ORIGINAL ARTICLE

Estimating the potential distribution of yellow spotted stink bug (*Erthesina fullo*) using ecological niche models

Gengping ZHU[1] , Tatiana N. BUSH[1], Katherine S. BURGSTAHLER[2],
Nathaniel GREEN[1], Hannah COOK[1], Emily RAMPONE[1], Salena HELMREICH[1],
Riley M. REED[1], Joshua M. MILNES[3] and David W. CROWDER[1]

[1]*Department of Entomology, Washington State University, Pullman, Washington, USA,* [2]*School of Environment, Washington State University, Pullman, Washington, USA* and [3]*Washington State Department of Agriculture, Yakima, Washington, USA*

Abstract

Invasive insect species threaten the productivity of ecosystems worldwide, and ecological niche models can be used to predict distributions of invaders and guide management efforts. Ecological niche models can also aid monitoring for invasive species that are globally distributed. One such species is the yellow spotted stink bug (*Erthesina fullo* Thunberg), a polyphagous pest native to Asia that has established in Europe and South America and threatens specialty crops. Here, we used ecological niche models to predict the potential distribution of *E. fullo*, and created a website to display predictions. We show that *E. fullo* has peak occurrence probability in areas with annual mean temperatures around 20°C, and that the occurrence probability increases as maximum monthly temperature reaches up to 38°C. The likelihood of occurrence decreased as annual precipitation increased, but increased with greater precipitation in the wettest and driest months. This suggests *E. fullo* is most suited to regions that are warm and dry and where most precipitation occurs across only a few months, such as southern North America, central and southern South America, southern Europe, southern Africa, and central and eastern Australia. Given that *E. fullo* is a highly mobile hitchhiking insect that travels through cargo and other containers to new areas, the potential spread of this species into new regions should be carefully monitored.

Key words: ecological niche model, ensemble model, potential distribution, response curve, risk assessment.

INTRODUCTION

Invasive insects represent a major threat to agriculture, biodiversity and ecosystem function by disrupting food webs, spreading pathogens, consuming crops and damaging infrastructure (Bradshaw *et al*. 2016). Many researchers have studied the likely impacts of climate change on invasive species and found that the changing environment will allow for range expansions of invasive species, and the number and impact of invasions are predicted to increase with climate change (Anton 2021).

The most efficient tactic to manage invasive species is to eradicate earlier established populations (Valentin *et al*. 2018), but we often lack information on where to monitor invasive species to identify incipient populations, especially for species that are not well studied. For those invasive species that have shown a capacity for establishment across continents or broad geographic regions, leveraging data from existing invasive populations into ecological niche models could aid in identifying other potential areas where invaders might occupy.

Early detection and eradication programs require spatial habitat suitability predictions to guide deployment of traps for monitoring and to manage interventions. Ecological niche models are common tools used in invasive species distributional modeling and risk assessment for these purposes (Jiménez-Valverde *et al*. 2011). The correlative niche models use environmental variables

that are associated with observation localities of an invasive species to characterize the realized niche, and then identify regions where the realized niche is located geographically (Peterson *et al.* 2011). The low-data-demand of ecological niche models make them useful in invasion risk assessment for species that are not well studied, and for species that have invaded parts of the globe and are considered threats to invade other regions. For example, for species such as yellow spotted stink bug, *Erthesina fullo* Thunberg, 1783 (Hemiptera: Pentatomidae) that are native to Asia but have invaded Europe and South America, models may explore other world regions that could potentially be invaded.

Erthesina fullo is a highly polyphagous insect pest species that is native to East Asia, but has recently become of global interest as it invaded Albania in Europe and Brazil in South America (Mi *et al.* 2020; Brugnera *et al.* 2022). It has been reported to feed on many economically important plants such as kiwifruit, cherries, pear and apple (Mi *et al.* 2020). The ability of *E. fullo* to overwinter across a range of conditions, combined with their broad dispersal potential and high polyphagy, has caused worry that populations could invade regions outside of Asia. As *E. fullo* shares many ecological traits with the devastating invader brown marmorated stink bug, *Halyomorpha halys* Stål, 1855 (Hemiptera: Pentatomidae), a species that has rapidly swept across Europe and North America and is causing extreme damage to specialty crops (Zhu *et al.* 2012), assessment of potential areas of invasion is warranted.

