



## ISSUES IN AGRICULTURAL AND FOREST ENTOMOLOGY

## Considerations regarding species distribution models for forest insects

Frank H. Koch 

USDA Forest Service, Southern Research Station, Eastern Forest Environmental Threat Assessment Center, 3041 East Cornwallis Road, Research Triangle Park, NC 27709, U.S.A.

- Abstract**
- 1 Species distribution models, or SDMs, have become important decision support tools by answering fundamental questions about where species, including invasive species, are likely to survive and thrive based on environmental conditions.
  - 2 For an inexperienced modeller or model reviewer, the terminology and technical aspects of SDMs can be overwhelming, and even well-trained modellers can struggle to understand the implications of various modelling choices.
  - 3 Here, I outline some key considerations with respect to SDMs, focusing on their application to forest insects. Foremost, I assert that a model should be developed and evaluated with attention to relationships between an insect and its hosts, as those relationships determine much about the places the insect may occupy.
  - 4 In my view, the most successful models are constructed carefully and incorporate honest assessments of their limitations, sources of error and uncertainty, and the degree of linkage between the model and the real-world circumstances it is meant to portray.

**Keywords** Forest insects, insect–host relationships, invasive species, species distribution models.

### Introduction

Insects, especially invasive alien insects, figure prominently in my research, which falls under the umbrella of forest health monitoring and analysis. During my career, I have worked on some of the most ecologically and economically significant forest insects introduced to North America. Most of that work has involved spatial modelling at a variety of scales: regional, continental, and sometimes global. Given my background, I will focus on forest insects in this commentary, although many of my points should apply similarly to other categories of insects. I envision the target audience as forest entomologists who have limited familiarity with spatial models, but who are asked to review modelling results or are consulted during the development of a model for an insect of interest.

The term ‘spatial model’ can refer to several types of models utilized for various purposes. One type that may come to mind is a model of insect dispersal or spread (*e.g.*, Rudd & Gandour, 1985). Another type of spatial modelling relevant to insects,

and the main subject here, is species distribution modelling. A wide variety of modelling techniques are used to explain or (as has become increasingly common) predict the geographic distributions of species (Elith & Leathwick, 2009; Guillera-Arroita *et al.*, 2015). The philosophy behind the techniques is consistent: the locations where a species may occur can be predicted with a model of where environmental conditions are suitable for its survival. The objective of the model may be interpolation, which is predicting the likelihood of distribution at unsampled locations within a species’ known range, or extrapolation, which is prediction in unsampled geographic areas outside its known range (Elith & Leathwick, 2009; Dormann *et al.*, 2012). With respect to insects and many other taxa, a large share of the literature on species distribution models (SDMs) has dealt with extrapolation to new areas, reflecting the models’ popularity for studies of invasive alien species. Indeed, SDMs are key tools in agricultural and forest biosecurity and in invasive species risk analysis more generally (Kriticos, 2012; Guillera-Arroita *et al.*, 2015). An essential task when assessing the ecological or economic risk presented by an invasive species, whether an insect or some other organism, is determining where it can establish successful populations.

Correspondence: Frank H. Koch. Tel.: +1 919 549 4006; fax: +1 919 549 4047; e-mail: frank.h.koch@usda.gov

SDMs are primarily climate-driven, meaning that the variables used to develop them typically portray climatic factors. This makes sense because climate is a chief driver of environmental suitability (Kriticos, 2012; Venette, 2017). Moreover, SDMs are at least regional and more likely continental or even global in scale, and climate data are among the most readily available geospatial data at larger spatial scales. As a logical extension of this, many SDM-based analyses in recent years have predicted species' geographic distributions during future time periods, based on climate projections developed under different general circulation models (GCMs) and emissions scenarios (e.g., Barredo *et al.*, 2015; Srivastava *et al.*, 2020). Nevertheless, SDMs that rely solely on climatic variables and ignore biotic factors such as habitat accessibility may not predict reliably in either current or future conditions (Araújo & Luoto, 2007; Sax *et al.*, 2007; Hill *et al.*, 2017; Liu *et al.*, 2020; Srivastava *et al.*, 2021). I will return to this point later, but first I will go over some central concepts and related technical aspects of SDMs.

### Species distribution modelling: competing paradigms?

There are two main analytical approaches for SDMs: mechanistic modelling and correlative modelling. Mechanistic or process-based models (e.g., CLIMEX) use *a priori* information about the biology and ecology of a species to define values for model parameters representing the processes that drive the species' response to environmental conditions (Yates *et al.*, 2018). This *a priori* information can be derived empirically (i.e., from controlled laboratory experiments or field observations) or, more frequently, through a combination of available data and input from species experts (Venette, 2017; Yates *et al.*, 2018). Notably, the parameters of mechanistic models have straightforward ecological interpretations. For instance, CLIMEX utilizes three types of parameters: parameters that characterize seasonal population growth of a species, parameters that define environmental stresses that may limit a species' distribution, and parameters that define constraints, such as the length of the growing season, that may exclude a species from some otherwise suitable locations (Kriticos *et al.*, 2015).

