Received: 13 December 2022 | Revised: 25 April 2023 | Accepted: 20 July 2023

DOI: 10.1111/gcb.16901

Global Change Biology WILEY

RESEARCH ARTICLE

Different facets of the same niche: Integrating citizen science and scientific survey data to predict biological invasion risk under multiple global change drivers

Mirko Di Febbraro[1] | Luciano Bosso[2] | Mauro Fasola[3] | Francesca Santicchia[4] | Gaetano Aloise[5] | Simone Lioy[6] | Elena Tricarico[7,8] | Luciano Ruggieri[9] | Stefano Bovero[10] | Emiliano Mori[8,11] | Sandro Bertolino[12]

[1]Environmetrics Lab, Department of Biosciences and Territory, University of Molise, Pesche, Isernia, Italy

[2]Department of Research Infrastructures for Marine Biological Resources, Stazione Zoologica Anton Dohrn, Naples, Italy

[3]Dipartimento Scienze della Terra e dell'Ambiente, Università di Pavia, Pavia, Italy

[4]Environment Analysis and Management Unit, Guido Tosi Research Group, Department of Theoretical and Applied Sciences, Università degli Studi dell'Insubria, Varese, Italy

[5]Museo di Storia Naturale e Orto Botanico, Università della Calabria, Rende, Cosenza, Italy

[6]Department of Agricultural, Forest and Food Sciences, University of Turin, Turin, Italy

[7]Department of Biology, University of Florence, Sesto Fiorentino, Italy

[8]National Biodiversity Future Center (NBFC), Palermo, Italy

[9]EBN Italia at Pro Natura, Turin, Italy

[10]"Zirichiltaggi" Sardinia Wildlife Conservation NGO, Sassari, Italy

[11]Consiglio Nazionale delle Ricerche, Istituto di Ricerca sugli Ecosistemi Terrestri, Florence, Italy

[12]Department of Life Sciences and Systems Biology, University of Turin, Turin, Italy

Correspondence
Luciano Bosso, Department of Research Infrastructures for Marine Biological Resources, Stazione Zoologica Anton Dohrn, Villa Comunale, 80121 Naples, Italy.
Email: luciano.bosso@szn.it

Funding information
NBFC; University of Florence; CNR–IRET; Italian Ministry of University and Research, Grant/Award Number: CN00000033; University of Turin

Abstract

Citizen science initiatives have been increasingly used by researchers as a source of occurrence data to model the distribution of alien species. Since citizen science presence-only data suffer from some fundamental issues, efforts have been made to combine these data with those provided by scientifically structured surveys. Surprisingly, only a few studies proposing data integration evaluated the contribution of this process to the effective sampling of species' environmental niches and, consequently, its effect on model predictions on new time intervals. We relied on niche overlap analyses, machine learning classification algorithms and ecological niche models to compare the ability of data from citizen science and scientific surveys, along with their integration, in capturing the realized niche of 13 invasive alien species in Italy. Moreover, we assessed differences in current and future invasion risk predicted by each data set under multiple global change scenarios. We showed that data from citizen science and scientific surveys captured similar species niches though highlighting exclusive portions associated with clearly identifiable environmental conditions. In terrestrial species, citizen science data granted the highest gain in environmental space to the pooled niches, determining an increased future biological invasion risk. A few aquatic species modelled at the regional scale reported a net loss in the pooled niches compared to their scientific survey niches, suggesting that citizen science data may also lead to contraction in pooled niches. For these species, models predicted a lower future biological invasion risk. These findings indicate that citizen science data may represent a valuable contribution to predicting future spread of invasive alien species, especially within national-scale programmes. At the same time, citizen science data collected on species poorly known to citizen scientists, or in strictly local contexts, may strongly affect the niche quantification of these taxa and the prediction of their future biological invasion risk.

KEYWORDS
alien species, biological invasions, citizen science, data science, ecological niche models, global change

1 | INTRODUCTION

Reducing impacts of invasive alien species on biodiversity and human well-being is a key target for the post-2020 Global Biodiversity Framework under the Convention on Biological Diversity. To reach this goal, it is pivotal to effectively manage the most significant pathways of introduction, regulate the most harmful alien species and reduce the impacts in the most vulnerable areas. The regulation of deleterious species trade and invasion pathways is required, for instance, by the European Union regulation on invasive alien species (Regulation EU 1143/2014; Tollington et al., 2017). Considering the growing number of introduced species worldwide (Seebens et al., 2017), it is essential to identify alien species that could establish and spread over large areas, producing negative impacts (Finnoff et al., 2007). This involves a prioritization process that can be done through horizon-scanning (Bertolino, Cerri, et al., 2020; Bertolino, Sciandra, et al., 2020; Roy et al., 2014) and risk assessments (Roy et al., 2018). Expert-based evaluation, required for alien species screening activities, could improve the availability of present and potential future ranges of target species and of information on their ecological niche. Ecological niche models (hereafter, ENMs) are widely used to obtain range maps, forecast future spread and depict the niche of alien species. These models combine occurrence data with geographic layers of environmental information to predict distributions across landscapes, with extrapolation in space and time (Elith et al., 2011). Therefore, the quality of models' output depends primarily on the availability and reliability of occurrence and environmental data.

Citizen science is the collection of data of scientific interest (e.g. species occurrences) by the general public as part of collaborative projects with professional scientists, who validate and elaborate the data (Wiggins & Crowston, 2011). This method proved effective and economical for gathering species occurrences where scientifically designed surveys would require time-consuming field efforts (Johnson et al., 2020), with successful applications also on invasive alien species (Grez et al., 2022; Maistrello et al., 2016; Werenkraut et al., 2020). In recent years, citizen science initiatives have increasingly been used by researchers as a source of occurrence data to model the distribution of species in both native (Arenas-Castro et al., 2022; Milanesi et al., 2020; Park et al., 2022; Stuber et al., 2022) and invaded ranges (Di Febbraro et al., 2019; Giuntini et al., 2022; Tran et al., 2022). The availability of abundant and widely distributed occurrence records represented a key characteristic behind the growing importance of citizen science initiatives in modelling studies, especially as providers of presence-only data (Fletcher et al., 2019; Johnston et al., 2022). Since presence-only data gathered from opportunistic surveys such as citizen science initiatives suffer from some key issues (e.g. sampling bias, imperfect detection, unavailability of absences, etc.), efforts have been made to combine these data with those provided by scientifically planned, structured surveys (Fletcher et al., 2019; Miller et al., 2019). Among the plethora of different data integration methods proposed (e.g. Broennimann & Guisan, 2008; Gallien et al., 2012; Robinson et al., 2020), the most recent approaches focused on hierarchically combining big, unstructured citizen science data sets with small, scientifically structured data sets, often generated at different spatial resolution, extent and data type (i.e. presence-only and presence/absence; Ahmad Suhaimi et al., 2021; Chevalier et al., 2021; Fletcher et al., 2019; Grabow et al., 2022; Ovaskainen et al., 2016; Stuber et al., 2022). Most of these studies explicitly compared the predictive performance of traditional versus integrated distribution models (Chevalier et al., 2021; Robinson et al., 2020; Zulian et al., 2021), without providing conclusive evidence to support the use of the latter (Ahmad Suhaimi et al., 2021; Simmonds et al., 2020). Only a small number of the most recent studies proposing data integration evaluated the contribution of this process to the effective sampling of species environmental niche and, consequently, its effect on model predictions on new areas/time intervals (see Chevalier et al., 2021; Scherrer et al., 2021).

The often incomplete coverage of species' environmental preferences inherent to small-scale, scientifically structured surveys introduces an issue of niche truncation that might affect predictions on novel environmental conditions (e.g. future climates) included in the species niche but uncaptured in the survey phase due to its limited extent (Chevalier et al., 2021). Data integration methods represent a promising approach for dealing with niche truncation problems, making it relevant to ascertain their effect in favouring niche gain (i.e. reducing niche truncation) and to predict future global change impacts on species distribution. Since invasive alien species often exhibit just a small fraction of their fundamental niche in the invaded regions (Bertolino, Cerri, et al., 2020; Bertolino, Sciandra, et al., 2020; Guisan et al., 2014), using occurrence data from these areas to model their distribution might likely introduce niche truncation. Accordingly, invasive alien species represent suitable candidates to test whether integrating data from different sources (i.e. citizen science and scientific surveys) might mitigate niche truncation and influence predictions of future biological invasion risk under global change drivers.

We propose a study aimed at comparing the ability of citizen science and scientific survey data, along with their pooling, to capture realized niche of a set of invasive alien species in Italy, and, consequently, assessing differences in predicted current and future biological invasion risk under global change scenarios. To cover the broadest possible habitat and taxonomic spectrum, we selected six freshwater and seven terrestrial species introduced in Italy in the last 10–75 years, representing seven orders and nine families (Table S1). We chose these species as they are highly invasive in Italy and are sufficiently known and sampled from both citizen and scientific survey programmes, thus providing an excellent experimental set for the study purpose. Since the occurrence records of some species were gathered from national-scale data sets while others were from regional-scale sources, we analysed the species at two different levels, that is, either national (Italy) or regional (Piedmont region). To corroborate the study aims, we started with two working hypotheses, that is, that pooling data from scientific surveys and citizen science initiatives always implied a net gain in the coverage of the environmental niche sampled in the invaded range, and that such niche gain is systematically paired with an increased biological invasion risk under future global change. These hypotheses were tested following five objectives: (i) calculating niche overlap and similarity between citizen science and scientific survey data, identifying the major environmental conditions differentiating citizen science and scientific survey realized niches; (ii) quantifying the percentage of environmental space that is gained/lost when integrating citizen science and scientific survey data; (iii) modelling current and 2100 species distributions in Italy or Piedmont region from citizen science, scientific survey and pooled data sets under climate, land cover and human population change scenarios; (iv) assessing differences in predictive accuracy and in range net change between the current time and 2100, as modelled from the three occurrence data sources; (v) testing for the relationship between niche gain generated by data pooling and range net change values.

2 | METHODS

2.1 | Analytical framework

We deployed an analytical framework structured around two separate pipelines: (i) niche overlap analyses coupled with machine learning classification models; and (ii) ENMs. The first line of analysis was set to investigate differences in the realized niches of 13 target species estimated using in turn data from scientific surveys, citizen science programmes or a pooled data set including both. We specifically focused on the non-overlapping niche portions, that is, the regions of the environmental space that are exclusively represented by a single data source (‘only citizen science niche’, ‘only scientific survey niche’ and ‘only pooled niche’) and their combinations (Figure 1). We extracted the environmental conditions associated with these non-overlapping regions and calibrated machine learning classification models to evaluate which environmental variables contributed the most in differentiating non-overlapping niche portions.

