Predicting the spatio-temporal pattern of range expansion under lack of equilibrium with climate

Victoria Formoso-Freire [a,*], A. Márcia Barbosa [b], Andrés Baselga [c,1], Carola Gómez-Rodríguez [a,1]

[a] CRETUS, Department of Functional Biology (Area of Ecology), Universidade de Santiago de Compostela, Santiago de Compostela, Spain
[b] CICGE (Centro de Investigação em Ciências Geo-Espaciais), Universidade do Porto, Porto, Portugal
[c] CRETUS, Department of Zoology, Genetics and Physical Anthropology, Universidade de Santiago de Compostela, Santiago de Compostela, Spain

ABSTRACT

Early detection of invasive species in regions under threat is key for biodiversity conservation. Here we conduct a retrospective study to assess whether correlative Species Distribution Models (SDMs) can predict the spatio-temporal range of expansion in an invasive species, the Asian hornet. Given that modelling invasive species distributions defies one of the main assumptions of SDMs, the equilibrium with climate, we also introduce a quantitative assessment of climatic disequilibrium in the invaded area based on hypervolume comparison between the native and invaded areas (both unoccupied and occupied territories). We evaluate the ability of three different modelling approaches (presence-only, presence-background and presence-absence) calibrated with data until 2013 to predict the species distribution observed the following years (2015, 2017, 2019, and 2021). Our results show that presence-only models based on the BIOCLIM algorithm can effectively predict the spatio-temporal pattern of invasion when predictions are rescaled based on percentiles (i.e. *ranked suitability*) and the model is calibrated with data from both native and invaded areas. These models predicted higher suitability values for sites that were invaded earlier in time and, importantly, they did not predict low suitability values for sites that were eventually occupied years later. Thus, they can be very useful for decision-makers and managers, as they provide a probabilistic prediction of both (i) the temporal dimension of species range expansion, and (ii) the potential distribution range in the long term. Such information would allow prioritizing monitoring efforts in the short term without losing sight of the potential risks in the long term.

1. Introduction

The early detection of invasive species in regions under threat is key for biodiversity conservation. Correlative Species Distribution Models (SDMs) have been widely used to identify where environmental conditions are suitable, and hence where invasive species are likely to occur (Baquero et al., 2021; Gallien et al., 2012; Pereira et al., 2020), but rarely to predict when this would happen. Previous attempts have used long-term temporal series of invasion progress (e.g., Jones et al., 2022) or process-based models (e.g., Meentemeyer et al., 2011) to anticipate the species' expansion towards new territories, but these are usually data-intensive approaches that cannot be applied to most invasive species (but see Botella et al., 2022). Therefore, it would be interesting to assess to which extent standard SDMs may be used to predict both the spatial and temporal dimensions of the invasion process. Remarkably,

using SDMs to predict where and when an invasive species will expand its range under current climatic conditions defies one of the main assumptions of SDMs (Foster et al., 2022), the equilibrium of the species distribution with climate (Guisan and Thuiller, 2005; Václavík and Meentemeyer, 2012), as the very question implies that some current absences will turn into future presences (Václavík and Meentemeyer, 2009). In other words, if models are calibrated with data from the invaded territory, many (pseudo)absence records or background points may not actually reflect unsuitable environmental conditions (but see Chefaoui and Lobo, 2008; Barbet-Massin et al., 2012), and such incompleteness of the species' niche in the study area is expected to reduce SDM performance (Boyd et al., 2023; Foster et al., 2022). Therefore, the lack of equilibrium with climate can be seen as a limitation compromising the use of SDMs for accurately predicting the range expansion of invasive species. Alternatively, as we aim here, the lack of

equilibrium with climate can also be regarded as an intrinsic feature of the invasive species distribution, and the different types of SDMs can be framed under this non-equilibrium paradigm. This opens the opportunity to assess the ability of different SDM methods to predict the spatio-temporal pattern of range expansion of invasive species (i.e. the temporal order in which different regions will be invaded) under the assumption of no dispersal limitation, and potentially take advantage of these predictions to inform where and when intensive surveillance would be required for early action (Ferraz et al., 2021).

While SDMs are regularly used to predict potential changes in species distributions with time (e.g., Baquero et al., 2021), these studies are usually motivated by a projected temporal change in abiotic conditions (e.g., climate warming or land-use change, Araújo and Peterson, 2012). However, in the case of biological invasions, temporal changes in species distributions would not only be driven by environmental changes, but are inherent to the process of invasion of suitable territories not yet occupied. Under this scenario, the rationale behind using SDMs to predict such temporal changes is that the most suitable territories are expected to be the ones likely to be invaded first. This prediction can be tested with a retrospective assessment of model predictions over time, to validate whether the suitability or the probability of presence predicted by SDMs constructed at different time periods also reflected the order in which different regions were sequentially invaded in the following years. Retrospective studies have proven useful for similar purposes in other fields, such as the study of ecosystem changes (e.g., Foster et al., 2003), but they are scarce regarding the temporal progression of species range expansions (but see Barbet-Massin et al., 2018) and, to our knowledge, they have not been applied to assess the ability of SDMs to predict the spatio-temporal pattern of invasion.

The assumption of climatic equilibrium is a cornerstone of many ecological studies, especially those regarding biodiversity patterns or species distributions (Václavík and Meentemeyer, 2012; Guisan et al., 2017). However, equilibrium is a conceptual hypothetical state that is never fully reached in nature. In the context of SDMs, the most widely used algorithms assume that species ranges are in equilibrium with environmental conditions, that is, that the species occupies all areas with suitable environments (Guisan et al., 2017). Despite the potential implications for the predictive accuracy of SDMs, the assumption that species are in equilibrium with their environment is seldom tested (but see Foster et al., 2022). In fact, in the case of invasive species, when a species is expanding its range, this assumption will likely bias the SDM, especially when considering absences or background data to infer the unsuitable environmental conditions for the species (Chefaoui and Lobo, 2008). Thus, although presence-absence models have been shown to better fit species distributions when equilibrium is a (nearly) realistic assumption (Elith et al., 2020), this may not be the case for invasive species. In contrast, simpler models that are not as severely affected by deviations from equilibrium with climate, such as true presence-only models, may still provide useful information (Booth et al., 2014). Presence-only models, such as BIOCLIM (Busby, 1991), are not widely used for species distribution modelling due to their simplicity (e.g., delimitation of a n-dimensional environmental "envelope" with suitable conditions), their generally poor predictive power, and their seemingly unrestrictive prediction of broad ranges of suitable conditions (Araújo and Peterson, 2012; Booth et al., 2014). Modern implementations of this method (Hijmans et al., 2020) provide quantitative predictions that are usually transformed into binary outcomes (suitable vs. unsuitable). This binary transformation is used due to the complexity of interpreting BIOCLIM's absolute quantitative predictions, as their values strongly decrease when the number of model variables increase (Beaumont et al., 2005). However, disregarding the quantitative values in favour of a binary approach assumes that all regions within the environmental envelope are equally suitable. This simplification overlooks that those regions towards the centre of the environmental envelope are expected to present more favourable conditions. Here we introduce a rescaling of BIOCLIM predictions into *ranked suitability* values, which provide more realistic predictions independent of the number of variables.

In the context of species invasions, we can quantify the n-dimensional space of climatic conditions occupied by the species (i.e. climatic hypervolumes, Blonder et al., 2014) to assess different facets of the equilibrium assumption and its potential effect on SDMs. We can consider three different aspects: (i) to which extent an invasive species has filled the climatic space available in the invaded area (i.e. climatic space filling), (ii) to which extent the climatic space available in the invaded area is within the climatic space occupied by the species in the native area (i.e. unoccupied, but suitable, climatic space), and (iii) to which extent the climatic space occupied in the invaded area is within the climatic space occupied by the species in the native area (i.e. climatic niche conservatism). The evidence of climatic niche conservatism would support the use of data from the native area in order to increase niche completeness for model calibration (Broennimann and Guisan, 2008), while the quantification of the unoccupied suitable climatic space will inform about the degree of equilibrium with climate in the invaded region and, hence, of the potential bias introduced when considering absence data as indicative of unsuitable conditions for the species. Thus, hypervolume comparisons between the native and invaded areas can be used as a quantitative estimate of climatic disequilibrium in the non-native range (Fig. 1).

