



A new chapter of the Japanese beetle invasion saga: predicting suitability from long-invaded areas to inform surveillance strategies in Europe

Leyli Borner^{1,*}, Davide Martinetti², Sylvain Poggi^{1,*}

¹ INRAE, Institut Agro, Univ Rennes, IGEPP, 35653, Le Rheu, France

² INRAE, UR 546 BioSP, Avignon, France

* Corresponding authors: leyli.borner@gmail.com; sylvain.poggi@inrae.fr

With 3 figures

Abstract: The Japanese beetle (*Popillia japonica*) has been listed as a priority pest in the EU since its introduction and gradual spread over the past decade. Our study presents a species distribution model that incorporates environmental factors relevant to the beetle's biology. Presence-only data from both native and long-invaded ranges (Japan, North America, and the Azores – Portugal) were used to train the model. These data were collected from both citizen-science platforms and standardised surveys. We adapted standard machine-learning techniques to account for the sampling bias inherent in opportunistic data. Suitability maps produced for the Northern Hemisphere show good agreement with actual beetle presence, especially in North America. They also show that Central Europe is generally suitable, while Northern European countries are at lower risk. A large cluster of high suitability is located in the currently invaded region, further confirming the robustness of our predictions, and extends eastwards towards the Balkans. Other clusters of moderate suitability are scattered across the central part of the continent, embedded in a matrix of low suitability. This picture highlights the importance of designing surveillance strategies considering both active insect dispersal and the possibility of hitchhiking to reach distant areas, and advocates for contingency planning at continental level.

Keywords: Citizen science; epidemic surveillance; machine learning; risk mapping; *Popillia japonica*; species distribution model

1 Introduction

The introduction and establishment of species outside their native range is a major issue due to their potential impact on ecosystems (Bellard et al. 2016), food security (Oerke 2006) and economic balances (Diagne et al. 2021). Globalization has greatly facilitated biological invasions, as movements of goods and people provide countless opportunities for pests to spread across different regions of the world (Hulme 2009). The Japanese beetle (*Popillia japonica* Newman; hereafter Pj) is a prime example of a successful biological invader that has spread outside its native range through human transport.

This species is native to Japan, and was first detected outside its native range in the United States of America in 1916 (Fleming 1972). Since this accidental introduction, the beetle has spread throughout North America. In 1974, it was detected on Terceira island in the Azores archipelago (Martins et al. 1988) and has now invaded all the islands except Santa Maria. More recently, Pj was detected in

Northern Italy in 2014 and in Switzerland in 2017 (EFSA Panel on Plant Health (PLH) et al. 2018). Interceptions have been observed in Central Europe, while stable populations are only found in the Italian regions of Lombardy, Piedmont, Valle d'Aosta and Emilia-Romagna; and in the Swiss region of Ticino (Poggi et al. 2022a). European plant health agencies classified Pj as a high-priority pest (Commission Delegated Regulation (EU) 2019), and Pj has been ranked second among quarantine pests in terms of economic, social and environmental impacts (Sanchez et al. 2019).

The challenge is to counter this invasion by containing established populations and by promoting early detection of new outbreaks to increase the chances of successful eradication. This brings the design of surveillance strategies to the forefront. The effectiveness of these strategies is critical, as management becomes disproportionately more costly and difficult as pest incidence increases (Parnell et al. 2014). To this end, several authors advocate the use of strategies based on the examination of risk maps, rather than systematic sam-

pling programmes that lead to suboptimal use of resources allocated to surveillance (Hyatt-Twynam et al. 2017).

The aim of the present article is to draw a reliable map of the suitability of Continental Europe for the Japanese beetle and to discuss implications for the design of surveillance and containment strategies. Pj suitability is estimated using a species distribution model (SDM) trained using a machine-learning approach based on presence-only data from native and long-invaded ranges (Japan, North America, Azores). The agreement of our suitability maps with current knowledge of invasion in long-invaded regions and official infestation status (NPB 2016) is reported in detail. We also discuss improvements over previous suitability predictions for Pj (Zhu et al. 2017; Kistner-Thomas 2019; Della Rocca & Milanese 2022a). We discuss how our predictions in Continental Europe can contribute to European policies and strategies for sustainable management of the invasive beetle.

2 Materials and methods

All data processing and analyses were performed using R version 4.1.2 (R Core Team 2021).

2.1 Predictors and presence-only data

2.1.1 Predictor variables

Previous studies have investigated the importance of predictors for the Japanese beetle's biology. Accordingly, we considered climatic variables and soil characteristics (Fleming 1972; Kistner-Thomas 2019; Simonetto et al. 2022), land use and human-related characteristics such as population and road densities (Zhu et al. 2017; Della Rocca & Milanese 2022b). As Pj's biological cycle is driven by seasons, we aggregated monthly climatic variables by season (Spring: March to May, Summer: June to August, Fall: September to November, Winter: December to February). A set of 133 variables was used in this study (Appendix 1), upscaled to a resolution of 4 km × 4 km (the coarsest resolution among predictors), averaging values contained within each cell where necessary.

2.1.2 Presence and pseudo-absence data

We considered georeferenced Pj presence data since 2010 from both citizen-science platforms (~37 000 observations) and standardised surveys (~11 000 observations) (Fig. 1A and Appendix 2). As reliable absence observations were not available in native and long-invaded areas, we generated pseudo-absences using the target-group method (Phillips et al. 2009). This method ensures that pseudo-absences suffer from the same sampling bias as the opportunistic citizen-science presence data. Thus, we considered ~715,000 opportunistic observations of *Coleoptera* (excluding *Popillia japonica*) collected between 2010 and 2021 (Fig. 1B and

Appendix 2). Data sources and pre-processing are described in Appendix 2.