Correlative models relate occurrence or abundance data for a species at known locations to a set of environmental predictors. Rather than focusing on the mechanisms that define a species' response to environmental conditions, a correlative model attempts to determine the conditions that are suitable for the species, as represented by a set of abiotic or biotic covariates, from which it is possible to predict its distribution. The underlying processes and mechanisms are thus implicit (Dormann *et al.*, 2012; Yates *et al.*, 2018). Numerous statistical and machine learning methods have been applied for correlative modelling, including random forest models, boosted regression trees, and generalized linear and additive models (Yates *et al.*, 2018). For practical explanations of Maxent (short for 'maximum entropy'), a widely used machine learning method, it is worth consulting work by Elith *et al.* (2011) and Merow *et al.* (2013).

I cannot say much about mechanistic and correlative models that have not been said before by others, and in greater detail. For perspectives on their relative merits and drawbacks,

I would point to articles by Webber *et al.* (2011), Dormann *et al.* (2012) and Venette (2017); in particular, the latter review focuses on the application of the approaches to invasive forest insects. Ultimately, both mechanistic and correlative modelling approaches have demonstrated utility for predicting species distributions. A positive aspect of correlative models is that most of the underlying modelling methods are freely accessible *via* open-source software, including the R statistical environment (e.g., Phillips *et al.*, 2017). Furthermore, correlative models do not have the *a priori* information requirements of mechanistic models, which are perhaps best suited to well-studied species (Castaño-Quintero *et al.*, 2020). Nevertheless, I must offer some words of warning about correlative models: they can be easy to misuse or misinterpret. Many correlative methods, particularly Maxent, can be applied to presence-only data, or more accurately, what is known as presence-background data (Phillips *et al.*, 2009; Elith *et al.*, 2010; Guillera-Aroita *et al.*, 2015; Iturbide *et al.*, 2015). The 'background' represents the geographic space being modelled (i.e., the region containing the presence locations), which is sampled to generate presumed absence locations, commonly known as pseudo-absences. These pseudo-absence locations can then be compared with the presence locations in terms of their climatic and other environmental characteristics (Guillera-Aroita *et al.*, 2015). Although this may seem fairly routine, the background can be challenging to define correctly (Phillips & Dudík, 2008; Phillips *et al.*, 2009; VanDerWal *et al.*, 2009; Rodda *et al.*, 2011; Iturbide *et al.*, 2015; Jarnevič *et al.*, 2017). For example, there is substantial risk that the presence data come from a biased sample, either because detection methods are imperfect (common with insects) or because sampling effort is uneven. Nevertheless, there are various methods to account for such bias when sampling the background (Merow *et al.*, 2013; Venette, 2017; Yates *et al.*, 2018; Chapman *et al.*, 2019). Any presence-only correlative model that fails to address the background issue in some way should be viewed sceptically.

Another potential mistake with correlative models is violation of the assumption of independence among the covariates (Jarnevič *et al.*, 2015). If two predictors are highly correlated, they should not both be included in a model. The same is probably true for covariates exhibiting high degrees of spatial dependence (i.e., spatial autocorrelation), although there are techniques, such as spatial cross-validation, that account for the influence of spatial autocorrelation on model evaluation metrics (Jarnevič *et al.*, 2015). As with the background issue, if a correlative model is presented without evaluation of possible dependence between covariates, it should be treated with scepticism. Also problematic with correlative models is misinterpretation of the model predictions as true probabilities, when in fact they are only relative likelihoods of occurrence (Guillera-Aroita *et al.*, 2015). Although this may be a relatively minor concern in practice, it is still important for modellers to recognize that true probabilities are only calculable with actual absence data.

A concern with all SDMs, but especially those applied to invasive species, is transferability (sometimes called spatial transferability). Briefly, a modeller is attempting to 'transfer' a model based on a species' known range (i.e., the model's training domain) to a novel environmental space. This is challenging with respect to invasive species because they are rarely at equilibrium

with this new environment, meaning they usually have not invaded all locations that are suitable for them. Consequently, a model trained using data from a species' invaded range may underestimate its potential distribution (Elith *et al.*, 2010; Barbet-Massin *et al.*, 2018; Yates *et al.*, 2018; Liu *et al.*, 2020). Regardless, species–environment relationships seldom remain consistent across geographic space. This spatial inconsistency is known as nonstationarity. Mechanistic models are presumably more robust to transferability issues than correlative models, but they are still hindered by nonstationarity (Dormann *et al.*, 2012; Yates *et al.*, 2018).

