

Modelling for risk and biosecurity related to forest health

Emerging Topics in Life Sciences

Robinet, Christelle; Dool, Robbert; Collot, Dorian; Douma, Jacob C.

<https://doi.org/10.1042/ETLS20200062>

This publication is made publicly available in the institutional repository of Wageningen University and Research, under the terms of article 25fa of the Dutch Copyright Act, also known as the Amendment Taverne. This has been done with explicit consent by the author.

Article 25fa states that the author of a short scientific work funded either wholly or partially by Dutch public funds is entitled to make that work publicly available for no consideration following a reasonable period of time after the work was first published, provided that clear reference is made to the source of the first publication of the work.

This publication is distributed under The Association of Universities in the Netherlands (VSNU) 'Article 25fa implementation' project. In this project research outputs of researchers employed by Dutch Universities that comply with the legal requirements of Article 25fa of the Dutch Copyright Act are distributed online and free of cost or other barriers in institutional repositories. Research outputs are distributed six months after their first online publication in the original published version and with proper attribution to the source of the original publication.

You are permitted to download and use the publication for personal purposes. All rights remain with the author(s) and / or copyright owner(s) of this work. Any use of the publication or parts of it other than authorised under article 25fa of the Dutch Copyright act is prohibited. Wageningen University & Research and the author(s) of this publication shall not be held responsible or liable for any damages resulting from your (re)use of this publication.

For questions regarding the public availability of this publication please contact openscience.library@wur.nl

Review Article

Modelling for risk and biosecurity related to forest health

 **Christelle Robinet¹, Robbert van den Dool², Dorian Collot¹ and Jacob C. Douma²**

¹INRAE, URZF, F-45075 Orléans, France; ²Centre for Crop Systems Analysis, Wageningen University, Droevendaalsesteeg 1, 6708PB Wageningen, The Netherlands

Correspondence: Christelle Robinet (christelle.robinet@inrae.fr)

Modelling the invasion and emergence of forest pests and pathogens (PnPs) is necessary to quantify the risk levels for forest health and provide key information for policy makers. Here, we make a short review of the models used to quantify the invasion risk of exotic species and the emergence risk of native species. Regarding the invasion process, models tackle each invasion phase, e.g. pathway models to describe the risk of entry, species distribution models to describe potential establishment, and dispersal models to describe (human-assisted) spread. Concerning the emergence process, models tackle each process: spread or outbreak. Only a few spread models describe jointly dispersal, growth, and establishment capabilities of native species while some mechanistic models describe the population temporal dynamics and inference models describe the probability of outbreak. We also discuss the ways to quantify uncertainty and the role of machine learning. Overall, promising directions are to increase the models' genericity by parameterization based on meta-analysis techniques to combine the effect of species traits and various environmental drivers. Further perspectives consist in considering the models' interconnection, including the assessment of the economic impact and risk mitigation options, as well as the possibility of having multi-risks and the reduction in uncertainty by collecting larger fit-for-purpose datasets.

Introduction

Forests provide ecological, economic, social, and aesthetic services. In addition, they largely contribute to carbon sequestration and they could thus represent an important means of climate change mitigation [1]. However, forest health is threatened by various pests and pathogens (PnPs). Global change, including climate change but also other changes at global scale such as the intensification of international trade, has triggered the emergence of an increasing number of tree PnPs. The number of biological invasions of exotic species in new areas has been dramatically increasing [2,3], and climate change has been promoting the range expansion or population outbreaks of many native species (so-called emergent species; [4–7]). Both invasive exotic species and emerging native species thus represent important risks for forest health.

