Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Effects of modelling choices on early invasive risk assessments: lessons from the psyllid vectors of Citrus Huanglongbing

Pedro Nunes
CBGP, INRAE, CIRAD, Institut Agro, IRD, Univ Montpellier
Virginie Ravigné
PHIM, Univ. Montpellier, INRAE, CIRAD, Institut Agro, IRD
Nicolas Sauvion
PHIM, Univ. Montpellier, INRAE, CIRAD, Institut Agro, IRD
Christine N. Meynard
christine.meynard@inrae.fr

CBGP, INRAE, CIRAD, Institut Agro, IRD, Univ Montpellier

Article

Keywords: Data quality, Citrus greening disease, HLB, ecological niche modelling, quarantine species, species distribution modelling (SDM)

Posted Date: November 1st, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11233148/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Species distribution models (SDMs) are widely used in early invasion risk assessments and are increasingly critical for anticipating biological threats under global change. Yet, SDM outcomes can vary substantially depending on modelling choices, many of which remain poorly standardized. These variations have tangible consequences when model outputs guide regulatory and management decisions. Here, we systematically evaluated how five key modelling decisions affect the performance and spatial projections of early-stage invasion risk models, including strategies that reflect the most common practices: (i) quality and curation of occurrence data, (ii) use of native-only versus native and invasion occurrence data in model calibration, (iii) background point selection, (iv) environmental variable selection, and (v) ensemble versus single-model approaches. In addition to conventional whole-range evaluation metrics (AUC, TSS, Boyce index), we applied a complementary index focused on the invaded range: the invasive False Negative Rate (iFNR). Finally, differences in projected areas at risk were estimated using the Jaccard index of overlap. As a case study, we modelled two invasive insects with contrasting invasion histories and distributions —*Diaphorina citri* and *Trioza erytreae*— vectors of the bacteria responsible for Huanglongbing (HLB), the most devastating citrus disease worldwide. Our results demonstrate that expert-curated occurrence data and inclusion of validated invaded-range records consistently improved model performance in novel environments. Moreover, conventional classification metrics alone failed to distinguish between models producing widely divergent invasion forecasts. Incorporating invasion-focused metrics and comparing spatial projections was essential to discriminate between models and to support more robust, policy-relevant decisions in invasion risk assessments.

Introduction

Biological invasions are among the greatest global threats to biodiversity [1], and represent a source of substantial economic and societal damages [2, 3]. Recent estimates suggest annual costs exceeding €12.5 billion in Europe and $21 billion in the United States alone [4–6]. The growing urgency to prevent biological invasions thus aligns both conservation priorities and socio-economic imperatives [7].

Early detection, coupled with rapid and decisive action, is a cornerstone of effective invasion prevention and control strategies [8–10]. Two widely-used quantitative approaches are central to early warning systems: horizon scanning to identify potential future invaders [11–13] and evaluation of climatic suitability in the invaded range [14–16]. Among these, species distribution models (SDMs) are the most widespread risk analysis tools used in early invasion stages [12, 16–18]. By statistically relating species occurrence or abundance to environmental variables, SDMs allow the identification of areas where an introduced species could potentially establish and persist. Thanks to the availability of large-scale environmental and occurrence datasets via web platforms such as GBIF and climate portals, SDMs can be rapidly implemented across a wide variety of species with a relatively low effort [19]. However, SDM predictions and their management implications are highly sensitive to methodological choices [20–22], making it particularly important to set clear standards for their implementation.

While recent best-practice guidelines for SDMs converge on core principles—such as using high-quality occurrence and environmental data, ensemble modelling, and performance metrics tailored to study goals [19, 23–25]—such standards are seldom applied in early invasion risk analyses. Instead, empirical invasive risk assessments frequently rely on the limited data and expertise available at the onset of the emerging threat [16]. This frequently translates into reliance on broadly available online data repositories such as GBIF and iNaturalist [26, 27] and on easy-to-use algorithms such as MaxEnt [28], typically run under default settings. For example, a review of SDM studies conducted on the 20 top-priority quarantine pests of the European Union (see Supplementary Material S1) reveals that 84% of the 112 studies identified used a single model—66% of them using solely MaxEnt—while only 10% implemented ensemble modelling. Almost all of these studies relied entirely or predominantly on occurrence data downloaded from GBIF, and evaluated models using performance metrics such as AUC or TSS, calculated on presence-background data (rather than on presence-absence) without explicitly assessing performance in invaded versus native ranges. Although GBIF and similar platforms provide unprecedented data access, they are also prone to well-documented limitations, including geographic sampling bias, identification errors, and geolocation errors [29–31]. Standard data cleaning procedures usually eliminate the most common errors [32, 33]. However, no simple solution exists to correct all datasets [31]. Ultimately, expert curation remains critical to ensure data quality. These gaps between best-practice guidelines and empirical risk assessments often reveal the value of simplified procedures that are more likely to be applied by managers and practitioners on the ground.

Beyond these core principles, other aspects of the modelling process remain more contentious. Predictor selection is one example. Araújo et al., (2019) [19] emphasized that predictors should ideally reflect the environmental conditions that are critical to the species' ecology. However, while this approach may offer clearer ecological interpretations, it is difficult to implement for recently introduced or poorly studied species, for which biological knowledge is sometimes limited. In practice, many studies either use a large set of predefined bioclimatic variables to ensure broad environmental coverage (e.g. [34]), or implement an automated variable selection strategy based on model performance and predictor’s significance (e.g. [35, 36]). The treatment of absence data is another key issue. Most SDMs used for invasive species rely on presence-only data, requiring the generation of background or pseudo-absence points. Random background sampling within the calibration area [37] is probably the most widespread strategy, but alternatives—such as kernel density filtering or exclusion of probable presence areas—have been proposed [14, 38, 39]. Yet, without true absence data, conventional classification metrics may poorly reflect real-world performance [40–42]. Furthermore, confirmed occurrences are often much more common in the native range than in the invasion range, especially during early stages of invasion. This inadvertently produces an imbalance in the number of occurrences used during the validation phase that underrepresents performance success in the invasion range [16]. A final critical decision is whether to include invaded-range occurrences during model calibration. Calibrating models using only native-range data aims to capture the species' fundamental niche, free from the confounding effects of human-mediated dispersal, novel biotic interactions, or recent evolutionary adaptation [43]. However, native distributions are themselves constrained by historical dispersal barriers, predation,

competition, and other biotic factors that may not exist in invaded regions [44, 45]. As such, native-only models may systematically underestimate potential invaded distributions [16, 46, 47]. Therefore, using occurrence data from both native and invasion ranges may better capture the full range of favorable conditions for the species of interest. Clear guidelines for invasion risk assessments regarding these more contentious issues are long overdue, and could potentially translate into important differences in terms of management recommendations for new invasion threats.

In this study, we aimed at tackling how these key modelling decisions affect the performance and outcomes of early-stage invasion risk assessments. Specifically, we evaluated which modelling choices provide the best simplicity-to-performance gains to provide clear guidelines for SDM studies during early invasion risk assessments. We explored these questions using two invasive insect psyllids, considered as major agricultural pests: the Asian citrus psyllid, *Diaphorina citri* (Kuwayama, 1908), and the African citrus psyllid, *Trioza erytreae* (Del Guercio, 1918). Both species are vectors of the citrus greening disease (also known as Huanglongbing or HLB), which is considered the largest global threat to the citrus industry and a EU priority quarantine pest [48, 49]. *Diaphorina citri* is native to the warm low-elevation citrus belt of South-East Asia, where year-round leaf flushing and mean winter temperatures above ≈ 10°C prevail. *Diaphorina citri* has spread with global citrus production since the 1940s—first to South America, and then to the South-West Indian Ocean in the 1960s–70s, to the Caribbean and North America in the 1990s, and more recently to the Middle East, eastern and western Africa. *Trioza erytreae*, originally restricted to the cool, humid highlands of eastern and southern Africa, was detected in Madeira (1994) and the Canary Islands (2002), then in the Arabian Peninsula and north-west Iberia in the 2010s [50, 51]. For each psyllid species we evaluated SDM performance using traditional classification metrics (AUC, TSS), the Boyce Index (based solely on presence data), and a novel metric focused on errors in the invaded range (invasive False Negative Rate, iFNR). We also compared predicted distributions using a spatial overlap index to quantify geographic discrepancies between model outputs. To this aim, we sought to answer three specific questions (Fig. 1). First, how does the most common and simplest modelling strategy compare to the best practice alternative? Second, do the usual whole-range classification metrics reflect performance in the invaded range? And, finally, do differences in performance metrics reflect differences in geographic predictions of the potential range?