Various other factors can affect transferability, including input data quality and model complexity. Complex SDMs with large numbers of predictor variables risk being over-fitted to the training data and generating unreliable predictions as a result (Liu *et al.*, 2020). Theoretically, this is more of a problem with correlative models than mechanistic models. In any case, it is easier to trust a model that is accompanied by some evaluation of its transferability than one that is not. Regrettably, there is a lack of standard methods and metrics for performing such assessments (Yates *et al.*, 2018; Liu *et al.*, 2020).

Setting aside issues of transferability, either a correlative or mechanistic approach can be appropriate in most situations if a model is developed thoughtfully. Although not targeted at insects, the inaugural editorial of *Biology of Invasive Plants*, a series in the journal *Invasive Plant Science and Management*, laid out detailed technical guidelines that should be instructive for all species distribution modellers as well as potential reviewers (Kriticos *et al.*, 2020). To get around issues with specific modelling techniques, some researchers have explored the utility of ensemble models or blending of correlative and mechanistic models for deriving robust predictions (Iturbide *et al.*, 2015; Yates *et al.*, 2018; Hao *et al.*, 2020). Assuredly, these will be areas of continued investigation, but there is some peril that an ensemble or blended model conceals errors arising from one (or more) of its components (Elith *et al.*, 2011; Kriticos *et al.*, 2020).

## The importance of insect–host relationships

What is the ultimate objective of an SDM for a forest insect or any other species? If the model is intended to show where a species is likely to maintain persistent populations, the assumed goal is for someone to use that information to formulate a response, *i.e.*, for making decisions about how to manage the species or its habitat. I mention habitat purposely to highlight an aspect that I believe is essential to consider when modelling forest insects: the relationships with their host species. For an insect to become established in any setting, it must have utilizable hosts. Although environmental conditions may predict a species' potential geographic distribution broadly, the climatic factors typically used to represent those conditions are seldom adequate surrogates for factors like host availability; as suggested earlier, solely climate-based approaches are poor at predicting the naturalized distributions of many invasive species, including insects (Sax *et al.*, 2007; Hill *et al.*, 2017). This limitation is addressed hierarchically in the formal pest risk analysis process, where there are areas at risk (*i.e.*, areas that are environmentally suitable) as well as endangered areas, which are sub-regions

within at-risk areas where there is substantial risk of economic loss (Baker *et al.*, 2015). In short, this means where hosts are present.

I am not alone in emphasizing the importance of insect–host relationships in distribution modelling, particularly for forest insects (*e.g.*, see Dang *et al.*, 2021). It is worth delving into some of the nuances of this assertion. For a polyphagous forest insect (*i.e.*, a true generalist), assuming there are adequate hosts where climatic conditions are projected to be suitable may be appropriate (Venette, 2017). But there are examples that reveal the assumption's shortcomings. What if the insect is a specialist? For instance, the emerald ash borer (*Agrilus planipennis* Fairmaire) is essentially limited to hosts in the ash (*Fraxinus*) genus (but see Cipollini & Peterson, 2018; Olson & Rieske, 2019). A less specialized insect is the Asian longhorned beetle (*Anaplophora glabripennis* (Motschulsky)), for which the preferred hosts include maples (*Acer*), willows (*Salix*), and elms (*Ulmus*), while birch (*Betula*) and sycamore (*Platanus*) species are also common hosts (Haack *et al.*, 2010). With both of these insects, knowing the preferred hosts provides highly useful information for modelling. Both insects are native to Asia and have been introduced into North America and Europe; however, neither has a known host that is common in the Southern Hemisphere. Effectively, this rules out establishment in many places that are otherwise environmentally suitable, although such determinations are uncertain because host range expansion is always possible, even if emergence of an unforeseen but globally abundant major host seems unlikely (Haack *et al.*, 2010; Peterson *et al.*, 2020). Furthermore, where these insects' known hosts occur in the Southern Hemisphere, they appear sporadically and mostly in urban or peri-urban environments. Consequently, there is probably insufficient functional connectivity between hosts to permit either insect to establish there beyond a local scale. Note also that their known hosts are exclusively hardwoods. Absent significant shifts in the host species' ranges due to climate change, neither insect is likely to become widespread in boreal forests dominated by conifers.

Conspicuously, *A. planipennis* and *A. glabripennis* are wood borers with some degree of specialization. Polyphagous defoliators such as the Asian gypsy moth (*Lymantria dispar asiatica* Vnukovskij) present a different dilemma for modellers: distinguishing suitable hosts can sometimes require in-depth analyses. For instance, parts of Australia and New Zealand are thought to be suitable climatically for *L. dispar asiatica* (Paini *et al.*, 2018). In laboratory trials involving 59 plant species from seven families, Matsuki *et al.* (2001) found that larval performance on several native Australian eucalypt (*Eucalyptus*) species was similar to that on known preferred hosts (*Quercus robur* and *Q. pubescens*, both European oaks). By contrast, larval performance on most tree species native to New Zealand was poor, suggesting the insect is unlikely to become established in New Zealand's indigenous forests. This was echoed by Pitt *et al.* (2007), who stated that host availability was likely to be a major limiting factor in New Zealand despite suitable climatic conditions, particularly in the country's North Island.