In this paper, we used ecological niche models to identify where potential establishment of *E. fullo* could occur across the globe. We first explored whether the realized niche of *E. fullo* is similar in the native range to the two invaded ranges (Europe and South America). This analysis allowed us to predict whether adaptation has occurred in invasive populations and the degree of overlap in the realized niche in the native compared to the invaded ranges (Zhu *et al.* 2012). Next, responses of *E. fullo* to annual trends and extremes of temperature and precipitation were modelled using a fine-tuned Maxent model (Radosavljevic & Anderson 2014) with inflated response curves (Zurell *et al.* 2012). Finally, we used six types of ecological niche models and two environmental datasets, along with ensemble models (Zhu & Peterson 2017) that averaged predictions, to predict the potential distribution of *E. fullo* globally. Here we sought to identify the regions that would be most likely to support establishment and spread of *E. fullo* by searching suitable areas that were supported by multiple niche models. By comparing the variation among models in the distributional predictions, we identified global regions that should be targeted for

monitoring of *E. fullo* to prevent establishment of this highly mobile and devastating invader.

MATERIALS AND METHODS

Input data

Raw occurrence records in East Asia were gathered from multiple sources: (i) the Global Biodiversity Information Facility (http://gbif.org/); (ii) iNaturalist (http://inaturalist.org/); and (iii) idgibio (http://idigbio.org/) using the *spocc* package in the *R* statistical software (Owens *et al.* 2023; R Core Team 2023). After raw records were obtained, we cleaned the data by removing duplicate records and records that were beyond native distribution. We then enforced a distance of 100 km between records to reduce sampling bias with the *spThin* package in *R*, as ecological niche models are sensitive to sampling bias (Aiello-Lammens *et al.* 2015). This left 136 records that were retained for model calibration and evaluation (Fig. 1A). Six bioclimate variables (see below) that were associated with these occurrence records were extracted and plotted in three scatterplots, that is, annual precipitation against annual mean temperature (Fig. 1B), minimum temperature of the coldest month against maximum temperature of the hottest month (Fig. 1C) and precipitation of the driest month against that of the wettest month (Fig. 1D).

We assembled 12 environmental variables at a resolution of 2.5 min that summarized aspects of climate and vegetation for niche and distribution modeling (Table S1). We collected bioclimate variables associated with annual trends and extremes of temperature, precipitation and radiation from WorldClim (version 2.1; http://worldclim.org/); these predictors were derived from historical monthly temperature and precipitation data and averaged from 1970 to 2000. For vegetation data, we used the European Global Landcover 2000. We assessed annual maximum green vegetation fraction based on 12 years (2001–2012) normalized difference vegetation index data. Two datasets were prepared (Table S1): (i) dataset 1 included annual trends and extremes of temperature and precipitation, two abiotic factors that have been used to effectively predict the distribution of another invasive stink bug, *H. halys* (Zhu *et al.* 2012); and (ii) dataset 2 included all variables with Pearson correlation <0.7, including annual mean radiation, annual mean temperature, precipitation of the driest month, land cover and maximum green vegetation fraction (Table S1).

Modeling the realized niche of *E. fullo*

After obtaining occurrence data, we assembled six abiotic variables (i.e. dataset 1) to characterize the realized niche
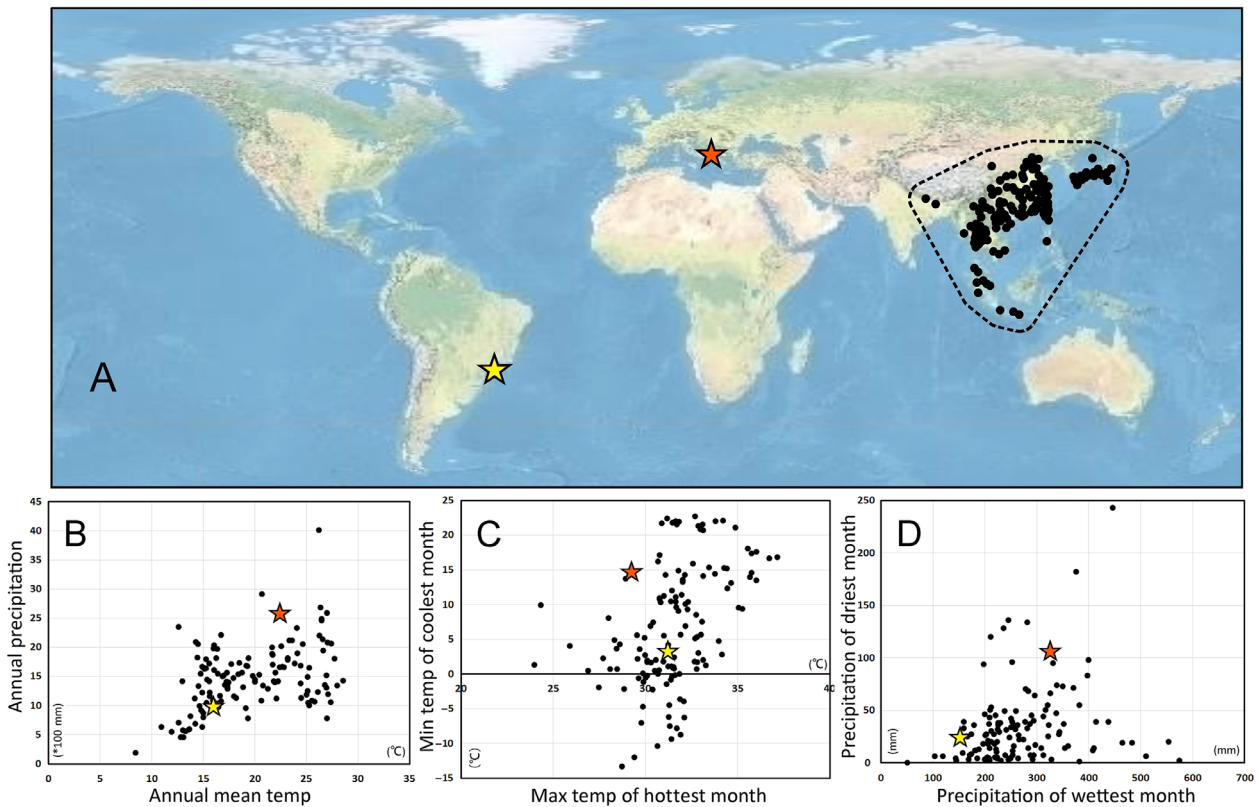
Figure 1 (A) Occurrence data of *Erthesina fullo* and the native accessible area used in native niche models. Six bioclimate variables associated with these occurrence records were plotted in scatterplots, showing values for (B) annual mean temperature and precipitation, (C) maximum temperature of the warmest month and minimum temperature of the coolest month and (D) precipitation of the wettest and driest months. In each scatterplot, values for all native (black circles) and invasive occurrence data (red and yellow stars) used in the models are shown.