The second pipeline was developed to assess possible discrepancies in the biological invasion risk (in terms of habitat suitability) of the 13 target species as predicted by citizen science, scientific survey and pooled data under an ENM framework. Accordingly, we modelled target species current and future distribution under climate, land cover and human population change scenarios. As the study area represented just a limited portion of the global distribution ranges of the analysed species, we trained ENMs according to a hierarchical framework (Gallien et al., 2012), that is, refining global predictions at the regional scale (Di Febbraro et al., 2018, 2019). Hence, a first group of models was calibrated considering the global species range and bioclimatic variables (i.e. global ENMs; Appendix S1). Then, we fitted a second set of models to the study area level (Italy or Piedmont region, i.e. regional ENMs) incorporating predictions from global ENMs (Appendix S1). Regional ENMs calibrated from citizen science, scientific survey and pooled data were compared in terms of predictive performances and range net change. Lastly, to test if an enlarged niche space generated by data pooling determined an increase in the future predicted biological invasion risk, we assessed the relationship between niche net gain (as emerging from the first modelling approach) and range net change values (as derived from the second line of analysis).

2.2 | Species occurrence data

Occurrence records for global ENMs calibration were gathered from both native and invaded ranges (Broennimann & Guisan, 2008). For native range, we converted the IUCN extent of occurrence maps for each species into grids with a resolution of 50 km (Roll et al., 2017) using the resulting cells as presence data (Appendix S1). We also added data from invaded ranges as extracted from the ‘Global Biodiversity Information Facility’ (GBIF) database (Strubbe et al., 2015; see Table S1). The accuracy of records gathered from GBIF was assessed by including only occurrences given to at least two decimal places (0.01 decimal degrees, corresponding to 1.11 km at the equator) and by removing duplicated and unrealistic records. As to regional ENMs, scientific survey data were provided by authors and represent the results of extensive and multi-year monitoring projects, whereas citizen science data were supplied by national organizations and were integrated with data from the citizen science online platform ‘iNaturalist’ (https://www.inaturalist.org/; see

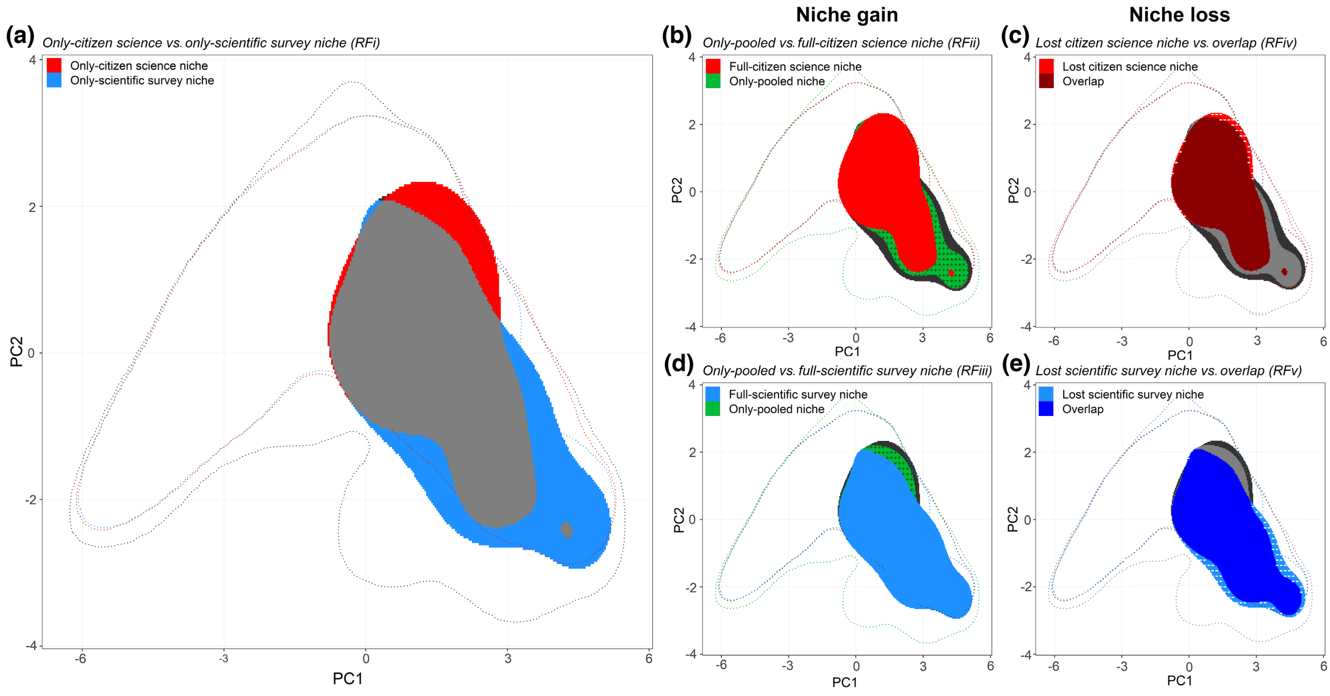

FIGURE 1 Conceptual framework used to compare and characterize non-overlapping niche portions as generated by citizen science, scientific survey and pooled data sets. (a) Depicts an example of only citizen science (red) and only scientific survey (blue) portions that are compared in the first group of RF classification models (i.e. RFi) to assess the environmental characteristics differing the most between the two. (b, d) Refer to the comparisons between only pooled (dark green crosses over a green background) and full citizen science or full scientific niche as performed in the second and third group of RF models (i.e. RFii and RFiii). (c, e) Depict the comparisons between overlapped niche (dark red and blue) and lost portions of citizen science or scientific survey niches (white dashes over red or blue backgrounds), as calculated in the fourth and fifth group of RF classification models (i.e. RFiv and RFv). Grey colours indicate niche portions not involved in RF models. RF, Random Forest. [Colour figure can be viewed at wileyonlinelibrary.com]

Table S2 for credits and single species data sources). Citizen science data were critically examined, retaining only the records that were fully reliable (i.e. with photos attached).

2.3 | Environmental variables

For global ENMs, we considered the 19 bioclimatic variables provided in the CHELSA database (Karger et al., 2017), which were rasterized at 50 km. Once checked for multicollinearity variance inflation factor (VIF ≤ 5; Zuur et al., 2010), the final predictor set was reduced to nine (Table S3). For regional ENMs calibration, we considered the 19 CHELSA variables, along with three topographical variables (elevation, slope and roughness; Danielson & Gesch, 2011), four natural land cover categories (Euclidean distance to barren areas, forests, grasslands and waterbodies extracted from the GeoSOS global database; Li et al., 2017), five variables describing human-modified land cover (Euclidean distance to farmlands, density and distance to urban areas from Li et al., 2017, and density and distance to roads from OpenStreetMap; https://www.openstreetmap.org) and two human population size predictors (i.e. in urban and rural areas; Gao, 2017). All variables were rasterized at a 1-km spatial resolution and checked for multicollinearity (VIF ≤ 5), retaining 21 final predictors (Table S3).

2.4 | Niche overlap analyses and classification models

For each species, we calculated the niche overlap between scientific survey and citizen science data according to the framework proposed by Broennimann et al. (2012), which has been widely adopted in ecological studies (Antell et al., 2021; Bosso et al., 2022; Collart et al., 2020; Di Febbraro et al., 2017; Lioy et al., 2023; Raia et al., 2020). Through this approach, principal component analysis (PCA) was used to decompose the environmental space defined for citizen science, scientific survey and pooled data sources (i.e. all the environmental conditions intersected by the occurrences and background environments). Occurrence records and environmental conditions were projected into this PCA space; then, their densities were calculated across the first two principal components using a kernel density smoother. Occurrence and background environment densities were then divided by the maximum number of occurrences in any cell of the environmental space and by the number of sites with the most common environment respectively (Broennimann et al., 2012). The process generated a density grid in the environmental space that was used to calculate niche overlap between scientific survey and citizen science data sources in terms of Schoener's *D* index (Schoener, 1970). We also performed niche similarity tests (Warren et al., 2008), to evaluate whether the two niches being compared (i.e. citizen science vs. scientific survey) are more similar/different than expected by chance. The test prescribes comparing the observed Schoener's *D* value to a null distribution of 100 overlap values, yielding a significant outcome if the observed overlap is higher ('niche conservatism' hypothesis) or lower ('niche divergence' hypothesis) than the 95th percentile of the null distribution ($p<.05$). For each species, we compared the width of full citizen science (i.e. only citizen science niche plus overlap), full scientific survey (i.e. only scientific survey niche plus overlap) and pooled niches, as well as of the only citizen science and only scientific survey niche portions in the environmental space (Figure 1).

Furthermore, to quantify the amount of gained niche space that is obtained by pooling citizen science and scientific survey data sets, we calculated the percentage of only pooled niche width versus the full citizen science or full scientific survey niche (Figure 1b,d). We also assessed the amount of niche margins lost in the pooling process (e.g. due to an overall shrink or barycentre shift) by calculating the percentage of lost environmental space from citizen science or scientific survey niches compared to the overlapping part (i.e. the part included in the pooled niche; Figure 1c,e).

For each species, we identified the most critical environmental conditions differentiating non-overlapping niche portions by implementing five Random Forest (RF; Breiman, 2001) classification models: (i) only citizen science versus only scientific survey niches (RFi; Figure 1a); (ii) only pooled versus full citizen science niches (RFii; Figure 1b); (iii) only pooled versus full scientific survey niche (RFiii; Figure 1d); (iv) lost citizen science niche versus overlap (RFiv; Figure 1c); and (v) lost scientific survey niche versus overlap (RFv; Figure 1e). For each model, environmental variables associated with each non-overlapping portion in the PCA environmental space were used as covariates. We calculated the classification performance as the out-of-bag accuracy rate and the variables contribution as the mean decrease in such accuracy (Liaw & Weiner, 2002). All the analyses were carried out using the 'ecospat' (Broennimann et al., 2016) and 'randomForest' (Liaw & Weiner, 2002) R packages.

2.5 | Ecological niche models

We calibrated global and regional ENMs using an ensemble forecasting approach as implemented by the 'biomod2' (Thuiller et al., 2009) R package. Specifically for global ENMs, we averaged models through committee averaging, which quantifies the percentage of agreement on the species occurrence among several model predictions (Thuiller et al., 2009). According to Gallien et al. (2012), background points created to calibrate regional ENMs were given a different weight depending on committee averaging values generated by global ENMs (Appendix S1). As to regional ENMs, we adopted a mixed strategy depending on the species number of occurrences (Di Febbraro et al., 2018; Lyu et al., 2022). In particular, we calibrated the so-called 'ensemble of small models' (Breiner et al., 2018) for those species reporting fewer than 50 occurrences in either scientific survey or citizen science data sets (Santini et al., 2021), that is, considering all possible combinations of the 21 environmental variables taken two at a time. Species with ≥50 occurrences for both scientific and citizen data sets were modelled with 'traditional' ensemble ENMs (i.e. including all 21 environmental variables at the same time). Pooled data sets were modelled alternatively through ensemble of small models or ENMs according to their sample size.