The invasion of the Asian hornet (*Vespa velutina* Lepeletier 1836) in Europe offers a great opportunity to evaluate the performance of SDMs under non-equilibrium conditions, while taking advantage of a large amount of data and a detailed record of the temporal progression of the invasion. The advance of this species across the European continent has been closely followed by active monitoring since its introduction in France in 2004 (Monceau et al., 2014). This facilitates the development of a comprehensive database of presence records, which will allow not only a good characterisation of the climatic space occupied by the species (Araújo et al., 2019), but also the development of a retrospective study for model validation, assessing how well the predicted probabilities yielded by models calibrated on the initial stages of the invasion reflected the subsequent gradual occupation of newly invaded areas. SDMs have been previously used to predict the invasion of *V. velutina* from ensembles of presence-absence models (Barbet-Massin et al., 2013; Fournier et al., 2017), using data from both the native and invaded areas (Villemant et al., 2011). The invasion of closely related species, such as *Vespa mandarinia* Smith, 1852, has also been recently modelled (Zhu et al., 2020), even accounting for the unreliability of data in the invaded area for model calibration (Jiménez and Soberón, 2022). Moreover, Barbet-Massin et al. (2018) pioneered a retrospective validation, assessing whether presence data from France until 2010 were able to predict the species distribution in France five years later. We here aim to go one step further and assess whether different SDM algorithms can predict the temporal pattern of invasion across Europe.

In this paper, we conduct a retrospective study to assess to which extent SDMs may be used under non-equilibrium conditions to predict both the spatial and temporal dimensions of range expansion of an invasive species, the Asian hornet, in Europe. With this aim, we compiled a comprehensive database of records of *V. velutina* in its invaded and native areas. We used these data to (i) characterize the climatic niche occupied by the species in its native and invaded areas, as well as the unoccupied climatic space in the invaded area, as the basis (ii) to quantify the deviation from equilibrium with climate in the invaded region, and (iii) to evaluate the ability of three different modelling approaches (presence-only, presence-background and presence-absence), calibrated with presence records until 2013, to predict the spatio-temporal pattern of invasion observed in the following years (2015, 2017, 2019, and 2021).

2. Methods

2.1. Database of climatic predictors and presences in the native and invaded areas

Data on *V. velutina* presence in the native and invaded areas were
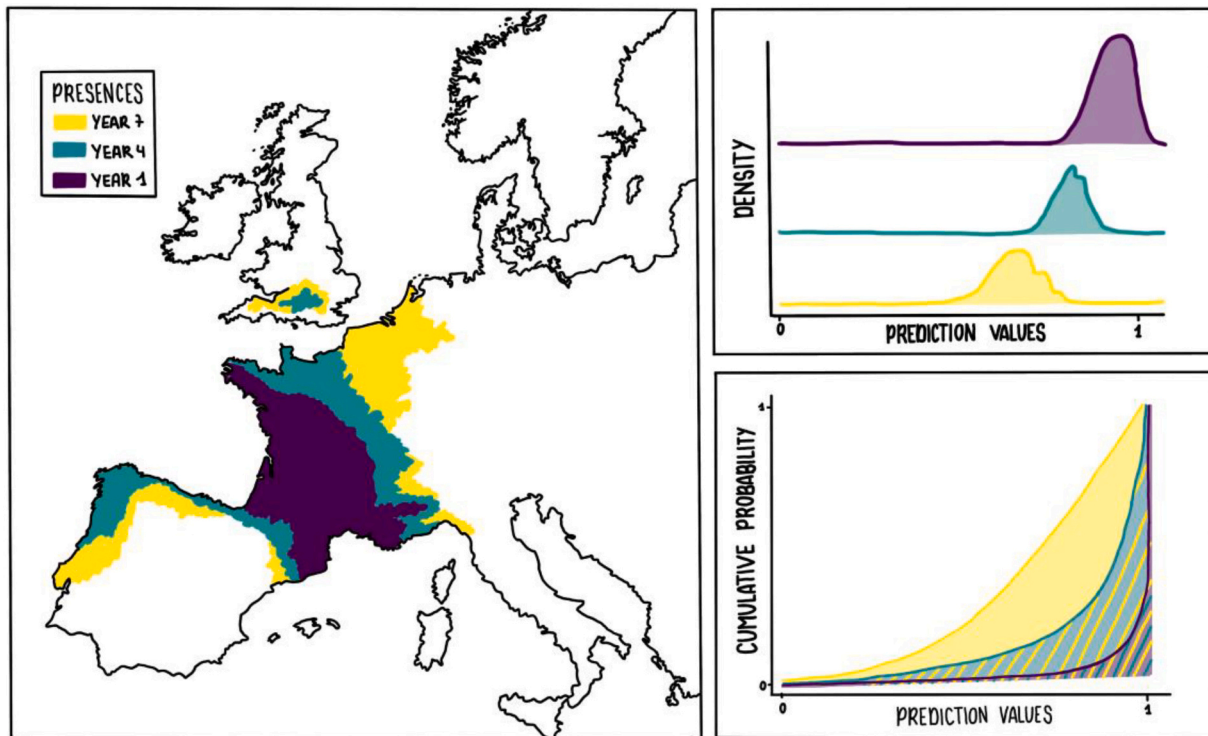
Fig. 1. Conceptual representation of the retrospective assessment of SDMs ability to predict the spatio-temporal pattern of invasion based on density curves and cumulative probabilities. (A) Map showing how range expansion over time would correspond to (B) density curves of model predictions if SDMs could be used to predict the temporal dimension of invasion (i.e. higher prediction values for territories that were early colonised, i.e. in year one). Note that models would be calibrated with data recorded in year one but independently validated against the distributions observed in the same and following years (e.g., year one, year four, year seven…). C) Representation of expected cumulative probability curves of prediction values for models validated with different temporal subsets. If SDMs could be used to predict the temporal dimension of invasion, we expect the area under the curve to increase with time given that territories invaded in the long term would be assigned lower prediction values than those invaded in the short term.

obtained from public databases, institutions and/or national organisations. Data on the invasion front (UK, Germany and the Netherlands) were complemented with records provided by experts. Most of the European records were reported for the European Environment Agency (EEA) equal-area 10 × 10 km^2 grid, which was the one used for our analysis. Records from other databases were cleaned and extracted to this grid (see Appendix S1 in Supplementary material for details). The year in which the first presence was reported in each grid cell informs about the temporal dimension of invasion in the period 2004–2021. Given that official data arise from reporting obligations of European Union countries, and given the conspicuousness and social alarm caused by *V. velutina*, monitoring programs have been very active across Europe, both through government institutions and citizen platforms. Therefore, we expect a relatively low degree of sampling incompleteness or spatial biases generated by uneven survey.

Climatic variables were sourced from the CHELSA (v2.1) database (Karger et al., 2017) at a 30-arc-second resolution and referred to the European 10 × 10 km^2 grid using the *extract* function of the 'raster' package (Hijmans et al., 2020) in R (R Core Team, 2019). To ensure that climatic conditions of non-invaded regions corresponded to territories that could, in principle, have been colonised in terms of space and time, we extracted climate data only for grid cells within countries that have reported *V. velutina* observations.