of *E. fullo*: (i) annual mean temperature; (ii) maximum temperature of the warmest month; (iii) minimum temperature of the coldest month; (iv) annual precipitation; (v) precipitation of the wettest month; and (vi) precipitation of the driest month. To assess the realized niche of *E. fullo*, we used Maxent to simulate the occurrence probability of *E. fullo* in the native as well as the two invaded ranges. For this analysis, we fine-tuned the Maxent model using the Akaike information criterion with the *SDMtune* package in R (Warren & Seifert 2011; Vignali *et al.* 2020), and we used inflated response curves to explore species–environment relationships (i.e. the realized niche of *E. fullo*). These inflated response curves demonstrate the effect of a variable on the response while accounting not only for the average effects of the other variables but also for mean, minimum, median, maximum and quartile values (Zurell *et al.* 2012).

Distributional modeling of *E. fullo*

We used a classical model approach, where ecological niche models were calibrated based on occurrence

records in the native East Asia region and transferred onto the globe to predict the potential distribution of *E. fullo* (Zhu *et al.* 2012); this approach assumes a shared realized niche in the invaded and native ranges. We used six common ecological niche models to predict the distribution of *E. fullo*: (i) generalized additive; (ii) boosted regression tree; (iii) generalized linear; (iv) Maxent; (v) random forest; and (vi) support vector machine. Models were built using the *sdm* package in R, and this package was used to average model predictions to create an ensemble model (Naimi & Araújo 2016); the ensemble models have been proposed as a method to reduce uncertainty given that variations exist for predictions of any individual model (Araújo & New 2007). Ensemble models (see below) were calibrated using the two environmental datasets separately, results of which were averaged for the final distribution predictions.

For each model, a native model was built using occurrence data from accessible areas in East Asia, which was delimited by buffering minimum convex polygons of observed points at 500 km. Following

established best practices for Maxent (Araújo *et al.* 2019), we used the fine-tuned settings and a "random" method to select 10,000 pseudo-absence records from the "accessible" areas (Barve *et al.* 2011), given that real absence data was not available. For the other models, we selected 680 pseudo-absence records from the "accessible" areas, which is five times the number of occurrence records used to create the models (Barbet-Massin *et al.* 2012). Pseudo-absence records represent background landscapes or areas that are not predicted to have the species; these types of records are used in models when data on the real absence of species is lacking, as in this study. Each of the models, and the ensemble models, were fit in the accessible areas of East Asia and then transferred on the globe to estimate potential habitat suitability.

For model validation, 70% of the records were used for training and 30% for evaluation. We used the area under the curve (AUC) of the receiver operating characteristic plot as an effective discrimination measure in concert for evaluating model performance (Lobo *et al.* 2008; Peterson *et al.* 2008) rather than simply reporting the AUC value. Finally, we used the multivariate environmental similarity surface metric (MESS) (Elith *et al.* 2010), to measure the nonanalogue environmental conditions for the two datasets in model transferred areas (i.e., areas of the globe where distributional predictions occurred in areas not reflected by conditions in the training area of East Asia).