We used the following four modelling algorithms: Generalized linear models, generalized additive models, generalized boosted models and RF. For each species, we identified all the WWF Terrestrial Ecoregions (Olson et al., 2001) including species records as the background area (Barve et al., 2011; Di Febbraro et al., 2019; Guisan et al., 2014), where we randomly placed a pool of 10,000 background points. In particular, background points were geographically placed according to the density of the occurrence data pooled among all the species (Chauvier et al., 2021), so that there are more background points where presences are denser (Mondanaro et al., 2021; Roy-Dufresne et al., 2019).

To evaluate model predictive accuracy, we performed a block cross-validation approach (Muscarella et al., 2014; Roberts et al., 2017), that is, splitting data into four geographically non-overlapping bins of equal occurrence number, corresponding to each corner of the entire geographical space. Model accuracy was evaluated by measuring the area under the receiver operating characteristic curve (AUC; Hanley & McNeil, 1982) and the Continuous Boyce Index (CBI; Hirzel et al., 2006). To avoid using poorly calibrated models, only projections from models with an AUC ≥ 0.7 were considered in further analyses. Model averaging was performed by weighting the individual model projections by their AUC values and averaging the result (Marmion et al., 2009).

Models were projected to year 2100 under the worst-case scenarios for climate (RCP.85; IPCC, 2013), land cover (A1B; Li et al., 2017) and human population size (SSP5; Gao, 2017) change (Table S4). Since different global circulation models may lead ENMs to predict diverging climate change effects (Buisson et al., 2010), we considered five alternative versions for the RCP8.5 scenario, generated by the GFDL-ESM4, IPSL-CM6A-LR, MPI-ESM1-2-HR, MRI-ESM2-0 and UKESM1-0-LL global circulation models (Sanderson et al., 2015). To account for the effect of model extrapolation on covariate values lying outside the calibration range, additional projections were also generated through environmental clamping (i.e. capping covariates at the limit values of the training range; Elith et al., 2011). Current and future model projections were binarized to obtain presence/absence maps according to three thresholding schemes, that is, 'equalize sensitivity and specificity', 'maximize TSS' and 'minimum training presence' (Liu et al., 2013), to take into account the effect of using different binarization approaches (Jamwal et al., 2022).

2.6 | Quantification of global change effects on species biological invasion risk

The impact of future global change on species biological invasion risk was assessed by calculating the range net change metric (in terms of gain/loss percentage between the current and the future range) on binary maps generated for each species, model and scenario (Franklin et al., 2013). To test for differences in model predictive accuracy and range net change among different ENM groups (i.e. citizen science, scientific survey or pooled), we alternatively regressed AUC, CBI and net change values against the data set type using linear mixed models (LMM; 'lme4' R package; Bates et al., 2015). For AUC and CBI, we set the data set type as fixed term and species as random effect within a random slope design, to allow the modelled response (i.e. the difference in mean AUC/CBI values among the three data sources) to vary at the species level. Similarly, LMM for range net change were fitted considering the data set type and the broad habitat (i.e. aquatic vs. terrestrial) along with their interaction, as fixed effects, whereas setting the species, the global circulation model and the binarization threshold as random slope terms. As a post hoc test for these LMM, we performed a pairwise comparison between marginal means (i.e. the difference in fitted means between each pair of data set types). The statistical significance was assessed through Kenward–Roger *p*-values (Bates et al., 2015). LMM were also used to test whether the niche gain generated by data pooling was associated with higher range net change values predicted by pooled ENMs with respect to citizen science or scientific survey ENMs (i.e. whether a reduction in niche truncation yields a wider range prediction in the future). In this case, we included the difference in range net change between pooled and citizen science or scientific survey ENMs as the response variable and the size of only pooled niche in the environmental space (Figure 1b,d) in interaction with the broad habitat, as the explanatory one. Also in these models, we set species, global circulation model and binarization threshold as random slope terms. All the LMMs for range net change were also fitted on net change values calculated on clamped predictions, to exclude possible effects by ENM extrapolation.

3 | RESULTS

3.1 | Niche overlap analyses and classification models

In nine species, the pooled niche was wider than both the full citizen science and full scientific survey niches, while in the four remaining species, the full scientific survey niche was the largest one (Figure 2a). The Schoener's *D* values representing the overlap degree between full citizen science and full scientific survey niches ranged between 0.272 and 0.804, with 10 out of 13 species reporting statistically significant niche similarity tests (except for *Cyprinus carpio*, *Misgurnus anguillicaudatus* and *Procambarus clarkii*; Figure 2b).

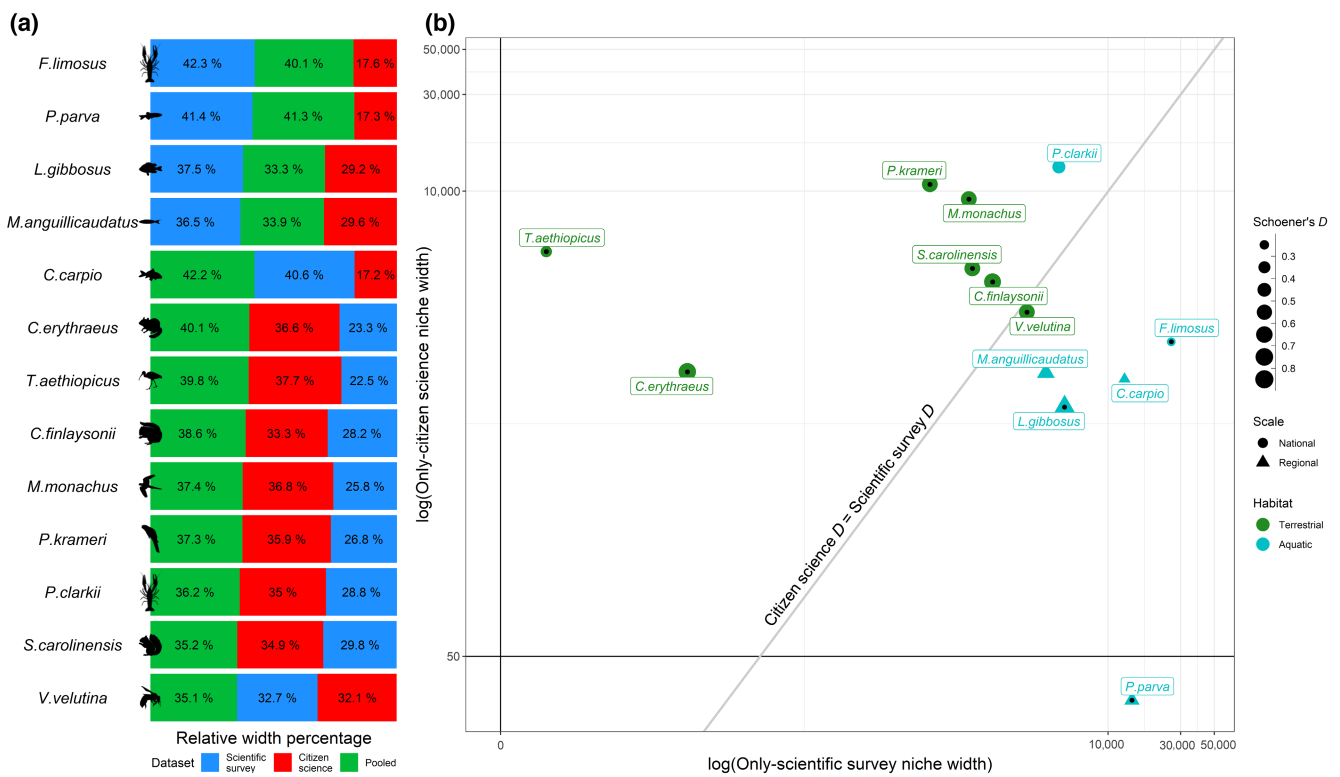

FIGURE 2 (a) Depicts the relative width percentage of full citizen science (red), full scientific survey (blue) and pooled (green) niches. (b) Illustrates the comparison of niche width values calculated for non-overlapping (i.e. only citizen science and only scientific survey) niche portions. Dot size indicates niche overlap degree (in terms of Schoener's *D* values) for each species between full citizen science and full scientific survey niches. Symbols filled with a small black dot refer to species reporting significant niche similarity tests. Terrestrial species are shown in green, while aquatic species are coloured in cyan. Circles refer to species modelled at the national scale, while triangles depict species modelled at the regional scale. [Colour figure can be viewed at wileyonlinelibrary.com]

As to non-overlapping niche portions, seven of 13 species showed that only citizen science niche is wider than only scientific survey niche, most being terrestrial species modelled at the national scale. The remaining six species, that is, mostly aquatic species modelled at the regional scale, reported the opposite pattern (Figure 2b). All five RF models achieved excellent classification performances, with a mean out-of-bag accuracy rate of 92.9% (SD = ±3.7%). RFi model evidenced that only citizen science and only scientific survey niches maximally diverged in terms of topography (i.e. elevation and slope), distance from human-modified land cover categories (both showing higher values in only scientific survey niche) and human population size (higher values in only citizen science niche; Figure 3). Although with a lower magnitude, temperature also differentiated only citizen science and only scientific survey niches, with higher values in the former (Figure 3).

The environmental space provided by scientific survey niches to the pooled niches led these to moderately increase their width compared to the full citizen science niches, with a median percentage gain equal to 16.4% (range = 0.2%–148.6%; Figure 4a). Only three species, that is, *C. carpio*, *Pseudorasbora parva* and *Faxonius limosus*, reported a percentage increase higher than 10% above the median (i.e. 16.4%), with most of the other species scoring relatively lower values. Niche space margins lost in the pooling process are relatively scarce (median = 4.3%; range = 0%–8%) and far smaller than the margins gained (Figure 4a). According to RFii models, scientific survey niches provided pooled niches with environments characterized by higher topography, precipitation and distance from human-modified land cover categories (Figure 4b). On the contrary, RFiii models reported that the contribution of the variables characterizing the lost space from full citizen science niches was almost negligible and mostly pertaining to a loss of sites distant from natural land cover (Figure 4b).

