SDM including multiple predictors related to vectors of introduction and dispersal and correcting for potential error sources (Sillero and Barbosa 2021). We examined the distribution of four aquatic alien species from four different taxonomic groups (fish, reptiles, amphibians and crustaceans) deliberately introduced through different routes (i.e., ornamental trade, aquaculture, pet release). The aims were: i) to identify the main drivers of dispersal of alien freshwater species in Great Britain using detailed predictors that can give light into the invasion process and ii) to assess the power of SDMs to predict the invasion process using two types of evaluations including: a) the common cross-validation approach but using a temporal data splitting instead of random selection which is more realistic to recreate the invasion process and b) an independent data evaluation using the range shift of the species with the absolute newest records (during the last 6 years) to retrospectively predict current invasion patterns.

Methods

Study area and species

We used Great Britain as study area as, being an island, it represents a closed system that allows the study of invasion processes without the possible influx of freshwater species from adjacent areas. Great Britain was divided into grids with a 5 Km² resolution as in previous studies of freshwater species in large areas (Gallardo and Aldridge 2013; Fletcher et al. 2016; Rodrigues et al. 2016). The study area was therefore composed of 8,735 grid cells after removing cells with less than 70% of the grid area to avoid the inclusion of those cells with a high sea cover.

We included signal crayfish (*Pacifastacus leniusculus*) among the arthropods, the marsh frog (*Pelophylax ridibundus*) among the amphibians, the red-eared slider (*Trachemys scripta*) as a reptile, and the pike-perch (*Sander lucioperca*) amongst the fish. The species had 982, 94, 125 and 164 grid occurrences respectively. These four species represented various taxonomic groups that were deliberately introduced in Great Britain at the end of the 19th century or beginning of the 20th century except for signal crayfish, with a contemporaneous introduction but also a high range expansion. Species distribution data were obtained from the National Biodiversity Network (NBNAtlas) using records between the date of introduction of each species and 2015 for the modelling process; therefore, including a time frame of more than 150 years depending on the species arrival (Suppl. material 1: tables S1, S2).

To control for sampling bias, we applied “systematic sampling” consisting in pooling occurrences in grids to avoid the effect of overrepresentation of records in oversampled areas, showing the higher performance among other procedures (Fourcade et al. 2014). Also, we assessed the sampling bias in the study area using the Target Group Sampling method (Phillips et al. 2009) by examining the grid cells with at least an occurrence of a non-native species under study or a native species within the same taxonomic group (i.e., fish, amphibians) to ensure that the grids included in our study had been sampled. Pseudo-absences and background locations are then more certain, for instance, if an area is considered as well sampled but the target species was missing (Phillips et al. 2009; Merow et al. 2013). For this purpose, we compared the distribution of the marsh frog with the distribution of the native *Rana temporaria* and *Bufo bufo* as well as the pike-perch with the *Salmo trutta*, *Salmo salar*, *Cobitis taenia*, *Cottus gobio* and *Alosa alosa*. The slider and the crayfish could not be reliably compared since there is no similar species of reptiles in the study area and the native crayfish (*Austropotamobius pallipes*) is threatened over the study area and its distribution is currently restricted.

Predictors

We used a combination of both environmental and anthropogenic variables, to assess the relative importance of the predictors for the alien species distribution as in other studies, (Gallardo and Aldridge 2015; Rodríguez-Rey et al. 2019; Chapman et al. 2020). Models were calibrated only in the invaded area as it has been considered the best approach for the study species in the study area (Rodríguez-Rey et al. 2019) avoiding extrapolation issues which have important implications for SDM methodologies (Yates et al. 2018). We included bioclimatic predictors due to the known effect of climate on the distribution of alien freshwater species (Britton et al. 2010; Bellard et al. 2013) (Suppl. material 1: tables S3, S4) and land use predictors within a 50 m buffer strip from each riverbank, using the hydrography map OS Open Rivers from the Ordnance Survey (www.ordnancesurvey.co.uk). Worldclim was used as climate database because it has shown higher performance in comparison with other databases (Datta et al. 2020; Jiménez-Valverde et al. 2021), it is the most widely used in macroecological studies (Poggio et al. 2018; Marchi et al. 2019) and it includes climate information from 1970 to 2000, covering a large number of occurrences in the study, dating back to 1935 (Suppl. material 1: table S2). Regarding the anthropogenic predictors, we included the distance to the closest city and population density as an indicator of human presence and pressure and road density as an indicator of human accessibility. We included predictors related to activities that are sources of introduction, dispersal and propagule pressure as the distance to the nearest port, to the nearest boat launch ramp, the distance to aquaculture facilities and to garden centres, as most of the species had an economical interest in aquaculture, sport fishing or as pets (Padilla and Williams 2004). We also included the Euclidean distance to the locality where the species were firstly introduced to account for any spatially correlated pattern of dispersal and the history of the invasion (Rodríguez-Rey et al. 2013). We extracted the mean values of all the variables in each grid cell using the zonal statistic tool in QGIS (QGIS Development Team 2016) and accounted for predictors' collinearity by including only predictor with a Variance Inflation Factor (VIF) smaller than 10 (Dormann et al. 2013) (Suppl. material 1: table S4).

Model calibration and evaluation 1

For each species, we analysed the distribution using four different SDM algorithms including Generalised Additive Models [GAM (Hastie and Tibshirani 1990)], MaxEnt (Phillips et al. 2006), Generalized Linear Models [GLM (Hosmer and Lemeshow 2000)], Random Forest models [RF (Cutler et al. 2007)] and an ensemble model combining all of the preceding algorithms (Thuiller et al. 2009). GAMs were fitted using a logit link function (binomial family) and thin plate penalised regression splines (Wood 2017) using the 'mgcv' package (Wood 2017) for R. Maxent was implemented using 'dismo' package in R (Hijmans et al. 2017) and GLMs were fitted using the logit link function (binomial family) with the base function in R (R Development Core Team 2018). RF we implemented using regression with 500 trees and a "mtry" of 6 using the randomForest (Liaw and Wiener 2002) and 'caret' packages (Kuhn et al. 2020) in R. Ensemble models including the four algorithms were created using 'biomod2' package in R (Thuiller et al. 2016) and the ensemble was calculated by averaging model predictions weighted by ROC and TSS. Due to the variety of study species, the multitude of possible combinations of hyperparameters' values (Zeng et al. 2016) and the 1000 simulations performed for the null models (see below), the modelling resulted

highly computationally demanding, so we did not tune hyperparameters. However, we made sure the convergence and meaning of all the models. The use of multiple algorithms intends to reduce the uncertainty associated with the model selection and outputs (Marmion et al. 2009). One of the most important improvements of our methodological approach in comparison with other studies was the temporal validation (Schatz et al. 2017; Rodríguez-Rey et al. 2019). We split data as in cross-validation with 70% and 30% for training and testing, respectively but trained the models with the oldest records (summing up to 70% of the total distribution data) and evaluated them using the most recent records (the remaining 30% of records) to reconstruct the invasion pathways followed by the species (Rodríguez-Rey et al. 2013; Barbet-Massin et al. 2018) (Suppl. material 1: table S2). We also accounted for sampling bias (as above) and applied sorting bias correction by pairwise distance sampling (Hijmans 2012) using ‘dismo’ package in R to avoid the spatial autocorrelation between training and test data known to inflate evaluation metrics (Sillero and Barbosa 2021). Pseudo-absences (for GAM, GLM and RF) and background data (for MaxEnt) were randomly selected but with a high certainty of being real absences after detecting that the sampling effort was high with over 70% of the grid cells being sampled.