To quantify the current and projected risk levels, and to provide key information for decision-support experts and policy makers, modelling the invasion and emergence dynamics of these forest PnPs is necessary. Pest risk mapping is particularly useful to support strategic and tactical decisions [8]. Improving methods contributing to risk mapping has thus become an important challenge during the last few years [9]. Quantitative pest risk assessment allows for a higher transparency, the assessment of uncertainty, and the exploration of various risk reduction options [10]. To assess the invasion or emergence risk of a species, one may develop and parameterize a model for that species (so-called 'species-specific models'). Such an approach requires a lot of information on the not-yet-introduced/emerged species. On the contrary, 'generic models' utilize what is known about the invasion or emergence process of other species by describing species by their traits and feeding those traits into the model. This is an important advantage because the estimation of various parameters has been proven

Received: 10 June 2020
 Revised: 12 August 2020
 Accepted: 13 August 2020

Version of Record published:
 4 September 2020

difficult for not-yet-introduced species, and for those that previously did not cause any damage. Generic models can thus be applied more easily and more rapidly to a newly arriving or emerging species, but their predictive value should be proven across a range of forest PnPs.

In this short review, we make a synthesis of models developed to assess the invasion risk of exotic forest PnPs, and the emergence risk of native forest PnPs. We also discuss the ways to quantify uncertainty and the role of machine learning. For each part, we also identify promising directions.

Modelling the invasion risk of exotic forest pests and pathogens

The invasion process can be described as a sequence of four phases: entry (transport from the country of origin and arrival in the new area), establishment, spread, and impact [11]. The factors determining the outcome of each phase can be grouped into three broad categories related to: (i) the vectors carrying the PnPs (understood widely to include animal vectors, abiotic factors such as wind and weather, and human actors), (ii) the environment the pest or pathogen interacts with during the first three phases, and (iii) the characteristics of the pest or pathogen [12]. Modelling entails capturing only the relevant processes and factors to produce tractable and valuable insights for risk assessors and managers. To assess the risk posed to biosecurity, models are needed that tackle each invasion phase [9] (Figure 1).

Modelling species entry

Different types of pathway models including epidemic networks and gravity models have been developed so far [13]. They track the PnP from their source area to their destination area where transfer to a suitable host may take place [13–15]. Such models help to identify locations where high propagule pressure (i.e. a set of PnP individuals) is expected (e.g. ports, trucks, ships, or factories) and to explore the effects of phytosanitary measures to prevent entry. The drawback of pathway models is the parameterization. They request many parameters and data for both the species and their vectors on for instance interception and transportation. This information is often scarce, inconsistent, and variable in time. Sampling methodology, frequency, and reporting of intercepted PnPs may actually differ per commodity, per country, and per year. In risk assessment, this issue is often solved by simplifying the pathway models [10], and/or using expert knowledge elicitation (EKE) to derive sensible parameter values [16].

Modelling species establishment

The risk of PnP establishment is often modelled using species distribution models (SDMs). These models are generally based on a correlation between presence points (and if available absence points) of the species and climate in the native or already invaded areas [17]. Models are generally purely correlative (e.g. MaxEnt [8]) but more process-based niche models can also be used (e.g. CLIMEX [19]). Caution is needed with such methods, as the resulting potential distribution is a ‘worst case’ estimate, assuming that the PnP actually arrived in that area. These methods require sufficient spatial data on the target species’ presence and eventually

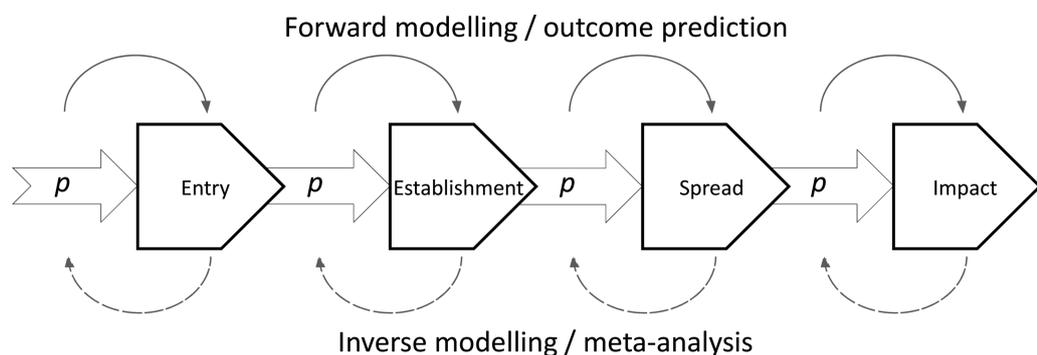


Figure 1. Forward modelling and inverse modelling.