The percentage of gained niche space granted by citizen science niches was overall higher than that generated by scientific survey ones (median = 18.9%; range = 0.3%–76.7%). Six species showed an increase >10% above the median, that is, *Threskiornis aethiopicus*,

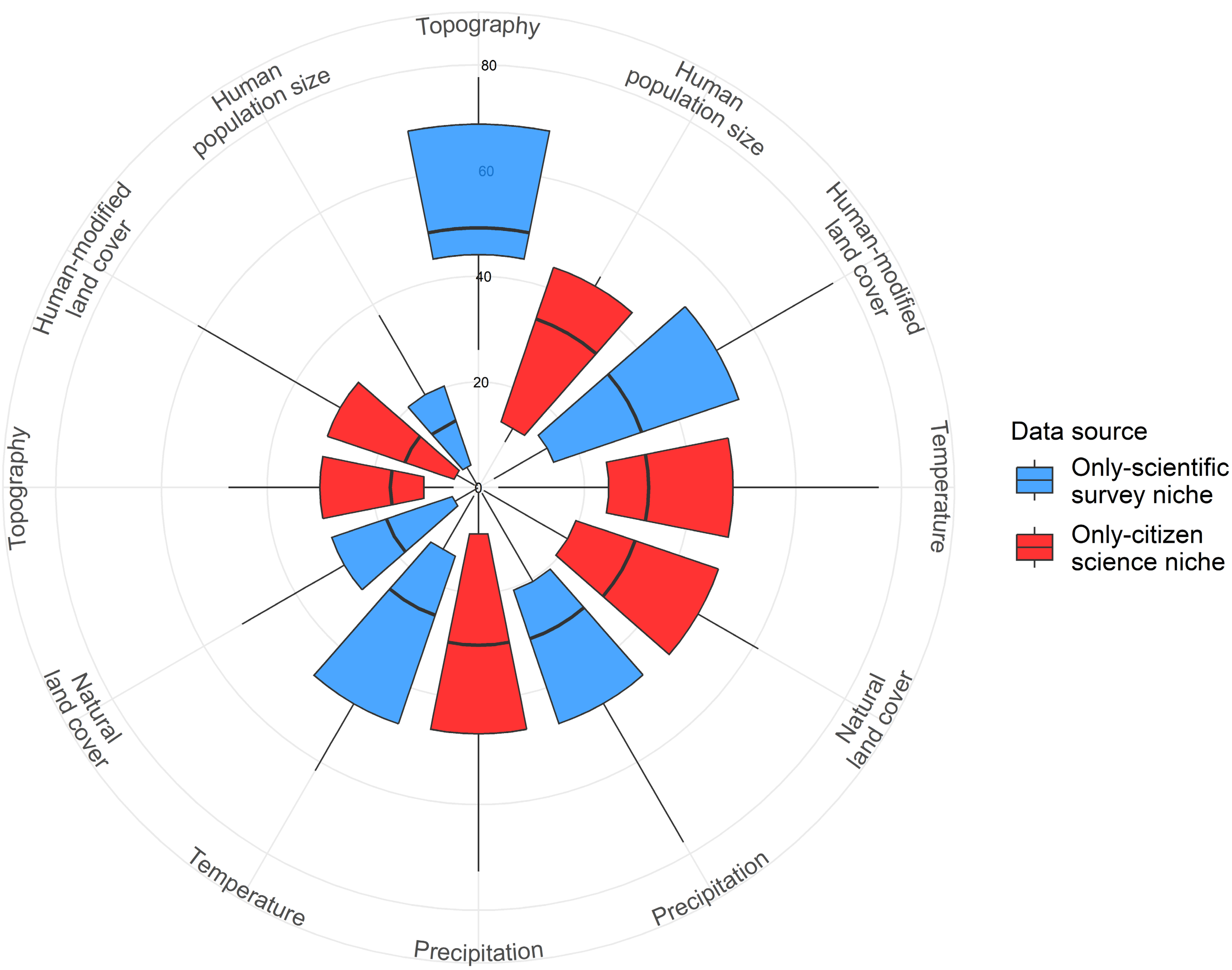

FIGURE 3 Variables importance values as emerged in the first group of Random Forest classification models (i.e. RFi). Blue boxes group the variables contribution scores for those species where predictor values are higher in only scientific survey than in only citizen science niches, while red boxes indicate the opposite pattern. [Colour figure can be viewed at wileyonlinelibrary.com]

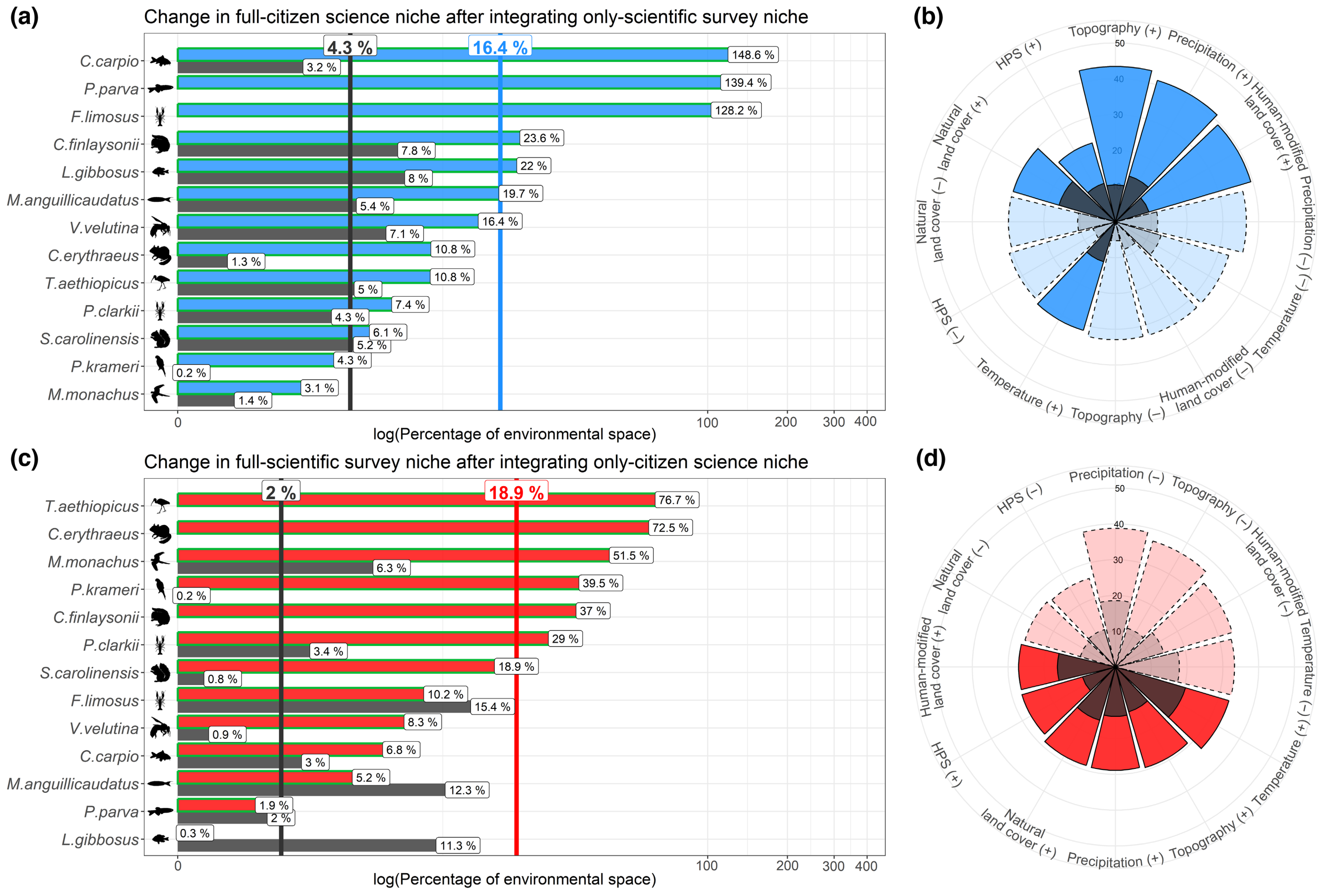

FIGURE 4 Horizontal bar plots in (a) depict the percentage gain (blue) and loss (dark grey) in environmental space of the full citizen science niches after the integration of only scientific survey niches. Similarly, (c) shows the percentage gain (red) and loss (dark grey) in environmental space of the full scientific survey niches after the integration of only citizen science niches. In these plots, vertical lines refer to median gain/loss values. Radial bar plots (b, d) show the mean contribution of the variables characterizing the gained (blue/red) and lost (dark grey) niche space. In particular, solid bars indicate that data integration led to an increase in the given variable (a '+' is placed along the variable name), while transparent bars refer to a decrease (a '–' is placed along the variable name). [Colour figure can be viewed at wileyonlinelibrary.com]

Callosciurus erythraeus and *Myiopsitta monachus*, *Psittacula krameri*, *C. finlaysonii* and *P. clarkii* (Figure 4c). The percentage of lost space from full scientific survey niches was lower than that emerged from full citizen science niches (median = 2%; range = 0.2%–15.4%), though four species reported a higher percentage of lost than gained niche space in data integration (Figure 4c). As shown by RFiv models, citizen science niches enriched pooled niches with sites at lower precipitation, topography and distance from human-modified land cover categories, as well as with an extended thermal range (Figure 4d). The contribution of the variables characterizing the lost niche space from full scientific survey niches was slightly higher than that emerged from full citizen science niches, and mostly pertaining to a loss of sites with high temperatures and low precipitation (Figure 4d).

3.2 | Ecological niche models

Depending on the species, global ENMs achieved fair to excellent predictive performances (sensu Swets, 1988), with a mean AUC = 0.876 (range = 0.732–0.951) and a mean CBI = 0.867 (0.629–0.953). As to regional ENMs (see Figure S1 for suitability predictions), predictive performances were fair to excellent depending on species and data source. Models calibrated with citizen science data achieved a mean AUC = 0.901 (range = 0.718–0.993) and a mean CBI = 0.652 (range = −0.202–0.996; Table S5). The corresponding figure for models relying on scientific survey data reported a mean AUC = 0.886 (range = 0.746–0.999) and a mean CBI = 0.748 (range = 0.632–0.934; Table S5). Lastly, pooled ENMs achieved a mean AUC = 0.903 (range = 0.756–0.991) and a mean CBI = 0.756 (range = 0.524–0.968; Table S5). There was no significant difference in AUC or CBI values among the three data sources (Table S6).