Model quality was assessed using True Skills Statistic [TSS (Allouche et al. 2006)] and the Area Under the Curve based on the Receiver Operating Characteristic (ROC) curve [AUC, (Fielding and Bell 1997)] measures of accuracy commonly used in SDM. Since a low model performance is expected after bias correction, we applied null models for significance testing of presence-only SDM which is being increasingly applied to improve evaluation and to reduce the over-optimistic nature of cross-validation (Lima et al. 2022; Osborne et al. 2022). Therefore, in addition to these commonly applied discrimination metrics, we also built null models to test the differences in the accuracy measures between the real and null models of the SDMs assuming that their performance is random (Raes and ter Steege 2007; Rodríguez-Rey et al. 2019). For each species and algorithm combination, we obtained 1000 null models, each one with a different rearrangement of the predictors obtained by permutation. Then, we calculated the differences in the discrimination measures between the null and the real models by subtracting real model statistic value minus the null model 95 CI maximum value (thereafter ‘effect size’). Hence, positive values indicated better than null models (i.e., not random), whereas negative values indicated worse than null or random results.

Variable importance

We calculated the relative contribution of the predictor variables in the best model for each species. For MaxEnt, variable importance corresponded to the resulting contribution percentage supported by jack-knife variable importance analysis (Phillips 2017); for GAM, RF and GLM, we applied the VarImp function from the ‘caret’ package (Kuhn et al. 2020) and for the ensemble, the average of all the models was used and then the variables_importance function available in ‘biomod2’ package was applied (Thuiller et al. 2016). Partial-dependence plots were used to graphically examine relationships between the predictors and the distribution of each species.

Model predictive power (evaluation 2)

We evaluated models obtained in the previous steps using the range shift of the species in the same study area after six years (from January 2016 to December 2021) to investigate if the most recent spread of the alien species occurred in the

areas predicted by the models. This constitutes another improvement in the evaluation approach using independent data. For this purpose, we evaluated the model predictions with the data corresponding to the real range expansion of the species (i.e., new distribution records) during the last six years published from the same source (NBNAtlas).

We calculated the accuracy of the models to predict the most recent spread of the species by calculating the area under curve AUC using 'modEvA' package in R (Barbosa et al. 2021). We also examined the probability values of the recently occupied localities by calculating the quartile of values corresponding to the sites of new occurrences to examine the expected correspondence of new invasions in areas of higher predicted suitability as in Barbet-Massin et al. (2018).

Results

Species distribution models

The best models explaining the species distribution according to the effect size (i.e., the difference between the real model and the highest 95CI values of the null models for both discrimination statistics) were obtained with GAM for the signal crayfish, the red-eared slider and pike-perch, and with MaxEnt for the marsh frog. (Suppl. material 1: table S5). Signal crayfish occurrences were explained by the relative contribution of the distance to ports (16%) followed by road density (12%) both with a positive relationship (Fig. 1 and Suppl. material 1: fig. S2.1). Temperature seasonality and precipitation seasonality (10%), referred to the standard deviation of the temperature and precipitation and were the most important climatic predictors for the signal crayfish, with 11% and 10% of relative contribution to the model, both with positive relationships. Distance to the first introduction record (8% relative contribution) was negatively related to the signal crayfish presence. Altitude positively affected the probability of occurrence (7.5%), until above 400 meters where the effect was negative (Suppl. material 1: fig. S2.1).

For the red-eared slider, isothermality (i.e., mean diurnal range divided by mean annual range of temperature) and temperature in the wettest quarter had the highest relative importance within the environmental variables (19% and 11% respectively) (Fig. 1). The increase of both decreased the probability of the alien turtle presence. Road density positively contributed to the probability of occurrence by 15% as well as the proximity to the introduction point by 14% (Suppl. material 1: fig. S2.2). The distance to garden centres negatively affected the slider occurrence with a 10% of relative importance.

The marsh frog had a distribution mainly marked by the distance to the first introduction site followed by the population density, with a relative importance of 62% and 14% and negative and positive relationships, respectively (Fig. 1 and Suppl. material 1: fig. S2.3). The remaining variance was explained by a small contribution from most of the rest of the variables.

For the pike-perch, distance to the first record had a relative importance of 23% followed by precipitation seasonality with 14% of importance, both decreasing the probability of its presence (Suppl. material 1: fig. S2.4). In contrast, temperature seasonality positively affected the presence of pike-perch with a relative importance of 13%. Among the anthropogenic variables, population density (11% of importance) and distance to garden centres (9.5%) affected the pike-perch's presence with negative and positive signs, respectively. Other variables such as altitude and grassland cover in the riverside also contributed to explain the pike-perch distribution.

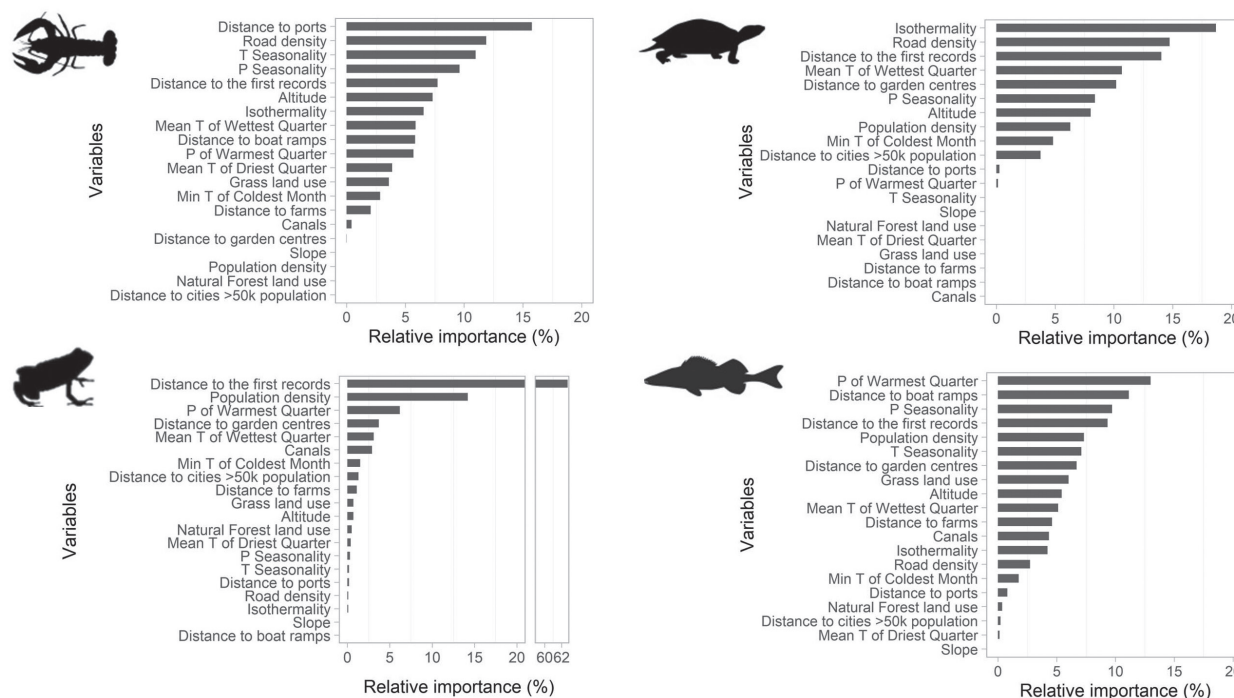


Figure 1. Relative importance of environmental and anthropic predictors for the best model for each species. The crayfish icon represents the signal crayfish (*Pacifastacus leniusculus*), the turtle icon represents the red-eared slider (*Trachemys scripta*), the frog icon corresponds to marsh frog (*Pelophylax ridibundus*) and the fish icon represents the pike-perch (*Sander lucioperca*).

Model predictive power (evaluation 2)

Based on the predicted suitability for each species (Fig. 2), the prediction rate curves of AUC values of the models evaluated using the most recent range expansion of the species (i.e., from 2016 to 2021) were 0.826 for the signal crayfish, 0.853 for the red-eared slider, 0.935 for the marsh frog and 0.797 for the pike-perch (Fig. 3). Signal crayfish highly increased its distribution between 2016 and 2021 with 404 new occurrences and 71% of the new registers were recorded in areas with predicted values of suitability over the third quartile. Red-eared slider and the pike-perch had most of their new records (80% and 82% respectively) in areas with predicted probability values within the third quartile for the whole study area. The marsh frog, increased its distribution with new 61 records with 92% of them falling in areas with suitabilities over or within the third quartile.