RESULTS

Realized niche modelling of *E. fullo*

The fine-tuned Maxent model simulations showed that *E. fullo* has the greatest probability of occurrence in habitats with temperatures ≈20°C; suitability based on annual temperatures showed a relative normal distribution around this value (Fig. 2A). Model predictions show *E. fullo* has the highest probability of occurrence in regions with relatively high maximum temperatures in the warmest month (up to 38°C); occurrence probability of *E. fullo* was also predicted to increase in regions with minimum temperatures of the coldest month of up to 25°C (Fig. 2B,C). Results from Maxent models show that the likelihood of occurrence for *E. fullo* was highest in regions with relatively low annual precipitation (Fig. 2D), but also where most precipitation falls in only a few months (Fig. 2E,F). For each predictor, inflated response curves (i.e. Fig. 2, gray lines) that accounted for values of predictors generally followed the partial response curves that did not account for other predictors (i.e. Fig. 2, blue lines).

This suggests that the predictors largely were observed to have independent and additive effects on probability of occurrence.

When comparing climatic conditions in native and invasive ranges, climatic conditions occupied by invasive populations of *E. fullo* overlap with native populations, supporting our assumption of a shared niche (Fig. 1). This suggests that invasive populations have not adapted to unique climatic conditions, or invasive populations have not yet reached regions with climatic conditions not found in the native range.

Ecological niche modelling of the potential distribution of *E. fullo*

All six ecological niche model approaches showed good discrimination ability in model interpolation evaluations with either dataset 1 or 2 as input data, with AUC values for the 70% training data ranging from 0.70 to 1.0 and the 30% validation data ranging from 0.72 to 0.83 (Figs S1,S2). The random forest model was particularly efficient in reclassifying presence and absence distributions and showed better performance in training data evaluation than the other five models with either predictor dataset (Figs S1,S2). There was considerable variability in predictions across the six model types that were unequally distributed across the globe (Figs S3,S4). The greatest congruence among the six models occurred in regions north of 40°N latitude, in northern and western South America, and in parts of central Africa; these predictions were consistent with either predictor dataset (Figs S3,S4). In contrast, the greatest variability observed across the six models occurred in areas south of 40°N latitude in dataset 1 (but not dataset 2) based models, including the majority of Africa, the southern United States, central America, eastern South America and Australia. Based on our analysis of areas outside the training area, our model predictions were robust throughout most of the globe except in parts of northern Canada and Russia (Figs S3,S4), as these two regions have one or more environmental variables outside the range present in training data (i.e. MESS analysis; Figs S3,S4); however, these regions were predicted to have almost no suitability for *E. fullo*.