3.3 | Biological invasion risk under future scenarios

Models calibrated on pooled data predicted an average positive range net change for terrestrial species, significantly higher than values resulting from both scientific survey and citizen science ENMs (both negative; Figure 5; Table S7). For aquatic species, only

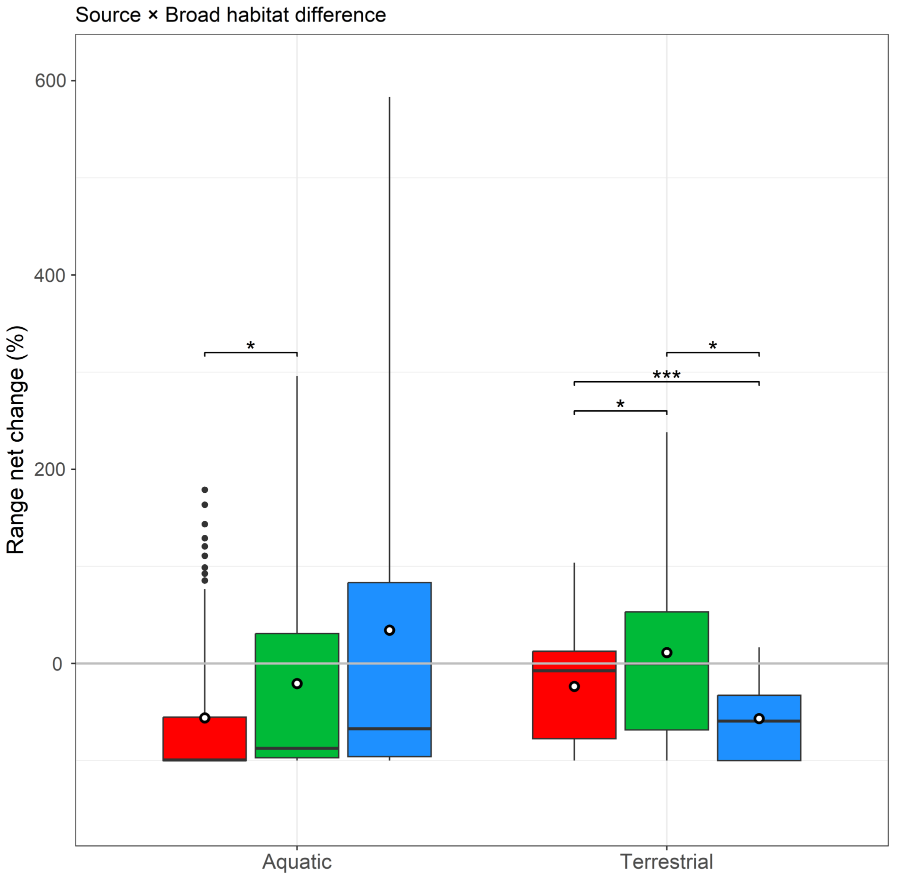

FIGURE 5 Comparison of range net change predicted by citizen science (red), scientific survey (blue) and pooled (green) models between current time and 2100 under climate, land cover and human population change scenarios. White dots indicate the average values for each group. Statistical significance of data set comparison is expressed by the horizontal whiskers and asterisks (*$p<.05$, ***$p<.001$). The variation depicted in each box plot refers to range net change values as generated by the five global circulation models and the four binarization thresholds. [Colour figure can be viewed at wileyonlinelibrary.com]

scientific survey ENMs predicted a mean positive range net change, which was higher than the values predicted by pooled and citizen science models (Figure 5; Table S7). This pattern also holds true with clamped predictions, although it is statistically significant only for terrestrial species (Figure S2; Table S8).

The niche net gain provided by scientific survey niches to pooled niches is not significantly correlated to the higher range net change in pooled ENMs with respect to scientific survey models, neither for aquatic nor terrestrial species (Figure 6a; Table S9). On the contrary, the niche net gain granted by citizen science niches to pooled niches significantly explains the higher range net change predicted by pooled ENMs compared to citizen science ENMs for both aquatic and terrestrial species (Figure 6b; Table S9). As for clamped predictions, both scientific survey and citizen science data granted a niche net gain to the pooled niches that significantly explain the higher range net change values predicted by pooled ENMs, with this pattern holding true for both aquatic and terrestrial species (Figure S3a,b; Table S10).

4 | DISCUSSION

In our work, we provided evidence that data collected by citizens and scientists generate species niches with notably large exclusive (i.e. non-overlapping) portions in most of the analysed species, with associated and clearly identifiable environmental characteristics. Among the environmental factors differentiating full citizen from full scientific survey niches, topography, human population size, distance from human-modified land cover and temperature exert the highest contribution. Most species underwent a net gain in their pooled niches after integrating citizen science and scientific survey data, substantially mitigating the niche truncation and incomplete gradient coverage inherent in both data sources. For terrestrial species modelled at the national scale, citizen science data granted the most apparent gain in environmental space to pooled niches, which determined an increased biological invasion risk under global change drivers as predicted by pooled models. A few aquatic species reported a net loss in the pooled niches with respect to their full scientific survey niches, suggesting that citizen science data may also lead to shrinkage and barycentre shift in pooled niches. Consequently, these species showed a lower biological invasion risk predicted by pooled models than by scientific survey models.

4.1 | Scientific survey and citizen science niches. How similar, how different?

Niches generated from scientific survey and citizen science data of all the analysed species exhibited moderate to large non-overlapping regions, indicating that alternative data sources captured different portions of the environmental gradients. This potentiality of scientific survey and citizen science data to provide complementary,

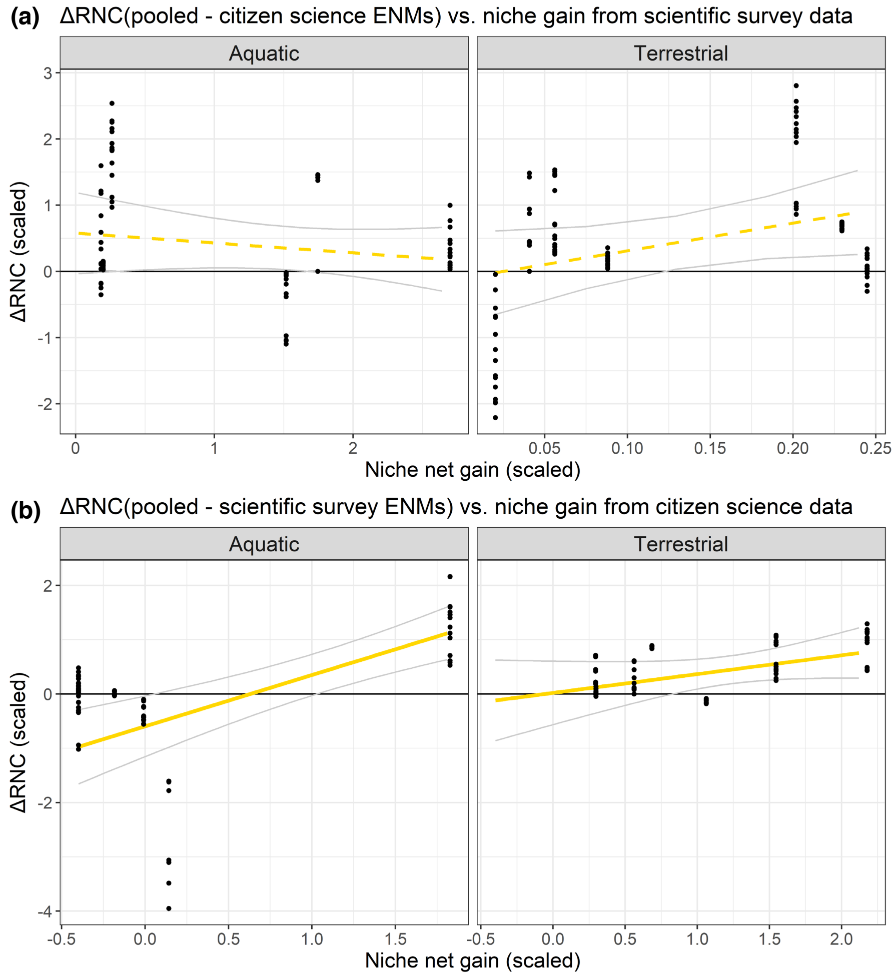

FIGURE 6 Marginal plots showing the relationship between niche net gain determined by the inclusion of scientific survey data in the pooled niches and the difference in range net change (ΔRNC) between pooled and citizen science ENMs (a). (b) Depicts the effect of niche gain granted by citizen science data versus the difference in range net change between pooled and scientific survey ENMs. Solid, thick lines indicate a significant relationship, while dashed lines refer to non-significant ones. Grey, thin lines delimit 95% confidence intervals around the mean. ENM, ecological niche model. [Colour figure can be viewed at wileyonlinelibrary.com]

non-substitutable information about species environmental preferences has been observed in several other studies (Galván et al., 2021; Grabow et al., 2022; Stuber et al., 2022). Moreover, it is in accordance with the evidence provided by Perret and Sax (2022) that both structured, scientifically driven sampling and even large collaborative databases capture the full spectrum of species' environmental niches only limitedly (i.e. 22%–45%).

Interestingly, citizen science and scientific survey non-overlapping niche portions are rather asymmetrical in all the analysed species. Most of the terrestrial species exhibit moderate to large exclusive niche portions generated from citizen science data, indicating that this data source is able to capture a wider portion of the environmental gradients in such species. The opposite figure is found for aquatic taxa, where only scientific survey niches tend to be wider than only citizen science ones. This pattern likely reflects a well-described taxonomic imbalance inherent in most citizen science initiatives. In fact, fish species are the least sampled vertebrate group by citizen science programmes worldwide, whereas they surpass amphibian and reptile species among the taxa surveyed by professional scientists (Theobald et al., 2015). Probably, this evidence appeared even exacerbated in our study context as we focused on invasive alien species outside of their native range, with likely fewer volunteers skilled and/or predisposed towards alien fishes and crustaceans than those surveying birds and mammals, that is, the two most sampled vertebrate taxa by volunteers worldwide (Lloyd et al., 2020). By the way, citizen scientists also reported an overall lower interest in alien species (Petersen et al., 2021; but see Price-Jones et al., 2022).

As to the main environmental drivers differentiating the niches generated by scientific survey and citizen science data, topography emerged as the most important and recurrent factor among species. We found a strong tendency by citizen science data to concentrate in less elevated and steep environments than scientific surveys, namely sites with a lower imperviousness degree (Grabow et al., 2022). This 'accessibility bias' (Petersen et al., 2021) has been described in a variety of contexts and taxa, with volunteers preferring to survey lowland sites (Tang et al., 2021), and close to roads (Petersen et al., 2021; Zhang, 2020) or summer residences (Millar et al., 2019). The accessibility bias can also explain another outcome of our study, that is, that citizen science niches mostly include sites with a higher human population density and closer to human-modified land cover categories (i.e. cities, roads and farmlands) than scientific survey ones (Geldmann et al., 2016; Mahecha et al., 2021; Planillo et al., 2021). Besides the above-mentioned accessibility bias, this pattern can also reflect the increased importance of urban ecology during the last 10 years (Petersen et al., 2021), which has mostly focused on detecting alien species spread in urban areas (Gaertner et al., 2017). Differently from all the above-mentioned environmental drivers, the higher temperatures found in citizen exclusive niches are likely unrelated to any specific observer behaviour. That said, temperature, along with other important variables differentiating scientific survey and citizen science niches (i.e. human-modified land cover and population density), is usually involved in future global change scenarios. In light of that, ENMs which include these variables and are trained exclusively on either scientific survey or citizen science data will suffer from niche truncation and will likely lead to biased predictions, since none of the two data sources provides a comprehensive coverage of the variability of these predictors.