Two different approaches (forward modelling and inverse modelling) could be used to describe the risk of invasion from entry to impact.

absence at large scale. Note that even when a PnP may establish in an area, sufficient propagule pressure is needed to overcome Allee and demographic stochasticity effects occurring at low densities [20].

Modelling species spread

Once arrived and established, PnPs can disperse in the new area. Spread involves population growth, individual dispersal and population establishment further in the new area. Spread models can be used separately or as part of a combined model with entry and establishment [21,22]. Generally, they focus on the dispersal component only, and describe dispersal through kernels in case of natural (unassisted and assisted by natural vectors) dispersal or with network/correlative models in case of human-assisted spread, or combinations of the above [23]. Most spread models are pest-specific, requiring detailed information on population and dispersal characteristics [24–26]. Even simpler models with few parameters may still be difficult to parameterize for new biosecurity risks [27].

Promising approaches

Using models for assessing invasion risk is challenging because of information scarcity on not-yet-introduced species. As a result, one has to rely on existing data of related species — if available, or EKEs, introducing unquantifiable uncertainty in predictions. Because scientific studies in the field of invasion science often focus on a single well-characterized pest species, invasion phase, area, or vector, it is unclear whether generalization to species with similar traits, propagule pressure, and habitat characteristics is possible [12,25]. Some studies have been performed on finding generalities in the establishment and spread characteristics of species [28–30,31]. These studies do not focus on pest-specific drivers, but try to unveil general patterns by going beyond species. For example, Fahrner and Aukema [31] reported in a meta-analysis across 147 studies a four-fold difference in spread rates of univoltine and multivoltine species. Meta-analyses, such as the one mentioned are often static analyses, but they could be combined with dynamic models. A recent example is the generic spread model by Hudgins et al. [28]. Hudgins et al. used a spread model with a negative exponential dispersal kernel with its parameters being a function of covariates such as species traits and environmental traits. Optimal parameters for these covariates were obtained by running the model inversely (‘inverse modelling’ Table 1, Figure 1). For each species, the spread was simulated by running the model from the moment of the establishment to the final time of the observed distribution. They compared this final predicted distribution to the observed one of all species to get the best fits for the model parameters and structure. Mech et al. [32] also used this kind of approach to compare the impact of currently established pest insects with insect and host traits and evolutionary relationships between native and novel hosts and insects to show that the evolutionary history of a pest’s host species may be a good predictor for impact. Using meta-analysis approaches to generalize results across species invasions, by identifying the factors that drive invasions and exploring how these factors interact with species traits, could be the way forward to build models that are fit for purpose. When combined with model selection techniques [33], it is ensured that models do not become more complex than warranted by available data.

Table 1. Overview of model types for each invasion phase from entry to spread

Model type	Phase	Modelling approach	Limitations
Pathway	Entry Establishment	Forward	Data on entry quite limited, this approach has high uncertainty
Species distribution	Establishment Spread	Inverse (mostly)	Result is worst-case scenario, no temporal dynamics, limited by occurrence data quality
Mechanistic spread	Spread	Forward	Often requires detailed PnPs information
Network model/ stochastic spread	Spread	Inverse/Forward	Limited by occurrence data quality
Meta-analysis	Any	Inverse	Needs sufficient data on PnPs and predictor variables for a meaningful model

Modelling the emergence risk of native forest pests and pathogens

Native species or species introduced a long time ago can also cause high damage. As these species are present in the environment for a long time, they have reached an equilibrium (in population size and spatial distribution) and they are generally not subject to specific regulation or official control. Therefore, pest risk assessment and generic approaches are not as well developed as for newly invasive species. However, a change in environmental conditions can break this equilibrium and lead to an increase in the damage incurred by this species. Recently some species, such as *Thaumetopoea pityocampa* [34] and *Lymantria monacha* [35] have expanded their distribution, while others, such as *Operophtera brumata* [36], *Biscogniauxia mediterranea* [37], or *Dendrolimus pini* [38] have made outbreaks. In the first case, spreading species invade territories that were previously not colonized (mainly involving dispersal and growth processes), while in the second case, outbreaking species suddenly increase in population density in territories (either newly colonized or not) where they have been previously present but at a lower density (mainly involving growth processes). Hereafter, we review models that describe such spread and outbreak dynamics (Figure 2).