Another interesting case is the spotted lanternfly (*Lycorma delicatula* (White)), a phloem-feeding insect with a long history as a pest in China that recently has invaded South Korea, Japan, and the eastern USA (Wakie *et al.*, 2020). Although highly

polyphagous – *L. delicatula* has been reported to feed on more than 100 plant taxa across at least 33 families – it has a strong host preference for tree of heaven (*Ailanthus altissima*), which itself is an aggressive invader found on every continent except Antarctica (Barringer & Cifré, 2020). Although *L. delicatula* may not require tree of heaven to complete its life cycle, its affinity for the species means that the insect cannot be easily categorized as a generalist (Derstine *et al.*, 2020). A global SDM for *L. delicatula* suggested that portions of Australia, Africa, and South America exhibit moderate to high suitability for the insect (Wakie *et al.*, 2020), but currently tree of heaven is less prevalent in these regions than in North America and Europe, where it is pervasive in many urban and peri-urban settings (Sladonja *et al.*, 2015). Because tree of heaven is recognized globally as a high-risk invader, there is some hope that management activities to limit its expansion will also curtail establishment of *L. delicatula* outside of its current invaded range. On the other hand, tree of heaven is naturalized in at least 51 countries and problematically invasive in at least 23 countries (Walker *et al.*, 2017), so it is probably unrealistic to expect its distribution to constrain *L. delicatula* in the long term.

The preceding examples are meant to illustrate how insect–host relationships can be highly meaningful when characterizing the potential distributions of forest insects. These relationships may be especially relevant when attempting to project distributions at a global scale and under future climates (Araújo & Luoto, 2007). Admittedly, incorporating host data does not always improve predictions over models built from climatic factors alone (*e.g.*, Silva *et al.*, 2014), and sometimes species' host preferences are simply too poorly known to model effectively. Nonetheless, I would question any distribution model for a forest insect species that did not feature some appraisal of its host relationships, including potential relationships.

I acknowledge that I am saying this from a privileged position, in a part of the world where detailed forest inventories and diverse ancillary data enable generation of high-resolution tree distribution maps at the species or genus level. Unfortunately, disparity persists between data-rich regions (*e.g.*, North America, Europe, and Australia) and comparatively data-sparse regions (*e.g.*, central Africa and southeast Asia) with respect to tree species distributions, despite the expanded reach of occurrence data aggregators such as the Global Biodiversity Information Facility (Serra-Diaz *et al.*, 2017). Still, there are simple options that should work almost anywhere. Although a remote-sensing-derived map of forest land cover developed at a continental scale may be insufficient to depict the distributions of a specific set of hosts, it should be feasible to filter the forested cells based on general biogeographic characteristics gleaned from scientific or even popular literature.

### Thinking critically about modelling results

As I suggested earlier, the overarching purpose of an SDM is to inform decision-making with respect to a species of interest. Naturally, other modellers may be keenly interested in the methodological choices and underlying assumptions of an SDM, whether it is developed for a forest insect or any other organism. These aspects can be highly informative when developing a new model (or revising an existing one). Yet, a decision-maker

who is using the model result is unlikely to be focused solely on the species of interest. If the target species is a forest insect or some other herbivorous insect, the decision-maker is probably thinking about a critical resource (*i.e.*, the hosts) and the degree of pressure the insect might place on it. This is on top of the many other factors that must be accounted for when managing a resource. Guisan *et al.* (2013) discussed this issue eloquently, arguing that SDMs should be developed as part of a structured decision-making process, where there are open lines of communication between modellers and decision-makers.

An SDM is essentially a hypothesis, albeit one that is difficult to test because it cannot be validated with independent occurrence data (*i.e.*, external validation) for places where no such data exist (Dormann *et al.*, 2012; Venette, 2017). Of course, reliable predictions for these places may be the outcome most desired by a decision-maker. But if a model cannot be truly validated, why should a decision-maker embrace it? Venette (2017) broached the idea of 'model fatigue' among journal reviewers of forest insect SDMs. Commonly, editors identify reviewers who are either experts in the target species or the modelling approach. The species experts may not be able to perceive the value in another model if prior models exist for the target species, while modelling experts may not see novelty in the work if it uses established methods. I would extend the notion of model fatigue to decision-makers, who I perceive as mostly akin to species experts. In some instances, decision-makers or species expert reviewers may be able to critique technical details, but this is probably not the norm.