2.2. Quantitative estimation of climatic disequilibrium

To assess if *V. velutina* is in equilibrium with climate in Europe, we compared the occupied climatic space (i.e. its realised climatic niche) in the invaded area with the one in the native area and, more importantly, the occupied climatic space in the native area with the European

unoccupied climatic space. These three climatic spaces are relevant because equilibrium with climate in the invaded area would be inferred if (i) the occupied climatic space in Europe is nested within (or equals) the occupied climatic space in the native area, and (ii) the unoccupied climatic space in Europe does not overlap with the occupied climatic space in the native area, as any overlap would suggest that *V. velutina* has not filled its suitable climatic space in Europe. Conversely, deviations from equilibrium with climate would be inferred if (iii) the unoccupied climatic space in Europe is nested within the occupied climatic space in the native area, or (iv) the occupied climatic space in the invaded area does not overlap with the occupied climatic space in the native area, thus evidencing the existence of unknown suitable climatic conditions outside the species native range (i.e. niche shift, Broennimann et al., 2021). We define a new measure of climatic disequilibrium (D) as the ratio between the suitable but unoccupied climatic space (U) and the available suitable climatic space in the invaded area (A, which equals the sum of the occupied and unoccupied suitable climatic spaces in the invaded area).

To model the species niche along the main climatic gradients, we performed a Principal Components Analysis (PCA) of climatic data using the *principal* function with varimax rotation in the *psych* R package (Revelle, 2022) in R. For each climatic dimension, we selected the variable with the highest correlation with each principal component to characterize the climatic space (hypervolume) of three subsets of data (i. e. presences in native area, presences in invaded area, and absences or unoccupied space in Europe) and the intersections among them, using the functions provided by the *hypervolume* package (Blonder, 2022) in R (see details in Supplementary Material Appendix S2).

2.3. Retrospective assessment of invasion risk prediction

We subset the species presence data according to the year of first presence to produce five different calibration datasets: all presences recorded until (i) 2013, (ii) 2015, (iii) 2017, (iv) 2019, and (v) 2021. We built SDMs for each calibration dataset and compared the predicted probabilities (i.e., invasion risk) with the observed distribution of the species in the following years. Note that only records from continental Europe were considered for model calibration, given the early stage of invasion of the British Islands.

We used five different SDM approaches: (i) BIOCLIM (Busby, 1991), a true presence-only model that creates a set of rectilinear envelopes in the environmental space defined only by the presences of the species; (ii) MAXENT (Phillips et al., 2006), a maximum entropy model that uses presence data and the environmental conditions in the entire study area (background data) to compute the environmental suitability for the species; (iii) Generalised Linear Model (GLM) (McCullagh and Nelder, 1989), a presence-absence method that can fit a combination of linear, quadratic and/or cubic parametric terms; (iv) Generalised Additive Model (GAM) (Hastie and Tibshirani, 1987), a presence-absence method that fits non-parametric terms as non-linear functions defined by smoothers; and (v) Bayesian Additive Regression Trees (BART) (Chipman et al., 2010), a machine learning method that uses presence-absence data to estimate the probability of presence based on a cumulative set of decision trees using Bayesian formulation.

2.3.1. SDM calibration and prediction

To determine the minimal adequate model for each statistical approach, we followed an automated stepwise procedure based on Akaike's Information Criterion (AIC), using the *stepAIC* function from the R package *MASS* (Ripley et al., 2022) in the case of GLM and GAM and, in the case of BART, the automated procedure implemented with the *bart.step* function in the *embarcadero* R package (Carlson, 2020). Previously, we had used the *corSelect* function of the R package *fuzzySim* (Barbosa, 2015) to identify pairs of highly correlated variables ($R > 0.8$) and select, within each pair, the one with a smaller Variance Inflation Factor (VIF) for MAXENT, GLM and GAM models. BIOCLIM does not require a variable reduction step, as it does not penalise the use of redundant predictors and, instead, aims to reconstruct the most informative environmental envelope (Booth et al., 2014). Once the predictors were selected, we computed BIOCLIM, GLM and MAXENT models using the *dismo* R package (Hijmans et al., 2020), GAM using the *gam* R package (Hastie, 2020), and BART using the *dbarts* R package (Dorie et al., 2020). Models that require absence or background data (GLM, GAM, MAXENT and BART) were calibrated only with data from the invaded area, in order to use relatively reliable absence or unoccupied background cells, as the scarcity of records in the native area indicates it has been under-surveyed. On the contrary, presence-only models (BIOCLIM) were separately built with two datasets: (i) only data from the invaded area, for comparison with the other modelling approaches, and (ii) data from both the native and the invaded areas, to define an environmental envelope with the most complete available dataset (Guisan and Thuiller, 2005) (see Appendix S3 in Supplementary Material for details).

We projected all models onto the European grid to create an invasion risk cartography for each calibration subset and SDM approach. We removed the effect of sample prevalence from presence-absence models (GLM, GAM and BART) to obtain a value of environmental favourability (Real et al., 2006) in each grid cell. In the case of BIOCLIM, we here introduce the rescaling of model predictions into percentiles, using the 'quantReclass' function (Appendix S4 in Supplementary Material, and also implemented in the *modEvA* R package (Barbosa et al., 2023)). This function rescales the predictions into a ranked suitability value for each grid cell, thus informing about the suitability of climatic conditions in relative terms and hence which locations are more suitable than others. Current implementations of BIOCLIM compute a percentile distribution of the values of each environmental variable at known presence locations. Then, the closer to the 50th percentile (the median), the more suitable each location is according to that variable (Hijmans et al., 2020). However, the more variables are included in the model, the less suitable any location is considered, because it is less likely to be close to the median for all variables, and more likely to be outside the range observed at known presences for at least one variable. Our proposed rescaling of Bioclim predictions removes the dependence of the potential distribution on the number of variables included, and it provides more realistic predictions (see Results).

2.3.2. Model validation

To evaluate the ability of SDMs to predict the potential distribution of *V. velutina* in the European continent, we validated each model with the most recent dataset (all presences recorded until 2021). This informs about whether range expansion of an invasive species can be predicted by models calibrated with the distribution observed several years ago (e. g., all records until 2013). We used two threshold-independent evaluation metrics that assess different aspects of model performance: i) Boyce's index (Hirzel et al., 2006) and ii) AUC (Fielding and Bell, 1997). In the case of models calibrated with all data (i.e. 2021-model calibration), we also performed a k-fold spatial block cross-validation (Valavi et al., 2019) to validate the current prediction of *V. velutina* expansion (i.e. with the most up-to-date dataset) while ensuring that we did not use the same dataset for calibration and evaluation (see Appendix S3 in Supplementary material for details).

To assess if SDMs could predict the spatio-temporal pattern of invasion (i.e. which localities are likely to be invaded first), we assessed how the distribution of predictions of models calibrated with 2013 data shifted from sites invaded earlier to sites invaded later in time, and to sites not invaded (yet). The reasoning we follow is that a model accurately accounts for the spatio-temporal pattern of invasion if predicted values are higher in locations invaded earlier, lower in those invaded later, and even lower in those not invaded yet (see Fig. 1, and Appendix S3 in Supplementary Material for details). To quantify this, we compared (i) kernel density and (ii) cumulative probability curves of model predictions for sites invaded at different times (i.e. 2013, 2015, 2017, 2019 and 2021). These curves should shift regularly as a function of the year of invasion if the model were actually predicting the spatio-temporal pattern of invasion.

Presence records of *V. velutina* are available from the Dryad Digital Repository doi:https://doi.org/10.5061/dryad.931zcrjr8. R code is available as Supplementary Material (Appendix S5).

3. Results

3.1. Database of presences in the native and invaded areas

We assembled a total of 159,010 records of *V. velutina* in the European continent, collected from 2004 to 2021 (Appendix S6 and Table S1 in Supplementary material). These records were transferred to a total of 6286 cells in a 10×10 km^2 grid. We also compiled 524 records in the native area, that corresponded to 275 cells in a 10×10 km^2 grid (Fig. S1 in Supplementary material).