Discussion

Our models were highly capable to predict the last range expansion of the species (last six years). Independent data for evaluation is desirable to measure the ability of the predicted areas to be occupied. Model calibration and evaluation 1 aimed to replicate the standard practice in SDM evaluation (cross-validation using 70% and 30% of the data for training and testing, respectively) but splitting datasets using the chronological information of the range expansion instead of a random selection. Our models provided better than random models but with relatively low values of performance. The reason for low values in model performance arose from the sorting bias correction. Spatial sorting bias is known to inflate evaluation metrics (Sillero and Barbosa 2021; Illanas et al. 2022) and generate model over-estimation (Veloz 2009; Hijmans 2012). Therefore, sorting bias correction overall reduced model performance since uncorrected models showed a strong bias in

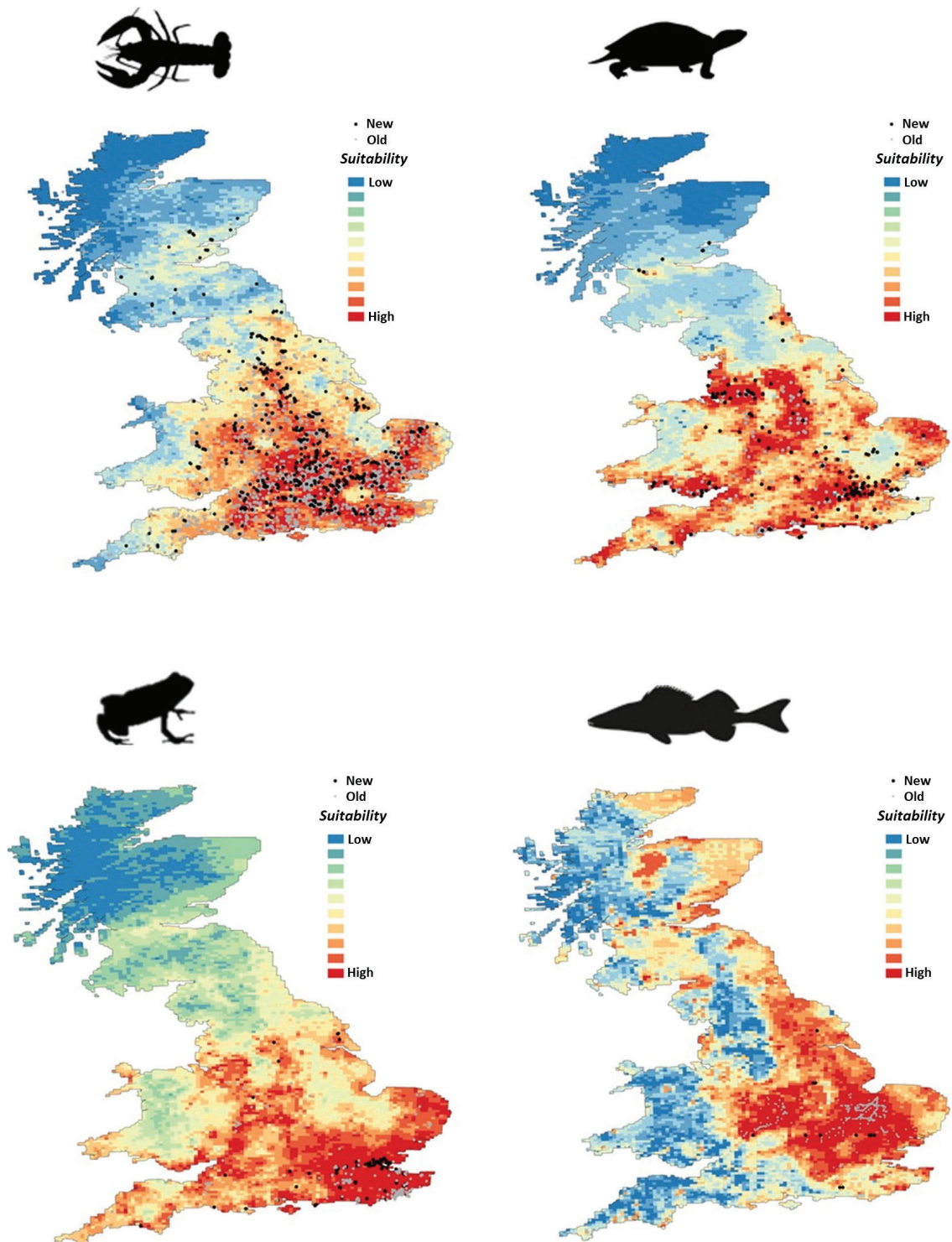


Figure 2. Suitability maps for the best model for each species showing the occurrences until 2015 used for model training and testing and the occurrences from 2016 to 2021 to posterior evaluation of the range expansion. The crayfish icon represents the signal crayfish (*Pacifastacus leniusculus*), the turtle icon represents the red-eared slider (*Trachemys scripta*), the frog icon corresponds to marsh frog (*Pelophylax ridibundus*) and the fish icon represents the pike-perch (*Sander lucioperca*).

their predicted species patterns (Hertzog et al. 2014; Freeman et al. 2022). Nevertheless, the most relevant evaluation in this study is the one based on the real range expansion of the species after 6 years (evaluation 2) showing that models' predictions were highly accurate after independent data validation. This change in the