Ideally, a decision-maker or species expert reviewer could rely on a modeller's proficiency and safely assume that appropriate methods were employed in developing an SDM. However, I imagine that few people would be comfortable with this assumption. As a practical matter, most modellers will provide justification of their choices and present some diagnostic measures alongside the model results. Alas, a statement like 'The area under the curve (AUC) value was 0.91' may not be especially meaningful to a nonmodeller, and at any rate, this is merely one diagnostic measure out of many. Ultimately, the burden is on the modeller to explain and interpret a comprehensive set of diagnostics. Furthermore, if another SDM already exists for the target species, the modeller must articulate the comparative advantages of their model.

I appreciate that, even when given such information, a decision-maker or species expert reviewer may feel unequipped to evaluate an SDM. I maintain that technical proficiency is unnecessary to gauge whether a model included earnest efforts to address the critical modelling aspects noted earlier in this commentary. In fact, I will mention another modelling mistake that should be readily observable: when a modelling result fails to characterize the uncertainty in the predictions. Accounting for uncertainty in SDMs is a topic I cannot cover adequately here, although Beale and Lennon (2012) provided a helpful overview. But more simply, if a model description omits any discussion of uncertainty, then this likely indicates that the modeller is overconfident in the outputs (Dormann *et al.*, 2012) and is definitely a cause for doubt.

Irrespective of our training, one qualification that we share is an ability to think critically about information put in front of us. It is reasonable, maybe even advisable, to evaluate a model

result at a basic level. Setting aside all other considerations, does the predicted distribution for a forest insect seem plausible? Often, we can be confident about a large portion of a predicted distribution because environmental conditions are optimal and suitable hosts are prevalent. Conversely, issues with model plausibility tend to emerge where conditions are projected to be marginal for a species. This is another context where insect–host relationships can be pertinent. For instance, we tend to think of insects as being primarily limited by temperature, and temperature-related variables are among the strongest predictors of potential distribution (Messenger, 1959; Peacock *et al.*, 2006). But moisture is still a requirement, and further, is a critical requirement for hosts. If I was asked to evaluate an SDM that predicted, for example, the survival of a forest insect in an arid or semi-arid ecosystem, I would investigate whether the ecosystem receives enough rainfall on average to sustain populations of its host trees. Indeed, I might do this even if host availability was incorporated as a model constraint unless I was satisfied with the modeller's description of this constraint.

Focusing on the marginal areas of a predicted distribution might also reveal plausibility issues in an SDM applied to future climatic conditions. To illustrate, an SDM might predict that a forest insect currently restricted to the temperate biome will encounter suitable environmental conditions in the boreal zone 50 years into the future. Additionally, a complementary model of the future distributions of its hosts might also predict that they could survive in some parts of the boreal zone 50 years from now. Nonetheless, latitudinal range shifts of tree species are slow, perhaps on the order of 20–40 km per century (Davis & Shaw, 2001). Will the insect's hosts migrate far enough in 50 years to allow the insect to exploit many new areas within the boreal zone?

Although I emphasized the importance of hosts in these latter examples and several previous ones, I want to be clear that not every forest insect SDM must address host relationships explicitly. As alluded to earlier, environmental suitability can be relevant by itself in some cases (*e.g.*, with a generalist insect or an insect that has demonstrated frequent host-switching behaviour). I also want to avoid presenting a straw man argument. An exclusively climate-driven model can only reveal so much about the ecological niche, fundamental or realized, of a species (Jiménez-Valverde *et al.*, 2011), and expecting it to accomplish more is unfair. Still, there are potential benefits to incorporating hosts when possible. First, doing so provides something salient for nonmodellers to evaluate: an insect's host relationships and whether they are represented appropriately. More significantly, it may increase the chance of buy-in from decision-makers, who have a fuller set of information on which to base the decisions that probably motivated model development in the first place.

## Acknowledgements

This work was supported by funding from the U.S. Department of Agriculture, Forest Service.

## Data availability statement

Data sharing is not applicable as no new data were created or analyzed for this article.