Material and Methods

General approach

Our general approach is outlined in Fig. 1. In brief, we compared the outcomes of seven modelling strategies differing in classical modelling choices for data quality, calibration data, treatment of absence points, variable selection and modelling algorithms. We first implemented a highest-standard reference strategy (called REF in Fig. 1) which implements all major recommendations for SDMs to reach ecological relevance [19, 34]. REF relies on an ensemble modelling approach with the settings recommended by Valavi et al., (2022) [34], including five modelling algorithms (MaxEnt, GAM, LASSO, BRT, and Random Forests). REF was calibrated with a high-quality occurrence dataset compiled and

verified by expert knowledge, including all verified native and invasion occurrences. At the opposite end of the spectrum, we also fitted a default modelling strategy (DEF) with quicker and simpler options for all modelling choices (Fig. 1): use of occurrences downloaded from GBIF including only a basic cleaning stage without input from experts, a pre-determined set of predictors, and MaxEnt-only modelling. Five variants of REF (Fig. 1, lower-right box), including all high-standard modelling choices but one, were then implemented: strategy OCC replaced the high-quality occurrence dataset with the basic occurrence dataset downloaded from GBIF; strategy NAT used only native-range presence records, as filtered from the full high-quality dataset, for model calibration; strategy BACK replaced the pseudo-absence sampling with a random background point sampling; strategy AUTO replaced the pre-determined set of predictors used in REF with an automated variable selection; and strategy MAX used only MaxEnt instead of an ensemble modelling strategy. All modelling strategies were compared to REF in terms of classification error and projected distributions (Fig. 1).

We provide further details below, and a full modelling report is also available in Online Resource 2 following the ODMAP reporting protocol [52]. Modelling was conducted in R v4.2.1 [53]. Scripts were adapted from [34, 54] are available in a repository (https://tinyurl.com/yc29szae).

Occurrence datasets

Two datasets were used, corresponding to two different levels of data cleaning and compilation investment (Fig. 2 and Online Resource 3). The first dataset, hereafter named the *basic occurrence dataset*, consists of occurrence records compiled only from GBIF (https://www.gbif.org), followed by a standard cleaning procedure (i.e. elimination of duplicates, records that fall in visibly unlikely places such as oceans, records without coordinates, with high uncertainty in the coordinates, and with a mismatch between coordinates and country of origin). We included occurrences of the two psyllid species of interest, as well as the three bacterial species responsible for HLB, which are only vectored by the studied species [55], and four parasitoid species, all strict specialists of the studied species [56–59]. Neither the vectored bacteria nor the parasitoids can occur in the absence of these psyllids, therefore providing for an excellent proxy for the insect's presence.

The second dataset, hereafter called the *high-quality dataset* (Online Resource 3) involved a search for additional records as well as consultation with psyllid experts and local researchers, especially in the invaded range, to identify potential identification errors and unlikely records. This search also included other public biodiversity repositories (iNaturalist, EPPO), official reports, and scientific articles. This process helped us identify incorrect occurrence errors (false positives) in previous studies, and to uncover new records not included in earlier research datasets (Online Resource 3).

Both occurrence datasets were up-scaled to match the environmental layers, eliminating spatial replicates within individual grid cells. After upscaling to the modelling resolution, the basic occurrence dataset contained 241 unique occurrences (64 for *T. erytreae* and 177 for *D. citri*), while the high-quality

dataset contained 1,194 unique occurrences (355 for *T. erytreae* and 839 for *D. citri*). The high-quality dataset was used for all modelling efforts except for DEF (Fig. 1).

Environmental variables

Climatic predictors were obtained from CHELSA V2.1, which provides mean annual conditions averaged over 1990–2010 at a resolution of 30 arc seconds (~ 1 km) [60]. All variables were up-scaled to a resolution 20 times coarser than the original (i.e., 10 min resolution, ~ 18.5 km) using mean values for modelling at the global scale.

We included 19 bioclimatic variables (Bio1–Bio19), representing annual, seasonal, and extreme temperature and precipitation [60], four variables related to thermal accumulation (growing degree days at 0°C, 5°C, and 10°C, i.e., GDD0, GDD5, GDD10) and a variable describing vegetation productivity (Net Primary Productivity, NPP).

The strategy that used automated variable selection (AUTO, Fig. 1) considered this set of predictors as the starting point. Variable selection was then performed using a data snooping procedure [36, 61], applied to GAM following [35]. This approach consists in a forward selection based on the predictors that maximize deviance and that are not highly correlated with previously selected predictors at each step, until no other predictor with low correlation or with a significant effect is available among the remaining variables considered [35].

All other modelling strategies used a pre-determined subset of predictors that were chosen following two main criteria: (i) scanning the literature for the bioclimatic variables most associated with the studied species [62, 63], and (ii) including at least one variable representing the average, seasonality, and extremes (e.g., in the coldest quarter, in the driest quarter) of both temperature and precipitation (Online Resource 4). Pairwise correlations among these variables were assessed within the calibration dataset of each modelling strategy. A variable was removed if a high correlation with the other predictors was present (Spearman $r > 0.7$).

Pseudo-absences and background points

For each model, 10,000 background points were randomly generated within a 1,000 km buffer around all occurrences used in the calibration phase, to represent historically accessible area for each species [64–66] as recommended by Barve et al. (2011). Random background points therefore included areas where the species are potentially present.

For all strategies except DEF and BACK (Fig. 1), pseudo-absences were used, i.e., the random background points lying within a 100-km buffer around occurrence points were eliminated, thereby avoiding pseudo-absences near confirmed presences. The 100-km buffer corresponds to the potential individual dispersal abilities, and was aimed at excluding areas where the psyllids could potentially be

present given their dispersal range and the distribution of confirmed occurrences. This is the only main change implemented with respect to [34] in the high-standard strategies, and the comparison of BACK with REF enables checking its effect on model outcomes. During the modelling phase, a weight was added to the different modelling strategies to balance presences and background or pseudo-absences to a 50% prevalence [34].

Modelling performance

We implemented a fivefold spatial cross-validation scheme using 100 spatial blocks (10 × 10) generated with the "blockCV" R package [68]. These blocks were randomly assigned to folds across 50 balanced iterations to find the optimal spatial blocks. This method allows testing the models on a geographically distinct block, different from the ones used to calibrate them, decreasing the spatial dependence on performance statistics [68]. Model performance was assessed using a suite of commonly used metrics: the Area Under the ROC Curve (AUC), the True Skill Statistic (TSS), and the Boyce Index (based on presence-only data) [69]. Additionally, we introduced the invasive False Negative Rate (iFNR), calculated as the rate of omission errors in the invaded range (using only confirmed occurrences in the invaded range), to capture how often models fail to identify known invaded areas. Binary presence-absence classification of predictions, when needed, was based on the threshold that maximized the sum of sensitivity and specificity [70]. To compare spatial projections, we used the Jaccard index [40, 71, 72] which quantifies overlap between each model's binary prediction map and the reference model. Values range from 0 (no spatial overlap) to 1 (identical projections), providing a measure of the difference in terms of predicted range over the whole range.

Results

Model performance

In the cross-validation assessment of model performance, the standard metrics based on whole-range presences were seemingly satisfactory for both species for all modelling strategies (Fig. 3). Mean AUC values were all > 0.82, TSS > 0.55, and Boyce index > 0.69 (Fig. 3a-f). Considering these three classification indices, the reference modelling strategy was always among the best ones. MaxEnt-only (MAX), automated variable selection (AUTO), and background points (BACK) also had consistently good performances that were not significantly lower than the reference strategy (Fig. 3a-f). In contrast, the default modelling strategy (DEF), and the basic occurrence modelling strategy (OCC) always had significantly lower performances compared to the reference (REF). The native-only strategy (NAT) also showed significantly lower performance compared to REF, except for *D. citri* using TSS (Fig. 3d). Although the Boyce index (Fig. 3e-f) produced slightly different results with respect to the significance levels, general trends and model rankings were the same as using AUC or TSS. Overall, based on whole-range metrics, REF, MAX, AUTO and BACK come out as performing equally well, whereas DEF, OCC and NAT ranked lower in performance.