Modelling species spread

Many spread models have been developed [24] but very few of them jointly describe population growth, individual dispersal and population establishment beyond the previous species distribution. Although dispersal is the key process in the spread of invasive species, population growth, and establishment in new areas are also very important processes in the spread of native species. We can distinguish three main approaches: (1) SDMs or ecological niche models, (2) dispersal models, and (3) spread models in changing environment (Figure 2A). SDMs are widely used to project the potential establishment area under climate change [39]. Although they do not simulate spread explicitly, they predict the future potential distribution based on climate change scenarios [19,40]. A drawback of these models is that they ignore dispersal capability and the population dynamics, elements that could limit the extent to which they can track the shift of the climate envelope. In turn, dispersal models, such as reaction–diffusion models or dispersal kernels generally ignore changes in habitat and climate conditions. Only a few spread models describing both population growth and dispersal in a changing

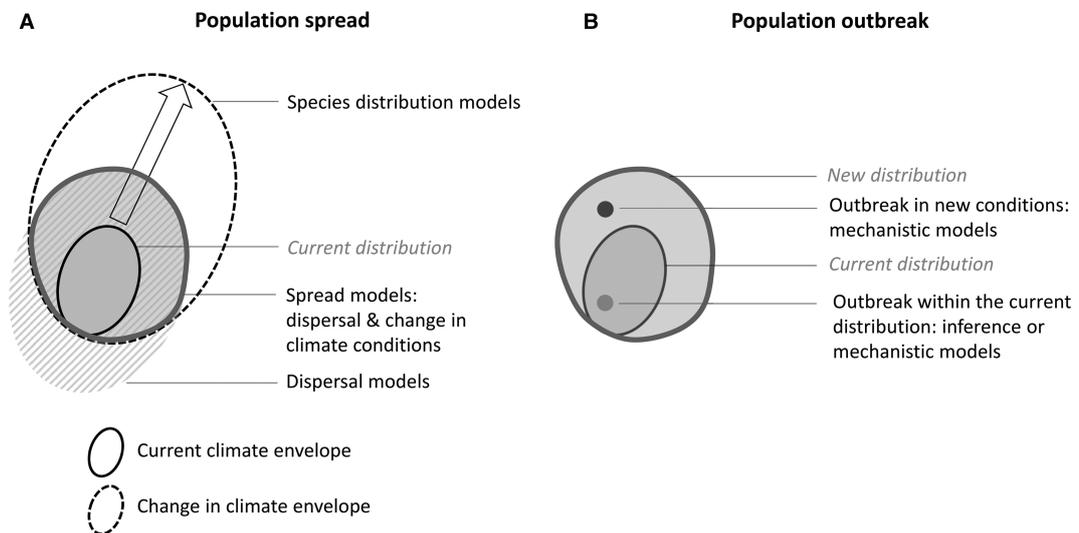


Figure 2. Models for emergence risk.

Models describing population spread (A) and outbreak (B) of emerging species. The dark grey area indicates the current species distribution and light grey area the change in the species distribution with climate change. (A) The climate envelope extends from the full black curve to the projected envelope (dotted black curve) due to climate change in the direction indicated by the arrow. Three types of models could describe spread: they take into account climate suitability, dispersal capability, and eventually both together. (B) Two kind of models could describe outbreak: they are either based on statistical correlations (inference models — appropriate within the current range) or mechanistic processes (more robust in new conditions).

environment have been developed [41]. These models could be seen as a mix of SDMs (or other models considering a change in climate conditions on species establishment or survival) and dispersal models (Figure 2A). However, these models are generally tailor-made for a given species and require many parameters to estimate [42]. They are thus not easily applicable to other species.