## References

- Araújo, M.B. & Luoto, M. (2007) The importance of biotic interactions for modelling species distributions under climate change. *Global Ecology and Biogeography*, **16**, 743–753. <https://doi.org/10.1111/j.1466-8238.2007.00359.x>.
- Baker, R., Eyre, D., Brunel, S., Dupin, M., Reynaud, P. & Jarosík, V. (2015) Mapping endangered areas for pest risk analysis. *Pest Risk Modelling and Mapping for Invasive Alien Species* (ed. by R. C. Venette), pp. 18–34. CABI, U.K. <https://doi.org/10.1079/9781780643946.0018>.
- Barbet-Massin, M., Rome, Q., Villemant, C. & Courchamp, F. (2018) Can species distribution models really predict the expansion of invasive species? *PLoS ONE*, **13**, 1–14. <https://doi.org/10.1371/journal.pone.0193085>.
- Barredo, J.I., Strona, G., de Rigo, D., Caudullo, G., Stancanelli, G. & San-Miguel-Ayanz, J. (2015) Assessing the potential distribution of insect pests: case studies on large pine weevil (*Hylobius abietis* L.) and horse-chestnut leaf miner (*Cameraria ohridella*) under present and future climate conditions in European forests. *EPPO Bulletin*, **45**, 273–281. <https://doi.org/10.1111/epp.12208>.
- Barringer, L. & Ciafré, C.M. (2020) Worldwide feeding host plants of spotted lanternfly, with significant additions from North America. *Environmental Entomology*, **49**, 999–1011. <https://doi.org/10.1093/ee/nvaa093>.
- Beale, C.M. & Lennon, J.J. (2012) Incorporating uncertainty in predictive species distribution modelling. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **367**, 247–258. <https://doi.org/10.1098/rstb.2011.0178>.
- Castaño-Quintero, S., Escobar-Luján, J., Osorio-Olvera, L., Townsend Peterson, A., Chiappa-Carrara, X., Martínez-Meyer, E. & Yañez-Arenas, C. (2020) Supraspecific units in correlative niche modeling improves the prediction of geographic potential of biological invasions. *PeerJ*, **8**, 1–16. <https://doi.org/10.7717/peerj.10454>.
- Chapman, D., Pescott, O.L., Roy, H.E. & Tanner, R. (2019) Improving species distribution models for invasive non-native species with biologically informed pseudo-absence selection. *Journal of Biogeography*, **46**, 1029–1040. <https://doi.org/10.1111/jbi.13555>.
- Cipollini, D. & Peterson, D.L. (2018) The potential for host switching via ecological fitting in the emerald ash borer-host plant system. *Oecologia*, **187**, 507–519. <https://doi.org/10.1007/s00442-018-4089-3>.
- Dang, Y.Q., Zhang, Y.L., Wang, X.Y., Xin, B., Quinn, N.F. & Duan, J.J. (2021) Retrospective analysis of factors affecting the distribution of an invasive wood-boring insect using native range data: the importance of host plants. *Journal of Pest Science*, **94**, 981–990. <https://doi.org/10.1007/s10340-020-01308-5>.
- Davis, M.B. & Shaw, R.G. (2001) Range shifts and adaptive responses to quaternary climate change. *Science*, **292**, 673–679. <https://doi.org/10.1126/science.292.5517.673>.
- Derstine, N.T., Meier, L., Canlas, I. *et al.* (2020) Plant volatiles help mediate host plant selection and attraction of the spotted lanternfly (Hemiptera: Fulgoridae): a generalist with a preferred host. *Environmental Entomology*, **49**, 1049–1062. <https://doi.org/10.1093/ee/nvaa080>.
- Dormann, C.F., Schymanski, S.J., Cabral, J. *et al.* (2012) Correlation and process in species distribution models: bridging a dichotomy. *Journal of Biogeography*, **39**, 2119–2131. <https://doi.org/10.1111/j.1365-2699.2011.02659.x>.
- Elith, J. & Leathwick, J.R. (2009) Species distribution models: ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics*, **40**, 677–697. <https://doi.org/10.1146/annurev.ecolsys.110308.120159>.