The aggregation rule for presence and pseudo-absence data was defined as follows: a cell (4 km × 4 km) is labelled as "presence" if at least one observation of the Japanese beetle is found within the cell, whereas it is considered as "absence" if at least one *Coleoptera* and no Japanese beetle have been reported in the cell. This results in 6844 presence cells and 49010 pseudo-absence cells.

2.2 Modelling species distribution using a machine learning approach

2.2.1 Algorithm choice

Random Forest (hereafter RF) has been identified as one of the top-performing models for presence-only data in a recent review of SDM algorithms (Valavi et al. 2022). RF has also been demonstrated to be more robust to collinearity between predictors (Freeman et al. 2016). The RF algorithm also provides an estimate of variable importance.

2.2.2 Presence-only dataset for model fitting

In the case of biological invasions, it is recommended to use data from both native and invaded ranges in SDM, as shifts in species niche may occur during the invasion process (Broennimann & Guisan 2008). However, in recently invaded ranges, species distributions are likely to be impacted by dispersal limitations (Elith et al. 2010). We therefore deliberately excluded data from Continental Europe during model fitting, as they represent a recent and ongoing invasion process (less than 10 years since first detection) and considered only data from native (Japan) and long-invaded areas (North America and the Azores) (Fig. 2A–B).

2.3 Model fitting and evaluation

We fitted SDMs and measured their performance on data from 2010–2019. In total, 53210 cells are available for model fitting, of which 4200 represent presence data (c.a. 7%) (Fig. 2A–B). Such unbalanced datasets are known to affect RF predictions (Barbet-Massin et al. 2012). Therefore, we used repeated down-sampling (Japkowicz & Stephen 2002) and repeatedly fitted the model on a subset of the original data containing all presences and a subsample of pseudo-absences of similar size. Model evaluation using standard cross-validation based on random sampling of observations is known to underestimate prediction error when applied to ecological data with spatial structure (Roberts et al. 2017). Thus, we split our presence and pseudo-absence dataset into blocks based on environmental dissimilarity among observations (Fig. 2C–D), and used this structure both in down-sampling and in an environmental block cross-validation strategy. The construction of the environmental blocks is described in detail in Appendix 3, and ultimately resulted in the definition of 7 environmental blocks (Fig. 2C–D).

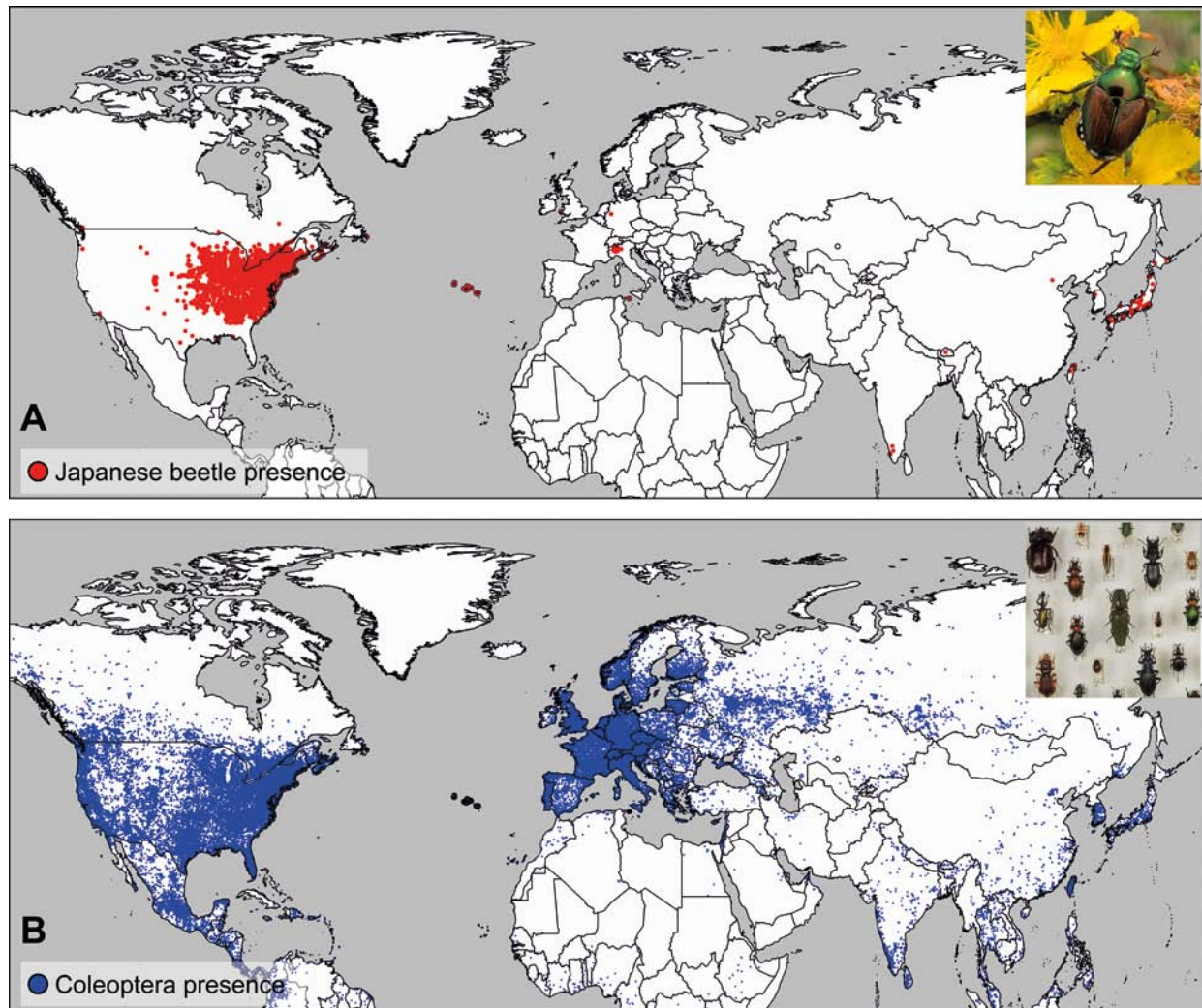


Fig. 1. Global distributions of Japanese beetle presence data and *Coleoptera* presence data. **A** Worldwide distribution of geo-referenced Japanese beetle presence data (in red), including interceptions, collected from citizen science platforms and standardised surveys over the period 1917–2021; **B** Worldwide distribution of *Coleoptera* presence data (target-group, in blue) extracted from the GBIF citizen science platform over the period 2010–2021. Insert photos: **A** Adult individual of *Popillia japonica*, © Leyli Borner, INRAE, **B** *Coleoptera* species from Franck Duval's private collection, © Stéphane Jumel, INRAE.