4.2 | Niche truncation and future biological invasion risk predictions

Overall, the integration of citizen science and scientific survey data sets reduced niche truncation in most analysed taxa primarily through the contribution of citizen science data, with this pattern resulting particularly evident in terrestrial species modelled at the national scale. In these contexts, the potentiality of citizen science data to capture wider portions of the environmental gradients is far higher than that of scientific survey data, leading to a significant increase in the predictions of future biological invasion risk. Some aquatic species modelled at the regional scale showed a reduction in niche width after data integration compared to their full scientific survey niches, substantially failing our first working hypothesis. In these species, the inclusion of citizen science data in the pooled niches seems to determine a niche shrinkage and barycentre shift, particularly eroding margins related to high temperatures and low precipitations. These environments emerged among the most recurrent in those niche portions lost from full scientific survey niches after the inclusion of citizen science data (Figure 4d). In light of that, the peculiar pattern exhibited by aquatic species at the regional scale also affected their predicted future biological invasion risk. While terrestrial species modelled at the national scale reported a significantly higher biological invasion risk predicted by pooled ENMs than by either citizen science or scientific survey ones, in keeping with other studies (Scherrer et al., 2021), pooled ENMs for most aquatic species forecasted a lower biological invasion risk than that predicted by scientific survey ENMs. This evidence suggests that the loss of warmer and drier environments due to the inclusion of citizen science data in the pooled niche hampered the model's ability to account for the tolerance of these species to such extreme conditions. This reduced ability also seems to suggest that either the hierarchical modelling approach by Gallien et al. (2012) partly fails in mitigating niche truncation in this context or the species do not experience the above-mentioned extreme conditions even in their native range. Whatever the case, the predicted biological invasion risk for these species in future environments where extreme temperature and precipitation conditions will become more frequent (European Environment Agency, 2019) results lower (see also Capinha et al., 2013).

Besides the idea that data integration does not automatically imply a reduction in niche truncation, our outcomes suggest that pooled ENMs are not necessarily more accurate than models calibrated with separate data sets, since we did not find any significant difference among the evaluation metrics achieved by the three ENMs groups. About this outcome, there is inconsistent evidence in literature, with some authors underlining the superior predictive abilities of integrated modelling approaches (Robinson et al., 2020; Zulian et al., 2021), while others reported no significant differences (Chevalier et al., 2022), or even contexts where integration approaches might only be a suboptimal choice (Ahmad Suhaimi et al., 2021; Simmonds et al., 2020). At any rate, what is firmly supported by our findings is that a net reduction in niche truncation (whenever it happens) significantly explains an increase in future biological invasion risk, substantially confirming our second working hypothesis. In particular, this evidence emerged more markedly when the reduction in niche truncation derives from the inclusion of citizen rather than scientific survey data (Figure 6). However, after accounting for the extrapolation effect, this pattern holds true also for niche net gain determined by the inclusion of scientific survey data (Figure S3), differently from the findings of Chevalier et al. (2022).

4.3 | Conclusions

The potentiality of scientific survey data to affect future predictions of biological invasion risk is rather small, since the novel environments they add are fewer than citizen science data or not involved among the global change drivers (i.e. a high topography is the most recurrent environment provided by scientific survey data to pooled niches). By counterpart, citizen science data showed a more pervasive effect within the integration process, both from a positive and a negative perspective. This double-faced role played by citizen science data suggests that they may represent a valuable contribution to monitoring the distribution of alien species (e.g. terrestrial taxa) and predicting their future spread, especially when they are gathered within national-scale initiatives. At the same time, citizen science data collected on species less common among citizen scientists, or in strictly local contexts, may strongly affect the realized niche quantification of these taxa in the invasive areas, as well as the prediction of their future biological invasion risk.

Citizen science is a growing activity with hundreds of projects aimed at monitoring single taxa or the whole biodiversity, including alien species, at large spatial scales (Price-Jones et al., 2022). Data gathered through these projects, alone or integrated with records from scientific surveys, are pivotal for modelling approaches aimed at predicting the biological invasion risk of introduced species. Data collected by citizen scientists may include not only the target species but also other taxa, such as native species impacted by alien species, shedding light on species interactions (Groom et al., 2021; Gurnell et al., 2014). A close partnership between citizens and professional scientists would increase the data quality and provide broad educational benefits, increasing public awareness of alien species.

ACKNOWLEDGEMENTS

We are grateful to Luciano Ruggieri from EBN Italia, Associazione Teriologica Italiana (ATIt), Ornitho.it and the LIFE STOPVESPA project for providing us with some of the data used in this work. See also Table S2 for credits. E. Basset (University of Birmingham) kindly took the time to revise our manuscript and to improve the English language, grammar and syntax. ET and EM acknowledge the support of NBFC to the University of Florence and CNR–IRET, funded by the Italian Ministry of University and Research, PNRR, Missione 4 Componente 2, 'Dalla ricerca all'impresa', Investimento 1.4, Project CN00000033. This study was supported by the "CitAlien" departmental project funded by the Department of Biosciences and Territory, University of Molise.

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no conflict of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from Zenodo at https://doi.org/10.5281/zenodo.8185610.

ORCID

Mirko Di Febbraro https://orcid.org/0000-0001-8898-7046
Luciano Bosso https://orcid.org/0000-0002-9472-3802
Mauro Fasola https://orcid.org/0000-0002-7366-5893
Francesca Santicchia https://orcid.org/0000-0003-4814-9632
Gaetano Aloise https://orcid.org/0000-0002-2938-4357
Simone Lioy https://orcid.org/0000-0003-2939-988X
Elena Tricarico https://orcid.org/0000-0002-7392-0794
Emiliano Mori https://orcid.org/0000-0001-8108-7950
Sandro Bertolino https://orcid.org/0000-0002-1063-8281

REFERENCES

Ahmad Suhaimi, S. S., Blair, G. S., & Jarvis, S. G. (2021). Integrated species distribution models: A comparison of approaches under different data quality scenarios. *Diversity and Distributions*, *27*(6), 1066–1075. https://doi.org/10.1111/ddi.13255

Antell, G. S., Fenton, I. S., Valdes, P. J., & Saupe, E. E. (2021). Thermal niches of planktonic foraminifera are static throughout glacial-interglacial climate change. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(18), 105118. https://doi.org/10.1073/pnas.2017105118

Arenas-Castro, S., Regos, A., Martins, I., Honrado, J., & Alonso, J. (2022). Effects of input data sources on species distribution model predictions across species with different distributional ranges. *Journal of Biogeography*, *49*(7), 1299–1312. https://doi.org/10.1111/jbi.14382

Barve, N., Barve, V., Jiménez-Valverde, A., Lira-Noriega, A., Maher, S. P., Peterson, A. T., Soberón, J., & Villalobos, F. (2011). The crucial role of the accessible area in ecological niche modeling and species distribution modeling. *Ecological Modelling*, *222*(11), 1810–1819. https://doi.org/10.1016/j.ecolmodel.2011.02.011

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bertolino, S., Cerri, J., Ancillotto, L., Bartolommei, P., Gasperini, S., Benassi, G., Capizzi, D., Lucchesi, M., Mori, E., Scillitani, L., Sozio, G., Falaschi, M., Ficetola, G. F., Genovesi, P., Carnevali, L., Monaco, A., & Loy, A. (2020). A framework for prioritising present and potentially invasive mammal species for a national list. *NeoBiota*, *62*, 31–54. https://doi.org/10.3897/neobiota.62.52934

Bertolino, S., Sciandra, C., Bosso, L., Russo, D., Lurz, P. W. W., & Di Febbraro, M. (2020). Spatially explicit models as tools for implementing effective management strategies for invasive alien mammals. *Mammal Review*, *50*(2), 187–199. https://doi.org/10.1111/mam.12185

Bosso, L., Smeraldo, S., Russo, D., Chiusano, M. L., Bertorelle, G., Johannesson, K., Butlin, R. K., Danovaro, R., & Raffini, F. (2022). The rise and fall of an alien: Why the successful colonizer *Littorina saxatilis* failed to invade the Mediterranean Sea. *Biological Invasions*, *24*(10), 3169–3187. https://doi.org/10.1007/s10530-022-02838-y

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

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., Di Cola, V., Petitpierre, B., Breiner, F., D'Amen, M., Randin, C., Engler, R., Hordijk, W., Pottier, J., Di Febbraro, M., Pellissier, L., Pio, D., Mateo, R. G., Dubuis, A., Scherrer, D., Maiorano, L., Psomas, A., Ndiribe, C., Salamin, N., … Guisan, A. (2016). *ecospat: Spatial ecology miscellaneous methods*. http://cran.r-project.org/package=ecospat

Broennimann, O., Fitzpatrick, M. C., Pearman, P. B., Petitpierre, B., Pellissier, L., Yoccoz, N. G., Thuiller, W., Fortin, M., Randin, C., Zimmermann, N. E., Graham, C. H., & Guisan, A. (2012). Measuring ecological niche overlap from occurrence and spatial environmental data. *Global Ecology and Biogeography*, *21*(4), 481–497. https://doi.org/10.1111/j.1466-8238.2011.00698.x

Broennimann, O., & Guisan, A. (2008). Predicting current and future biological invasions: Both native and invaded ranges matter. *Biology Letters*, *4*(5), 585–589. https://doi.org/10.1098/rsbl.2008.0254

Buisson, L., Thuiller, W., Casajus, N., Lek, S., & Grenouillet, G. (2010). Uncertainty in ensemble forecasting of species distribution. *Global Change Biology*, *16*(4), 1145–1157. https://doi.org/10.1111/j.1365-2486.2009.02000.x

Capinha, C., Larson, E. R., Tricarico, E., Olden, J. D., & Gherardi, F. (2013). Effects of climate change, invasive species, and disease on the distribution of native european crayfishes. *Conservation Biology*, *27*(4), 731–740. https://doi.org/10.1111/cobi.12043

Chauvier, Y., Zimmermann, N. E., Poggiato, G., Bystrova, D., Brun, P., & Thuiller, W. (2021). Novel methods to correct for observer and sampling bias in presence-only species distribution models. *Global Ecology and Biogeography*, *30*(11), 2312–2325. https://doi.org/10.1111/geb.13383

Chevalier, M., Broennimann, O., Cornuault, J., & Guisan, A. (2021). Data integration methods to account for spatial niche truncation effects in regional projections of species distribution. *Ecological Applications*, *31*(7), e02427. https://doi.org/10.1002/eap.2427

Chevalier, M., Zarzo-Arias, A., Guélat, J., Mateo, R. G., & Guisan, A. (2022). Accounting for niche truncation to improve spatial and temporal predictions of species distributions. *Frontiers in Ecology and Evolution*, *10*, 944116. https://doi.org/10.3389/fevo.2022.944116

Collart, F., Hedenäs, L., Broennimann, O., Guisan, A., & Vanderpoorten, A. (2020). Intraspecific differentiation: Implications for niche and distribution modelling. *Journal of Biogeography*, *48*(2), 415–426. https://doi.org/10.1111/jbi.14009

Danielson, J. J., & Gesch, D. B. (2011). *Global multi-resolution terrain elevation data 2010 (GMTED2010)*. US Department of the Interior, US Geological Survey.