3.2. Quantitative estimation of climatic disequilibrium

V. velutina was recorded over a wide climatic range, both in the native and invaded areas. For instance, Annual Mean Temperature (bio1) ranged from -6.3 °C to 26.9 °C in the native area and from -7.8 °C to 19.2 °C in the invaded area (see Table S2 in Supplementary material). Four main climatic dimensions were identified by the PCA analysis, with the first component corresponding to variables associated with temperature (RC1, explained variance = 42 %), the second component to precipitation (RC2, explained variance = 22 %), the third component to temperature range (RC3, explained variance = 17 %), and

the fourth component to isothermality (RC4, explained variance = 9 %, see Table S3 in Supplementary Material). Thus, we selected the following variables for the comparison of climatic niches: mean annual daily mean air temperatures averaged over one year (°C, bio1), precipitation amount of the wettest month (kg m^{-2}, bio13), annual range of air temperature (°C, bio7) and isothermality, the ratio of diurnal variation to annual variation in temperatures (°C, bio3).
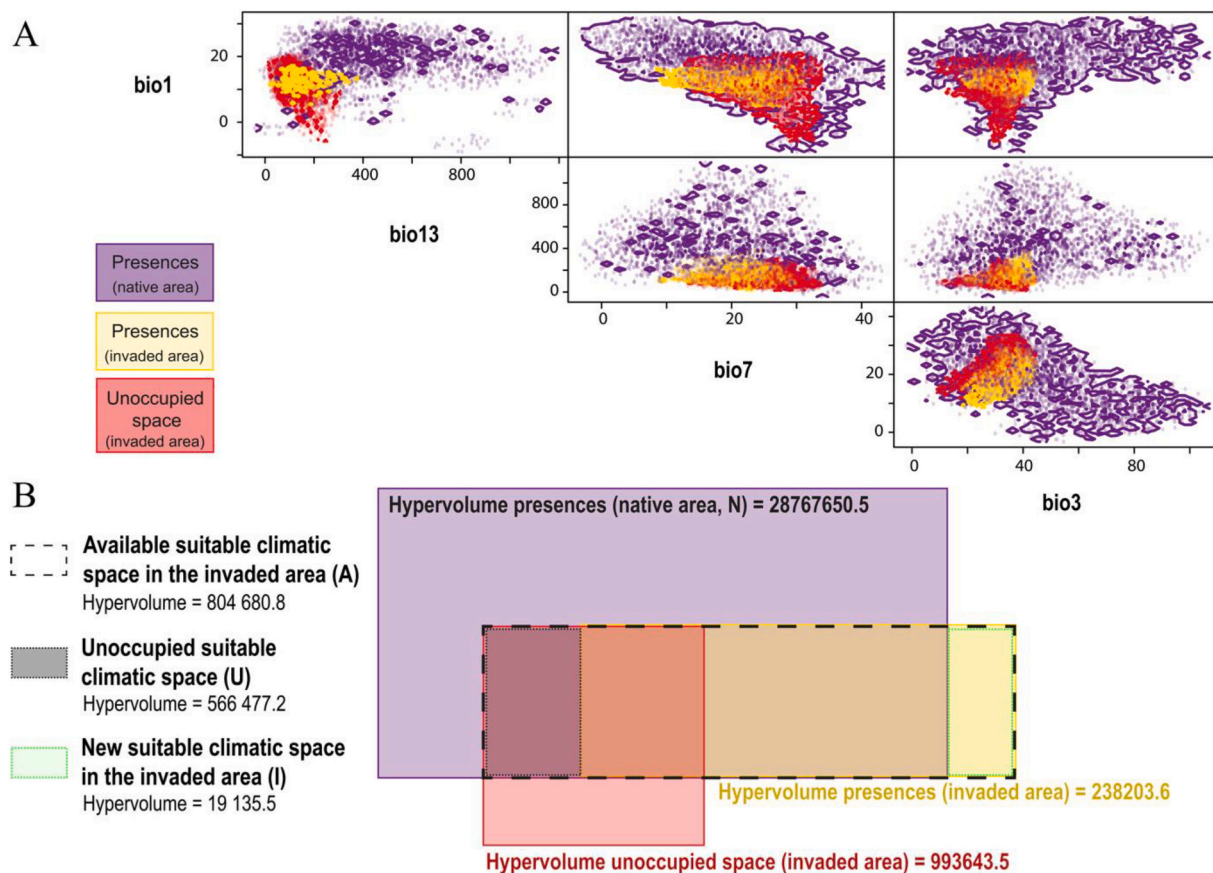
When we compared the climatic space in the native and invaded areas (Fig. 2), we observed that the unoccupied climatic space in Europe is mostly nested within the occupied climatic space in the native area, and that the occupied climatic space in the invaded area is nested within both (Fig. 2-A). This result points to a deviation from equilibrium with climate in Europe, given that there are regions with similar climatic conditions as those in the native area that have not been invaded yet. In Fig. 2, this corresponds to the climatic space where the red and purple points overlap. With the overlap information and the hypervolumes of each of the study areas (see Fig. S2 in Supplementary material), we can estimate the degree of climatic disequilibrium of *V. velutina* in Europe as follows: First, we estimate the hypervolume of the available climatic niche in Europe (available suitable climatic space, A = 804,680.8) as the sum of the fractions of occupied and unoccupied climatic spaces that are nested within the realised niche of the species. Second, we estimate the

hypervolume of the unoccupied climatic space that is nested within the climatic conditions already occupied by the species in the native area (unoccupied suitable climatic space, U = 566,477.2). Third, we estimate the hypervolume of the occupied climatic space in the invaded area that does not overlap with the climatic space occupied in the native area (new climatic space in the invaded area, I = 19,135.5). In this way, we introduce a measure of disequilibrium (D) as the ratio between unoccupied suitable climatic space and the available suitable climatic space (D = U/A). Finally, we measure the niche shift (N_S) as the ratio between the new climatic space in the invaded area, and the climatic space occupied in the native area (N) (N_S = I/N). For *V. velutina* in Europe, we found a disequilibrium of D = 0.70, which suggests that the species is still far from occupying all suitable climatic space in the invaded area. We also observed a N_S = 0.0007, which suggests practically no niche shift so far.

3.3. Retrospective assessment of invasion risk predictions

3.3.1. SDM calibration and prediction

The number of predictors included in the final models ranged from five (BART 2021) to nine (BART 2013). Only isothermality (bio3) and precipitation seasonality (bio15) were included in all models (see Table


Fig. 2. A) Representation of the climatic space occupied in Europe, in the native area and the unoccupied climatic space in Europe. The climatic space has been summarized into four main climatic gradients, identified by a Principal Component Analysis: mean annual daily mean air temperatures averaged over one year (bio1), precipitation amount of the wettest month (bio13), annual range of air temperature (bio7) and isothermality, the ratio of diurnal variation to annual variation in temperatures (bio3). Yellow and red represent the occupied niche and the unoccupied climatic space in the invaded area, and violet represents the climatic niche of the species in its native area. B) Conceptual representation of the calculations of the suitable climatic space (occupied and unoccupied) in Europe, which is used for the estimation of climatic disequilibrium and niche shift in the invaded area. Disequilibrium (D) is the ratio between the unoccupied suitable climatic space (U) and the available suitable climatic space in the invaded area (A). Niche shift (N_S) is the ratio between the new suitable climatic space in the invaded area (I) and the climatic space occupied in the native area (N). Areas in panel B are not to scale, so they are not proportional to the actual climatic hypervolumes. Please also note that hypervolume values depend on the random sampling of points for kernel estimation by function hypervolume(). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

S4 in Supplementary material for details). Nevertheless, all models included variables related to temperature, precipitation, seasonality and isothermality.

All presence-absence and presence-background models calibrated with 2021-data predicted similar potential distributions, including territories where the species is well established, such as northern and western Iberia, France or Belgium, as well as wider distributions in recently or still scarcely invaded territories, such as the Netherlands and north-western Germany. In addition, all models agreed that the Italian peninsula and the British Isles are favourable for the establishment of the species. In contrast, presence-only models (BIOCLIM), especially the one calibrated with data from both the invaded and native areas, predicted wider potential distributions, mostly towards the east and the north of the continent. Thus, presence-absence and presence-background predictions based on the invaded area were nested within the presence-only predictions based on native and invaded areas, with GLM, GAM and MAXENT predicting approximately 60 % of the potential distribution area predicted by BIOCLIM (invaded and native areas), while BART predicted 30 % of it.