Predictions of habitat suitability from the ensemble model suggest there are vast areas of suitable habitat in global regions south of 40°N latitude (Fig. 3). In the United States, a country that has not yet been invaded, there is predicted suitable habitat in California and the Pacific Northwest, two regions that are relatively warm and dry (Fig. 3A), as well as much of the southeastern United States (Fig. 3B). In South America, the majority
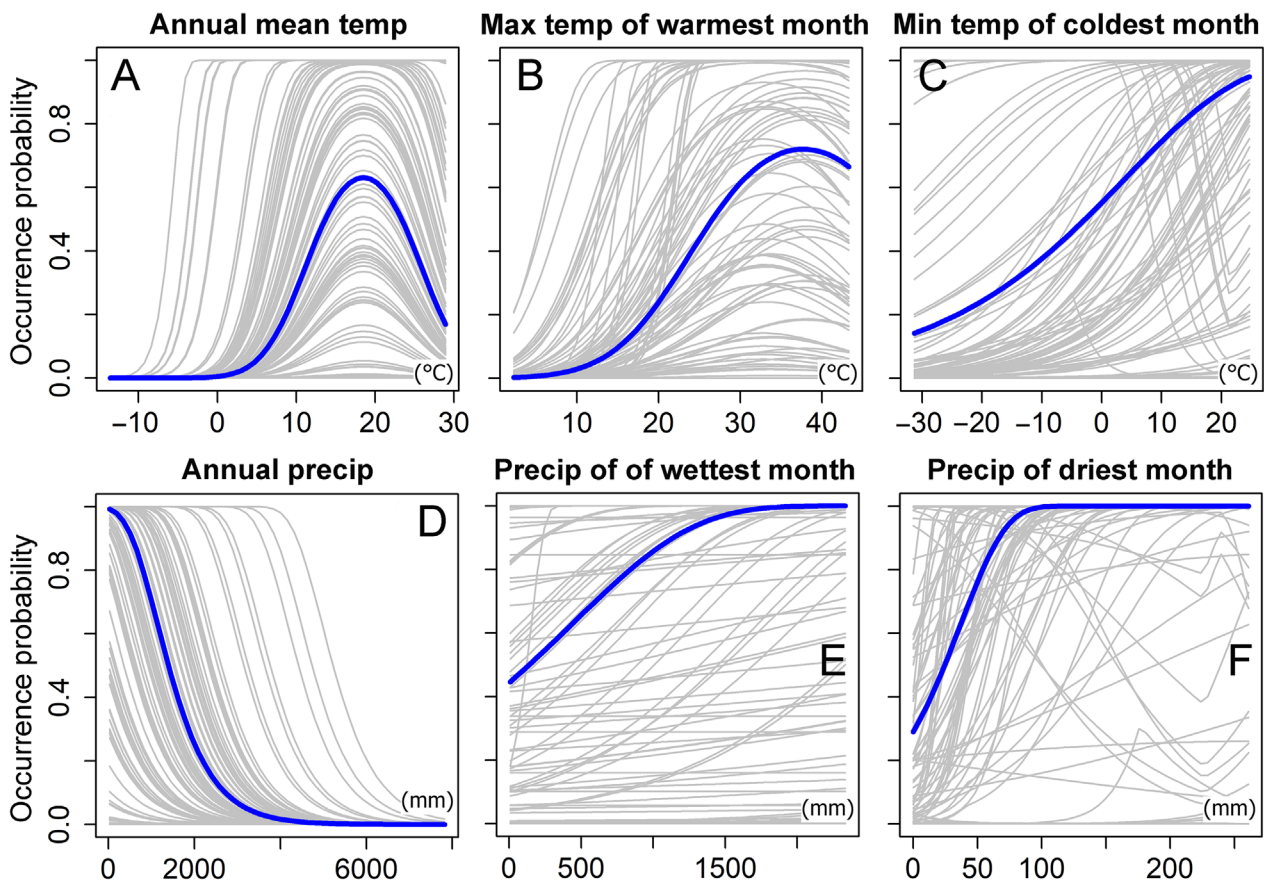
Figure 2 Simulations of *Erthesina fullo* response to climate factors with Maxent models and inflated response curves. For each panel, blue lines denote partial response curves for how *E. fullo* occurrence probability is expected to change when holding other predictors at their mean values. Gray lines denote inflated response curves when keeping other predictors at their minimum, median, maximum and quartile values.

of the continent is predicted to have medium to high suitability. Indeed, the region of Brazil that has been invaded is predicted to have extremely high suitability (Fig. 3C). Similarly, most of southern Europe, including Spain, Italy and Greece, are predicted to have extremely high suitability and these regions include the invaded area (Fig. 3D). To facilitate application, our ensemble model prediction was deposited in a website (https://gpzhu.github.io/YSSB/YSSB.html) for interactive visualization.

DISCUSSION

Realized niche modelling suggests global areas that are relatively warm and dry are most suitable for *E. fullo*, which tracks well with their observed native range in East and Southeast Asia. Prior studies show that *E. fullo* can complete development from 15°C to 30°C, with a hatch rate exceeding 97.5% across these temperatures (Mi *et al.* 2020); however, hatch rates