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

Di Febbraro, M., D'Amen, M., Raia, P., De Rosa, D., Loy, A., & Guisan, A. (2018). Using macroecological constraints on spatial biodiversity predictions under climate change: The modelling method matters. *Ecological Modelling*, *390*, 79–87. https://doi.org/10.1016/j.ecolmodel.2018.10.023

Di Febbraro, M., Menchetti, M., Russo, D., Ancillotto, L., Aloise, G., Roscioni, F., Preatoni, D. G., Loy, A., Martinoli, A., Bertolino, S., & Mori, E. (2019). Integrating climate and land-use change scenarios in modelling the future spread of invasive squirrels in Italy. *Diversity and Distributions*, *25*(4), 644–659. https://doi.org/10.1111/ddi.12890

Elith, J., Phillips, S. J., Hastie, T., Dudík, M., Chee, Y. E., & Yates, C. J. (2011). A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions*, *17*(1), 43–57. https://doi.org/10.1111/j.1472-4642.2010.00725.x

European Environment Agency. (2019). *The European environment—State and outlook 2020: Knowledge for transition to a sustainable Europe*. Publications Office of the European Union https://www.eea.europa.eu/publications/soer-2020

Finnoff, D., Shogren, J. F., Leung, B., & Lodge, D. (2007). Take a risk: Preferring prevention over control of biological invaders. *Ecological Economics*, *62*(2), 216–222. https://doi.org/10.1016/j.ecolecon.2006.03.025

Fletcher, R. J., Hefley, T. J., Robertson, E. P., Zuckerberg, B., McCleery, R. A., & Dorazio, R. M. (2019). A practical guide for combining data to model species distributions. *Ecology*, *100*(6), 1–15. https://doi.org/10.1002/ecy.2710

Franklin, J., Davis, F. W., Ikegami, M., Syphard, A. D., Flint, L. E., Flint, A. L., & Hannah, L. (2013). Modeling plant species distributions under future climates: How fine scale do climate projections need to be? *Global Change Biology*, *19*(2), 473–483. https://doi.org/10.1111/gcb.12051

Gaertner, M., Wilson, J. R. U., Cadotte, M. W., MacIvor, J. S., Zenni, R. D., & Richardson, D. M. (2017). Non-native species in urban environments: Patterns, processes, impacts and challenges. *Biological Invasions*, *19*(12), 3461–3469. https://doi.org/10.1007/s10530-017-1598-7

Gallien, L., Douzet, R., Pratte, S., Zimmermann, N. E., & Thuiller, W. (2012). Invasive species distribution models – How violating the equilibrium assumption can create new insights. *Global Ecology and Biogeography*, *21*(11), 1126–1136. https://doi.org/10.1111/j.1466-8238.2012.00768.x

Galván, S., Barrientos, R., & Varela, S. (2021). No bird database is perfect: Citizen science and professional datasets contain different and complementary biodiversity information. *Ardeola*, *69*(1), 97–114. https://doi.org/10.13157/arla.69.1.2022.ra6

Gao, J. (2017). *Downscaling global spatial population projections from 1/8-degree to 1-km grid cells*. https://opensky.ucar.edu/islandora/object/technotes:553

Geldmann, J., Heilmann-Clausen, J., Holm, T. E., Levinsky, I., Markussen, B., Olsen, K., Rahbek, C., & Tøttrup, A. P. (2016). What determines spatial bias in citizen science? Exploring four recording schemes with different proficiency requirements. *Diversity and Distributions*, *22*(11), 1139–1149. https://doi.org/10.1111/ddi.12477

Giuntini, S., Ancillotto, L., Falaschi, M., Viviano, A., Palagi, E., & Mori, E. (2022). Alien parakeets as a potential threat to the common noctule *Nyctalus noctula*. *Biodiversity and Conservation*, *31*, 3075–3092. https://doi.org/10.1007/s10531-022-02476-9

Grabow, M., Louvrier, J. L. P., Planillo, A., Kiefer, S., Drenske, S., Börner, K., Stillfried, M., Hagen, R., Kimmig, S., Straka, T. M., & Kramer-Schadt, S. (2022). Data-integration of opportunistic species observations into hierarchical modeling frameworks improves spatial predictions for urban red squirrels. *Frontiers in Ecology and Evolution*, *10*, 1–16. https://doi.org/10.3389/fevo.2022.881247

Grez, A. A., Zaviezo, T., Orellana, I., Pino, L., Roy, H. E., Torres, F., & Rebolledo, C. (2022). Citizen science and phytosanitary surveillance systems are complementary tools to follow the invasion of *Harmonia axyridis*. *Neotropical Entomology*, *52*(1), 46–56. https://doi.org/10.1007/s13744-022-01014-0

Groom, Q., Pernat, N., Adriaens, T., de Groot, M., Jelaska, S. D., Marčiulynienė, D., Martinou, A. F., Skuhrovec, J., Tricarico, E., Wit, E. C., & Roy, H. E. (2021). Species interactions: Next-level citizen science. *Ecography*, *44*(12), 1781–1789. https://doi.org/10.1111/ecog.05790

Guisan, A., Petitpierre, B., Broennimann, O., Daehler, C., & Kueffer, C. (2014). Unifying niche shift studies: Insights from biological invasions. *Trends in Ecology & Evolution*, *29*(5), 1–10. https://doi.org/10.1016/j.tree.2014.02.009

Gurnell, J., Lurz, P. W. W., & Bertoldi, W. (2014). The changing patterns in the distribution of red and grey squirrels in the North of England and Scotland between 1991 and 2010 based on volunteer surveys. *Hystrix, the Italian Journal of Mammalogy*, *25*(2), 83–89.

Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, *143*(1), 29–36. https://doi.org/10.1148/radiology.143.1.7063747

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

IPCC. (2013). *Climate change 2013: The physical science basis*. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on climage Change (T. F. Stocker, D. Qin, G. K. Plattner, M. Tignor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, & P. M. Midgley, Eds.). Cambridge University Press.

Jamwal, P. S., Di Febbraro, M., Carranza, M. L., Savage, M., & Loy, A. (2022). Global change on the roof of the world: Vulnerability of Himalayan otter species to land use and climate alterations. *Diversity and Distributions*, *28*, 1635–1649. https://doi.org/10.1111/ddi.13377

Johnson, B. A., Mader, A. D., Dasgupta, R., & Kumar, P. (2020). Citizen science and invasive alien species: An analysis of citizen science initiatives using information and communications technology (ICT) to collect invasive alien species observations. *Global Ecology and Conservation*, *21*, e00812. https://doi.org/10.1016/j.gecco.2019.e00812

Johnston, A., Matechou, E., & Dennis, E. B. (2022). Outstanding challenges and future directions for biodiversity monitoring using citizen science data. *Methods in Ecology and Evolution*, *14*, 103–116. https://doi.org/10.1111/2041-210X.13834

Karger, D. N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R. W., Zimmermann, N. E., Linder, H. P., & Kessler, M. (2017). Climatologies at high resolution for the earth's land surface areas. *Scientific Data*, *4*, 170122. https://doi.org/10.1038/sdata.2017.122

Li, X., Chen, G., Liu, X., Liang, X., Wang, S., Chen, Y., Pei, F., & Xu, X. (2017). A new global land-use and land-cover change product at a 1-km resolution for 2010 to 2100 based on human–environment interactions. *Annals of the American Association of Geographers*, *107*(5), 1040–1059. https://doi.org/10.1080/24694452.2017.1303357

Liaw, A., & Weiner, M. (2002). Classification and regression by randomForest. *R News*, *2*(3), 18–22.

Lioy, S., Carisio, L., Manino, A., & Porporato, M. (2023). Climatic niche differentiation between the invasive hornet *Vespa velutina nigrithorax* and two native hornets in Europe, *Vespa crabro* and *Vespa orientalis*. *Diversity*, *15*(4), 495. https://doi.org/10.3390/d15040495

Liu, C., White, M., & Newell, G. (2013). Selecting thresholds for the prediction of species occurrence with presence-only data. *Journal of Biogeography*, *40*(4), 778–789. https://doi.org/10.1111/jbi.12058

Lloyd, T. J., Fuller, R. A., Oliver, J. L., Tulloch, A. I., Barnes, M., & Steven, R. (2020). Estimating the spatial coverage of citizen science for monitoring threatened species. *Global Ecology and Conservation*, *23*, e01048. https://doi.org/10.1016/j.gecco.2020.e01048

Lyu, L., Leugger, F., Hagen, O., Fopp, F., Boschman, L. M., Strijk, J. S., Albouy, C., Karger, D. N., Brun, P., Wang, Z., Zimmermann, N. E., & Pellissier, L. (2022). An integrated high-resolution mapping shows congruent biodiversity patterns of Fagales and Pinales. *New Phytologist*, *235*, 759–772. https://doi.org/10.1111/nph.18158

Mahecha, M. D., Rzanny, M., Kraemer, G., Mäder, P., Seeland, M., & Wäldchen, J. (2021). Crowd-sourced plant occurrence data provide a reliable description of macroecological gradients. *Ecography*, *44*(8), 1131–1142. https://doi.org/10.1111/ecog.05492

Maistrello, L., Dioli, P., Bariselli, M., Mazzoli, G. L., & Giacalone-Forini, I. (2016). Citizen science and early detection of invasive species: Phenology of first occurrences of *Halyomorpha halys* in Southern Europe. *Biological Invasions*, *18*(11), 3109–3116. https://doi.org/10.1007/s10530-016-1217-z

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

Milanesi, P., Mori, E., & Menchetti, M. (2020). Observer-oriented approach improves species distribution models from citizen science data. *Ecology and Evolution*, *10*(21), 12104–12114. https://doi.org/10.1002/ece3.6832

Millar, E. E., Hazell, E. C., & Melles, S. J. (2019). The 'cottage effect' in citizen science? Spatial bias in aquatic monitoring programs. *International Journal of Geographical Information Science*, *33*(8), 1612–1632. https://doi.org/10.1080/13658816.2018.1423686

Miller, D. A. W., Pacifici, K., Sanderlin, J. S., & Reich, B. J. (2019). The recent past and promising future for data integration methods to estimate species' distributions. *Methods in Ecology and Evolution*, *10*(1), 22–37. https://doi.org/10.1111/2041-210X.13110

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

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

Olson, D. M., Dinerstein, E., Wikramanayake, E. D., Burgess, N. D., Powell, G. V. N., Underwood, E. C., D'amico, J. A., Itoua, I., Strand, H. E., Morrison, J. C., Loucks, C. J., Allnutt, T. F., Ricketts, T. H., Kura, Y., Lamoreux, J. F., Wettengel, W. W., Hedao, P., & Kassem, K. R. (2001). Terrestrial ecoregions of the world: A new map of life on Earth. *Bioscience*, *51*(11), 933. https://doi.org/10.1641/0006-3568(2001)051[0933:TEOTWA]2.0.CO;2