Similar results were observed for SDMs calibrated with 2013 data (Fig. 3). BIOCLIM models predicted a larger potential distribution area, mostly when calibrated with data from both the native and invaded areas. However, the differences between BIOCLIM and the rest of the models were bigger than in 2021, mainly due to the scarcity of records in 2013, so GLM, GAM and BART predicted 20 % of the area predicted by BIOCLIM (invaded and native areas), while MAXENT predicted 15 % of it.
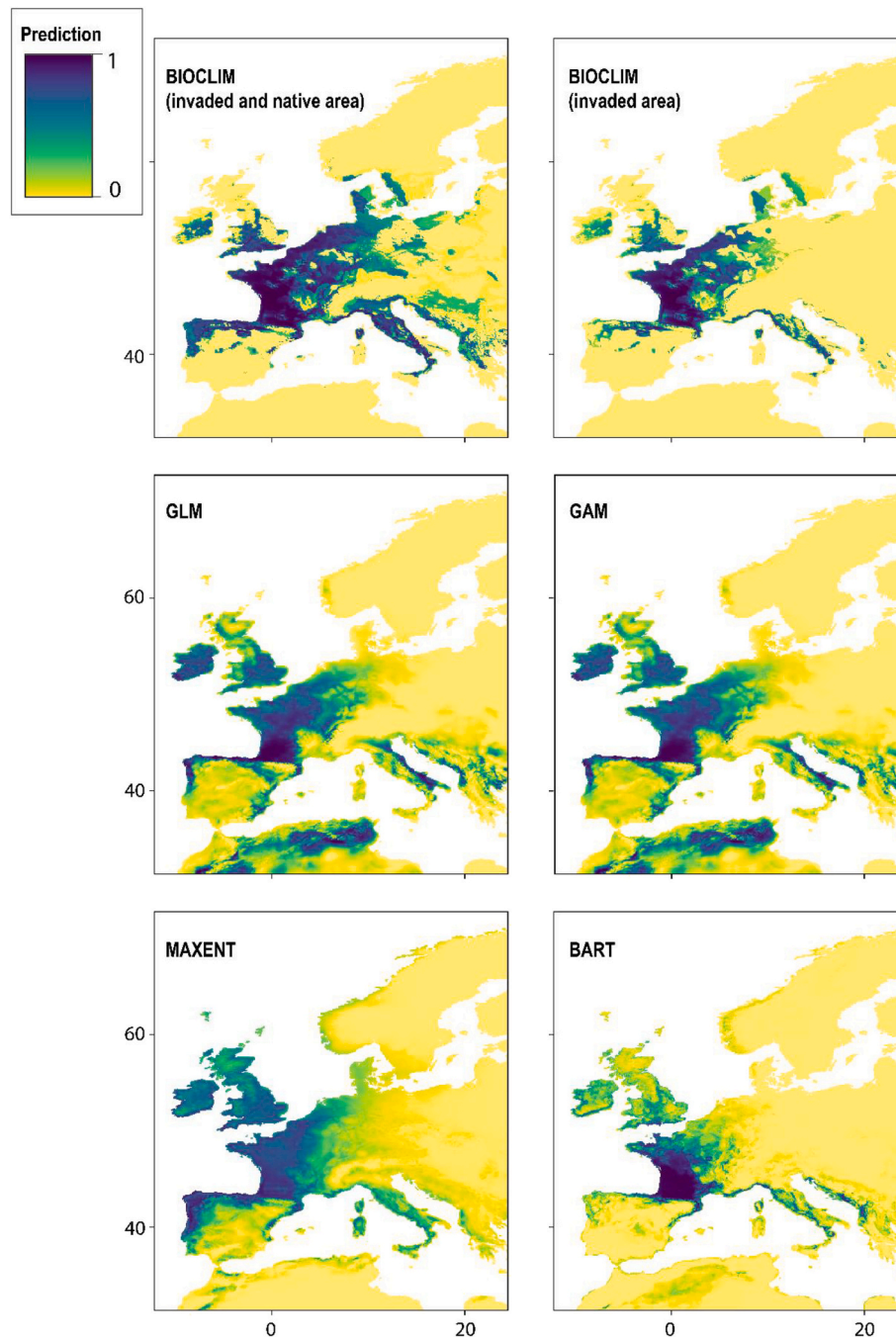
Fig. 3. Prediction of *V. velutina* distribution according to different correlative SDMs calibrated with presences observed until 2013. BIOCLIM prediction has been converted into *ranked suitability* values and GLM, GAM, and BART into favourability values according to the species prevalence.

3.3.2. Model validation

To validate the SDMs against the species distribution in 2021, we measured the AUC and Boyce indices (Table S5 in Supplementary material). All SDMs discriminated well between presences and absences (AUC ≈ 0.70 or higher). However, there was a tendency towards lower AUC values when models were calibrated with more recent data, after range expansion, which is consistent with the AUC being negatively correlated with prevalence or with relative occurrence area (e.g., Lobo et al., 2008). In fact, the largest AUC values for each modelling approach were observed when calibrating with 2013 data, except in the case of MAXENT and BIOCLIM (invaded area). Similarly, predictions differed markedly from a random prediction (Boyce ≈1, Table S5 and Fig. S3 in Supplementary material). This result supports the ability of all models to predict areas that will be invaded several years later (i.e. eight years, from 2013 to 2021). We also performed a *k-fold spatial block cross-validation* (Valavi et al., 2019) for models calibrated with 2021 data, to validate the current prediction of *V. velutina* expansion (i.e. with the most up-to-date dataset) while ensuring that we did not use the same dataset for calibration and evaluation. In this case, presence-absence models showed higher AUC values than presence-only and presence-background models (Table S5).

To assess whether SDMs could predict the spatio-temporal pattern of invasion, we measured the shift in the density curves of model predictions according to the year in which the presence was first observed at a locality, as well as to those in which the species has never been recorded to the moment. The density curves of BIOCLIM, GLM and GAM correctly reflected the spatio-temporal pattern of invasion, with the density curves shifting to the left with time and the cumulative probability curves shifting downwards from recently to formerly invaded areas (Fig. 4). However, the BIOCLIM model calibrated with data from

the native and invaded areas was the only one that had a relatively flat curve in the lower range of prediction values (i.e. from 0 to 0.5), evidencing that fewer presences observed in later years had been given low predicted suitability values. This is also observed with the cumulative probability curves, as these probabilities were low and sequentially increased from 2013 to 2021 (see Fig. 4). In contrast, in the presence-absence and presence-background model predictions, the spatio-temporal pattern is not clearly discernible and, importantly, in sites with confirmed presences in later years they predict very low values of favourability/suitability (see Fig. S4), virtually identical to sites that were never occupied.