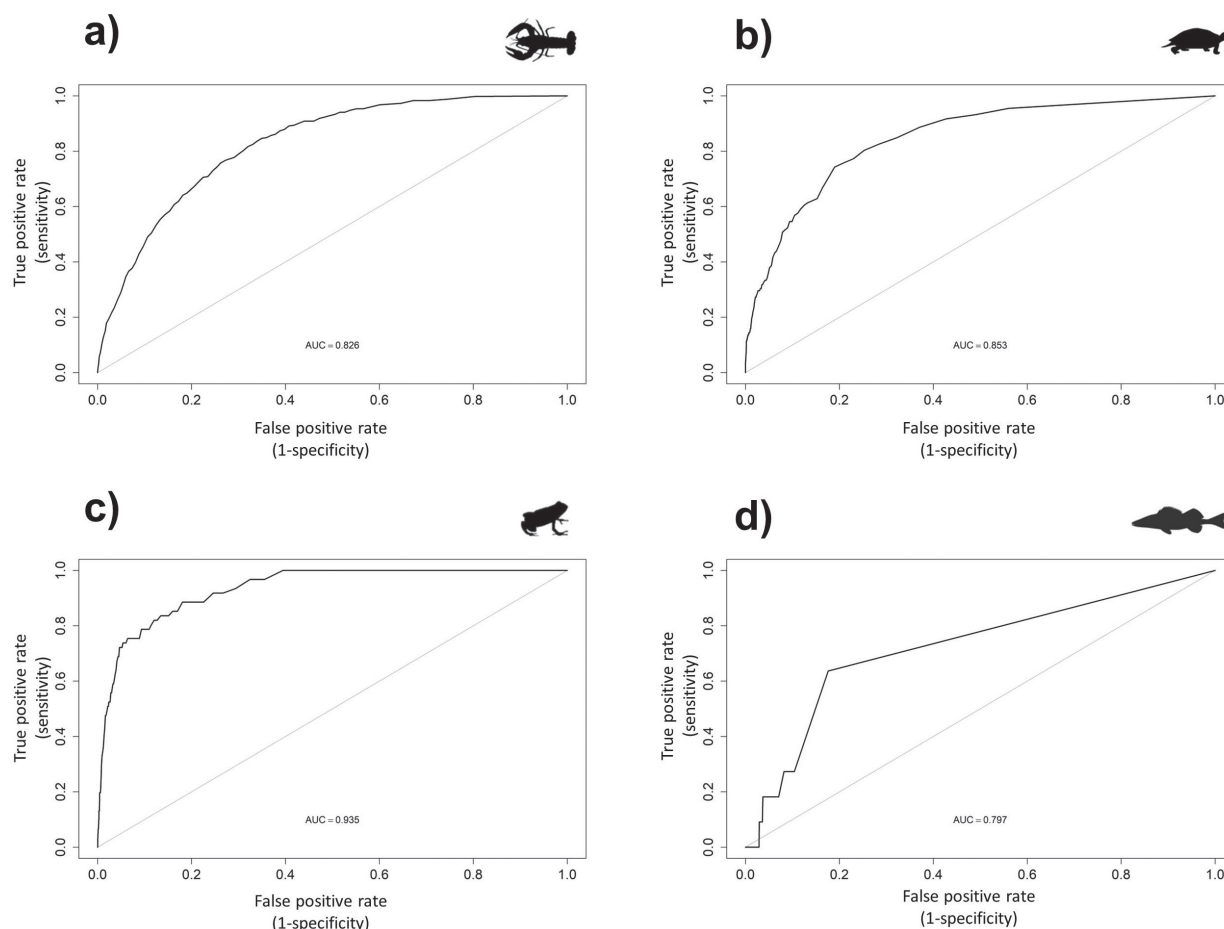


Figure 3. ROC plot of True Positive Rate (sensitivity) and False Positive Rate (1-specificity) with the AUC values calculated using the model predicted probabilities and the recent invasion of the a) signal crayfish (*Pacifastacus leniusculus*), b) red-eared slider (*Trachemys scripta*), c) marsh frog (*Pelophylax ridibundus*) and d) pike-perch (*Sander lucioperca*).

predictive ability is in agreement with Hertzog et al. (2014) that reported average differences in the AUC between sorting bias uncorrected and corrected models of 0.4, and the model with the lowest AUC but that was bias corrected, obtained the best performance after independent field data evaluation.

The distribution of alien species is affected by both climatic and anthropogenic variables. Human population density can be an indicator of non-native species propagule pressure (Copp et al. 2010), as confirmed by our results for marsh frog, pike-perch and red-eared slider occurrences, all favoured by high population density. Population density was the second most important predictor for the marsh frog distribution preceded by the distance to the introduction point, which indicates that the most populated areas not only promote their introduction but might also facilitate their establishment and expansion, for example via garden ponds (Vimercati et al. 2017), increasing the connectivity between populations in the absence of streams (Atobe et al. 2014). Population density as a predictor should be investigated in detail for management purposes, to identify the main points of entry or activities, as per se does not provide specific information on the activities or places that need to be regulated. Population density and the Human Footprint Index seem to shape the spatial distribution of biodiversity and also the presence of invasive species (Di Marco et al. 2013; Aronson et al. 2014; Gallardo et al. 2015; Dimitrakopoulos et al. 2017) which highlights the need to identify the specific vectors of secondary spread, such as flows of people, cars, trains or boats (Gilbert et al. 2004;

Rodríguez-Rey et al. 2021) or defining spatial proxies of human activities promoting species invasion (Chapman et al. 2020). This can be achieved by prioritising risk assessment approaches that rely less on invasion history (e.g., historical sources of species) and focus more on environmental change and the consequences of expanding trade routes or aquatic activities, that increase source pools of alien species (Seebens et al. 2018). The release of animals to the wild is more frequent in urban than in rural areas (Olden et al. 2005), a good example being the red-slider, which was in our study particularly abundant close to cities with more than 50k inhabitants. Yet, although urban areas host many alien species, and act as foci from which they can spread and invade surrounding natural areas (McLean et al. 2017), most of the distribution patterns of the species in our study were unrelated to distance to cities.

Road density and proximity to garden centres favoured the spread of the red-slider suggesting that this species was introduced deliberately as pets (Copp et al. 2005) in the areas most accessible to people (Hulme 2009). Distance to garden centres has previously been found to promote the introduction of imported turtles, which increases propagule pressure (Duggan et al. 2006; Copp et al. 2010). Another species potentially favoured by road density was the signal crayfish, whose presence was also driven by climatic conditions. Its ability to adapt to multiple climatic conditions has allowed this species to adapt to areas with higher temperature and precipitation seasonality, as well as to large rivers at low altitudes, which tends to facilitate the invasion of freshwater species (Gebauer et al. 2018). However, altitude might limit its distribution due to high flows associated with steep gradients in headwaters. This species tends to be located far from ports pointing to the species' tolerance to water salinity. Although ports are also located in estuarine waters, signal crayfish can only survive with salinity conditions below 14 PSU (Holdich et al. 1997) rarely common in British estuaries (van der Meer et al. 2016; Ownsworth et al. 2019).

Precipitation seasonality had a negative effect on pike-perch's presence, which indicates that this species is unable to cope with variable flow regimens. In fact, for this species, seasonal weather patterns have been reported to affect spawning success, egg survival and post-hatching survival (Kanno et al. 2016) and changes in flow reduce fish recruitment (Dutterer et al. 2013). High precipitation in the warmest months benefited pike-perch's presence which might be explained by the big size of the species, originally from large streams, that might be more affected than smaller fish by decreases in flow and habitat availability through an increase in competition (Rolls et al. 2012). Also, pike-perch prefers warm water which is why increasing temperature in the wettest months increased the presence of the species (Keskinen 2008). The influence of environmental drivers in the distribution of pike-perch is also reflected by its presence in less populated areas and far from garden centres, which might reflect a preference for more natural areas.

Most of the alien species we studied were more likely to be found closer to the places where they were first introduced, especially the marsh frog. This might be due to their limited ability to disperse due to climatic reasons. In this sense, further analysis is required to account for the type of dispersal followed by the different taxa. According to the importance of this predictor (i.e., "distance to the first records"), implementing management on areas surrounding the locations of the first introduction resulted highly important for all study species and supports the fact that effective early detection will facilitate the first management steps (Muha et al. 2017). Monitoring alien species over space and time provides data to develop time independent evaluations of the species' models, which is one of the biggest limitations in SDM (Araújo et al. 2005) due to the dependency between training and testing data, that leads to prediction errors (Roberts et al. 2017; Liu et al. 2020). The modelling procedure implemented in this study used the spatial records of

the species at earlier stages of invasion to calibrate the models and recent invasion records to evaluate them, can be applied with promising results (Barbet-Massin et al. 2018). In addition, we further conducted another evaluation using the range expansion in the last six years to realistically evaluate the predictive capacity of our models with independent data which constitutes a novel step in SDM for alien species. For all the species with a range expansion, the models were accurate in terms of AUC and correctly predicted the newly occupied localities, therefore, highlighting the usefulness of this approach.

Considering the most important anthropogenic predictors driving each species expansion together with the predictive risk maps (i.e., presented suitability maps, which help to identify those localities most vulnerable to be invaded next) can guide decision making to allocate resources and prioritising management actions to prevent the arrival of the species to new locations, especially to those places of conservation importance, with endangered species or in protected areas (Johovic et al. 2020), with biosecurity risks (Gallardo et al. 2022) or where high economic impacts are expected (Haubrock et al. 2021). Predicting species distributions to prioritise management is fundamental (Guillera-Arroita et al. 2015; Wood et al. 2018) and non-native species monitoring in many regions promote the availability of records for a convenient evaluation using SDM. Future steps should aim for detecting more specific human activities at spatial scales and with a higher impact to expand alien species.

Funding

This study was funded by the call “Ayudas para la realización de proyectos de investigación de la Universidad de Alcalá” (PIUAH22/CC-054) to M.R-R, and Marie Skłodowska-Curie ITN (AQUAINVAD-ED; Grant no. 642197) to S.C. M.R-R was supported by “Maria Zambrano” fellowship from the Spanish Ministry of Universities and Next Generation-EU.

Author contributions

M.R-R. and C.G.L. conceived the idea. M.R-R. compiled and curated the data and M.R-R. and C.G.L. analysed the data. All authors interpreted the outputs, contributed to manuscript writing, gave approval for publication and agree to be accountable for any question related to this work.

Availability of data and material and Code availability

All data used is derived from public domain resources. A reference script is available as supporting information.

Acknowledgements

We are grateful to the anonymous reviewers for their constructive and helpful comments on this manuscript. This paper was a contribution to the RIP4FISH (PIUAH22/CC-054) project financed by the University of Alcalá.