Ovaskainen, O., Roy, D. B., Fox, R., & Anderson, B. J. (2016). Uncovering hidden spatial structure in species communities with spatially explicit joint species distribution models. *Methods in Ecology and Evolution*, *7*(4), 428–436. https://doi.org/10.1111/2041-210X.12502

Park, I., Borzée, A., Park, J., Min, S., Zhang, Y., Li, S., & Park, D. (2022). Past, present, and future predictions on the suitable habitat of the Slender racer (*Orientocoluber spinalis*) using species distribution models. *Ecology and Evolution*, *12*(8), e9169. https://doi.org/10.1002/ece3.9169

Perret, D. L., & Sax, D. F. (2022). Evaluating alternative study designs for optimal sampling of species' climatic niches. *Ecography*, *2022*(1), 1–12. https://doi.org/10.1111/ecog.06014

Petersen, T. K., Speed, J. D. M., Grøtan, V., & Austrheim, G. (2021). Species data for understanding biodiversity dynamics: The what, where and when of species occurrence data collection. *Ecological Solutions and Evidence*, *2*(1), 1–17. https://doi.org/10.1002/2688-8319.12048

Planillo, A., Fiechter, L., Sturm, U., Voigt-Heucke, S., & Kramer-Schadt, S. (2021). Citizen science data for urban planning: Comparing different sampling schemes for modelling urban bird distribution. *Landscape and Urban Planning*, *211*, 104098. https://doi.org/10.1016/j.landurbplan.2021.104098

Price-Jones, V., Brown, P., Adriaens, T., Tricarico, E., Farrow, R. A., Cardoso, A.-C., Gervasini, E., Groom, Q., Reyserhove, L., Schade, S., Tsinaraki, C., & Marchante, E. (2022). Eyes on the aliens: Citizen science contributes to research, policy and management of biological invasions in Europe. *NeoBiota*, *78*, 1–24. https://doi.org/10.3897/neobiota.78.81476

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

Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., Hauenstein, S., Lahoz-Monfort, J. J., Schröder, B., Thuiller, W., Warton, D. I., Wintle, B. A., Hartig, F., & Dormann, C. F. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, *40*(8), 913–929. https://doi.org/10.1111/ecog.02881

Robinson, O. J., Ruiz-Gutierrez, V., Reynolds, M. D., Golet, G. H., Strimas-Mackey, M., & Fink, D. (2020). Integrating citizen science data with expert surveys increases accuracy and spatial extent of species distribution models. *Diversity and Distributions*, *26*(8), 976–986. https://doi.org/10.1111/ddi.13068

Roll, U., Feldman, A., Novosolov, M., Allison, A., Bauer, A. M., Bernard, R., Bohm, M., Castro-Herrera, F., Chirio, L., Collen, B., Colli, G. R., Dabool, L., Das, I., Doan, T. M., Grismer, L. L., Hoogmoed, M., Itescu, Y., Kraus, F., LeBreton, M., … Meiri, S. (2017). The global distribution of tetrapods reveals a need for targeted reptile conservation. *Nature Ecology & Evolution*, *1*, 1677–1682. https://doi.org/10.1038/s41559-017-0332-2

Roy, H. E., Peyton, J., Aldridge, D. C., Bantock, T., Blackburn, T. M., Britton, R., Clark, P., Cook, E., Dehnen-Schmutz, K., Dines, T., Dobson, M., Edwards, F., Harrower, C., Harvey, M. C., Minchin, D., Noble, D. G., Parrott, D., Pocock, M. J. O., Preston, C. D., … Walker, K. J. (2014). Horizon scanning for invasive alien species with the potential to threaten biodiversity in Great Britain. *Global Change Biology*, *20*(12), 3859–3871. https://doi.org/10.1111/gcb.12603

Roy, H. E., Rabitsch, W., Scalera, R., Stewart, A., Gallardo, B., Genovesi, P., Essl, F., Adriaens, T., Bacher, S., Booy, O., Branquart, E., Brunel, S., Copp, G. H., Dean, H., D'hondt, B., Josefsson, M., Kenis, M., Kettunen, M., Linnamagi, M., … Zenetos, A. (2018). Developing a framework of minimum standards for the risk assessment of alien species. *Journal of Applied Ecology*, *55*(2), 526–538. https://doi.org/10.1111/1365-2664.13025

Roy-Dufresne, E., Saltré, F., Cooke, B. D., Mellin, C., Mutze, G., Cox, T., & Fordham, D. A. (2019). Modeling the distribution of a wide-ranging invasive species using the sampling efforts of expert and citizen scientists. *Ecology and Evolution*, *9*(19), 11053–11063. https://doi.org/10.1002/ece3.5609

Sanderson, B. M., Knutti, R., & Caldwell, P. (2015). A representative democracy to reduce interdependency in a multimodel ensemble. *Journal of Climate*, *28*(13), 5171–5194. https://doi.org/10.1175/JCLI-D-14-00362.1

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

Scherrer, D., Esperon-Rodriguez, M., Beaumont, L. J., Barradas, V. L., & Guisan, A. (2021). National assessments of species vulnerability to

climate change strongly depend on selected data sources. *Diversity and Distributions*, *27*(8), 1367–1382. https://doi.org/10.1111/ddi.13275
Schoener, T. W. (1970). Nonsynchronous spatial overlap of lizards in patchy habitats. *Ecology*, *51*(3), 408–418. https://doi.org/10.2307/1935376
Seebens, H., Blackburn, T. M., Dyer, E. E., Genovesi, P., Hulme, P. E., Jeschke, J. M., Pagad, S., Pyšek, P., Winter, M., Arianoutsou, M., Bacher, S., Blasius, B., Brundu, G., Capinha, C., Celesti-Grapow, L., Dawson, W., Dullinger, S., Fuentes, N., Jäger, H., … Essl, F. (2017). No saturation in the accumulation of alien species worldwide. *Nature Communications*, *8*(1), 1–9. https://doi.org/10.1038/ncomms14435
Simmonds, E. G., Jarvis, S. G., Henrys, P. A., Isaac, N. J. B., & O'Hara, R. B. (2020). Is more data always better? A simulation study of benefits and limitations of integrated distribution models. *Ecography*, *43*(10), 1413–1422. https://doi.org/10.1111/ecog.05146
Strubbe, D., Beauchard, O., & Matthysen, E. (2015). Niche conservatism among non-native vertebrates in Europe and North America. *Ecography*, *38*(3), 321–329. https://doi.org/10.1111/ecog.00632
Stuber, E. F., Robinson, O. J., Bjerre, E. R., Otto, M. C., Millsap, B. A., Zimmerman, G. S., Brasher, M. G., Ringelman, K. M., Fournier, A. M. V., Yetter, A., Isola, J. E., & Ruiz-Gutierrez, V. (2022). The potential of semi-structured citizen science data as a supplement for conservation decision-making: Validating the performance of eBird against targeted avian monitoring efforts. *Biological Conservation*, *270*, 109556. https://doi.org/10.1016/j.biocon.2022.109556
Swets, J. A. (1988). Measuring the accuracy of diagnostic systems. *Science*, *240*(4857), 1285–1293.
Tang, B., Clark, J. S., & Gelfand, A. E. (2021). Modeling spatially biased citizen science effort through the eBird database. *Environmental and Ecological Statistics*, *28*(3), 609–630. https://doi.org/10.1007/s10651-021-00508-1
Theobald, E. J., Ettinger, A. K., Burgess, H. K., DeBey, L. B., Schmidt, N. R., Froehlich, H. E., Wagner, C., HilleRisLambers, J., Tewksbury, J., Harsch, M. A., & Parrish, J. K. (2015). Global change and local solutions: Tapping the unrealized potential of citizen science for biodiversity research. *Biological Conservation*, *181*, 236–244. https://doi.org/10.1016/j.biocon.2014.10.021
Thuiller, W., Lafourcade, B., Engler, R., & Araújo, M. B. (2009). BIOMOD – A platform for ensamble forecasting of species distribution. *Ecography*, *32*, 369–373. https://doi.org/10.1111/j.1600-0587.2008.05742.x
Tollington, S., Turbé, A., Rabitsch, W., Groombridge, J. J., Scalera, R., Essl, F., & Shwartz, A. (2017). Making the EU legislation on invasive species a conservation success. *Conservation Letters*, *10*(1), 112–120. https://doi.org/10.1111/conl.12214
Tran, T. T., Carter, B. E., & Castillo Vardaro, J. A. (2022). Predicted threats to a native squirrel from two invading species based on citizen science data. *Biological Invasions*, *24*, 3539–3553. https://doi.org/10.1007/s10530-022-02859-7
Warren, D. L., Glor, R. E. R. E., & Turelli, M. (2008). Environmental niche equivalency versus conservatism: Quantitative approaches to niche evolution. *Evolution*, *62*(12), 2868–2883. https://doi.org/10.1111/j.1558-5646.2008.00482.x
Werenkraut, V., Baudino, F., & Roy, H. E. (2020). Citizen science reveals the distribution of the invasive harlequin ladybird (*Harmonia axyridis* Pallas) in Argentina. *Biological Invasions*, *22*(10), 2915–2921. https://doi.org/10.1007/s10530-020-02312-7
Wiggins, A., & Crowston, K. (2011). From conservation to crowdsourcing: A typology of citizen science. *Proceedings of the annual Hawaii international conference on system sciences*. Kauai, HI. https://doi.org/10.1109/HICSS.2011.207
Zhang, G. (2020). Spatial and temporal patterns in volunteer data contribution activities: A case study of eBird. *ISPRS International Journal of Geo-Information*, *9*(10), 597. https://doi.org/10.3390/ijgi9100597
Zulian, V., Miller, D. A. W., & Ferraz, G. (2021). Integrating citizen-science and planned-survey data improves species distribution estimates. *Diversity and Distributions*, *27*(12), 2498–2509. https://doi.org/10.1111/ddi.13416
Zuur, A. F., Ieno, E. N., & Elphick, C. S. (2010). A protocol for data exploration to avoid common statistical problems. *Methods in Ecology and Evolution*, *1*(1), 3–14. https://doi.org/10.1111/j.2041-210X.2009.00001.x

SUPPORTING INFORMATION

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

How to cite this article: Di Febbraro, M., Bosso, L., Fasola, M., Santicchia, F., Aloise, G., Lioy, S., Tricarico, E., Ruggieri, L., Bovero, S., Mori, E., & Bertolino, S. (2023). Different facets of the same niche: Integrating citizen science and scientific survey data to predict biological invasion risk under multiple global change drivers. *Global Change Biology*, *29*, 5509–5523. https://doi.org/10.1111/gcb.16901