precipitously decline below 15°C and above 60% relative humidity. These findings are in line with our results showing increasing annual precipitation lowers the occurrence probability (Fig. 2). Our models suggest that the occurrence probability of *E. fullo* decreases as annual precipitation increases, but the probability increases with greater precipitation in the wettest and driest months. These simulations are consistent with ensemble model predictions that show *E. fullo* is most suited to regions that are warm and dry and where most precipitation occurs across only a few months (Fig. 3). Our suitable predictions were also observed in distinct dotted areas in the eastern United States (Fig. 3B) and Europe (Fig. 3D). This might be due to the urban heat island effect (Deilami *et al.* 2018) as these dotted areas are the homes of cities, for example, Columbus, Cleveland, and Dayton in Ohio, Pittsburgh and Philadelphia in Pennsylvania (Fig. 3B), Paris in France and London in the UK (Fig. 3D); the high temperatures there would help to provide suitable habitat
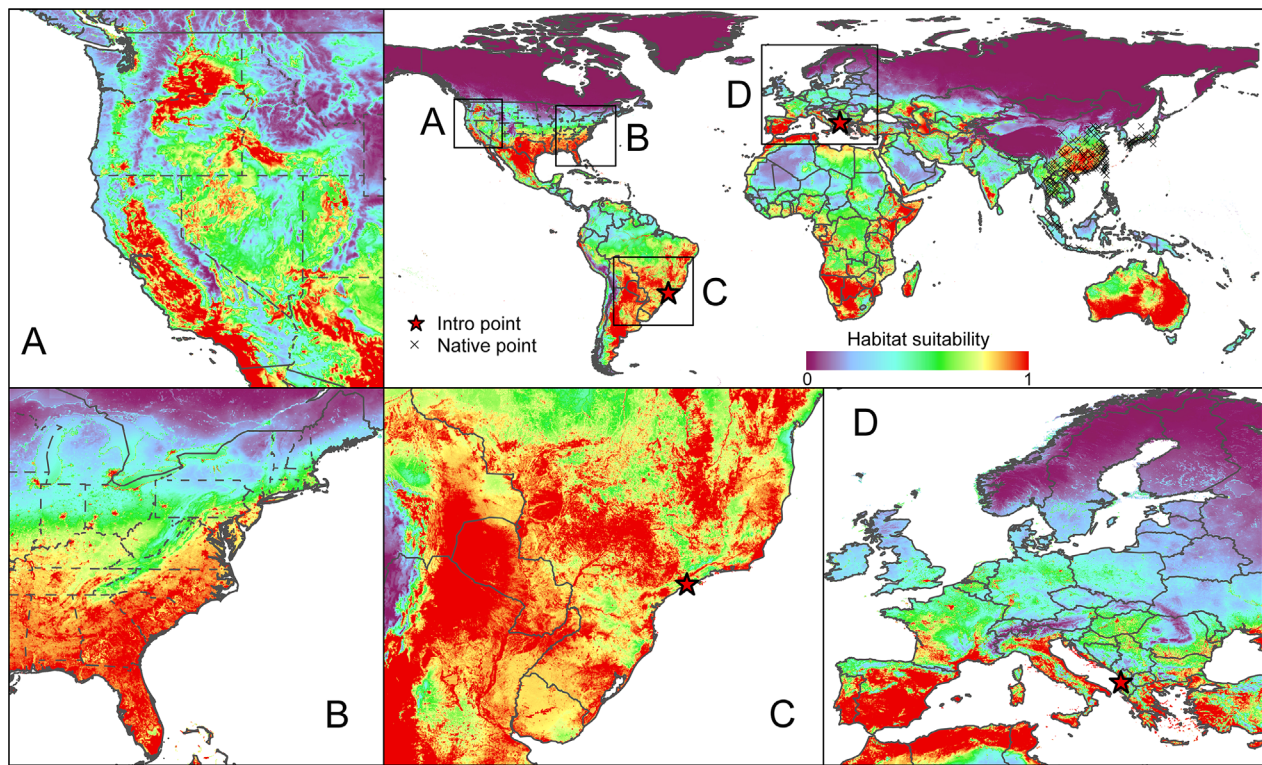
Figure 3 Predicted habitat suitability of *Erthesina fullo* based on the ensemble ecological niche models. Panels represent regional predictions for (A) the western United States, (B) the eastern United States, (C) South America and (D) Europe, along with a combined global map (top right).

for *E. fullo*. Of course, regions or areas predicted to have highly suitable habitat are not guaranteed to be invaded, as the ultimate distribution of species like *E. fullo* is determined not only by abiotic factors but also by biotic factors as well as dispersal (Peterson *et al.* 2011). However, *E. fullo* is known to be a hitchhiking insect that can travel through cargo and other containers to reach new areas. For example, in 2014, *E. fullo* was detected in New Zealand but did not establish; similar detections were confirmed to be established populations in Albania in 2017 and Brazil in 2020 (Mi *et al.* 2020; Brugnera *et al.* 2022).

Like every ecological model, correlative ecological niche models are sensitive to input occurrence data and model approaches (Thibaud *et al.* 2014). Selecting the best model for invasion risk assessment remains controversial, and there is no agreement in ideal metrics to evaluate model performances (Lobo *et al.* 2008; Leroy *et al.* 2018). It is therefore important to explore both suitability and uncertainty in model predictions, and to explore predictions with variable model techniques and input parameters. Uncertainty is common in both correlative (e.g. Maxent) and process (e.g. CLIMEX) based model predictions, and in some cases more

robust predictions can be achieved if ensemble models are produced appropriately (Araújo & New 2007). We used ensemble models, as well as a mixture of environmental datasets that included various predictor variables, to generate predictions on habitat suitability that also showcased variability in predictions across models. Our approach allowed us to show general congruence among model approaches, which increases the robustness of our findings.

Our ensemble model predictions are generally consistent with a recent model prediction that was based on CLIMEX (Santos *et al.* 2023). Our predictions were based on correlative niche approaches and provide continuous suitability predictions from 0 to 1 (Fig. 3), whereas the CLIMEX model related insect physiological tolerance data with grid data and was developed with low and high suitability (Santos *et al.* 2023). Some mechanistic approaches to model species distribution tend to require large numbers of parameters for model calibration, including information on morphology, physiology and behavior of species (Peterson *et al.* 2015), whereas most of these data were not available for *E. fullo*. This represents a significant barrier to the use of mechanistic models in estimating its niche

and distribution. However, CLIMEX models generally require fewer parameters and are more parsimonious than other mechanistic models, which allowed us to compare our ecological niche models with prior work using CLIMEX.