For each replicate (100 replicates total), we sampled within each block approximately the same number of pseudo-absences as the number of presences in the block, using a weighting scheme that prioritised observations further away from presences according to environmental dissimilarity (see Appendix 3). This ensures that data are balanced both within and between blocks, and that pseudo-absences (for which we have less confidence) are sampled at some environmental distance from presence data. We fitted a RF model to the subset and retrieved the variable importance as returned by the RF algorithm. We ranked predictors by importance, retaining only those that ranked higher than three randomly simulated variables added to the predictor set (Stoppiglia et al. 2003) and therefore contributed more than expected by chance. For

each replicate, we applied a 7-fold cross-validation strategy using the environmental blocks as folds. We calculated three model performance indices on the held-out fold: sensitivity, Area Under the ROC Curve (AUC) and the Continuous Boyce Index (CBI). CBI was designed to evaluate SDM with presence-only data and reflects the agreement between predicted occurrence probability and observed proportions of occupied sites (Hirzel et al. 2006). Finally, we also retrieved the predicted presence probability for all cells corresponding to Northern Hemisphere landmasses (8148045 cells).

2.4 Suitability maps

Our suitability maps were constructed as median RF predictions (100 replicates) expressed as probability of *Pj* presence

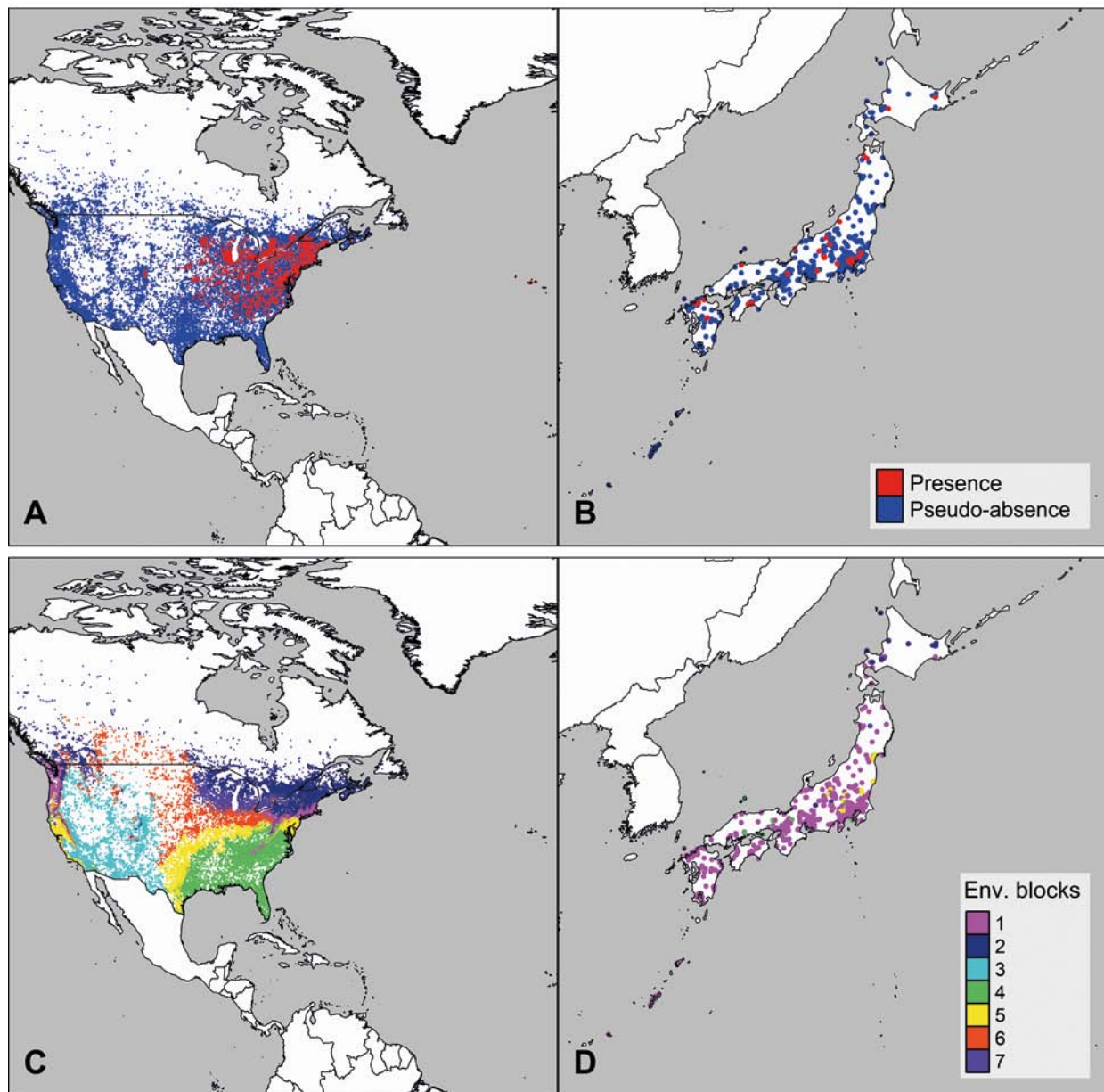


Fig. 2. Distribution of presence and pseudo-absence data used in model fitting. Distribution of Japanese beetle presence data (in red) and pseudo-absences (*Coleoptera* target-group, in blue) **A** in North America and **B** in Japan. Distribution of presence and pseudo-absence data among the environmental blocks used in model fitting ($N = 7$) **C** in North America and **D** in Japan.