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: 1223–1232. <https://doi.org/10.1111/j.1365-2664.2006.01214.x>

- Araújo MB, Pearson RG, Thuiller W, Erhard M (2005) Validation of species–climate impact models under climate change. *Global Change Biology* 11: 1504–1513. <https://doi.org/10.1111/j.1365-2486.2005.01000.x>
- Aronson MF, La Sorte FA, Nilon CH, Katti M, Goddard MA, Lepczyk CA, Warren PS, Williams NS, Cilliers S, Clarkson B (2014) A global analysis of the impacts of urbanization on bird and plant diversity reveals key anthropogenic drivers. *Proceedings of the Royal Society B: Biological Sciences* 281: 20133330. <https://doi.org/10.1098/rspb.2013.3330>
- Atobe T, Osada Y, Takeda H, Kuroe M, Miyashita T (2014) Habitat connectivity and resident shared predators determine the impact of invasive bullfrogs on native frogs in farm ponds. *Proceedings of the Royal Society B: Biological Sciences* 281: 20132621. <https://doi.org/10.1098/rspb.2013.2621>
- Barbet-Massin M, Jiguet F, Albert CH, Thuiller W (2012) Selecting pseudo-absences for species distribution models: how, where and how many? *Methods in Ecology and Evolution* 3: 327–338. <https://doi.org/10.1111/j.2041-210X.2011.00172.x>
- Barbet-Massin M, Rome Q, Villemant C, Courchamp F (2018) Can species distribution models really predict the expansion of invasive species? *PLoS ONE* 13: e0193085. <https://doi.org/10.1371/journal.pone.0193085>
- Barbosa A, Brown J, Jimenez-Valverde A, Real R, Barbosa MAM (2021) Package ‘modEvA’.
- Bellard C, Thuiller W, Leroy B, Genovesi P, Bakkenes M, Courchamp F (2013) Will climate change promote future invasions? *Global Change Biology* 19: 3740–3748. <https://doi.org/10.1111/gcb.12344>
- Bomford M, Kraus F, Barry SC, Lawrence E (2008) Predicting establishment success for alien reptiles and amphibians: a role for climate matching. *Biological Invasions* 11: 713. <https://doi.org/10.1007/s10530-008-9285-3>
- Britton JR, Cucherousset J, Davies GD, Godard MJ, Copp GH (2010) Non-native fishes and climate change: predicting species responses to warming temperatures in a temperate region. *Freshwater Biology* 55: 1130–1141. <https://doi.org/10.1111/j.1365-2427.2010.02396.x>
- Cardona L, Hereu B, Torras X (2008) Juvenile bottlenecks and salinity shape grey mullet assemblages in Mediterranean estuaries. *Estuarine, Coastal and Shelf Science* 77(4): 623–632. <https://doi.org/10.1016/j.ecss.2007.10.018>
- Chapman DS, Gunn ID, Pringle HE, Siriwardena GM, Taylor P, Thackeray SJ, Willby NJ, Carvalho L (2020) Invasion of freshwater ecosystems is promoted by network connectivity to hotspots of human activity. *Global Ecology and Biogeography* 29(4): 645–655. <https://doi.org/10.1111/geb.13051>
- Copp G, Wesley K, Vilizzi L (2005) Pathways of ornamental and aquarium fish introductions into urban ponds of Epping Forest (London, England): the human vector. *Journal of Applied Ichthyology* 21: 263–274. <https://doi.org/10.1111/j.1439-0426.2005.00673.x>
- Copp G, Stakénas S, Davison P (2006) The incidence of non-native fishes in water courses: example of the United Kingdom. *Aquatic Invasions* 1(2): 72–75. <https://doi.org/10.3391/ai.2006.1.2.3>
- Copp G, Vilizzi L, Gozlan R (2010) The demography of introduction pathways, propagule pressure and occurrences of non-native freshwater fish in England. *Aquatic Conservation: Marine and Freshwater Ecosystems* 20: 595–601. <https://doi.org/10.1002/aqc.1129>
- Cuthbert RN, Pattison Z, Taylor NG, Verbrugge L, Diagne C, Ahmed DA, Leroy B, Angulo E, Briski E, Capinha C, Catford JA, Dalu T, Essl F, Gozlan RE, Haubrock PJ, Kourantidou M, Kramer AM, Renault D, Wasserman RJ, Courchamp F (2021) Global economic costs of aquatic invasive alien species. *Science of The Total Environment* 775: 145238. <https://doi.org/10.1016/j.scitotenv.2021.145238>
- Cutler DR, Edwards TC, Beard KH, Cutler A, Hess KT, Gibson J, Lawler JJ (2007) Random Forest for Classification in Ecology. *Ecology* 88: 2783–2792. <https://doi.org/10.1890/07-0539.1>
- Darwall WRT, Holland RA, Smith KG, Allen D, Brooks EGE, Katarya V, Pollock CM, Shi Y, Clausnitzer V, Cumberlidge N, Cuttelod A, Dijkstra K-DB, Diop MD, García N, Seddon MB, Skelton PH, Snoeks J, Tweddle D, Vié JC (2011) Implications of bias in conservation research and investment for freshwater species. *Conservation Letters* 4: 474–482. <https://doi.org/10.1111/j.1755-263X.2011.00202.x>

- Datta A, Schweiger O, Kühn I (2020) Origin of climatic data can determine the transferability of species distribution models. *NeoBiota* 59: 61–76. <https://doi.org/10.3897/neobiota.59.36299>
- Di Marco M, Rondinini C, Boitani L, Murray KA (2013) Comparing multiple species distribution proxies and different quantifications of the human footprint map, implications for conservation. *Biological Conservation* 165: 203–211. <https://doi.org/10.1016/j.biocon.2013.05.030>
- Dimitrakopoulos PG, Koukoulas S, Galanidis A, Delipetrou P, Gounaridis D, Touloumi K, Ariannoutsou M (2017) Factors shaping alien plant species richness spatial patterns across Natura 2000 Special Areas of Conservation of Greece. *Science of The Total Environment* 601–602: 461–468. <https://doi.org/10.1016/j.scitotenv.2017.05.220>
- Dormann CF, Elith J, Bacher S, Buchmann C, Carl G, Carré G, Marquéz JRG, Gruber B, Lafourcade B, Leitão PJ, Münkemüller T, McClean C, Osborne PE, Reineking B, Schröder B, Skidmore AK, Zurell D, Lautenbach S (2013) Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography* 36: 27–46. <https://doi.org/10.1111/j.1600-0587.2012.07348.x>
- Duggan IC, Rixon CAM, MacIsaac HJ (2006) Popularity and Propagule Pressure: Determinants of Introduction and Establishment of Aquarium Fish. *Biological Invasions* 8: 377–382. <https://doi.org/10.1007/s10530-004-2310-2>
- Dutterer A, Mesing C, Cailteux R, Allen M, Pine W, Strickland P (2013) Fish recruitment is influenced by river flows and floodplain inundation at Apalachicola River, Florida. *River Research and Applications* 29: 1110–1118. <https://doi.org/10.1002/rra.2604>
- Faurby S, Araújo MB (2018) Anthropogenic range contractions bias species climate change forecasts. *Nature Climate Change* 8: 252. <https://doi.org/10.1038/s41558-018-0089-x>
- Fielding AH, Bell JF (1997) A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation* 24: 38–49. <https://doi.org/10.1017/S0376892997000088>
- Fletcher D, Gillingham P, Britton J, Blanchet S, Gozlan RE (2016) Predicting global invasion risks: a management tool to prevent future introductions. *Scientific Reports* 6: srep26316. <https://doi.org/10.1038/srep26316>
- Fourcade Y, Engler JO, Rödder D, Secondi J (2014) Mapping species distributions with MAXENT using a geographically biased sample of presence data: A performance assessment of methods for correcting sampling bias. *PLoS ONE* 9: e97122. <https://doi.org/10.1371/journal.pone.0097122>
- Freeman MS, Dick JT, Reid N (2022) Dealing with non-equilibrium bias and survey effort in presence-only invasive Species Distribution Models (iSDM); predicting the range of muntjac deer in Britain and Ireland. *Ecological Informatics* 69: 101683. <https://doi.org/10.1016/j.ecoinf.2022.101683>
- Gallardo B, Aldridge DC (2013) The ‘dirty dozen’: socio-economic factors amplify the invasion potential of 12 high-risk aquatic invasive species in Great Britain and Ireland. *Journal of Applied Ecology* 50: 757–766. <https://doi.org/10.1111/1365-2664.12079>
- Gallardo B, Aldridge DC (2015) Is Great Britain heading for a Ponto–Caspian invasional meltdown? *Journal of Applied Ecology* 52: 41–49. <https://doi.org/10.1111/1365-2664.12348>
- Gallardo B, Clavero M, Sánchez M, Vilà M (2016) Global ecological impacts of invasive species in aquatic ecosystems. *Global Change Biology* 22: 151–163. <https://doi.org/10.1111/gcb.13004>
- Gallardo B, Sutherland WJ, Martin P, Aldridge DC (2022) Applying fault tree analysis to biological invasions identifies optimal targets for effective biosecurity. *Journal of Applied Ecology* 59: 2553–2566. <https://doi.org/10.1111/1365-2664.14256>
- Gallardo B, Zieritz A, Aldridge DC (2015) The importance of the human footprint in shaping the global distribution of terrestrial, freshwater and marine invaders. *PLoS ONE* 10: e0125801. <https://doi.org/10.1371/journal.pone.0125801>
- Gebauer R, Divišek J, Buřič M, Večeřa M, Kouba A, Drozd B (2018) Distribution of alien animal species richness in the Czech Republic. *Ecology and Evolution* 8: 4455–4464. <https://doi.org/10.1002/ece3.4008>