Erthesina fullo and *H. halys* are two sympatric species in East Asia that share many biological characteristics such as polyphagy, a primarily r-selected life history, reproductive diapause, high dispersal capacity, aggregation behavior and association with human-modified ecosystems (Mi *et al*. 2020). These characteristics have made *H. halys* a successful global invader that has established populations in North and South America as well as Europe and has become a key agricultural pest causing considerable damage and losses to tree fruit, small fruit, nuts and vegetables (Zhu *et al*. 2012). There is little doubt that *E. fullo* has similar invasive potential and may be able to establish in many of the suitable areas identified here if transported to those regions. Like *H. halys*, invasions by *E. fullo* could have large economic impacts on a range of important horticultural crops, such as stone fruit, apples and kiwifruit (Mi *et al*. 2020). Our study shows that *E. fullo* could become a globally distributed invader and its potential establishment in new suitable regions should be carefully monitored and coordinated across the globe. More broadly, our study shows that using a range of ecological niche models and ensemble models can not only identify areas of potential suitability for an invasive species but also assess variability in predictions. For future invasive species, combining model techniques and explanatory variables of interest could be the optimal way to mitigate potential invaders before they occur or become widely established.

ACKNOWLEDGMENTS

REFERENCES

Aiello-Lammens ME, Boria RA, Radosavljevic A, Vilela B, Anderson RP (2015) spThin: an R package for spatial thinning of species occurrence records for use in ecological niche models. *Ecography* 38, 541–545. https://doi.org/10.1111/ecog.01132

Anton A (2021) How many alien species will there be in 2050? *Global Change Biology* 27, 968–969. https://doi.org/10.1111/gcb.15406

Araújo MB, New M (2007) Ensemble forecasting of species distributions. *Trends in Ecology and Evolution* 22, 42–47. https://doi.org/10.1016/j.tree.2006.09.010

Araújo MB, Anderson RP, Barbosa AM *et al*. (2019) Standards for distribution models in biodiversity assessments. *Science Advances* 5, eaat4858. https://doi.org/10.1126/sciadv.aat4858

Barbet-Massin M, Jiguet F, Albert CH, Thuiller W (2012) Selecting pseudo-absences for species distribution models: how, where and how many? *Methods in Ecology and Evolution* 3, 327–338. https://doi.org/10.1111/j.2041-210X.2011.00172.x

Barve N, Barve V, Jiménez-Valverde A *et al*. (2011) The crucial role of the accessible area in ecological niche modeling and species distribution modeling. *Ecological Modelling* 222, 1810–1819. https://doi.org/10.1016/j.ecolmodel.2011.02.011

Bradshaw CJA, Leroy B, Bellard C *et al*. (2016) Massive yet grossly underestimated global costs of invasive insects. *Nature Communications* 7, 12986. https://doi.org/10.1038/ncomms12986

Brugnera R, Lima Y, Grazia J, Schwertner CF (2022) Occurrence of the yellow-spotted stink bug *Erthesina fullo* (Thunberg) (Hemiptera: Pentatomidae) in Brazil, a polyphagous species from Asia. *Neotropical Entomology* 51, 325–329. https://doi.org/10.1007/s13744-021-00924-9

Deilami K, Kamruzzaman M, Liu Y (2018) Urban heat Island effect: a systematic review of spatio-temporal factors, data, methods, and mitigation measures. *International Journal of Applied Earth Observation and Geoinformation* 67, 30–42. https://doi.org/10.1016/j.jag.2017.12.009

Elith J, Kearney M, Phillips S (2010) The art of modelling range-shifting species. *Methods in Ecology and Evolution* 1, 330–342. https://doi.org/10.1111/j.2041-210x.2010.00036.x

Jiménez-Valverde A, Peterson AT, Soberón J, Overton JM, Aragón P, Lobo JM (2011) Use of niche models in invasive species risk assessments. *Biological Invasions* 13, 2785–2797. https://doi.org/10.1007/s10530-011-9963-4

Leroy B, Delsol R, Hugueny B *et al*. (2018) Without quality presence–absence data, discrimination metrics such as TSS can be misleading measures of model performance. *Journal of Biogeography* 45, 1994–2002. https://doi.org/10.1111/jbi.13402

Lobo JM, Jiménez-Valverde A, Real R (2008) AUC: A misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography* 17, 145–151. https://doi.org/10.1111/j.1466-8238.2007.00358.x

Mi Q, Zhang J, Gould E, Chen J, Sun Z, Zhang F (2020) Biology, ecology, and management of *Erthesina fullo* (Hemiptera: Pentatomidae): a review. *Insects* 11, 346. https://doi.org/10.3390/insects11060346