(real values between 0 and 1). Following Hirzel et al. (2006), we used the plot of Boyce Predicted to Expected ratio against suitability (P/E curves) generated for each model replicate, to transform the continuous $[0,1]$ probability interval into a discrete number of classes. These classes should reflect model calibration and more objectively represent levels of suitability. This method provides a more honest discretization of the unit interval than an arbitrary choice of thresholds (see Appendix 4 for more details).

We computed the Multivariate Environmental Similarity Surface (Elith et al. 2010) to identify regions of the Northern Hemisphere where model predictions represent an extrapolation due to greater differences between local predictors and those used in model fitting (Fig. 3A and Appendix 4: Fig. A4.2).

Finally, we used a test dataset of presence data set aside from model training to provide an evaluation of the suitability maps by contrasting the predicted suitability with the

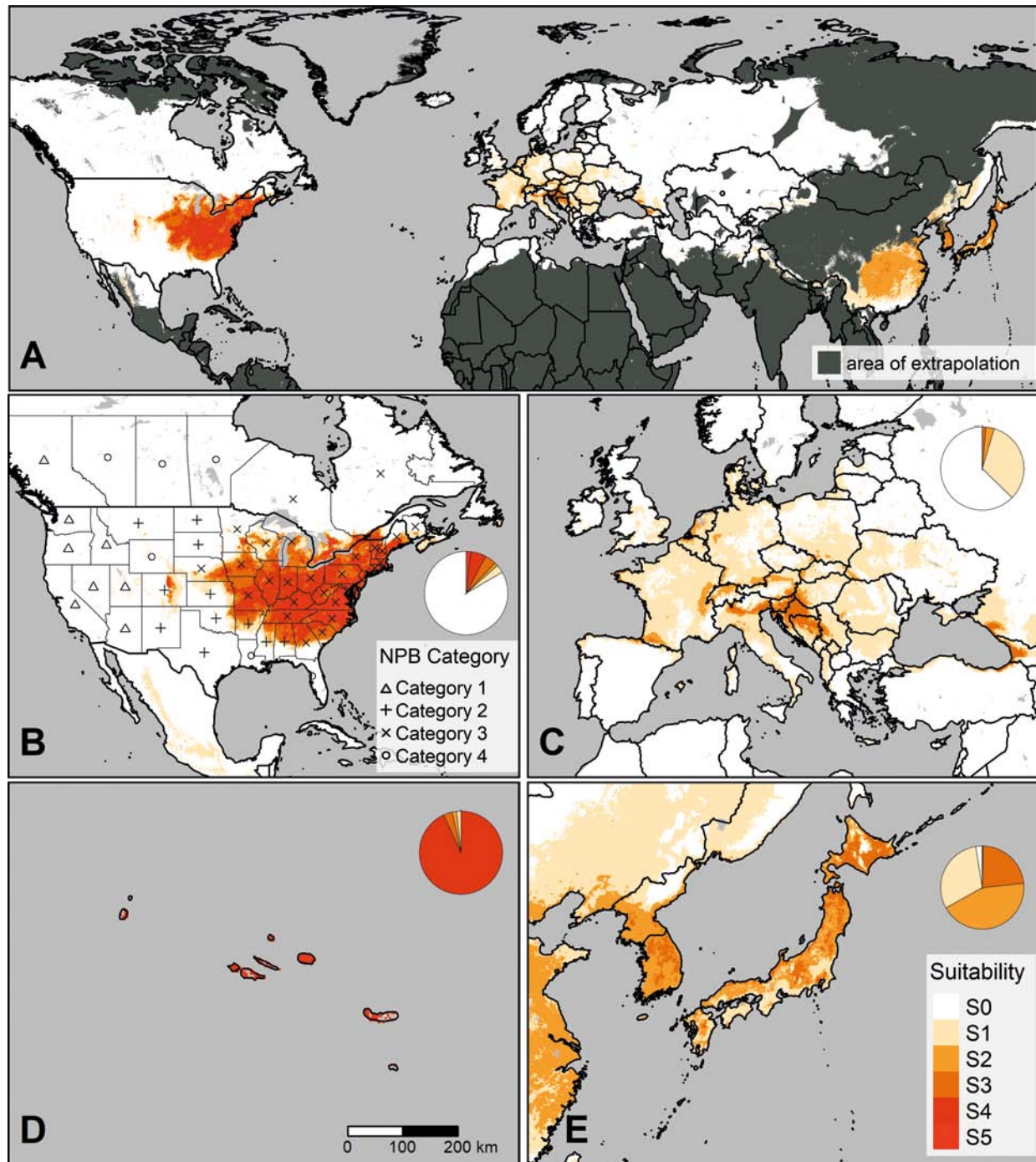


Fig. 3. Suitability maps for the Japanese beetle and composition of predicted suitability. **A** Northern Hemisphere, displaying areas of extrapolation in dark grey (Appendix 4), **B** North America, Japanese beetle-infested areas reported by the NPB (2016) are displayed by categories, **C** Continental Europe, **D** Azores archipelago, **E** Japan. In each panel, a pie chart depicts the composition of suitability, i.e. the proportion of total area covered by each suitability class. Darker colours correspond to greater suitability for the Japanese beetle. Pj infested areas categories according to NPB (2016): Category 1 = "Uninfested/Quarantine Pest", Category 2 = "Uninfested or Partially Infested", Category 3 = "Partially or Generally Infested" and Category 4 = "Historically not known to be infested / Unlikely to Become Established".

distribution of presence data. This test dataset includes all Japanese beetle presence data in Continental Europe, as well as presence data collected independently in the years 2020–2021 in North America, Japan and the Azores (4098 cells).