- Gilbert M, Grégoire JC, Freise JF, Heitland W (2004) Long-distance dispersal and human population density allow the prediction of invasive patterns in the horse chestnut leafminer *Cameraria ohridella*. *Journal of Animal Ecology* 73: 459–468. <https://doi.org/10.1111/j.0021-8790.2004.00820.x>
- Glen AS, Pech RP, Byrom AE (2013) Connectivity and invasive species management: towards an integrated landscape approach. *Biological Invasions* 15: 2127–2138. <https://doi.org/10.1007/s10530-013-0439-6>
- Guillera-Arroita G, Lahoz-Monfort J, Elith J, Gordon A, Kujala H, Lentini P, McCarthy M, Tingley R, Wintle B (2015) Is my species distribution model fit for purpose? Matching data and models to applications. *Global Ecology and Biogeography* 24: 276–292. <https://doi.org/10.1111/geb.12268>
- Guisan A, Tingley R, Baumgartner JB, Naujokaitis-Lewis I, Sutcliffe PR, Tulloch AIT, Regan TJ, Brotons L, McDonald-Madden E, Mantyka-Pringle C, Martin TG, Rhodes JR, Maggini R, Setterfield SA, Elith J, Schwartz MW, Wintle BA, Broennimann O, Austin M, Ferrier S, Kearney MR, Possingham HP, Buckley YM (2013) Predicting species distributions for conservation decisions. *Ecology Letters* 16: 1424–1435. <https://doi.org/10.1111/ele.12189>
- Hastie TJ, Tibshirani RJ (1990) Generalized additive models. Chapman and Hall/CRC Press, 335 pp.
- Haubrock PJ, Turbelin AJ, Cuthbert RN, Novoa A, Taylor NG, Angulo E, Ballesteros-Mejia L, Bodey TW, Capinha C, Diagne C (2021) Economic costs of invasive alien species across Europe. *NeoBiota* 67: 153–190. <https://doi.org/10.3897/neobiota.67.58196>
- Helmstetter NA, Conway CJ, Stevens BS, Goldberg AR (2021) Balancing transferability and complexity of species distribution models for rare species conservation. *Diversity and Distributions* 27: 95–108. <https://doi.org/10.1111/ddi.13174>
- Hertzog LR, Besnard A, Jay-Robert P (2014) Field validation shows bias-corrected pseudo-absence selection is the best method for predictive species-distribution modelling. *Diversity and Distributions* 20: 1403–1413. <https://doi.org/10.1111/ddi.12249>
- Hijmans RJ (2012) Cross-validation of species distribution models: removing spatial sorting bias and calibration with a null model. *Ecology* 93: 679–688. <https://doi.org/10.1890/11-0826.1>
- Hijmans RJ, Phillips S, Leathwick J, Elith J (2017) dismo: Species distribution modeling. R package version 1.
- Holdich D, Harlioglu M, Firkins I (1997) Salinity adaptations of crayfish in British waters with particular reference to *Austropotamobius pallipes*, *Astacus leptodactylus* and *Pacifastacus leniusculus*. *Estuarine, Coastal and Shelf Science* 44: 147–154. <https://doi.org/10.1006/ecss.1996.0206>
- Hosmer DW, Lemeshow S (2000) Special topics. *Applied Logistic Regression*, 2nd edn., 260–351. <https://doi.org/10.1002/0471722146>
- Huang J, Tang S, Cai F, Lin Y, Wu Z (2017) Microsatellite evidence of dispersal mechanism of red swamp crayfish (*Procambarus clarkii*) in the Pearl River basin and implications for its management. *Scientific Reports* 7: 8272. <https://doi.org/10.1038/s41598-017-08552-3>
- Hulme PE (2009) Trade, transport and trouble: managing invasive species pathways in an era of globalization. *Journal of Applied Ecology* 46: 10–18. <https://doi.org/10.1111/j.1365-2664.2008.01600.x>
- Illanas S, Croft S, Acevedo P, Fernández-López J, Vicente J, Blanco-Aguilar JA, Pascual-Rico R, Scandura M, Apollonio M (2022) Update of model for wild ruminant abundance based on occurrence and first models based on hunting yield at European scale. *EFSA Supporting Publications* 19: 7174E. <https://doi.org/10.2903/sp.efsa.2022.EN-7174>
- 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: 498–507. <https://doi.org/10.1111/j.1466-8238.2011.00683.x>
- Jiménez-Valverde A, Rodríguez-Rey M, Peña-Aguilera P (2021) Climate data source matters in species distribution modelling: the case of the Iberian Peninsula. *Biodiversity and Conservation* 30: 67–84. <https://doi.org/10.1007/s10531-020-02075-6>
- Johnson LE, Ricciardi A, Carlton JT (2001) Overland dispersal of aquatic invasive species: a risk assessment of transient recreational boating. *Ecological Applications* 11: 1789–1799. [https://doi.org/10.1890/1051-0761\(2001\)011\[1789:ODOAIS\]2.0.CO;2](https://doi.org/10.1890/1051-0761(2001)011[1789:ODOAIS]2.0.CO;2)