The iFNR, which specifically assessed model performance in predicting verified species occurrences in the invaded range, yielded slightly different results (Fig. 3g-h). The REF modelling strategy performed very well (mean iFNR was 7.7 ± 2.5% for *D. citri*, and 1.8 ± 0.4% for *T. erytreae*), with no significant differences compared to AUTO and BACK modelling strategies, although AUTO outperformed REF for *D. citri*. The MAX modelling strategy also performed well for *T. erytreae*, but significantly worse than REF for *D. citri* (Fig. 3g-h). At the other extreme, the default model (DEF) presented a very poor performance in the invaded range (mean iFNR of 25.4 ± 4.8% for *D. citri* and 8.7 ± 2.0% for *T. erytreae*), meaning that the model failed to predict, on average, more than 25% of the confirmed species occurrences in the invaded range for *D. citri*. For both species, OCC and NAT models also showed very poor iFNR values. NAT and DEF strategies also produced high variability in model performance metrics across cross-validation blocks (spread of points in NAT and DEF in Fig. 3a-h).

Finally, the Jaccard index (Fig. 3i-j) revealed substantial spatial discrepancies in range predictions compared to the REF modelling strategy. For example, for *D. citri*, BACK, AUTO and MAX showed a Jaccard index of around 0.75–0.80 with the REF modelling strategy (Fig. 3j), whereas for *T. erytreae* the range overlap decreased sharply for AUTO and MAX (Fig. 3i). This spatial variation was further highlighted by visual comparisons of the projected potential ranges globally (Fig. 4). For example, while all modelling strategies provided similar projections in the species native ranges (i.e., Africa for *T. erytreae* and Asia for *D. citri*), they showed great differences in the potential invaded range (Fig. 5). For example, OCC and NAT did not predict suitable conditions in wide areas of Portugal for *T. erytreae*, or South America (southern Brazil and northern Argentina) for *D. citri*, where there are many confirmed invasion areas already (Fig. 5). The uncertainty remains high in region where the species has not invaded yet. For example, in South America, where *T. erytreae* has not been recorded, the DEF, along with NAT and OCC modelling strategies predict large areas of potential range in areas where the REF, BACK or AUTO modelling strategies did not (Fig. 4). At the opposite, *D. citri* was predicted by the DEF strategy in vast areas, especially in the Sahel region in Africa, where it has not been recorded, while no other strategy predicts these areas as suitable. Overall, these discrepancies suggest that all model projections differ enormously in the projected distributions, irrespective of whether the differences in performance metrics are small or large.

Discussion

The present study showed the significant impact that methodological choices can have on the outcomes of SDMs used in early invasive risk assessments. The most influential factors were related to the occurrence data (both quality and spatial extent). To a lesser extent, the use of a single modelling algorithm, as opposed to ensemble modelling, and the use of automated variable selection as opposed to a predefined set of variables, led to substantial differences in the predicted invasion ranges. Hence modelling choices may imply very different recommendations regarding surveillance and monitoring. Below we answer our three initial questions (Fig. 1) in light of these results, and provide clear guidelines for early invasive risk assessments using SDMs.

Best practice modelling strategies compared to common practice

Early-invasion risk modelling with SDMs is often constrained by scarce occurrence data, limited or no records outside the species' native range, minimal biological information, along with the fact that absence data in the invasion range is not reliable given the expanding nature of the invasion front [14–16]. While these limitations are substantial, particularly as data quality heavily influences model performance, our analyses show that even under such constraints, SDMs can still yield informative results. The high-standards modelling strategy (REF) correctly predicts, on average, 90% to 96% of the confirmed occurrences in the invaded range (Fig. 3g-h), therefore providing for an excellent identification of high-risk areas. This result is important *per se* for moderate- to high-risk regions that have not yet been invaded, such as large parts of Australia and both the northern and southern coasts of the Mediterranean Basin, which are identified as suitable for both vector species by the REF strategy. At the opposite end of the spectrum, the default modelling strategy (DEF) performs worse than REF and all the considered variants, both in terms of overall classification rates, and in terms of invasive range predictions (Fig. 3). DEF misses on average between 8% to 25% of the occurrences in the invasion area, meaning that the actual identification of high-risk areas is compromised under the quick and simple approach (Figs. 4 & 5). This confirms recommendations that have already been made regarding modelling settings and the potential of SDMs to inform risk analysis in situations as complex as early invasion stages.

The most important modelling choices for proper model predictions in the invasive range were related to the occurrence dataset used during calibration. Two aspects were explored and had major impacts on modelling outcomes: data extent (native-only *versus* native and invasive, comparison between NAT and REF) and data quality (basic dataset *versus* a high-quality dataset, comparison between OCC and REF). Both using basic data (OCC) and using only native data (NAT) for calibration produced poorer whole-range statistics for both species, and significantly higher misclassification of confirmed invaded occurrences (Fig. 3). The use of native-only data for calibration is still a common practice in early invasion risk modelling (e.g. [63] for *T. erytreae*), making of this an important result. On this point our study is in line with Young et al. (2025) [16], who found that models including the invaded range during early invasion stages significantly improve model predictions. In later invasion stages, when a large number of distinct occurrences are available, models may even benefit from focusing on the invaded range rather than on native-only or native and invasion occurrences [16]. There are two main hypotheses that could help explain this result. The first one is that the species of interest occupies a broader environmental range in the invaded area because the dispersal constraints or biological interactions that restrict its range in their native area do not exist in the invaded range [73, 74]. The second one is that there is rapid evolution of environmental preferences in the invaded range, allowing an expansion of the environmental niche during invasion [75]. While testing both hypotheses would involve comparing native and invaded populations, the consequences in terms of modelling are the same: including occurrences in the invaded area is the only way to consider the full range of environmental conditions that the

species can tolerate, which has been one of the main recommendations coming from the modelling community [76]. In contrast, the exclusion of invasion occurrences does not seem to bring any potential benefits to the early-stage risk assessment results.

Differences in projections found in the literature are further confounded with errors in the occurrence datasets used during model calibration. During our intensive data validation process, we found occurrence points that had been used in previous SDM studies, including an incorrect isolated record for *D. citri* in South Africa [77], erroneous presences in central Spain for *T. erytreae* and for *D. citri* [78], as well as a false record in Madagascar [63], among many other errors (Online Resource 3). Of course, the influence of these errors will depend on how many, and how different from the other environments present in the correct occurrence records they are. Filtering these records involved checking in detail all the records, contacting psyllid experts, particularly in the invaded areas that seemed suspicious, and investing a significant amount of time in a rather tedious task. To enhance our occurrence dataset, we also incorporated records of highly specific parasitoid species and disease presence alongside known vector species. While rare in the literature, this strategy parallels studies that use vector or host species distributions to model pathogen risk [79]. This approach is valuable when data is scarce because there is a strong ecological specificity. However, we do not recommend this approach for generalist parasitoids, as their broader host range could lead to false inferences of target species' presence. Overall, the results show that this additional investment in data quality paid off, since the prediction success in the invaded area significantly increased.

In contrast to modelling choices affecting calibration data, modelling strategies differing from highest-standards in other methodological choices sometimes produced similar, and occasionally better, performance statistics. Not surprisingly, variant strategies that are comparable to REF involved modelling choices for which there is not necessarily a consensus on best practices [19, 25]. This is the case for the use of background data (BACK) rather than pseudo-absences, or using an automated variable selection (AUTO) rather than a pre-determined set of predictors. This is also consistent with other studies suggesting that neither variable selection [80] nor background sampling [39] are among the main factors affecting model performance, as long as the data is representative of the species environmental tolerances.

The case of the use of MaxEnt-only versus ensemble modelling is worth considering more carefully. For whole-range performance metrics, the MaxEnt-only strategy (MAX) did not differ significantly compared to the REF. However, MAX does show significantly lower performance in terms of iFNR for one of our model species (*D. citri*), suggesting that the use of a single model may occasionally produce higher uncertainty in the predicted invasion range. The principle of ensemble modelling is to take into account the variability generated by different modelling algorithms [81, 82]. This is particularly important when potential distribution areas are projected into novel environments [19]. Here we only considered two species, but Valavi et al. (2022), using a much more extensive pool of test taxa and regions, clearly showed that the ensemble modelling approach outperformed the MaxEnt-only modelling in a general presence-only SDM framework. Although Valavi et al. (2022) [34] was based on whole-range

performance statistics and was not focused on invasive species, it is likely that ensemble modelling provides for an additional guarantee to take modelling uncertainty into account. Nowadays, there are multiple packages that allow routinely integrating ensemble modelling into SDM usage, and published analysis pipelines that make such applications more accessible to the more lay-back users and even to researchers who do not have access to cluster or other computer-intensive resources [83]. The use of ensemble modelling rather than single models therefore seems like a low-cost and easy-to-implement solution for improving the reliability of risk assessments related to invasive species.