3 Results

3.1 Model performance and calibration

All measures of model performance indicate a well calibrated model with consistent and robust outputs across all 100 replicates. The AUC varied with a mean of 0.88 (SD = 0.07), reflecting a good discrimination capacity, while sensitivity reflected a good classification capacity with a mean of 0.82 (SD = 0.20). Furthermore, the Continuous Boyce Index (CBI) varied with a mean of 0.99 (SD = 0.01), reflecting a very good calibration of the models.

The P/E curve method identified six classes of increasing suitability, hereafter referred to as S0 to S5 (Appendix 4).

The proportion of presences from the test dataset increased across classes of increasing predicted suitability, and 93% of presences are located in the three highest suitability classes: 13.5% in S3, 27.5% in S4 and 52% in S5, respectively. Only 0.6% were located in cells of class S0.

3.2 Suitability maps

Model predictions are shown in Fig. 3. In North America we observe a rather sharp division between suitable and least suitable areas, with more than 80% of the area classified as least suitable (S0), and the remaining cells evenly distributed across increasing suitability classes. Much of north-eastern USA and adjacent areas of south-eastern Canada are predicted to be highly suitable. Another cluster of high suitability is found in the foothills of the Front Range, east of the Rocky Mountains, and is connected to the eastern cluster of high-suitability by a corridor of moderate suitability across the Great Plains. Smaller and scattered areas of low suitability are found west of the Rockies. The rest of the continent is dominated by the S0 class. In Japan, we observe a gradient from moderate (S1) to high (S2 and S3) suitability towards the northern part of the archipelago. The two highest suitability classes (S4 and S5) are not observed, while only 3% of cells are considered as least suitable (S0). South Korea and eastern China were also predicted to be moderately suitable. The Azores islands are predicted to be highly suitable, with 93% of the cells distributed in the highest suitability class S5. In Continental Europe, 62.7% of the area is classified as least suitable (S0), with no highest suitability classes (S4 and S5). Central Europe is almost entirely suitable (at least class S1, covering 32.7% of Continental Europe), with a continuous cluster of higher suitability (S2 and S3, 3.2% and 1.4% of cells respectively) extending from the foothills of Western Italian Alps to the

northern Balkans and the western Pannonian Basin. Another cluster of higher suitability is found along the south-eastern shores of the Black Sea. Smaller and scattered areas of moderate suitability are found in the northern foothills of the Alps (France, Germany, Austria, Switzerland and Slovenia), west of the Pyrenees and in the coastal Low Lands (Belgium and the Netherlands).

3.3 Variable importance

The variable importance associated with each of the 133 predictors is shown in Appendix 5 (Fig. A5.1), 75 of which contribute more to model performance than expected by chance. Variable importance is not evenly distributed across variables: 50% (or 80%) of the cumulative importance is captured by the first 26 (or 58) most important variables. Three predictor groups contribute most to model predictions: climatic, human-related and land-use predictors.

4 Discussion

The main objective of this study was to provide robust and reliable predictions of the suitability of Europe for the Japanese beetle. We applied a machine-learning approach to study Pj's distribution, based on observations of the beetle in its native and long-invaded ranges, and produced an accurate model for predicting suitability in the Northern Hemisphere, with a particular focus on the European continent. There, our predictions highlight the presence of clusters of high suitability, most of which are currently not infested, embedded in a matrix of moderate suitability. This confirms that Pj's invasion of Europe could spread further from its current distribution with potentially high pest impact, reinforcing the need for a risk-based surveillance strategy.

4.1 Variable importance

We found that climatic and human-related variables are most relevant to the beetle's distribution, in agreement with previous studies (Zhu et al. 2017; Kistner-Thomas 2019; Della Rocca & Milanesi 2022a). Low precipitation seasonality (bio15, Appendix 5) and high temperature seasonality (bio4) were found to be favourable. Our results also highlight the importance of variables such as actual evapotranspiration and vapour pressure. We believe they could be a proxy for soil water availability, which is related to soil suitability for adult oviposition, availability of food resources for adults, and larval survival. Interestingly, the most relevant land-use variable is associated with broadleaved forest in proximity to open space, which corresponds to a landscape where multiple food resources (broadleaved trees and typically crop fields) are in close proximity. They also provide shaded areas along the edges, which can preserve soil moisture during summer and are therefore favourable for oviposition.

4.2 SDM – methodological considerations

4.2.1 Opportunistic data, sampling bias and target-group strategy for pseudo-absences

In the absence of a global and standardised database of *Pj* presence and absence, we used opportunistic citizen-science observations to build a presence-only database. While this ensured global coverage with tens of thousands of presence records, it could affect the quality of the analysis (Dobson et al. 2020). Opportunistic citizen-science data are likely to be geographically biased towards easily accessible locations, typically urban, agricultural and recreational areas. To compensate for sampling bias, we chose to generate pseudo-absence data using the target-group strategy (Phillips et al. 2009). This approach has been confirmed to be effective in addressing sampling bias when the target group includes generalist species and occurrences in the broadest environmental subspace associated with the study area (Botella et al. 2020). This is the case for our *Coleoptera* target-group, which is a very diverse group including many generalist species (Bouchard et al. 2017). When comparing the performance of random vs. target-group strategies in terms of variable importance (not shown here), we found a dramatic reduction in the relative importance of all human-related predictors. In terms of suitability maps, the random-sampling strategy produced a patchy distribution of high-suitability pockets around major cities surrounded by low suitability areas whereas the target-group approach produced a more evenly distributed suitability envelope with varying levels of suitability. Both results confirmed the presence of sampling bias in the data, the effect of which was mitigated by the target-group approach. However, human-related factors did not disappear from the list of important predictors. This could reflect the Japanese beetle's preference for anthropogenic environments, due to the presence of preferred food sources and irrigated turfs suitable for oviposition (Althoff & Rice 2022). This could also be due to residual sampling bias not accounted for by the target-group strategy.