4. Discussion

Our results show that presence-only models such as BIOCLIM can effectively predict the spatio-temporal pattern of invasion of *V. velutina* when predictions are rescaled based on percentiles and the model is calibrated with data from both the native and invaded areas. Adopting a non-equilibrium perspective is essential when modelling invasive species, because many absences are expected to turn into presences as the species expands its range towards new territories. While presence-only models would fulfil such requirement from a conceptual standpoint, they are not usually favoured in this type of studies (Elith et al., 2020; Franklin, 2023). However, the percentile-based rescaling procedure proposed here shows its potential usefulness in a temporal context, given that rescaled BIOCLIM models predicted higher suitability values for sites that were invaded earlier in time and, importantly, did not predict low suitability values for sites that were eventually occupied years later. The better performance of such presence-only models can be attributed to the fact that, by construction, these models are less biased by the lack
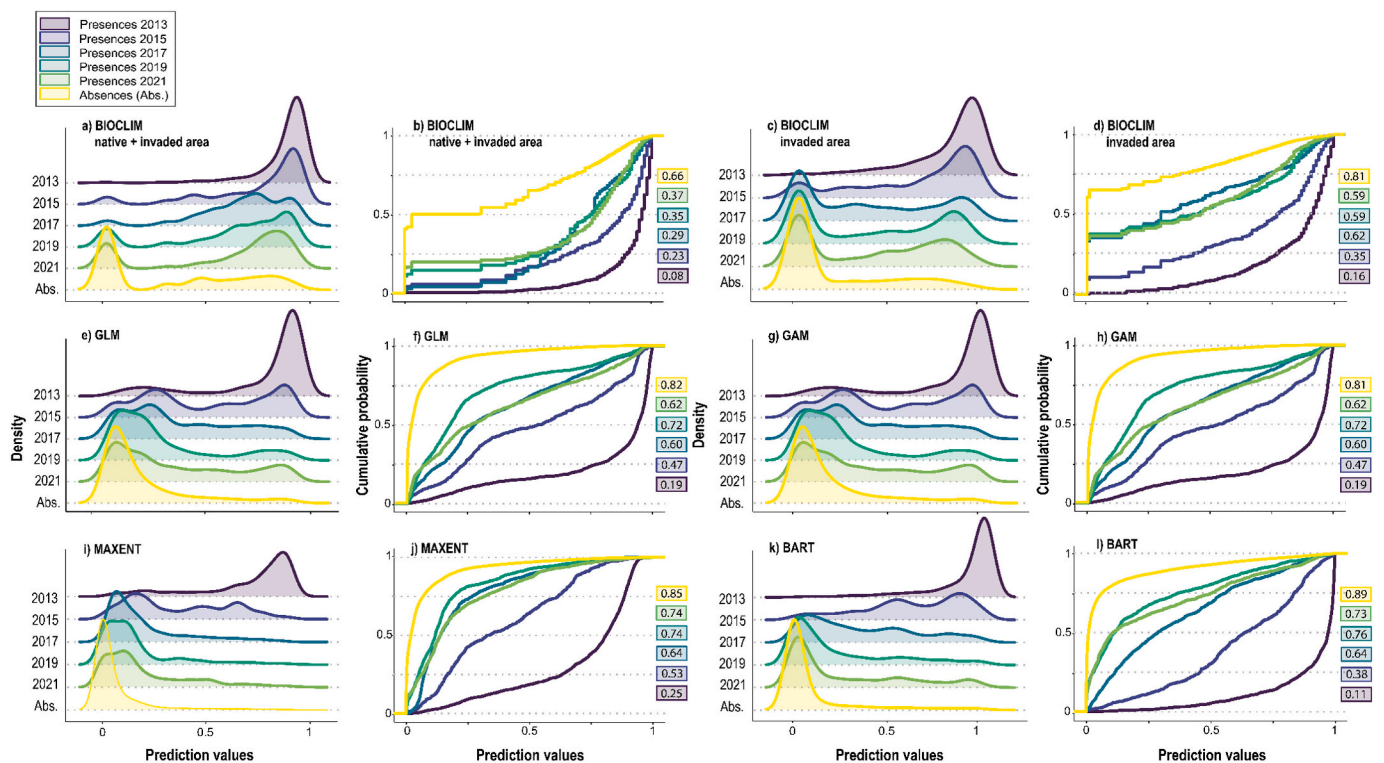
Fig. 4. Density plots of the distribution of 2013-model predictions for cells where presence data were observed in the following years (2013, 2015, 2017, 2019 and 2021) or were never colonised (absences). A curve is shown for absences and for each comparison year, which includes only data for the cells in which the presence was recorded, for the first time, that year. If models were able to predict the spatio-temporal pattern of invasion, a leftward shift with year and then to absences (from violet to yellow curves) is expected (panels a, c, e, g, i, k). Cumulative probability curves of 2013 predictions compared to presence data recorded from 2013 to 2021 and absences. If models were able to predict the spatio-temporal pattern of invasion, the cumulative probability curves are expected to shift downwards from absences to invaded areas recently and then to formerly invaded ones, i.e. from yellow to violet (panels b, d, f, h, j, l). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

of equilibrium, given that (i) presences from under-surveyed native areas can be also used to calibrate the model, thus identifying a wider range of climatic conditions that are suitable for the species, even if the fundamental niche may not be fully represented, and (ii) model performance is not compromised by the potential unreliability of absence data in the yet-to-be invaded areas (i.e. many suitable areas are likely unoccupied because the invasive species has not reached them yet, and not because they are unsuitable). The ability to predict the spatio-temporal pattern of invasion with presence-only models was possible by introducing a transformation of model predictions to *ranked suitability* values based on percentiles. This methodology can be very useful for decision-makers and managers, as it provides a probabilistic prediction of both (i) the temporal dimension of species range expansion, and (ii) the potential distribution range in the long term. Such information would allow prioritizing monitoring efforts in the short term, without losing sight of the potential risks in the long term.

Deviations from equilibrium between species distribution ranges and climatic conditions can be expected due to the very nature of the invasion process (Barbet-Massin et al., 2018; Pili et al., 2020), but they should still be assessed for each species under study (Boyd et al., 2023). Towards this goal, here we have used hypervolumes to compare the climatic space occupied in the native and invaded areas, as well as with the available, but not yet occupied, climatic space in the invaded area. From these hypervolumes, a newly introduced measure of disequilibrium (D) can be computed. We note that D is a relative measure, because its absolute value might depend on the selected climatic variables and/or the number of dimensions used to estimate the climatic space. However, despite of being relative, this measure of disequilibrium would still be useful for comparison among time periods, species and/or geographic regions. In this study, the lack of equilibrium was confirmed for *V. velutina* by the existence of climatic conditions in European regions not occupied by the species that are within the occupied climatic space in the native area. Deviations from equilibrium imply that absence data collected in the invaded area cannot be used to infer that such climatic conditions are unsuitable for the species (Václavík and Meentemeyer, 2009) and hence should be considered unreliable absences (Chefaoui and Lobo, 2008; Barbet-Massin et al., 2012). In this context, we would rather opt for the term "temporary absences", i.e., places where the species has not yet arrived. It should also be noted that our results currently support the hypothesis of climatic niche conservatism in the invaded range, as opposed to the presence of niche shifts (Broennimann et al., 2021) between the invaded and native areas. A similar result has been observed for a closely related invasive species, *V. mandarinia* (Zhu et al., 2020), and it supports the use of presence data from the native area in model calibration to reduce biases introduced by the lack of equilibrium with climate in the invaded range (Franklin, 2023).

Rescaled presence-only models calibrated with data from both the native and invaded areas outperformed other modelling approaches in terms of predicting the spatio-temporal dimension of invasion, as well as the species distribution range in the medium term. BIOCLIM models delimit the environmental space tolerable by a species as a set of n-dimensional "boxes" bounded by the extreme and core conditions endured by the species (Booth et al., 2014). Thus, the incorporation of climatic data from the native area improves the identification of such conditions that represent the boundaries of the species' climatic niche (Guisan and Thuiller, 2005; Broennimann and Guisan, 2008). It can then be assumed that the sites with climatic data within an n-dimensional box are suitable for the species, and the rest are unsuitable (Beaumont et al., 2005). This has led to presence-only predictions being often interpreted in such qualitative terms (suitable vs. unsuitable), thus giving equal weight to climatic conditions near the limits of the species climatic niche as to climatic conditions near its niche optimum. Current implementations of the BIOCLIM method (e.g., in R package 'dismo') provide a continuous prediction by computing the environmental similarity of a location to a percentile distribution of the environmental values at

species occurrence locations (Hijmans et al., 2020). However, the size of the predicted distributions is negatively correlated with the number of variables in the model. Thus, while including more variables should, in theory, improve the information in the model, it fatally results in smaller predicted ranges. We have overcome this limitation by reclassifying BIOCLIM predictions into percentiles, in what we call *ranked suitability*. This transformation allows a simple and intuitive method such as BIOCLIM, even when built from distributions far from equilibrium in the invaded area, to adequately predict the spatio-temporal pattern of species range expansion, with the great advantages of transparency, ease of understanding, and lack of need for (pseudo)absence or unoccupied background data.