Relationship between whole-range and invaded range performance metrics

By comparing model ranking using classical performance metrics based on the whole-range (AUC, TSS and Boyce) *versus* on confirmed presences in the invaded range (iFNR), our work emphasized the importance of using performance metrics tailored to the actual goal of the study. Specifically, AUC and Boyce result in the same ranking between modelling strategies, and almost identical significant comparisons with the REF modelling strategy. TSS differs slightly from these in that it ranks the OCC modelling strategy for *D. citri* lower than any of the others, and this is the only index where the DEF modelling strategy is not significantly worse than the REF modelling strategy. Despite many criticisms regarding these metrics [40, 41], they are still among the most commonly used performance metrics for ecological modelling. For example, all the SDM articles on the global distribution of HLB published to date have relied on AUC as the main index of model performance [62, 84, 85], except for one [63] which also used TSS and the Boyce index. Despite clear underperformance of modeling strategies such as NAT and OCC, obvious in both suitability maps (Fig. 5) and high false negative rates in the invasive range (Fig. 3, iFNR), whole-range metrics (Boyce Index, AUC, and TSS) remained relatively high for these strategies. This was particularly noticeable for *T. erytreae*, where the NAT strategy had a reduced predictive power in the invaded range, which was not captured at all by the Boyce index (Fig. 3). For both of our studied species, the number of occurrences in the invaded range is relatively small with respect to the number of occurrences in the native range, which inadvertently gives very little weight to performance in the invasion area [16], a pattern expected to be quite general in invasive species. This is especially true for *T. erytreae*, which has invaded a much smaller area as compared to *D. citri*. Disentangling the shortcomings of the indices themselves as compared to the shortcomings of the limited data would be worth pursuing. This type of exploration is difficult to accomplish with real-word datasets because a large dataset of both invaded and native ranges would be needed (but see Young et al., 2025). Although outside of the scope of this study, the use of virtual species [76] is a promising route to further explore these issues in depth.

Our study illustrated the interest of the false negative rate in invaded regions (iFNR) as a valuable metric for assessing model performance in invasion biology contexts. It effectively revealed the shortcomings of lower-quality strategies (OCC, NAT, and DEF) that were not highlighted by standard evaluation approaches. This metric also aligns with the core principle of biosecurity, which prioritizes minimizing false negatives to avoid underestimating potential threats. Furthermore, iFNR is simple to interpret and

grounded in the most reliable component of presence-only modeling: the prediction of presences (and thus, false negatives). However, iFNR can only be calculated once the invasion is already confirmed, since it is based on occurrences in the invasion range. Furthermore, care should be taken to consider only occurrences where the species of interest has already established. Indeed, in early invasion stages, some populations may be observed sporadically and then become extinct, indicating that the environment is not favorable to a long-term survival of the species.

Main early risk assessment guidelines and perspectives

Although our findings were based on two species only, the insights presented here provide valuable guidelines for enhancing SDM reliability in early invasive species assessments. Rigorous occurrence data curation and validation, inclusion of the invaded range data in model calibration when available, and the use of ensemble modeling to reduce individual algorithm biases are among the simplest and most effective recommendations that we can make. Another important component, often under-represented in SDM studies, is the use of model evaluation metrics that are adapted to the data and objectives, such as iFNR for invasion-focused applications when the invasion is already under way. The fact that the most commonly used performance indices such as AUC and TSS do not necessarily reflect predictive success in the invasion range, combined with very large differences in model projections, calls for further exploration of model performance metrics that are better adapted to early invasion. At the very least, a reiterative modelling process akin to an iterative near-term ecological forecasting [86] that can easily incorporate newly-available occurrence records in the invasion front should become standard practice to allow timely adaptation of management actions.

Declarations

Additional Information (including a Competing Interests Statement)

We do not have any competing interests to declare.

Funding

This research was funded by the project ANR BEYOND (ANR-20-PCPA-0002). VR also benefitted from support the projects ECOPHYTO II+/OFB IRIS (convention n° 23-1316).

Author Contribution

C.N.M., N. S. and V.R. obtained the funding. All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by P.N., who also wrotte the first draft

of the manuscript. All authors commented and edited subsequent versions of the manuscript. All authors read and approved the final version.

Acknowledgement

Authors wish to particularly thank David Ouvrard for identification advice on doubtful psyllid mentions and Philippe Ryckewaert for his expertise on psyllid diversity worldwide. We acknowledge the use of ChatGPT to rephrase some portions of the manuscript to improve English and to improve readability of the text. This research was funded by the project ANR BEYOND (ANR-20-PCPA-0002). VR also benefitted from support the projects ECOPHYTO II+/OFB IRIS (convention n° 23-1316).

Data Availability

Scripts, along with all data required to run them, are available in the following archive: https://tinyurl.com/yc29szae

References

1. Brook, B. W., Sodhi, N. S. & Bradshaw, C. J. A. Synergies among extinction drivers under global change. *Trends Ecol. Evol.* **23**, 453–460 (2008).
2. Bradshaw, C. J. A. et al. Damage costs from invasive species exceed management expenditure in nations experiencing lower economic activity. *Ecol. Econ.* **220**, 108166 (2024).
3. Soto, I. et al. Using species ranges and macroeconomic data to fill the gap in costs of biological invasions. *Nat. Ecol. Evol.* **9**, 1021–1030 (2025).
4. Diagne, C. et al. vol 592, 571,. High and rising economic costs of biological invasions worldwide *NATURE* 608, E35–E35 (2022). (2021).
5. Fantle-Lepczyk, J. E. et al. Economic costs of biological invasions in the United States. *Sci. Total Environ.* **806**, 151318 (2022).
6. Pyšek, P. et al. Scientists' warning on invasive alien species. *Biol. Rev. Camb. Philos. Soc.* **95**, 1511–1534 (2020).
7. Bridgewater, P. & Schmeller, D. S. Nature abhors a vacuum: invasive themes and other outputs from the tenth plenary of the Intergovernmental Platform on Biodiversity and Ecosystem Services (IPBES-10). *Biodivers. Conserv.* **33**, 867–873 (2024).
8. Branco, S. et al. Eradication programs against non-native pests and pathogens of woody plants in Europe: which factors influence their success or failure? *NEOBIOTA* 84, 281–317 (2023).
9. Liebhold, A. M. et al. Eradication of Invading Insect Populations: From Concepts to Applications. in ANNUAL REVIEW OF ENTOMOLOGY, VOL 61 (ed Berenbaum, M. R.) vol. 61 335–352 (Annual Reviews, Palo Alto, (2016).

10. Meynard, C. N., Lecoq, M., Chapuis, M. P. & Piou, C. On the relative role of climate change and management in the current desert locust outbreak in East Africa. *Glob. Change Biol.* **26**, 3753–3755 (2020).
11. Fournier, A., Penone, C., Grazia Pennino, M. & Courchamp, F. Predicting future invaders and future invasions. *Proc. Natl. Acad. Sci. U. S. A.* 116, 7905–7910 (2019).
12. Ivison, K. et al. Horizon scanning of potential invasive alien plant species and their distribution in Norway under a changing climate. *Ecography* e07604 (2025). (2025).
13. Pili, A. N. et al. Forecasting potential invaders to prevent future biological invasions worldwide. *Glob. Change Biol.* **30**, e17399 (2024).
14. Elith, J., Kearney, M. & Phillips, S. The art of modelling range-shifting species. *Methods Ecol. Evol.* **1**, 330–342 (2010).
15. Lake, T. A., Runquist, B., Moeller, D. A. & R. D. & Predicting range expansion of invasive species: Pitfalls and best practices for obtaining biologically realistic projections. *Divers. Distrib.* **26**, 1767–1779 (2020).
16. Young, N. E., Williams, D. A., Shadwell, K. S., Pearse, I. S. & Jarnevich, C. S. How to model a new invader? US-invaded range models outperform global or combined range models after 100 occurrences. *Ecol. Appl.* **35**, e70010 (2025).
17. Reaser, J. K. et al. The early detection of and rapid response (EDRR) to invasive species: a conceptual framework and federal capacities assessment. *Biol. Invasions*. **22**, 1–19 (2020).
18. Soubeyrand, S. et al. Building integrated plant health surveillance: a proactive research agenda for anticipating and mitigating disease and pest emergence. *CABI Agric. Bioscience*. **5**, 72 (2024).
19. Araújo, M. B. et al. Standards for distribution models in biodiversity assessments. *Sci. Adv.* **5**, eaat4858 (2019).
20. Kumar, S., Yee, W. L. & Neven, L. G. Mapping Global Potential Risk of Establishment of Rhagoletis pomonella (Diptera: Tephritidae) Using MaxEnt and CLIMEX Niche Models. *J. Econ. Entomol.* **109**, 2043–2053 (2016).
21. Srivastava, V., Lafond, V. & Griess, V. C. Species distribution models (SDM): applications, benefits and challenges in invasive species management. *CABI Reviews* 1–13 (2019). (2019).
22. Xia, Y. et al. Geographical distribution of two major quarantine fruit flies (*Bactrocera minax* Enderlein and *Bactrocera dorsalis* Hendel) in Sichuan Basin based on four SDMs. *PeerJ* **12**, e16745 (2024).
23. Feng, X. et al. A checklist for maximizing reproducibility of ecological niche models. *Nat. Ecol. Evol.* **3**, 1382–1395 (2019).
24. Sillero, N. et al. Want to model a species niche? A step-by-step guideline on correlative ecological niche modelling. *Ecol. Model.* **456**, 109671 (2021).
25. Soley-Guardia, M., Alvarado-Serrano, D. F. & Anderson, R. P. Top ten hazards to avoid when modeling species distributions: a didactic guide of assumptions, problems, and recommendations. *Ecography* e06852 (2024). (2024).