Modelling species outbreaks

The presence of forest PnPs does not necessarily imply damage on forest trees. Their population size should exceed a given threshold to be considered as an outbreak [43]. Various categories of outbreaks can be distinguished [44]. Populations can suddenly increase in size due to a change in the environment. If the population size is correlated with an environmental factor, the outbreak is called ‘gradient’, while if the population size depends on a threshold of an environmental factor, the outbreak is called ‘eruptive’. Besides, some pests show periodic outbreaks, which are then called ‘cyclic’. Whatever the outbreak category, two kinds of models can be used to describe these outbreaks: models focusing on the PnPs’ dynamics (mechanistic models) and models describing the probability that an outbreak occurs (inference models) (Figure 2B).

The first type of models focuses on the species traits. Using data on the biology and the phenology of the species, these mechanistic models can predict population dynamics in a large set of conditions, even if these conditions have not been observed yet. These models can describe the dynamics of the species alone for instance with a logistic growth function [45], the effects of environmental factors such as predators or parasites in prey–predator systems [46,47], or the effects of the host with an alike epidemiological susceptible–infected–recovering–susceptible model [48]. Natural enemies or host plants are indeed often used to explain cyclic outbreaks [49]. The complexity of these models can increase until it encompasses a precise description of the species traits and microhabitat [50]. However, these models require data about the population size and good knowledge on the species’ life cycle. The more complex a model is, the more realistic it could be, but it will also require more data to infer the parameters and generate high uncertainties in the model output. These models also require knowledge of the life-history traits, which are difficult to estimate, even for well-studied species. Therefore, such models can fail to identify a less-known species that could outbreak.

The second type of models consist in inferring the probability of an outbreak based on previous outbreaks in the same conditions. In these models, the description of the species is not required as they only use environmental data (e.g. temperature, precipitations, density of host or dead wood) related to the previous outbreak to identify the drivers of the outbreak. These data are analyzed using inference methods such as a regression [51] or random forest [52] to identify the correlation between past outbreaks and environmental factors and then to determine the probability of a next outbreak. However, the probability model is calibrated on a set of observed conditions. Extrapolating the probability beyond these conditions can lead to errors since the relationship between the outbreak likelihood and the environmental conditions could be different outside the observed range.

Promising approaches

Most of spread and outbreak models for native species are not generic so far. The challenge for spread models is to combine both dispersal capabilities and the effects of changing climate conditions in time and in space on population survival and growth, as a function of the species traits. The challenge for outbreak models is to determine the drivers in a way that the model could be easily applied to other conditions, as a function of the species traits. Such outbreak models do not necessarily describe the temporal dynamics of the population density (which would be data-demanding) but rather a probability of outbreaking as a function of environmental conditions. The development of generic spread and outbreak models needs to account for the process drivers, the species traits and eventually the effect of their interaction on the simulated process.

Parameterizing the models and quantifying uncertainty

Parameterization

To generate predictions, invasion and emergence models need to be parameterized. Typically, three ways of parameterization are used or combined thereof: (1) collecting information on a parameter by setting up experiments or doing field measurements (e.g. by rearing the pest in the laboratory and running flight-mill experiments to determine the average flight distance [53–54]; (2) using existing information from literature, sometimes on related species or vectors (e.g. [15,28,32,55]); and (3) the use of EKE on plausible parameter values. Different EKE methods can be used to estimate the most likely value for the parameters, such as fixed and variable interval, and

the roulette method [16]. For example, in the variable interval method, the expert can be asked to determine the parameter value at which the cumulative probability (quantile) is 0.05, 0.5, and 0.95.

Uncertainty

Each model output is associated with a given level of uncertainty or confidence level, which could rise from various sources, and notably from: (1) model choice, (2) parametrization, and (3) stochastic processes. It is thus crucial to communicate the uncertainty related to the model outputs, all the more when these results are used for decision-making. In particular, in the frame of pest risk analyses, a large range of questions should be answered regarding entry, establishment, spread, and impact. For each question, risk assessors have to provide both a rating regarding the related likelihood and a score of uncertainty [56].