- 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>.
- Elith, J., Phillips, S.J., Hastie, T., Dudík, M., Chee, Y.E. & Yates, C.J. (2011) A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions*, **17**, 43–57. <https://doi.org/10.1111/j.1472-4642.2010.00725.x>.
- Guillera-Arroita, G., Lahoz-Monfort, J.J., Elith, J. *et al.* (2015) Is my species distribution model fit for purpose? Matching data and models to applications. *Global Ecology and Biogeography*, **24**, 276–292. <https://doi.org/10.1111/geb.12268>.
- Guisan, A., Tingley, R., Baumgartner, J.B. *et al.* (2013) Predicting species distributions for conservation decisions. *Ecology Letters*, **16**, 1424–1435. <https://doi.org/10.1111/ele.12189>.
- Haack, R.A., Hérard, F., Sun, J. & Turgeon, J.J. (2010) Managing invasive populations of Asian longhorned beetle and citrus longhorned beetle: a worldwide perspective. *Annual Review of Entomology*, **55**, 521–546. <https://doi.org/10.1146/annurev-ento-112408-085427>.
- Hao, T., Elith, J., Lahoz-Monfort, J.J. & Guillera-Arroita, G. (2020) Testing whether ensemble modelling is advantageous for maximising predictive performance of species distribution models. *Ecography*, **43**, 549–558. <https://doi.org/10.1111/ecog.04890>.
- Hill, M.P., Gallardo, B. & Terblanche, J.S. (2017) A global assessment of climatic niche shifts and human influence in insect invasions. *Global Ecology and Biogeography*, **26**, 679–689. <https://doi.org/10.1111/geb.12578>.
- Iturbide, M., Bedia, J., Herrera, S., del Hierro, O., Pinto, M. & Gutiérrez, J.M. (2015) A framework for species distribution modelling with improved pseudo-absence generation. *Ecological Modelling*, **312**, 166–174. <https://doi.org/10.1016/j.ecolmodel.2015.05.018>.
- Jarnevich, C.S., Stohlgren, T.J., Kumar, S., Morissette, J.T. & Holcombe, T.R. (2015) Caveats for correlative species distribution modeling. *Ecological Informatics*, **29**, 6–15. <https://doi.org/10.1016/j.ecoinf.2015.06.007>.
- Jarnevich, C.S., Talbert, M., Morissette, J. *et al.* (2017) Minimizing effects of methodological decisions on interpretation and prediction in species distribution studies: an example with background selection. *Ecological Modelling*, **363**, 48–56. <https://doi.org/10.1016/j.ecolmodel.2017.08.017>.
- Jiménez-Valverde, A., Peterson, A.T., Soberón, J., Overton, J.M., Aragón, P. & Lobo, J.M. (2011) Use of niche models in invasive species risk assessments. *Biological Invasions*, **13**, 2785–2797. <https://doi.org/10.1007/s10530-011-9963-4>.
- Kriticos, D.J. (2012) Regional climate-matching to estimate current and future sources of biosecurity threats. *Biological Invasions*, **14**, 1533–1544. <https://doi.org/10.1007/s10530-011-0033-8>.
- Kriticos, D.M., Maywald, G.F., Yonow, T., Zurcher, E.J., Herrmann, N.I. & Sutherst, R. (2015) *CLIMEX Version 4: Exploring the Effects of Climate on Plants, Animals, and Diseases*. CSIRO, Australia.
- Kriticos, D.J., Clements, D.R. & DiTommaso, A. (2020) Biology of Invasive Plants: a new series within *Invasive Plant Science and Management*. *Invasive Plant Science and Management*, **13**, 115–119. <https://doi.org/10.1017/imp.2020.25>.
- Liu, C., Wolter, C., Xian, W. & Jeschke, J.M. (2020) Species distribution models have limited spatial transferability for invasive species. *Ecology Letters*, **23**, 1682–1692. <https://doi.org/10.1111/ele.13577>.
- Matsuki, M., Kay, M., Serin, J., Floyd, R. & Scott, J.K. (2001) Potential risk of accidental introduction of Asian gypsy moth (*Lymantria dispar*) to Australasia: effects of climatic conditions and suitability of native plants. *Agricultural and Forest Entomology*, **3**, 305–320. <https://doi.org/10.1046/j.1461-9555.2001.00119.x>.
- Merow, C., Smith, M.J. & Silander, J.A. (2013) A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. *Ecography*, **36**, 1058–1069. <https://doi.org/10.1111/j.1600-0587.2013.07872.x>.
- Messenger, P.S. (1959) Bioclimatic studies with insects. *Annual Review of Entomology*, **4**, 183–206. <https://doi.org/10.1146/annurev.en.04.010159.001151>.
- Olson, D.G. & Riesecke, L.K. (2019) Host range expansion may provide enemy free space for the highly invasive emerald ash borer. *Biological Invasions*, **21**, 625–635. <https://doi.org/10.1007/s10530-018-1853-6>.
- Paini, D.R., Mwebaze, P., Kuhnert, P.M. & Kriticos, D.J. (2018) Global establishment threat from a major forest pest via international shipping: *Lymantria dispar*. *Scientific Reports*, **8**, 1–7. <https://doi.org/10.1038/s41598-018-31871-y>.
- Peacock, L., Worner, S. & Sedcole, R. (2006) Climate variables and their role in site discrimination of invasive insect species distributions. *Environmental Entomology*, **35**, 958–963. <https://doi.org/10.1603/0046-225X-35.4.958>.
- Peterson, D.L., Slager, B., Anulewicz, A.C. & Cipollini, D. (2020) Feeding, survival, and fecundity of adult emerald ash borer (Coleoptera: Buprestidae) on foliage of two novel hosts and implications for host range expansion. *Environmental Entomology*, **49**, 709–716. <https://doi.org/10.1093/ee/nvaa046>.
- Phillips, S.J. & Dudík, M. (2008) Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography*, **31**, 161–175. <https://doi.org/10.1111/j.0906-7590.2008.5203.x>.
- Phillips, S.J., Dudík, M., Elith, J., Graham, C.H., Lehmann, A., Leathwick, J. & Ferrier, S. (2009) Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecological Applications*, **19**, 181–197. <https://doi.org/10.1890/07-2153.1>.
- Phillips, S.J., Anderson, R.P., Dudík, M., Schapire, R.E. & Blair, M.E. (2017) Opening the black box: an open-source release of Maxent. *Ecography*, **40**, 887–893. <https://doi.org/10.1111/ecog.03049>.
- Pitt, J.P.W., Régnière, J. & Worner, S. (2007) Risk assessment of the gypsy moth, *Lymantria dispar* (L), in New Zealand based on phenology modelling. *International Journal of Biometeorology*, **51**, 295–305. <https://doi.org/10.1007/s00484-006-0066-3>.
- Rodda, G.H., Jarnevich, C.S. & Reed, R.N. (2011) Challenges in identifying sites climatically matched to the native ranges of animal invaders. *PLoS ONE*, **6**, e14670. <https://doi.org/10.1371/journal.pone.0014670>.
- Rudd, W.G. & Gandour, R.W. (1985) Diffusion model for insect dispersal. *Journal of Economic Entomology*, **78**, 295–301. <https://doi.org/10.1093/jee/78.2.295>.
- Sax, D.F., Stachowicz, J.J., Brown, J.H. *et al.* (2007) Ecological and evolutionary insights from species invasions. *Trends in Ecology and Evolution*, **22**, 465–471. <https://doi.org/10.1016/j.tree.2007.06.009>.
- Serra-Diaz, J.M., Enquist, B.J., Maitner, B., Merow, C. & Svenning, J.C. (2017) Big data of tree species distributions: how big and how good? *Forest Ecosystems*, **4**, 30. <https://doi.org/10.1186/s40663-017-0120-0>.
- Silva, D.P., Gonzalez, V.H., Melo, G.A.R., Lucia, M., Alvarez, L.J. & De Marco, P. (2014) Seeking the flowers for the bees: integrating biotic interactions into niche models to assess the distribution of the exotic bee species *Lithurgus huberi* in South America. *Ecological Modelling*, **273**, 200–209. <https://doi.org/10.1016/j.ecolmodel.2013.11.016>.
- Sladonja, B., Sušek, M. & Guillermic, J. (2015) Review on invasive tree of heaven (*Ailanthus altissima* (Mill.) Swingle) conflicting values: assessment of its ecosystem services and potential biological threat. *Environmental Management*, **56**, 1009–1034. <https://doi.org/10.1007/s00267-015-0546-5>.
- Srivastava, V., Griess, V.C. & Keena, M.A. (2020) Assessing the potential distribution of Asian gypsy moth in Canada: a comparison of two methodological approaches. *Scientific Reports*, **10**, 1–10. <https://doi.org/10.1038/s41598-019-57020-7>.
- Srivastava, V., Roe, A.D., Keena, M.A., Hamelin, R.C. & Griess, V.C. (2021) Oh the places they'll go: improving species distribution modelling for invasive forest pests in an uncertain world.