26. La Sorte, F. A. & Somveille, M. Survey completeness of a global citizen-science database of bird occurrence. *Ecography* **43**, 34–43 (2020).
27. Nelson, G. & Ellis, S. The history and impact of digitization and digital data mobilization on biodiversity research. *Philosophical Trans. Royal Soc. B: Biol. Sci.* **374**, 20170391 (2018).
28. Phillips, S. J., Anderson, R. P. & Schapire, R. E. Maximum entropy modeling of species geographic distributions. *Ecol. Model.* **190**, 231–259 (2006).
29. Beck, J., Böller, M., Erhardt, A. & Schwanghart, W. Spatial bias in the GBIF database and its effect on modeling species' geographic distributions. *Ecol. Inf.* **19**, 10–15 (2014).
30. Morales-Barbero, J. & Vega-Alvarez, J. Input matters matter: Bioclimatic consistency to map more reliable species distribution models. *Methods Ecol. Evol.* **10**, 212–224 (2019).
31. Zizka, A. et al. No one-size-fits-all solution to clean GBIF. *PeerJ* **8**, e9916 (2020).
32. Aiello-Lammens, M. E., Boria, R. A., Radosavljevic, A., Vilela, B. & Anderson, R. P. spThin: an R package for spatial thinning of species occurrence records for use in ecological niche models. *Ecography* **38**, 541–545 (2015).
33. Zizka, A. et al. Standardized cleaning of occurrence records from biological collection databases. *Methods Ecol. Evol.* **10**, 744–751 (2019). CoordinateCleaner.
34. Valavi, R., Guillera-Arroita, G., Lahoz-Monfort, J. J. & Elith, J. Predictive performance of presence-only species distribution models: a benchmark study with reproducible code. *Ecol. Monogr.* **92**, e01486 (2022).
35. Meynard, C. N. et al. Climate-driven geographic distribution of the desert locust during recession periods: Subspecies' niche differentiation and relative risks under scenarios of climate change. *Glob. Change Biol.* **23**, 4739–4749 (2017).
36. Verdon, V. et al. Can we accurately predict the distribution of soil microorganism presence and relative abundance? *Ecography n/a*, e07086 (2024).
37. Barbet-Massin, M., Jiguet, F., Albert, C. H. & Thuiller, W. Selecting pseudo-absences for species distribution models: how, where and how many? *Methods Ecol. Evol.* **3**, 327–338 (2012).
38. Hattab, T. et al. A unified framework to model the potential and realized distributions of invasive species within the invaded range. *Divers. Distrib.* **23**, 806–819 (2017).
39. Rausell-Moreno, A., Galiana, N., Naimi, B. & Araújo, M. B. Improving species distribution models by optimising background points: Impacts on current and future climate projections. *Ecol. Model.* **507**, 111177 (2025).
40. Leroy, B. et al. Without quality presence-absence data, discrimination metrics such as TSS can be misleading measures of model performance. *J. Biogeogr.* **45**, 1994–2002 (2018).
41. Lobo, J. M., Jiménez-Valverde, A. & Real, R. AUC: a misleading measure of the performance of predictive distribution models. *Glob. Ecol. Biogeogr.* **17**, 145–151 (2008).
42. Yackulic, C. B. et al. Presence-only modelling using MAXENT: when can we trust the inferences? *Methods Ecol. Evol.* **4**, 236–243 (2013).

43. Araújo, M. B. & Peterson, A. T. Uses and misuses of bioclimatic envelope modeling. *Ecology* **93**, 1527–1539 (2012).
44. Liu, H. & Stiling, P. Testing the enemy release hypothesis: a review and meta-analysis. *Biol. Invasions*. **8**, 1535–1545 (2006).
45. Soberón, J. & Peterson, A. T. Interpretation of Models of Fundamental Ecological Niches and Species' Distributional Areas. *Biodivers. Inf.* **2**, 1–10 (2005).
46. Broennimann, O. & Guisan, A. Predicting current and future biological invasions: both native and invaded ranges matter. *Biol. Lett.* **4**, 585–589 (2008).
47. Loo, S. E., Nally, R. M. & Lake, P. S. Forecasting New Zealand Mudsnail Invasion Range: Model Comparisons Using Native and Invaded Ranges. *Ecol. Appl.* **17**, 181–189 (2007).
48. Li, S., Wu, F., Duan, Y., Singerman, A. & Guan, Z. Citrus Greening: Management Strategies and Their Economic Impact. *HortScience* 55, 604–612 (2020).
49. Singerman, A. The Perfect Storm That Hit Florida Citrus. *Choices: Magazine Food Farm. Resource Issues* **39**, (2024).
50. Monzó, C. Los psílidos Diaphorina citri y Trioza erytreae como vectores de la enfermedad de cítricos Huanglongbing (HLB): reciente detección de T. erytreae en la Península Ibérica. (2015).
51. Pérez Otero, R. & Mansilla Vázquez, J. P. & Del Estal, P. Detección de la psila africana de los cítricos, Trioza erytreae (Del Guercio, 1918) (Hemiptera: Psylloidea: Triozidae), en la Península Ibérica. *Arquivos Entomolóxicos* 119–122 (2015).
52. Zurell, D. et al. A standard protocol for reporting species distribution models. *Ecography* **43**, 1261–1277 (2020).
53. R-Core-Team *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2022).
54. Meynard, C. N., Piou, C. & Kaplan, D. M. A theoretical framework for upscaling species distribution models. *Methods Ecol. Evol.* **14**, 2888–2899 (2023).
55. Gottwald, T. R. Current epidemiological understanding of citrus huanglongbing. *Annu. Rev. Phytopathol.* **48**, 119–139 (2010).
56. Bistline-East, A., Pandey, R., Kececi, M. & Hoddle, M. S. Host Range Testing of *Diaphorencyrtus aligarhensis* (Hymenoptera: Encyrtidae) for Use in Classical Biological Control of *Diaphorina citri* (Hemiptera: Liviidae) in California. *J. Econ. Entomol.* **108**, 940–950 (2015).
57. Hoddle, M. S. & Pandey, R. Host range testing of *Tamarixia radiata* (Hymenoptera: Eulophidae) sourced from the Punjab of Pakistan for classical biological control of *Diaphorina citri* (Hemiptera: Liviidae: Euphyllurinae: Diaphorinini) in California. *J. Econ. Entomol.* **107**, 125–136 (2014).
58. Pérez-Rodríguez, J. et al. Classical biological control of the African citrus psyllid *Trioza erytreae*, a major threat to the European citrus industry. *Sci. Rep.* **9**, 9440 (2019).
59. Urbaneja-Bernat, P. et al. Host range testing of *Tamarixia dryi* (Hymenoptera: Eulophidae) sourced from South Africa for classical biological control of *Trioza erytreae* (Hemiptera: Psyllidae) in Europe.

Biol. Control. **135**, 110–116 (2019).