To cope with uncertainty about the model to choose, it is possible: (i) to run various models describing the same mechanism (ensemble prediction; for instance various climate models are available to describe a single climate change scenario), (ii) to consider various scenarios describing possible mechanisms (for instance, various climate change scenarios are considered based on different greenhouse gas concentration trajectories hypotheses), or (iii) to build a consensus model to combine all the tested models (e.g. [57–59]).

To cope with uncertainty in the parameters' estimate for a given model, several approaches could be considered. First, to assess the role of each parameter on the model output, sensitivity analyses are generally conducted. Although it is possible to calculate sensitivity analytically (see [60] for local, and [61] for global analyses) especially for not-too-complex models, the sensitivity is preferentially estimated numerically. These estimations are made by using latin hypercube sampling on an arbitrary range of values (e.g. estimated value or bounds of uniform distributions $\pm 10\%$) for each parameter and then identifying the parameters that have the greatest impact on the model output using partial correlation (e.g. [62–63]). These parameters require a more precise estimate to reduce the uncertainty of the model outputs. Second, uncertainty analyses could be done. These consist in assessing the range of possible outputs when varying the parameter values within their confidence intervals. It could result in a confidence interval for the model output (e.g. for a probability of outbreak or probability of introduction), or a series of maps to highlight the best case, the likely case and the worst case (e.g. for potential spread; [27]).

To account for uncertainty arising from highly stochastic processes such as human-mediated dispersal in species potential spread, it is possible to include this uncertainty into the final risk estimates using second-degree stochastic dominance criteria [64].

Promising approaches

Since there is no common rule in the way to handle uncertainty in risk models, it is most important to at least report the known uncertainty in one way or another [8]. Transparency about the main source of uncertainty and the level of confidence of the model outputs is necessary for decision-making and for improving forest managers' trust in modelling approaches. In addition, the combined effects of all uncertainties should be assessed [65].

Role of machine learning

In classical models in ecology, relationships between variables and parameters should be determined, and a statistical algorithm only infers the values of the parameters. As the number of available data and the number of measured variables increase, the number of possible relationships to test increases as well, and it becomes more and more challenging to choose the most appropriate one(s). Machine learning, defined as 'a set of methods that computers use to make and improve predictions or behaviours based on data' [66], allows overcoming this problem. Algorithms can indeed infer the relevant relationships (e.g. the number of variables, the number of parameters, or the shape of the functions), and thus improve or facilitate the development of models describing invasion and emergence risks. The diversity of machine learning algorithms is too large to be described here. However, the next paragraph provides a brief overview of these methods with a focus on SDMs. More details on machine learning in SDMs can be found in [39,67–71].

Machine learning generally has two main components: (i) defining the structure of the model, and (ii) obtaining realistic predictions with this structure. Many types of structures can be chosen, among the most common, there are decision trees (as random forest, e.g. [52]), maps (SVM, e.g. [72]), or functions (as in neural networks, e.g. [73] or GLM [51]). The algorithm will then build a huge amount of the chosen structures by testing many different combinations of relationships. The second step is to obtain predictions from these structures. To do so, the algorithm can either optimize, i.e. find the structure with the predictions which are the closest to the observed data, or make a consensus of every available prediction (bagging). MaxEnt [18] is commonly used for SDMs. In

this method, the structure is a distribution and MaxEnt optimizes this by maximizing the entropy. By doing so, MaxEnt imposes a minimal number of constraints to fit the observed data. Another optimization method is the evolutionary algorithm. Each structure has a score, based on the accuracy of the prediction compared with the data. Different scores can be used for SDMs [74], but the most frequent are the ‘number of correctly classified instances’ and the AUC (area under the ROC curve), which take into account false positives and false negatives. The algorithm then mimics reproduction and natural selection in order to select the prediction with the highest score [67]. On the contrary, random forest [75] is a bagging method. Many different decision trees are built, and the prediction of the model is either the average prediction of every tree (for quantitative predictions) or the result of a vote (for qualitative predictions). More information can be found in [39,67–71].