It should be noted that when a species is far from equilibrium in the native area, presence data would still represent an incomplete fraction of the fundamental niche. In that situation, our ability to estimate the degree of equilibrium will be hampered and the performance of all models, including presence-only ones, will be limited. Moreover, if presence data is incomplete due to sampling biases, we expect an overestimation of disequilibrium and, again, a lower performance of all models, including presence-only ones. Regardless, BIOCLIM has several potential advantages over more complex methods which can make it particularly useful in invasive species modelling, such as 1) being able to produce models even when very limited occurrence data are available; 2) making fewer assumptions of constancy in the species-climate relationships; 3) not being biased by correlations among predictor variables, which may differ between native and invaded areas; and 4) its rectilinear shapes including climates that are part of the fundamental niche but may be unavailable in invaded areas, thus not overly restricting niche envelopes (Rödder et al., 2013; Guillory and Brown, 2021).

In conclusion, the ability to predict the spatio-temporal pattern of invasion using presence-only models can become a valuable tool for optimal surveillance of invasive species, as it would allow identifying which locations will be colonised first, thus optimizing early monitoring and timely action (Barbet-Massin et al., 2018). Therefore, predictions of invasion risk would inform not only about where climatic conditions are suitable for the establishment of the species, but also the sequential order in which the species will likely become established in those areas. This approach is particularly useful in non-equilibrium scenarios, in which many of the absence data can be regarded as temporary, as they will inevitably turn into presence data with time. Our study shows that, by rescaling BIOCLIM predictions, the predictive capacity of this simple and easy-to-understand model can be visibly improved, even accounting for the temporal dimension of invasion long before the species approaches climatic equilibrium.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.biocon.2023.110361.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Victoria Formoso-Freire reports financial support was provided by Interreg Atlantic Area Program.

Data availability

Presence records of *V. velutina* are available from the Dryad Digital

Repository (doi: https://doi.org/10.5061/dryad.931zcrjr8).

Acknowledgments

References

Araújo, M.B., Peterson, A.T., 2012. Uses and misuses of bioclimatic envelope modeling. Ecology 93 (7), 1527–1539. https://doi.org/10.1890/11-1930.1.

Araújo, M.B., Anderson, R.P., Márcia Barbosa, A., Beale, C.M., Dormann, C.F., Early, R., Rahbek, C., 2019. Standards for distribution models in biodiversity assessments. *Science*. Advances 5 (1), eaat4858. https://doi.org/10.1126/sciadv.aat4858.

Baquero, R.A., Barbosa, A.M., Ayllón, D., Guerra, C., Sánchez, E., Araújo, M.B., Nicola, G.G., 2021. Potential distributions of invasive vertebrates in the Iberian Peninsula under projected changes in climate extreme events. Diversity and Distributions 27 (11), 2262–2276. https://doi.org/10.1111/ddi.13401.

Barbet-Massin, M., Jiguet, F., Albert, C.H., Thuiller, W., 2012. Selecting pseudo-absences for species distribution models: how, where and how many?: how to use pseudo-absences in niche modelling? Methods Ecol. Evol. 3 (2), 327–338. https://doi.org/10.1111/j.2041-210X.2011.00172.x.

Barbet-Massin, M., Rome, Q., Muller, F., Perrard, A., Villemant, C., Jiguet, F., 2013. Climate change increases the risk of invasion by the yellow-legged hornet. Biol. Conserv. 157, 4–10. https://doi.org/10.1016/j.biocon.2012.09.015.

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

Barbosa, A.M., 2015. fuzzySim: applying fuzzy logic to binary similarity indices in ecology. Methods in Ecology and Evolution 6 (7), 853–858. https://doi.org/10.1111/2041-210X.12372.

Barbosa, A.M., Brown, J.L., Jiménez-Valverde, A., Real, R., 2023. Model evaluation and analysis (3.10). https://cran.r-project.org/web/packages/modeva.

Beaumont, L.J., Hughes, L., Poulsen, M., 2005. Predicting species distributions: use of climatic parameters in BIOCLIM and its impact on predictions of species' current and future distributions. Ecol. Model. 186 (2), 251–270. https://doi.org/10.1016/j.ecolmodel.2005.01.030.

Blonder, B., 2022. Hypervolume: High Dimensional Geometry, Set Operations, Projection, and Inference Using Kernel Density Estimation, Support Vector Machines, and Convex Hulls (3.0.4). https://CRAN.R-project.org/package=hypervolume.

Blonder, B., Lamanna, C., Violle, C., Enquist, B.J., 2014. The n-dimensional hypervolume. Glob. Ecol. Biogeogr. 23 (5), 595–609. https://doi.org/10.1111/geb.12146.

Booth, T.H., Nix, H.A., Busby, J., Hutchinson, M., 2014. BIOCLIM: the first species distribution modelling package, its early applications and relevance to most current MAXENT studies. Diversity and Distributions 20, 1–9. https://doi.org/10.1111/ddi.12144.

Botella, C., Bonnet, P., Hui, C., Joly, A., Richardson, D.M., 2022. Dynamic species distribution modeling reveals the pivotal role of human-mediated long-distance dispersal in plant invasion. Biology 11 (9), 1293. https://doi.org/10.3390/biology11091293.

Boyd, R.J., Harvey, M., Roy, D.B., Barber, T., Haysom, K.A., Macadam, C.R., Pescott, O.L., 2023. Causal inference and large-scale expert validation shed light on the drivers of SDM accuracy and variance. Divers. Distrib. 1-11 https://doi.org/10.1111/ddi.13698.

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

Broennimann, O., Petitpierre, B., Chevalier, M., González-Suárez, M., Jeschke, J.M., Rolland, J., Guisan, A., 2021. Distance to native climatic niche margins explains establishment success of alien mammals. Nat. Commun. 12 (1), 2353. https://doi.org/10.1038/s41467-021-22693-0.

Busby, J., 1991. BIOCLIM: a bioclimate analysis and prediction system. Plant Protection Quarterly 8, 6–9.

Carlson, C.J., 2020. embarcadero: species distribution modelling with Bayesian additive regression trees in r. Methods in Ecology and Evolution 11 (7), 850–858. https://doi.org/10.1111/2041-210X.13389.

Chefaoui, R.M., Lobo, J.M., 2008. Assessing the effects of pseudo-absences on predictive distribution model performance. Ecol. Model. 210 (4), 478–486. https://doi.org/10.1016/j.ecolmodel.2007.08.010.

Chipman, H.A., George, E.I., McCulloch, R.E., 2010. BART: Bayesian additive regression trees. Ann. Appl. Stat. 4 (1), 266–298. https://doi.org/10.1214/09-AOAS285.

Dorie, V., Chipman, H.A., McCulloch, R.E., Dadgar, A., 2020. dbarts: discrete Bayesian additive regression trees sampler. https://cran.r-project.org/web/packages/dbarts.

Elith, J., Graham, C., Valavi, R., Abegg, M., Bruce, C., Ford, A., Zimmermann, N.E., 2020. Presence-only and presence-absence data for comparing species distribution modeling methods. Biodivers. Inform. 15 (2), 69–80. https://doi.org/10.17161/bi.v15i2.13384.

Ferraz, K.M.P.M.B., Morato, R.G., Bovo, A.A.A., Costa, C.O.R., Ribeiro, Y.G.G., Paula, R.C., Traylor-Holzer, K., 2021. Bridging the gap between researchers, conservation planners, and decision makers to improve species conservation decision-making. Conservation Science and Practice 3 (2). https://doi.org/10.1111/csp2.330.