60. Karger, D. N. et al. Climatologies at high resolution for the earth's land surface areas. *Sci. Data*. **4**, 170122 (2017).
61. Dormann, C. F. et al. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography* **36**, 27–46 (2013).
62. Aidoo, O. F. et al. Predicting the potential global distribution of an invasive alien pest Trioza erytreae (Del Guercio) (Hemiptera: Triozidae). *Sci. Rep.* **12**, 20312 (2022).
63. Godefroid, M. Species distribution models predicting climate suitability for the psyllid *Trioza erytreae*, vector of citrus greening disease. *Crop Prot.* **168**, 106228 (2023).
64. Bayles, B. R., Thomas, S. M., Simmons, G. S. & Grafton-Cardwell, E. E. Daugherty, M. P. Spatiotemporal dynamics of the Southern California Asian citrus psyllid (Diaphorina citri) invasion. *PLOS ONE*. **12**, e0173226 (2017).
65. Grafton-Cardwell, E. E., Stelinski, L. L. & Stansly, P. A. Biology and Management of Asian Citrus Psyllid, Vector of the Huanglongbing Pathogens. *Ann. Rev. Entomol.* **58**, 413–432 (2013).
66. Nunes, P., Robinet, C., Branco, M. & Franco, J. C. Modelling the invasion dynamics of the African citrus psyllid: The role of human-mediated dispersal and urban and peri-urban citrus trees. *NeoBiota* **84**, 369–396 (2023).
67. Barve, N. et al. The crucial role of the accessible area in ecological niche modeling and species distribution modeling. *Ecol. Model.* **222**, 1810–1819 (2011).
68. Valavi, R., Elith, J., Lahoz-Monfort, J. J., Guillera-Arroita, G. & blockCV An r package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models. *Methods Ecol. Evol.* **10**, 225–232 (2019).
69. Fielding, A. H. & Bell, J. F. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environ. Conserv.* **24**, 38–49 (1997).
70. Liu, C., White, M. & Newell, G. Selecting thresholds for the prediction of species occurrence with presence-only data. *J. Biogeogr.* **40**, 778–789 (2013).
71. Jaccard, P. Nouvelles Recherches Sur la Distribution Florale. *Bull. de la. Societe Vaudoise des. Sci. Nat.* **44**, 223–270 (1908).
72. Li, W. & Guo, Q. How to assess the prediction accuracy of species presence-absence models without absence data? *Ecography* **36**, 788–799 (2013).
73. Gallien, L. et al. Contrasting the effects of environment, dispersal and biotic interactions to explain the distribution of invasive plants in alpine communities. *Biol. Invasions*. **17**, 1407–1423 (2015).
74. Liu, C., Bellard, C. & Jeschke, J. M. Understanding biological invasions through the lens of environmental niches. *Trends Ecol. Evol.* **40**, 385–394 (2025).
75. Liu, C., Wolter, C., Courchamp, F., Roura-Pascual, N. & Jeschke, J. M. Biological invasions reveal how niche change affects the transferability of species distribution models. *ECOLOGY* vol. 103 (2022).

76. Meynard, C. N., Leroy, B. & Kaplan, D. M. Testing methods in species distribution modelling using virtual species: what have we learnt and what are we missing? *Ecography* 42, 2021–2036 (2019).
77. Aidoo, O. F. et al. Climate-induced range shifts of invasive species (Diaphorina citri Kuwayama). *Pest Manag. Sci.* **78**, 2534–2549 (2022).
78. Souza, P. G. C. et al. Modelling the potential distribution of the Asian citrus psyllid *Diaphorina citri* (Hemiptera: Liviidae) using CLIMEX. *Int. J. Trop. Insect Sci.* **44**, 771–787 (2024).
79. Singleton, A. L. et al. Species distribution modeling for disease ecology: A multi-scale case study for schistosomiasis host snails in Brazil. *PLOS Global Public. Health*. **4**, e0002224 (2024).
80. Meynard, C. N. & Quinn, J. F. Predicting species distributions: a critical comparison of the most common statistical models using artificial species. *J. Biogeogr.* **34**, 1455–1469 (2007).
81. Araújo, M. B. & New, M. Ensemble forecasting of species distributions. *Trends Ecol. Evol.* **22**, 42–47 (2007).
82. Garcia, R. A., Burgess, N. D., Cabeza, M., Rahbek, C. & Araujo, M. B. Exploring consensus in 21st century projections of climatically suitable areas for African vertebrates. *Glob. Change Biol.* **18**, 1253–1269 (2012).
83. Kass, J. M. et al. Achieving higher standards in species distribution modeling by leveraging the diversity of available software. *Ecography* e07346 (2025). (2025).
84. Aidoo, O. F. et al. A machine learning algorithm-based approach (MaxEnt) for predicting invasive potential of *Trioza erytreae* on a global scale. *Ecol. Inf.* **71**, 101792 (2022).
85. Narouei-Khandan, H. A., Halbert, S. E., Worner, S. P. & van Bruggen, A. H. C. Global climate suitability of citrus huanglongbing and its vector, the Asian citrus psyllid, using two correlative species distribution modeling approaches, with emphasis on the USA. *Eur. J. Plant. Pathol.* **144**, 655–670 (2016).
86. Dietze, M. C. et al. Iterative near-term ecological forecasting: Needs, opportunities, and challenges. *Proceedings of the National Academy of Sciences* 115, 1424–1432 (2018).

Figures

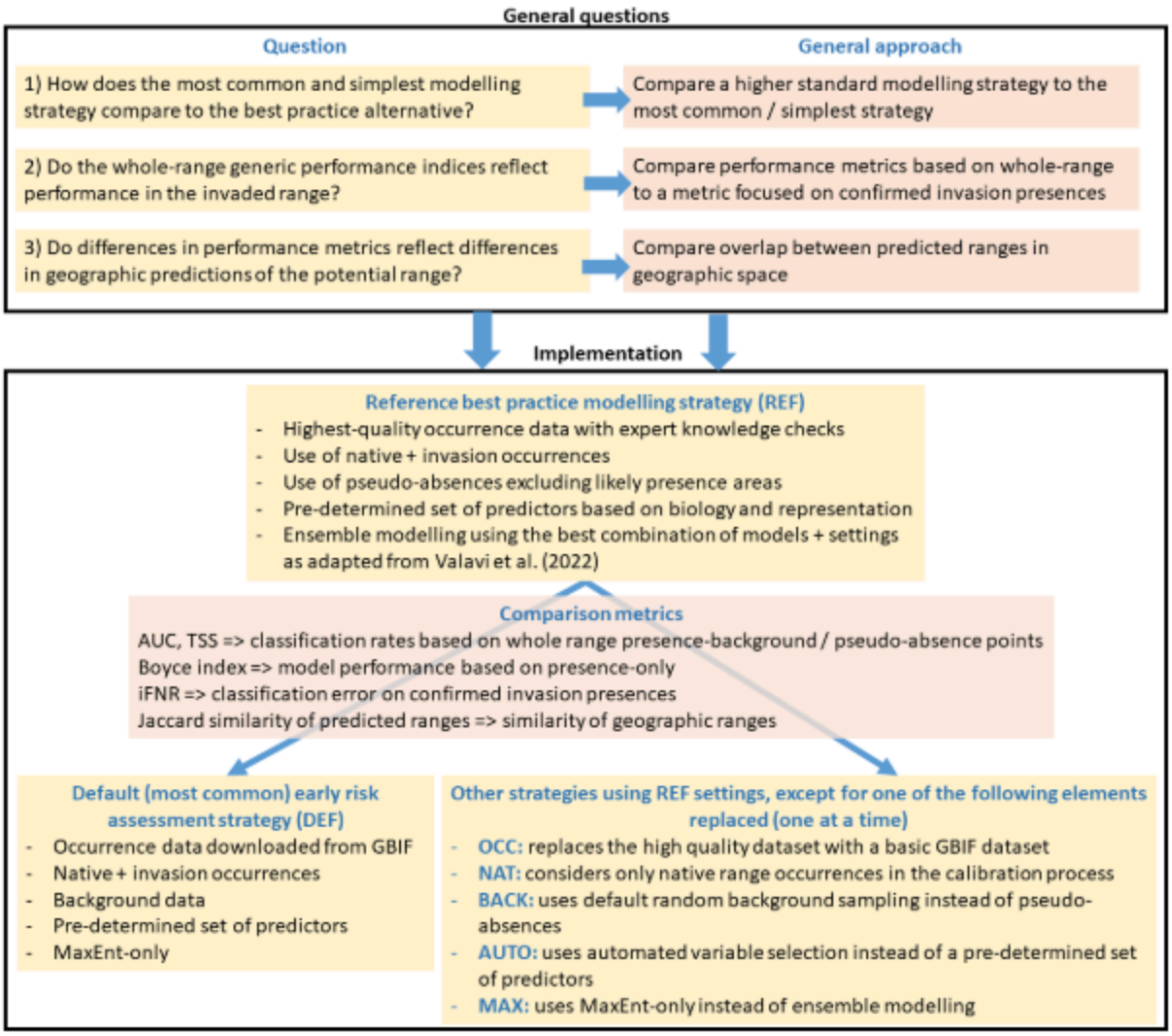


Figure 1

General questions addressed (upper box) and strategies implemented to address them (lower box). The questions were addressed through a comparison of a higher standard modelling strategy considered as a reference (REF) against several variants: a simple default strategy (DEF) using settings that are common in early risk assessments, and five variants of REF where an element was replaced one at a time (strategies in the lower-right box). To compare model performances, we used standard whole-range measures based on presence-absence (AUC, TSS) or presences-only (Boyce Index), as well as a measure focused on the invasion occurrences (iFNR) and an index of geographic range overlap (Jaccard index with respect to the REF model).