4.2.2 SDM of invasive species

A critical point in fitting an SDM is to build a database that is well representative of conditions in regions where the organism is present and absent, allowing the model to neatly separate conditions that are favourable from those that are not. Although previous studies included data from Continental Europe in model training, where the first detection was reported less than 10 years ago (Kistner-Thomas 2019; Della Rocca & Milanesi 2022a), our decision not to include these data is supported by recent literature. Indeed, Liu et al. (2020) showed that by fitting a SDM on invaded ranges with reduced residence time, the invaded niche appears smaller than the native niche with large unfilling, i.e. favourable conditions in the native niche were not yet occupied in the invaded niche.

4.3 Suitability predictions in native and long-invaded areas

Evaluating the quality of the predictions in Japan is difficult as there are few observations available. Nevertheless, most of the Japanese archipelago (97%) is classified in suitability classes S1, S2 and S3. Interestingly, very few areas are classified in the lowest and highest suitability classes (2.5% in S0, 0.3% in S4 and none in S5). This could be explained by a difference in crop types between the native and invaded niches, and the presence of natural predators and parasites in Japan that regulate *Pj* populations resulting in fewer observations (Clausen et al. 1927). In the Azores, the suitability predicted by the model is always very high (majority of classes S3, S4 and S5), which is supported by the invasion history and the difficulties reported by local authorities in eradication programmes (Martins et al. 1988).

The situation in North America, particularly the USA, warrants further examination as it is a well-documented, century-old invasion. The goodness-of-fit of our model in this area will reflect its transferability to Europe and the robustness of our predictions there. *Pj* is currently established in 28 US states, and interceptions have been reported in at least 13 additional US states and Canadian provinces, likely as a result of human-mediated dispersal (Althoff & Rice 2022). In 2016, the USA revised its Domestic Japanese Beetle Harmonization Plan (NPB 2016), which assigns each US state to one of four different categories based on *Pj* infestation status, risk of entry, expected impact, and need for quarantine certification protocols. To evaluate our suitability maps, we used this official classification as a reference for *Pj*'s current distribution in the USA. Recent studies also produced suitability maps for the Japanese beetle in the USA (Zhu et al. 2017; Kistner-Thomas 2019; Della Rocca & Milanesi 2022a), and key differences with our predictions are also discussed here.

For Category 4 states, “*Historically Not Known to Be Infested/Unlikely to Become Established*” (NPB 2016), we predicted Florida and Louisiana to be least suitable (S0) (Fig. 3B), which is consistent with this classification. This strongly contradicts previous studies, all of which predicted these states to be at least as suitable as the most infested areas in north-eastern USA (Zhu et al. 2017; Kistner-Thomas 2019; Della Rocca & Milanesi 2022a). In Wyoming, also a Category 4 state, our prediction of localised clusters of higher suitability (class S3) in a mostly low suitability matrix is supported by recent detections (Althoff & Rice 2022).

Category 1 states – “*Uninfested/Quarantine Pest*”, Western US states, were predicted as least-suitable (S0) with small and disconnected patches of low suitability (S1) (Fig. 3B). There, official pest risk assessments found “moderate to high pest impact in case of establishment that can be mitigated by applying quarantine protocols and eradication programs” (NPB 2016), which is not inconsistent with

our prediction of low suitability, but rather reflects high risk of entry and high pest impact that were not included in our analysis. Indeed, these states are at significant risk of accidental human-mediated introduction due to the presence of highly connected hubs. Pest impact would also be high due to the presence of high-value crops such as grapes and berries, amongst the preferred hosts of the Japanese beetle (Althoff & Rice 2022). Finally, successful eradication of previous introductions in California and Oregon supports our prediction of low suitability for these states (Althoff & Rice 2022). In contrast, previous studies predicted these states as suitable with multiple corridors connecting them to Eastern states (Zhu et al. 2017; Kistner-Thomas 2019; Della Rocca & Milanesi 2022a).

Category 2, – “*Uninfested or Partially Infested*”, includes both uninfested and partially infested states where NPB (2016) found that Pj could survive but pest impact would be low to moderate. In these states, we predicted mostly lowest suitability (82% of cells in class S0) and few clusters of high suitability (6% in classes S2 or higher). We believe that our results are largely consistent with this classification, as even if small local populations can maintain in managed areas (e.g. lawns), the majority of states are least suitable for the beetle.

Finally, we predicted the highest levels of suitability (73% classified as S2 or higher) for Category 3 states – “*Partially or Generally Infested*” located in Eastern USA (Fig. 3B).

Overall, we believe that our predictions are consistent with the NPB’s official classification (NPB 2016), and that observed differences ascribe to the fact that we did not account for risk of entry and pest impact. Therefore, we can confirm that our model is capable of producing reliable predictions of suitability for the Japanese beetle. Compared to previous studies, it is less prone to overestimation (i.e. tendency towards false positives, labelling as suitable areas that are not suitable), which could partly be due to the use of an environmental-block approach for pseudo-absence sampling rather than random down-sampling in model training (Roberts et al. 2017).

4.4 Suitability predictions in Continental Europe

Our predictions show a complex scenario for Europe (Fig. 3C), with two main clusters of high suitability: the first in the foothills of the Alps (both North and South) and in the Northern Balkans, and the second on the eastern shores of the Black Sea. Current distribution of Pj in Europe lies within the first main cluster of high suitability.