- Johovic I, Gama M, Banha F, Tricarico E, Anastácio PM (2020) A potential threat to amphibians in the European Natura 2000 network: Forecasting the distribution of the American bullfrog *Lithobates catesbeianus*. *Biological Conservation* 245: 108551. <https://doi.org/10.1016/j.biocon.2020.108551>
- Kanno Y, Pregler KC, Hitt NP, Letcher BH, Hocking DJ, Wofford JE (2016) Seasonal temperature and precipitation regulate brook trout young-of-the-year abundance and population dynamics. *Freshwater Biology* 61: 88–99. <https://doi.org/10.1111/fwb.12682>
- Keskinen T (2008) Feeding ecology and behaviour of pikeperch, *Sander lucioperca* (L.) in boreal lakes. *Jyväskylä Studies in Biological and Environmental Science* 190: 30–38.
- Kuhn M, Wing J, Weston S, Williams A, Keefer C, Engelhardt A, Cooper T, Mayer Z, Kenkel B, Team RC (2020) Package ‘caret’. *The R Journal* 223: 7.
- Liaw A, Wiener M (2002) Classification and Regression by randomForest. *R News* 2 (3): 18–22.
- Lima VP, de Lima RAF, Joner F, Siddique I, Raes N, ter Steege H (2022) Climate change threatens native potential agroforestry plant species in Brazil. *Scientific Reports* 12: 2267. <https://doi.org/10.1038/s41598-022-06234-3>
- Liu C, Wolter C, Xian W, Jeschke JM (2020) Species distribution models have limited spatial transferability for invasive species. *Ecology Letters* 23: 1682–1692. <https://doi.org/10.1111/ele.13577>
- Lobo JM, Jiménez-Valverde A, Real R (2008) AUC: a misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography* 17: 145–151. <https://doi.org/10.1111/j.1466-8238.2007.00358.x>
- Mainali KP, Warren DL, Dhileepan K, McConnachie A, Strathie L, Hassan G, Karki D, Shrestha BB, Parmesan C (2015) Projecting future expansion of invasive species: comparing and improving methodologies for species distribution modeling. *Global Change Biology* 21: 4464–4480. <https://doi.org/10.1111/gcb.13038>
- Mallez S, McCartney M (2018) Dispersal mechanisms for zebra mussels: population genetics supports clustered invasions over spread from hub lakes in Minnesota. *Biological Invasions* 20: 2461–2484. <https://doi.org/10.1007/s10530-018-1714-3>
- Marchi M, Sinjur I, Bozzano M, Westergren M (2019) Evaluating WorldClim Version 1 (1961–1990) as the baseline for sustainable use of forest and environmental resources in a changing climate. *Sustainability* 11: 3043. <https://doi.org/10.3390/su11113043>
- Marmion M, Parviainen M, Luoto M, Heikkinen RK, Thuiller W (2009) Evaluation of consensus methods in predictive species distribution modelling. *Diversity and Distributions* 15: 59–69. <https://doi.org/10.1111/j.1472-4642.2008.00491.x>
- Mazza G, Tricarico E (2018) *Invasive species and human health*. CABI, Wallingford, 210 pp. <https://doi.org/10.1079/9781786390981.0000>
- McLean P, Gallien L, Wilson JR, Gaertner M, Richardson DM (2017) Small urban centres as launching sites for plant invasions in natural areas: insights from South Africa. *Biological Invasions* 19: 3541–3555. <https://doi.org/10.1007/s10530-017-1600-4>
- Merow C, Smith MJ, Silander JA (2013) A practical guide to MaxEnt for modeling species’ distributions: what it does, and why inputs and settings matter. *Ecography* 36: 1058–1069. <https://doi.org/10.1111/j.1600-0587.2013.07872.x>
- Moustakas A, Voutsela A, Katsanevakis S (2018) Sampling alien species inside and outside protected areas: Does it matter? *Science of The Total Environment* 625: 194–198. <https://doi.org/10.1016/j.scitotenv.2017.12.198>
- Muha TP, Rodríguez-Rey M, Rolla M, Tricarico E (2017) Using Environmental DNA to Improve Species Distribution Models for Freshwater Invaders. *Frontiers in Ecology and Evolution* 5: 158. <https://doi.org/10.3389/fevo.2017.00158>
- Olden JD, Douglas ME, Douglas MR (2005) The human dimensions of biotic homogenization. *Conservation Biology* 19: 2036–2038. <https://doi.org/10.1111/j.1523-1739.2005.00288.x>
- Osborne OG, Fell HG, Atkins H, van Tol J, Phillips D, Herrera-Alsina L, Mynard P, Bodedi G, Gubry-Rangin C, Lancaster LT, Creer S, Nangoy M, Fahri F, Lupiyaningdyah P, Sudiana IM, Juliandi B, Travis JMJ, Papadopoulos AST, Algar AC (2022) Fauxcurrence: simulating multi-species

- occurrences for null models in species distribution modelling and biogeography. *Ecography* 2022: e05880. <https://doi.org/10.1111/ecog.05880>
- Owensworth E, Selby D, Ottley CJ, Unsworth E, Raab A, Feldmann J, Sproson AD, Kuroda J, Faidutti C, Bucker P (2019) Tracing the natural and anthropogenic influence on the trace elemental chemistry of estuarine macroalgae and the implications for human consumption. *Science of The Total Environment* 685: 259–272. <https://doi.org/10.1016/j.scitotenv.2019.05.263>
- Padilla DK, Williams SL (2004) Beyond ballast water: aquarium and ornamental trades as sources of invasive species in aquatic ecosystems. *Frontiers in Ecology and the Environment* 2: 131–138. [https://doi.org/10.1890/1540-9295\(2004\)002\[0131:BBWAAO\]2.0.CO;2](https://doi.org/10.1890/1540-9295(2004)002[0131:BBWAAO]2.0.CO;2)
- Phillips SJ (2017) A brief tutorial on Maxent. http://biodiversityinformatics.amnh.org/open_source/maxent/
- Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species geographic distributions. *Ecological Modelling* 190: 231–259. <https://doi.org/10.1016/j.ecolmodel.2005.03.026>
- Phillips SJ, Dudík M, Elith J, Graham CH, Lehmann A, Leathwick J, Ferrier S (2009) Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecological Applications* 19: 181–197. <https://doi.org/10.1890/07-2153.1>
- Poggio L, Simonetti E, Gimona A (2018) Enhancing the WorldClim data set for national and regional applications. *Science of The Total Environment* 625: 1628–1643. <https://doi.org/10.1016/j.scitotenv.2017.12.258>
- QGIS Development Team (2016) QGIS Geographic Information System. Open Source Geospatial Foundation Project. <http://qgis.osgeo.org>
- R Development Core Team (2018) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org>
- Raes N, ter Steege H (2007) A null-model for significance testing of presence-only species distribution models. *Ecography* 30: 727–736. <https://doi.org/10.1111/j.2007.0906-7590.05041.x>
- Rahel FJ, Olden JD (2008) Assessing the Effects of Climate Change on Aquatic Invasive Species. *Conservation Biology* 22: 521–533. <https://doi.org/10.1111/j.1523-1739.2008.00950.x>
- Roberts DR, Bahn V, Ciuti S, Boyce MS, Elith J, Guillera-Arroita G, Hauenstein S, Lahoz-Monfort JJ, Schröder B, Thuiller W (2017) Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography* 40: 913–929. <https://doi.org/10.1111/ecog.02881>
- Robertson PA, Mill A, Novoa A, Jeschke JM, Essl F, Gallardo B, Geist J, Jarić I, Lambin X, Musseau C, Pergl J, Pyšek P, Rabitsch W, von Schmalensee M, Shirley M, Strayer DL, Stefansson RA, Smith K, Booy O (2020) A proposed unified framework to describe the management of biological invasions. *Biological Invasions* 22: 2633–2645. <https://doi.org/10.1007/s10530-020-02298-2>
- Rodrigues JFM, Coelho MTP, Varela S, Diniz-Filho JAF (2016) Invasion risk of the pond slider turtle is underestimated when niche expansion occurs. *Freshwater Biology* 61: 1119–1127. <https://doi.org/10.1111/fwb.12772>
- Rodríguez-Rey M, Consuegra S, Börger L, de Leaniz CG (2019) Improving Species Distribution Modelling of freshwater invasive species for management applications. *PLoS ONE* 14: e0217896. <https://doi.org/10.1371/journal.pone.0217896>
- Rodríguez-Rey M, Consuegra S, Börger L, Garcia de Leaniz C (2021) Boat ramps facilitate the dispersal of the highly invasive zebra mussel (*Dreissena polymorpha*). *Biological Invasions* 23: 1487–1496. <https://doi.org/10.1007/s10530-020-02453-9>
- Rodríguez-Rey M, Jiménez-Valverde A, Acevedo P (2013) Species distribution models predict range expansion better than chance but not better than a simple dispersal model. *Ecological Modelling* 256: 1–5. <https://doi.org/10.1016/j.ecolmodel.2013.01.024>
- Rolls RJ, Leigh C, Sheldon F (2012) Mechanistic effects of low-flow hydrology on riverine ecosystems: ecological principles and consequences of alteration. *Freshwater Science* 31: 1163–1186. <https://doi.org/10.1899/12-002.1>
- Sala OE, Stuart Chapin F, III, Armesto JJ, Berlow E, Bloomfield J, Dirzo R, Huber-Sanwald E, Huenneke LF, Jackson RB, Kinzig A, Leemans R, Lodge DM, Mooney HA, Oesterheld Mn,