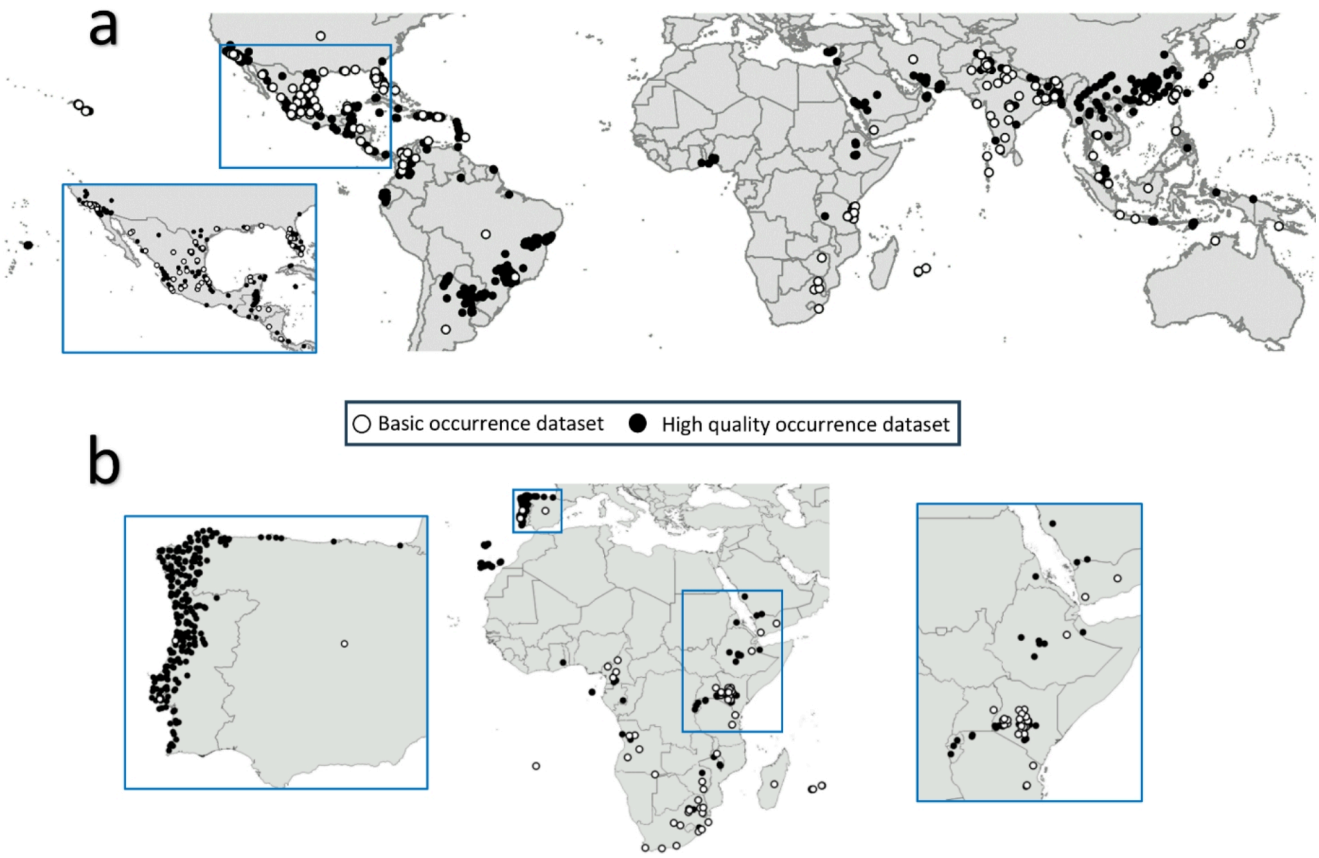


Figure 2

Worldwide distribution of occurrence data for the two species studied: (A) *Diaphorina citri* and (B) *Trioza erytreae*. White circles represent occurrences from a GBIF dataset cleaned with standard procedures, while black circles represent a high-quality occurrence dataset, which complemented GBIF occurrences with an extensive search and corrected through consultation with taxonomic and geographic expert knowledge. Insets were added in an invaded area for each species as examples showing the differences between the two datasets.

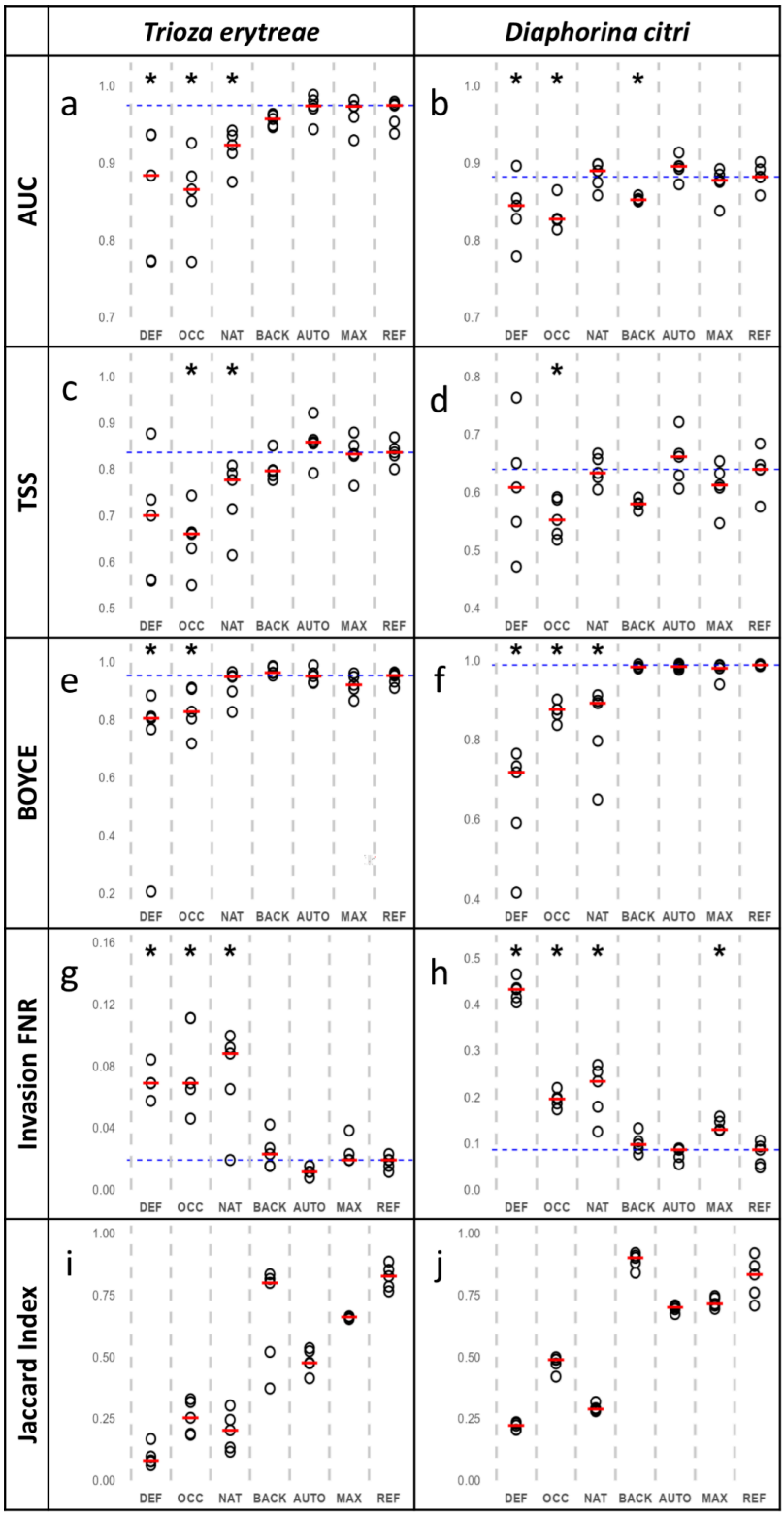


Figure 3

Performance and distribution overlap for each species and modelling strategy. Five metrics are represented (lines) for each species (columns). Each point represents the result of a single spatial block cross-validation; the red horizontal line represents the mean value across cross-validations. In the case of the Jaccard Index, the projected distribution for each cross-validation block is compared against the distribution predicted by REF using the full dataset. In all cases, an * indicates a significant difference (p

< 0.05) with the REF modelling strategy. Within each plot, the results are presented according to the modelling strategy. At the two extremes are DEF = default simplest strategy (2 in Figure 1) and REF = highest standard reference strategy (see also Figure 1); in between are the five variants of REF: OCC (basic occurrence dataset), NAT (native-only occurrences), BACK (use of random background), AUTO (automated variable selection), MAX (MaxEnt-only).