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

Foster, D., Swanson, F., Aber, J., Burke, I., Brokaw, N., Tilman, D., Knapp, A., 2003. The importance of land-use legacies to ecology and conservation. BioScience 53 (1), 77. https://doi.org/10.1641/0006-3568(2003)053[0077:TIOLUL]2.0.CO;2.

Foster, S.L., Kharouba, H.M., Smith, T.W., 2022. Testing the assumption of environmental equilibrium in an invasive plant species over a 130 year history. Ecography 2022 (12), e12933. https://doi.org/10.1111/ecog.06284.

Fournier, A., Barbet-Massin, M., Rome, Q., Courchamp, F., 2017. Predicting species distribution combining multi-scale drivers. Global Ecology and Conservation 12, 215–226. https://doi.org/10.1016/j.gecco.2017.11.002.

Franklin, J., 2023. Species distribution modelling supports the study of past, present and future biogeographies. J. Biogeogr. 50, 1533–1545. https://doi.org/10.1111/jbi.14617.

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: beyond the equilibrium assumption of SDMs. Glob. Ecol. Biogeogr. 21 (11), 1126–1136. https://doi.org/10.1111/j.1466-8238.2012.00768.x.

Guillory, W.X., Brown, J.L., 2021. A new method for integrating ecological niche modeling with phylogenetics to estimate ancestral distributions. Syst. Biol. 70 (5), 1033–1045. https://doi.org/10.1093/sysbio/syab016.

Guisan, A., Thuiller, W., 2005. Predicting species distribution: offering more than simple habitat models. Ecol. Lett. 8 (9), 993–1009. https://doi.org/10.1111/j.1461-0248.2005.00792.x.

Guisan, A., Thuiller, W., Zimmermann, N.E., 2017. Habitat Suitability and Distribution Models with Applications in R. Cambridge University Press.

Hastie, T., 2020. gam: generalized additive models (1.20.1). https://CRAN.R-project.org/package=gam.

Hastie, T., Tibshirani, R., 1987. Generalized additive models: some applications. J. Am. Stat. Assoc. 82 (398), 371–386. https://doi.org/10.1080/01621459.1987.10478440.

Hijmans, R.J., Phillips, S., Leathwick, J., Elith, J., 2020. dismo: species distribution modelling (1.3.5). https://CRAN.R-project.org/package=dismo.

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

Jiménez, L., Soberón, J., 2022. Estimating the fundamental niche: accounting for the uneven availability of existing climates in the calibration area. Ecol. Model. 464 (109823), 1–13. https://doi.org/10.1016/j.ecolmodel.2021.109823.

Jones, C., Skrip, M.M., Seliger, B.J., Jones, S., Wakie, T., Takeuchi, Y., Meentemeyer, R.K., 2022. Spotted lanternfly predicted to establish in California by 2033 without preventative management. Communications Biology 5 (1), 558. https://doi.org/10.1038/s42003-022-03447-0.

Karger, D.N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R.W., Kessler, M., 2017. Climatologies at high resolution for the earth's land surface areas. Scientific Data 4 (1). https://doi.org/10.1038/sdata.2017.122. Article 1.

Lobo, J.M., Jiménez-Valverde, A., Real, R., 2008. AUC: a misleading measure of the performance of predictive distribution models. Glob. Ecol. Biogeogr. 17 (2), 145–151. https://doi.org/10.1111/j.1466-8238.2007.00358.x.

McCullagh, P., Nelder, J.A., 1989. Generalized Linear Models, 2nd ed. Chapman & Hall, London.

Meentemeyer, R.K., Cunniffe, N.J., Cook, A.R., Filipe, J.A.N., Hunter, R.D., Rizzo, D.M., Gilligan, C.A., 2011. Epidemiological modeling of invasion in heterogeneous landscapes: spread of sudden oak death in California (1990–2030). Ecosphere 2 (2), art17. https://doi.org/10.1890/ES10-00192.1.

Monceau, K., Bonnard, O., Thiéry, D., 2014. *Vespa velutina*: a new invasive predator of honeybees in Europe. J. Pest. Sci. 87 (1), 1–16. https://doi.org/10.1007/s10340-013-0537-3.

Pereira, P.F., Barbosa, A.M., Godinho, C., Salgueiro, P.A., Silva, R.R., Lourenço, R., 2020. The spread of the red-billed leiothrix (*Leiothrix lutea*) in Europe: the conquest by an overlooked invader? Biol. Invasions 22 (2), 709–722. https://doi.org/10.1007/s10530-019-02123-5.

Phillips, S.J., Anderson, R.P., Schapire, R.E., 2006. Maximum entropy modeling of species geographic distributions. Ecol. Model. 190 (3), 231–259. https://doi.org/10.1016/j.ecolmodel.2005.03.026.

Pili, A.N., Tingley, R., Sy, E.Y., Diesmos, M.L.L., Diesmos, A.C., 2020. Niche shifts and environmental non-equilibrium undermine the usefulness of ecological niche models for invasion risk assessments. Sci. Rep. 10 (1), 7972. https://doi.org/10.1038/s41598-020-64568-2.

R Core Team, 2019. R: A Language and Environment for Statistical Computing. https://www.Rproject.org.

Real, R., Barbosa, A.M., Vargas, J.M., 2006. Obtaining environmental favourability functions from logistic regression. Environ. Ecol. Stat. 13 (2), 237–245. https://doi.org/10.1007/s10651-005-0003-3.

Revelle, W., 2022. psych: procedures for psychological, psychometric, and personality (2.2.5). https://CRAN.R-project.org/package=psych/.

Ripley, B., Venables, B., Bates, D.M., Hornik, K., Gebhardt, A., Firth, D., 2022. Supports functions and datasets for Venables and Ripple's MASS. (7.3-58). https://cran.r-project.org/web/packages/MASS.

Rödder, D., Lawing, A.M., Flecks, M., Ahmadzadeh, F., Dambach, J., Engler, J.O., Habel, J.C., Hartmann, T., Hörnes, D., Ihlow, F., Schidelko, K., Stiels, D., Polly, P.D.,

2013. Evaluating the significance of paleophylogeographic species distribution models in reconstructing Quaternary range-shifts of nearctic chelonians. PloS One 8 (10), e72855. https://doi.org/10.1371/journal.pone.0072855.

Václavík, T., Meentemeyer, R.K., 2009. Invasive species distribution modeling (iSDM): are absence data and dispersal constraints needed to predict actual distributions? Ecol. Model. 220 (23), 3248–3258. https://doi.org/10.1016/j.ecolmodel.2009.08.013.

Václavík, T., Meentemeyer, R.K., 2012. Equilibrium or not? Modelling potential distribution of invasive species in different stages of invasion. Diversity and Distributions 18 (1), 73–83. https://doi.org/10.1111/j.1472-4642.2011.00854.x.

Valavi, R., Elith, J., Lahoz-Monfort, J.J., Guillera-Arroita, G., 2019. BlockCV: an R package for generating spatially or environmentally separated 1 folds for k-fold cross-validation of species distribution models. Methods in Ecology and Evolution 10 (2), 225–232. https://doi.org/10.1111/2041-210X.13107.

Villemant, C., Barbet-Massin, M., Perrard, A., Muller, F., Gargominy, O., Jiguet, F., Rome, Q., 2011. Predicting the invasion risk by the alien bee-hawking yellow-legged hornet *Vespa velutina nigrithorax* across Europe and other continents with niche models. Biol. Conserv. 144, 2150–2152. https://doi.org/10.1016/j.biocon.2011.04.009.

Zhu, G., Gutierrez Illan, J., Looney, C., Crowder, D.W., 2020. Assessing the ecological niche and invasion potential of the Asian giant hornet. Proc. Natl. Acad. Sci. 117 (40), 24646–24648. https://doi.org/10.1073/pnas.2011441117.