A large part of Central Europe is at a low level of suitability (class S1), and all clusters of moderate and high suitability are embedded within this low suitability area, creating a continuum of suitable land across the continent. Although the overall picture is quite worrying, especially for

Central Europe, it does not compare with studies that predicted high suitability overall (Zhu et al. 2017; Della Rocca & Milanesi 2022a). This would be expected if these studies overestimated suitability in the training dataset (mostly North America). We suspect that this tendency would be maintained when projecting to Europe.

Overall, our map partially supports previous predictions (Bourke 1961; EFSA Panel on Plant Health et al. 2018; Kistner-Thomas 2019), while highlighting a more spatially refined and contrasting suitability scenario for Europe.

4.5 Implications for the control of the ongoing Japanese beetle invasion in Europe

Designing a surveillance strategy for Europe is crucial and mapping suitability is a major milestone. However, the dense European transport network may facilitate the introduction of the beetle into distant regions, independently of their suitability. We therefore call for surveillance strategies tailored to infested and highly reachable regions. In infested areas, efforts should focus on containing growing populations and tracking the invasion front. Quarantine protocols should be applied to regulated products from infested areas (fresh vegetables and fruits, plants and plant material) to ensure beetle-free shipments (Poggi et al. 2022b), especially during the adult flight period. Adapted protocols should also be applied to transport hubs, such as railway stations, airports and truck warehouses, while the public should be sensitized to the risk of unintentional movement of the beetle. Outside of the infested areas, the most reachable transport hubs should be prioritised for surveillance and prepared to respond to the potential arrival of the beetle. Quarantine and eradication protocols in areas of potential high pest impact in Western USA have shown that a timely response is critical to controlling the spread of the beetle (Althoff & Rice 2022). Communication campaigns across Europe could inform citizens and stakeholders about the identification and reporting of Japanese beetle sightings.

Acknowledgements: This research was supported by the IPM-Popillia project, funded by the European Union Horizon 2020 research and innovation programme under grant agreement No 861852. Authors would like to thank Servizio Fitosanitario Ticino, Plant Health Service – Piedmont Region, Secretaria Regional da Agricultura e do Desenvolvimento Rural do Governo dos Açores, and the Canadian Food Inspection Agency, for providing standardised survey presence data included in analyses. Authors would like to thank citizen science platforms and all citizens who participate in surveys, for their contribution to the advancement of research. *The datasets generated and/or analysed during the current study are available in the French Research Government repository*, <https://doi.org/10.57745/GM2YVL>