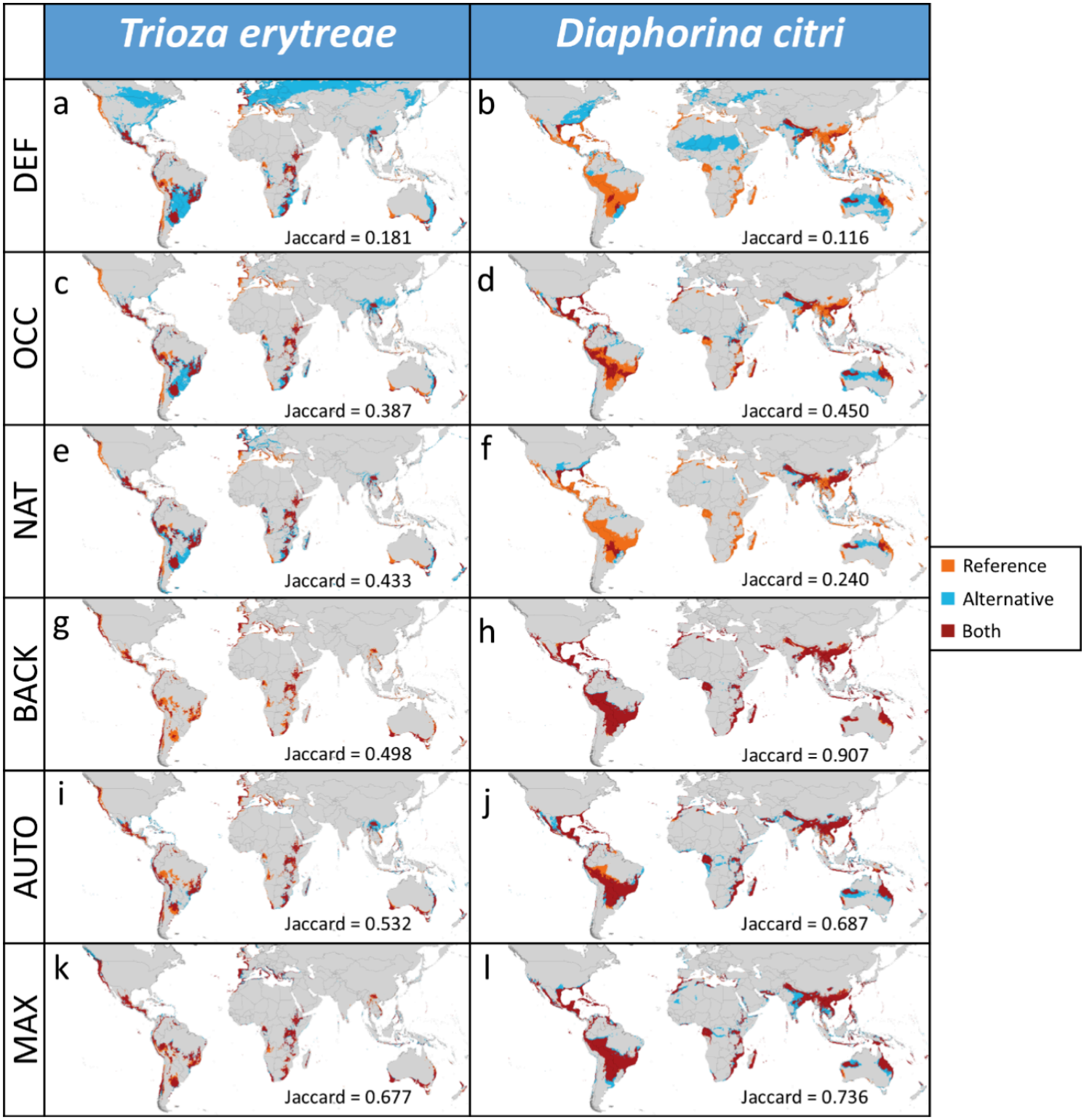


Figure 4

Comparison of predicted distributions against the REF modelling strategy. Each column corresponds to a species, while each line corresponds to a modelling strategy (see Figure 1 for abbreviations). Each map represents the area predicted as potential range by both the REF and the alternative model (dark red), the area that is predicted as potential range only by the reference model (orange), and the area that is predicted only by the alternative model (blue). The Jaccard index presented in each map represents the overlap between the given projection and that of the reference model, with an index closer to 1 indicating perfect overlap, and one closer to 0 indicating no overlap.

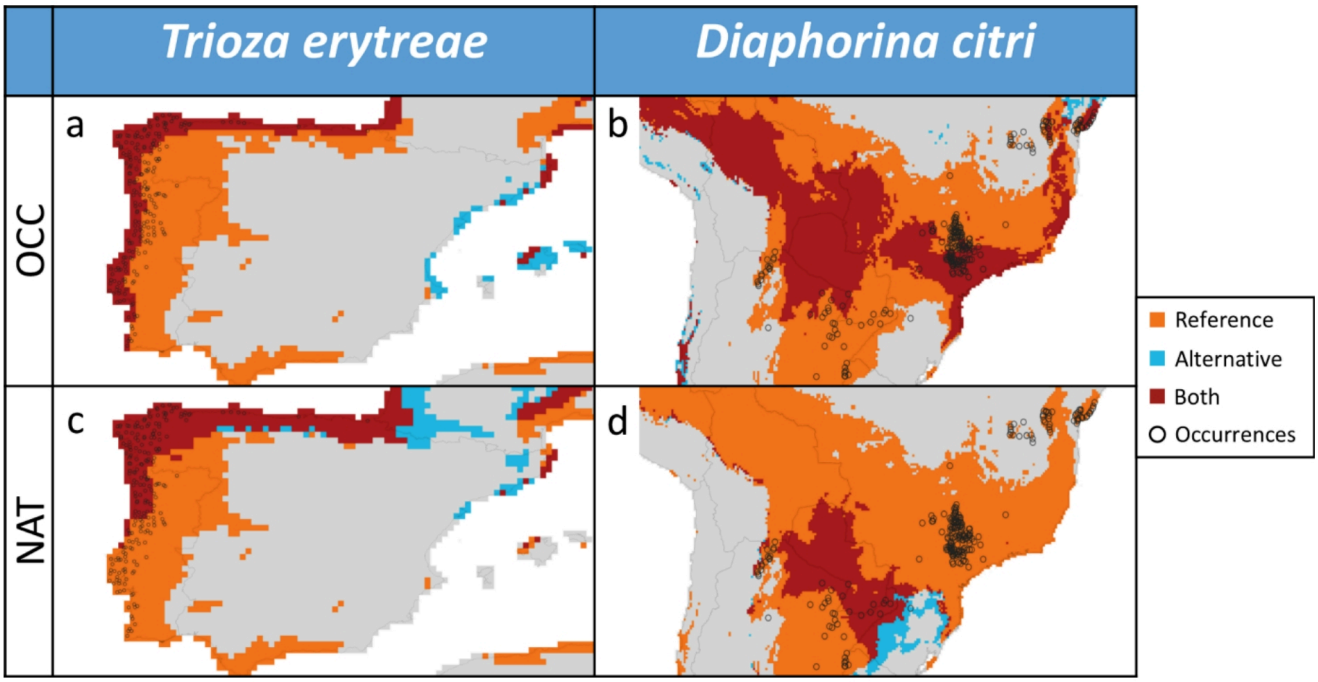


Figure 5

Zoom into one invaded area for each species, comparing only the two strategies that seem to produce the most differences with respect to the reference modelling strategy: OCC, which replaced the high-quality occurrence dataset used in REF with a basic occurrence dataset; and NAT, which uses only occurrences within the native range of each species in the calibration phase. Dark red represents areas where both REF and the alternative modelling strategy overlap; orange represents areas where only REF predicts potential range; blue is the area predicted only by the alternative model. In all cases, transparent points represent confirmed occurrences for each species.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplMaterials.pdf