- Poff NL, Sykes MT, Walker BH, Walker M, Wall DH (2000) Global Biodiversity Scenarios for the Year 2100. *Science* 287(5459): 1770–1774. <https://doi.org/10.1126/science.287.5459.1770>
- Schatz AM, Kramer AM, Drake JM (2017) Accuracy of climate-based forecasts of pathogen spread. *Royal Society Open Science* 4 (3): 160975. <https://doi.org/10.1098/rsos.160975>
- Seebens H, Blackburn TM, Dyer EE, Genovesi P, Hulme PE, Jeschke JM, Pagad S, Pyšek P, van Kleunen M, Winter M, Ansong M, Arianoutsou M, Bacher S, Blasius B, Brockerhoff EG, Brundu G, Capinha C, Causton CE, Celesti-Grapow L, Dawson W, Dullinger S, Economo EP, Fuentes N, Guénard B, Jäger H, Kartesz J, Kenis M, Kühn I, Lenzner B, Liebhold AM, Mosena A, Moser D, Nentwig W, Nishino M, Pearman D, Pergl J, Rabitsch W, Rojas-Sandoval J, Roques A, Rorke S, Rossinelli S, Roy HE, Scalera R, Schindler S, Štajerová K, Tokarska-Guzik B, Walker K, Ward DE, Yamanaka T, Essl F (2018) Global rise in emerging alien species results from increased accessibility of new source pools. *Proceedings of the National Academy of Sciences* 115(10): E2264–E2273. <https://doi.org/10.1073/pnas.1719429115>
- Signorile AL, Reuman DC, Lurz PWW, Bertolino S, Carbone C, Wang J (2016) Using DNA profiling to investigate human-mediated translocations of an invasive species. *Biological Conservation* 195: 97–105. <https://doi.org/10.1016/j.biocon.2015.12.026>
- Sillero N, Barbosa AM (2021) Common mistakes in ecological niche models. *International Journal of Geographical Information Science* 35: 213–226. <https://doi.org/10.1080/13658816.2020.1798968>
- Sousa R, Novais A, Costa R, Strayer DL (2014) Invasive bivalves in fresh waters: impacts from individuals to ecosystems and possible control strategies. *Hydrobiologia* 735: 233–251. <https://doi.org/10.1007/s10750-012-1409-1>
- Thuiller W, Georges D, Engler R, Breiner F, Georges MD, Thuiller CW (2016) Package ‘biomod2’.
- Thuiller W, Lafourcade B, Engler R, Araújo MB (2009) BIOMOD – a platform for ensemble forecasting of species distributions. *Ecography* 32: 369–373. <https://doi.org/10.1111/j.1600-0587.2008.05742.x>
- Tickner D, Opperman JJ, Abell R, Acreman M, Arthington AH, Bunn SE, Cooke SJ, Dalton J, Darwall W, Edwards G (2020) Bending the curve of global freshwater biodiversity loss: an emergency recovery plan. *BioScience* 70: 330–342. <https://doi.org/10.1093/biosci/biaa002>
- Tricarico E, Borrell Pichs YJ, García Vázquez E, Rico Ordás JM, Rech S (2017) Developing innovative methods to face aquatic invasions in Europe: the Aquainvad-ED project. *Management of Biological Invasions* 8(3): 403–408. <https://doi.org/10.3391/mbi.2017.8.3.13>
- Václavík T, Meentemeyer RK (2009) Invasive species distribution modeling (iSDM): are absence data and dispersal constraints needed to predict actual distributions? *Ecological Modelling* 220: 3248–3258. <https://doi.org/10.1016/j.ecolmodel.2009.08.013>
- Václavík T, Meentemeyer RK (2012) Equilibrium or not? Modelling potential distribution of invasive species in different stages of invasion. *Diversity and Distributions* 18: 73–83. <https://doi.org/10.1111/j.1472-4642.2011.00854.x>
- Van der Meer R, de Boer MK, Liebich V, ten Hallers C, Veldhuis M, Ree KJCM (2016) Ballast water risk indication for the North Sea. *Coastal Management* 44(6): 547–568. <https://doi.org/10.1080/08920753.2016.1233794>
- Vander Zanden MJ, Olden JD, Sciences A (2008) A management framework for preventing the secondary spread of aquatic invasive species. *Canadian Journal of Fisheries* 65(7): 1512–1522. <https://doi.org/10.1139/F08-099>
- Veloz SD (2009) Spatially autocorrelated sampling falsely inflates measures of accuracy for presence-only niche models. *Journal of Biogeography* 36: 2290–2299. <https://doi.org/10.1111/j.1365-2699.2009.02174.x>
- Vimercati G, Davies SJ, Hui C, Measey J (2017) Does restricted access limit management of invasive urban frogs? *Biological Invasions* 19: 3659–3674. <https://doi.org/10.1007/s10530-017-1599-6>
- Warren DL, Matzke NJ, Iglesias TL (2020) Evaluating presence-only species distribution models with discrimination accuracy is uninformative for many applications. *Journal of Biogeography* 47: 167–180. <https://doi.org/10.1111/jbi.13705>

- Wood KA, Stillman RA, Hilton GM (2018) Conservation in a changing world needs predictive models. *Animal Conservation* 21: 87–88. <https://doi.org/10.1111/acv.12371>
- Wood SN (2017) Generalized additive models: an introduction with R. Chapman and Hall/CRC Press, 496 pp. <https://doi.org/10.1201/9781315370279>
- Yates KL, Bouchet PJ, Caley MJ, Mengersen K, Randin CE, Parnell S, Fielding AH, Bamford AJ, Ban S, Barbosa AM (2018) Outstanding challenges in the transferability of ecological models. *Trends in Ecology and Evolution* 33(10): 790–802. <https://doi.org/10.1016/j.tree.2018.08.001>
- Zeng Y, Low BW, Yeo DC (2016) Novel methods to select environmental variables in MaxEnt: A case study using invasive crayfish. *Ecological Modelling* 341: 5–13. <https://doi.org/10.1016/j.ecolmodel.2016.09.019>
- Zięba G, Copp G, Davies G, Stebbing P, Wesley K, Britton J (2010) Recent releases and dispersal of non-native fishes in England and Wales, with emphasis on sunbleak *Leucaspis delineatus* (Heckel, 1843). *Aquatic Invasions* 5(2): 155–161. <https://doi.org/10.3391/ai.2010.5.2.04>

Supplementary material 1

Supplementary information

Authors: Marta Rodríguez-Rey, Sofia Consuegra, Carlos Garcia de Leaniz

Data type: docx

Explanation note: **figure S1**. Diagram of the workflow for Species Distribution Modelling in this study; **figure S2.1**. Response curves for signal crayfish (*Pacifastacus leniusculus*) according to GAM; **figure S2.2**. Response curves for red-eared slider (*Trachemys scripta*) according to GAM; **figure S2.3**. Response curves for marsh frog (*Pelophylax ridibundus*) according to MaxEnt. Negative values on the predictor, although unrealistic in some cases, resulted from the function projecting into the negative and positive range of values; **figure S2.4**. Response curves for pike-perch (*Sander lucioperca*) according to GAM; **table S1**. Characteristics of the species regarding the year and locality of introduction in the study area, the reason of their introduction, the way they commonly spread and their origin and functional type classification.; **table S2**. Number of occurrences and covered years used in each modelling step for the different species; **table S3**. Predictor variables used in the Species Distribution Models. Variables in bold had VIF scores smaller than 10 (Hairs et al. 1998) and were included in the Species Distribution Models; **table S4**. Variance Inflation Factor (VIF) for those predictors with VIF<10; **table S5**. True Skill Statistic (TSS) and Area Under the Curve (AUC) results for the models computed using four algorithms and the ensemble. Effect size corresponds to the difference between the real model and the null model 95 CI maximum value, calculated for both discrimination statistics.

Copyright notice: This dataset is made available under the Open Database License (<http://opendata-commons.org/licenses/odbl/1.0/>). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: <https://doi.org/10.3391/ai.2023.18.4.111481.suppl1>