- Biological Invasions*, **23**, 297–349. <https://doi.org/10.1007/s10530-020-02372-9>.
- VanDerWal, J., Shoo, L.P., Graham, C. & Williams, S.E. (2009) Selecting pseudo-absence data for presence-only distribution modeling: how far should you stray from what you know? *Ecological Modelling*, **220**, 589–594. <https://doi.org/10.1016/j.ecolmodel.2008.11.010>.
- Venette, R.C. (2017) Climate analyses to assess risks from invasive forest insects: simple matching to advanced models. *Current Forestry Reports*, **3**, 255–268. <https://doi.org/10.1007/s40725-017-0061-4>.
- Wakie, T.T., Neven, L.G., Yee, W.L. & Lu, Z. (2020) The establishment risk of *Lycorma delicatula* (Hemiptera: Fulgoridae) in the United States and globally. *Journal of Economic Entomology*, **113**, 306–314. <https://doi.org/10.1093/jee/toz259>.
- Walker, G.A., Gaertner, M., Robertson, M.P. & Richardson, D.M. (2017) The prognosis for *Ailanthus altissima* (Simaroubaceae; tree of heaven) as an invasive species in South Africa; insights from its performance elsewhere in the world. *South African Journal of Botany*, **112**, 283–289. <https://doi.org/10.1016/j.sajb.2017.06.007>.
- Webber, B.L., Yates, C.J., Le Maitre, D.C. *et al.* (2011) Modelling horses for novel climate courses: insights from projecting potential distributions of native and alien Australian acacias with correlative and mechanistic models. *Diversity and Distributions*, **17**, 978–1000. <https://doi.org/10.1111/j.1472-4642.2011.00811.x>.
- Yates, K.L., Bouchet, P.J., Caley, M.J. *et al.* (2018) Outstanding challenges in the transferability of ecological models. *Trends in Ecology and Evolution*, **33**, 790–802. <https://doi.org/10.1016/j.tree.2018.08.001>.

Accepted 1 June 2021