References

- Althoff, E. R., & Rice, K. B. (2022). Japanese Beetle (Coleoptera: Scarabaeidae) Invasion of North America: History, Ecology, and Management. *Journal of Integrated Pest Management*, 13(1), 2. <https://doi.org/10.1093/jipm/pmab043>
- Barbet-Massin, M., Jiguet, F., Albert, C. H., & Thuiller, W. (2012). Selecting pseudo-absences for species distribution models: How, where and how many? *Methods in Ecology and Evolution*, 3(2), 327–338. <https://doi.org/10.1111/j.2041-210X.2011.00172.x>
- Bellard, C., Cassey, P., & Blackburn, T. M. (2016). Alien species as a driver of recent extinctions. *Biology Letters*, 12(2), 20150623. <https://doi.org/10.1098/rsbl.2015.0623>
- Botella, C., Joly, A., Monestiez, P., Bonnet, P., & Munoz, F. (2020). Bias in presence-only niche models related to sampling effort and species niches: Lessons for background point selection. *PLoS One*, 15(5), e0232078. <https://doi.org/10.1371/journal.pone.0232078>
- Bouchard, P., Smith, A. B. T., Douglas, H., Gimmel, M. L., Brunke, A. J., & Kanda, K. (2017). Biodiversity of Coleoptera. In R. G. Foottit, & P. H. Adler (Eds.), *Insect Biodiversity* (pp. 337–417). John Wiley & Sons, Ltd. <https://doi.org/10.1002/9781118945568.ch11>
- Bourke, P. A. (1961). Climatic aspects of the possible establishment of the Japanese beetle in Europe. World Methodological Organization Technical Note 41.
- 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>
- Clausen, C. P., King, J. L., & Teranishi, C. (1927). *The parasites of Popillia japonica in Japan and Chosen (Korea), and their introduction into the United States*. Washington, D.C.: U.S. Dept. of Agriculture. <https://doi.org/10.5962/bhl.title.108287>
- Commission Delegated Regulation (EU) 2019/1702 of 1 August 2019 supplementing Regulation (EU) 2016/2031 of the European Parliament and of the Council by establishing the list of priority pests. [2019] L 260/8
- Della Rocca, F., & Milanese, P. (2022a). The New Dominator of the World: Modeling the Global Distribution of the Japanese Beetle under Land Use and Climate Change Scenarios. *Land (Basel)*, 11(4), 567. <https://doi.org/10.3390/land11040567>
- Della Rocca, F., & Milanese, P. (2022b). The Spread of the Japanese Beetle in a European Human-Dominated Landscape: High Anthropization Favors Colonization of *Popillia japonica*. *Diversity*, 14(8), 658. <https://doi.org/10.3390/d14080658>
- Diagne, C., Leroy, B., Vaissière, A.-C., Gozlan, R. E., Roiz, D., Jarić, I., ... Courchamp, F. (2021). High and rising economic costs of biological invasions worldwide. *Nature*, 592(7855), 571–576. <https://doi.org/10.1038/s41586-021-03405-6>
- Dobson, A. D. M., Milner-Gulland, E. J., Aebischer, N. J., Beale, C. M., Brozovic, R., Coals, P., ... Keane, A. (2020). Making Messy Data Work for Conservation. *One Earth*, 2(5), 455–465. <https://doi.org/10.1016/j.oneear.2020.04.012>
- EFSA Panel on Plant Health (PLH), Bragard, C., Dehnen-Schmutz, K., Di Serio, F., Gonthier, P., Jacques, M.-A., Jaques Miret, J. A., ... MacLeod, A. (2018). Pest categorisation of *Popillia japonica*. *EFSA Journal*, 16(11), e05438. <https://doi.org/10.2903/j.efsa.2018.5438>
- Elith, J., Kearney, M., & Phillips, S. (2010). The art of modelling range-shifting species. *Methods in Ecology and Evolution*, 1(4), 330–342. <https://doi.org/10.1111/j.2041-210X.2010.00036.x>
- Fleming, W. E. (1972) Biology of the Japanese beetle. *USDA Technical Bulletin*, 1449, 1–129
- Freeman, E. A., Moisen, G. G., Coulston, J. W., & Wilson, B. T. (2016). Random forests and stochastic gradient boosting for predicting tree canopy cover: Comparing tuning processes and model performance. *Canadian Journal of Forest Research*, 46(3), 323–339. <https://doi.org/10.1139/cjfr-2014-0562>
- 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>
- Hulme, P. E. (2009). Trade, transport and trouble: Managing invasive species pathways in an era of globalization. *Journal of Applied Ecology*, 46(1), 10–18. <https://doi.org/10.1111/j.1365-2664.2008.01600.x>
- Hyatt-Twynam, S. R., Parnell, S., Stutt, R. O. J. H., Gottwald, T. R., Gilligan, C. A., & Cunniffe, N. J. (2017). Risk-based management of invading plant disease. *The New Phytologist*, 214(3), 1317–1329. <https://doi.org/10.1111/nph.14488>
- Japkowicz, N., & Stephen, S. (2002). The class imbalance problem: A systematic study. *Intelligent Data Analysis*, 6(5), 429–449. <https://doi.org/10.3233/IDA-2002-6504>
- Kistner-Thomas, E. J. (2019). The Potential Global Distribution and Voltinism of the Japanese Beetle (Coleoptera: Scarabaeidae) Under Current and Future Climates. *Journal of Insect Science*, 19(2), 16. <https://doi.org/10.1093/jisesa/iez023>
- Liu, C., Wolter, C., Xian, W., & Jeschke, J. M. (2020). Most invasive species largely conserve their climatic niche. *Proceedings of the National Academy of Sciences of the United States of America*, 117(38), 23643–23651. <https://doi.org/10.1073/pnas.2004289117>
- Martins, A., Paiva, M. R., & Simões, N. (1988). Japanese Beetle: Monitoring in the Azores with Semiochemicals. *Ecological Bulletins*, 39, 101–103.
- NPB (National Plant Board) (2016). *U.S. Domestic Japanese Beetle Harmonization Plan*.
- Oerke, E.-C. (2006). Crop losses to pests. *Journal of Agricultural Science*, 144(1), 31–43. <https://doi.org/10.1017/S0021859605005708>
- Parnell, S., Gottwald, T. R., Riley, T., & van den Bosch, F. (2014). A generic risk-based surveying method for invading plant pathogens. *Ecological Applications*, 24(4), 779–790. <https://doi.org/10.1890/13-0704.1>
- Phillips, S. J., Dudík, M., Elith, J., Graham, C. H., Lehmann, A., Leathwick, J., & Ferrier, S. (2009). Sample selection bias and presence-only distribution models: Implications for background and pseudo-absence data. *Ecological Applications*, 19(1), 181–197. <https://doi.org/10.1890/07-2153.1>
- Poggi, S., Borner, L., Roche, J., Tayeh, C., & Martinetti, D. (2022a). *Biological invasion of the Japanese beetle in Continental Europe at a glance*. Recherche Data Gouv, V3. <https://doi.org/10.57745/R18NGL>
- Poggi, S., Desneux, N., Jactel, H., Tayeh, C., & Verheggen, F. (2022b). A nationwide pest risk analysis in the context of the ongoing Japanese beetle invasion in Continental Europe: The case of metropolitan France. *Frontiers in Insect Science*, 2, 1079756. <https://doi.org/10.3389/finsec.2022.1079756>

- R Core Team. (2021). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., ... 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>
- Sanchez, B., Barreiro-Hurle, J., Soto-Embodas, I., & Rodriguez-Cerezo, E. (2019). *The Impact Indicator for Priority Pests (I2P2): a tool for ranking pests according to Regulation (EU) 2016/2031*. Joint Research Centre (Seville site)
- Simonetto, A., Sperandio, G., Battisti, A., Mori, N., Ciampitti, M., Cavagna, B., ... Gilioli, G. (2022). Exploring the main factors influencing habitat preference of *Popillia japonica* in an area of recent introduction. *Ecological Informatics*, 70, 101749. <https://doi.org/10.1016/j.ecoinf.2022.101749>
- Stoppiglia, H., Dreyfus, G., Dubois, R., & Oussar, Y. (2003). Ranking a Random Feature for Variable and Feature Selection. *Journal of Machine Learning Research*, 3, 1399–1414.
- Valavi, R., Guillerá-Arroita, G., Lahoz-Monfort, J. J., & Elith, J. (2022). Predictive performance of presence-only species distribution models: A benchmark study with reproducible code. *Ecological Monographs*, 92(1), e01486. <https://doi.org/10.1002/ecm.1486>
- Zhu, G., Li, H., & Zhao, L. (2017). Incorporating anthropogenic variables into ecological niche modeling to predict areas of invasion of *Popillia japonica*. *Journal of Pest Science*, 90(1), 151–160. <https://doi.org/10.1007/s10340-016-0780-5>

Manuscript received: March 29, 2023

Revisions requested: May 16, 2023

Modified version received: September 14, 2023

Accepted: September 19, 2023

The pdf version (Adobe JavaScript must be enabled) of this paper includes an electronic supplement:
Appendix 1–5.