Naimi B, Araújo MB (2016) Sdm: A reproducible and extensible R platform for species distribution modelling. *Ecography* 39, 368–375. https://doi.org/10.1111/ecog.01881

Owens H, Barve V, Chamberlain S (2023) *spocc: Interface to Species Occurrence Data Sources*. Available from URL: http://docs.ropensci.org/spocc/

Peterson AT, Papeş M, Soberón J (2008) Rethinking receiver operating characteristic analysis applications in ecological niche modeling. *Ecological Modelling* **213**, 63–72. https://doi.org/10.1016/j.ecolmodel.2007.11.008

Peterson AT, Soberón J, Pearson RG *et al.* (2011) *Ecological Niches and Geographic Distributions*. Princeton University Press, New Jersey.

Peterson AT, Papeş M, Soberón J (2015) Mechanistic and correlative models of ecological niches. *European Journal of Ecology* **1**(**2**), 28–38. https://doi.org/10.1515/eje-2015-0014

R Core Team (2023) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna. Available from URL: https://www.R-project.org

Radosavljevic A, Anderson RP (2014) Making better Maxent models of species distributions: complexity, overfitting and evaluation. *Journal of Biogeography* **41**, 629–643. https://doi.org/10.1111/jbi.12227

Santos RC, Batista CV, Santos AA, da Silva RS, Picanço MC (2023) Predicting suitable areas for the invasive species *Erthesina fullo* (Hemiptera: Pentatomidae). *Crop Protection* **172**, 106297. https://doi.org/10.1016/j.cropro.2023.106297

Thibaud E, Petitpierre B, Broennimann O, Davison AC, Guisan A (2014) Measuring the relative effect of factors affecting species distribution model predictions. *Methods in Ecology and Evolution* **5**, 947–955. https://doi.org/10.1111/2041-210X.12203

Valentin RE, Fonseca DM, Nielsen AL, Leskey TC, Lockwood JL (2018) Early detection of invasive exotic insect infestations using eDNA from crop surfaces. *Frontiers in Ecology and the Environment* **16**, 265–270. https://doi.org/10.1002/fee.1811

Vignali S, Barras AG, Arlettaz R, Braunisch V (2020) *SDMtune*: an R package to tune and evaluate species distribution models. *Ecology and Evolution* **10**, 11488–11506. https://doi.org/10.1002/ece3.6786

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

Zhu G, Bu W, Gao Y, Liu G (2012) Potential geographic distribution of brown marmorated stink bug invasion (*Halyomorpha halys*). *PLoS ONE* **7**, e31246. https://doi.org/10.1371/journal.pone.0031246

Zhu GP, Peterson AT (2017) Do consensus models outperform individual models? Transferability evaluations of diverse modeling approaches for an invasive moth. *Biological Invasions* **19**, 2519–2532. https://doi.org/10.1007/S10530-017-1460-Y/FIGURES/5

Zurell D, Elith J, Schröder B (2012) Predicting to new environments: tools for visualizing model behaviour and impacts on mapped distributions. *Diversity and Distributions* **18**, 628–634. https://doi.org/10.1111/j.1472-4642.2012.00887.x

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the Supporting Information section at the end of the article.

Table S1. Candidate variables assembled. Candidate variables in dataset 1 have been used in *Halyomorpha haly*s.

Figure S1. The area under the curve (AUC) of receiver operating characteristic (ROC) plots for six ecological niche models using dataset 1: (i) generalized linear (GLM); (ii) generalized additive (GAM); (iii) boosted regression tree (BRT); (iv) random forest (RF); (v) support vector machine (SVM); and (vi) Maxent. Notice that RF is good at interpreting training data (i.e., high training AUC), a common phenomenon observed elsewhere.

Figure S2. The area under the curve (AUC) of receiver operating characteristic (ROC) plots for six ecological niche models based on dataset 2: (i) generalized linear (GLM); (ii) generalized additive (GAM); (iii) boosted regression tree (BRT); (iv) random forest (RF); (v) support vector machine (SVM); and (vi) Maxent. Notice that RF is good at interpreting training data (i.e., high training AUC), a common phenomenon observed elsewhere.

Figure S3. Mean (top) and standard deviation (SD, middle) of model predictions, and environmental similarity (MESS) in dataset 1 based models. In the spatial SD of model predictions, areas with higher values had less congruence across models. Areas in red have one or more variables outside the range present in training data and have less reliable distributional predictions than blue areas.

Figure S4. Mean (top) and standard deviation (SD, middle) of model predictions, and environmental similarity (MESS) in dataset 2 based models. In the spatial SD of model predictions, areas with higher values had less congruence across models. Areas in red have one or more variables outside the range present in training data and have less reliable distributional predictions than blue areas.