Promising approaches

A huge diversity of algorithms exists, with varying performance depending on the issue they have to address. The differences between the different approaches are mainly linked to the extrapolation to non-observed cases. Each model extrapolates differently and, as long as there is no data, all assumptions are equally true. Therefore, maintaining a diversity of models can be useful, as global changes can lead to situations not observed yet. Another issue is the interpretability of the model, which is not limited to machine learning in ecology [66,76]. Some models are based on assumptions and structure which are easier to interpret, as a decision tree, while others are based on complicated functions which are harder to interpret even if their predictions are accurate. A model easy to interpret is simpler to share with ‘non-modellers’ and can easily be compared with expert knowledge.

General perspectives

This review points out the need of making a quantitative synthesis of literature and data to integrate species traits and various drivers into models to allow the application to other species. Further model development, listed hereafter, could improve the risk assessment and the support to policy makers regarding forest biosecurity.

Models tackling biological invasions and emergence of native species are generally considered separately. The two processes rely on different mechanisms: the invasive species are in a transient state, as they are arriving and establishing in a new territory, while the native species are initially supposed to be in equilibrium with their environment. Changes in native species are, therefore, generally triggered by a change in the environment, while changes in invasive species are mostly associated with their own population traits (e.g. growth and dispersal capabilities) and human activities (for their entry and spread). However, some processes such as population dispersal are common to both types of species. In addition, the factors involved in the emergence (spread or outbreak) of native species, such as climate change, also apply to exotic species newly arrived in the territory. As a result, it could be interesting to explore the extent to which both modelling approaches could eventually be merged into a unified framework.

As previously explained, various models have already been developed to describe each phase of biological invasions. However, for pest risk analyses and decision-support, decision-makers need to have a global estimate of the risk for forest biosecurity. All these models need to be interconnected at the end. Since these models are often very different, it is very challenging to combine them. Therefore, it is of utmost importance to design these models so that they can be interconnected since the beginning of their development. Since pathway models are data-demanding and should cope with different types of uncertainties, considering both entry and establishment simultaneously could be a solution to determine the locations where invasive species could be first established. These locations could then be used as starting points for spread models. The outputs of spread models should then be appropriate inputs for models describing the economic impact of PnPs (not described here).

The assessment of the economic impact of forest PnPs is an extra layer that should be considered to have a full risk analysis. It should include not only the direct impact of PnP on the yield reduction and the quality loss, but the indirect impacts such as changes in prices, the demand and supply [77,78]. In addition, this economic impact should also rely on the costs and effectiveness of the risk mitigation options. These final cost-benefit analyses are useful for decision-makers to choose the best measure(s). However, this is the final step of a long and complex modelling work, and the uncertainty accumulated along all these interconnected models should be clearly highlighted. Furthermore, the economic impact is only one side of all possible impacts of these forest PnPs. Social and environmental impacts are also of high importance but very difficult to estimate so far, and the total cost is usually underestimated [79].

Lastly, decision-makers do not face a single threat. Modelling researches in forest biosecurity should consider as far as possible multi-risks in forest stands, e.g. simultaneous attacks of several PnPs on the same forest stand,

or attack of a given PnP combined with an extreme climate event impacting the tree health. Identifying the cause of a tree decline is all the more difficult when different factors are impacting the tree. Modelling can help to analyse threats with multiple factors and to explore the best risk reduction options. A huge amount of data is, therefore, needed to continuously monitor all these risks and reduce the uncertainty of the model. The use of large datasets coming from citizen science (e.g. for species occurrence points) could complement available datasets and thus contribute to the refinement of model parameterization [80,81]. However, some specific data can only be collected by specific specialists or official bodies (e.g. inspections at the country border), and standardization is actually needed even for this type of data collection [82].

Summary

- Developing generic models to describe the invasion of exotic forest pests and pathogens (PnPs) and the emergence of native ones is necessary to aid rapid assessment of risks and provide support to policy makers.
- Each invasion phase can be described by a set of models but data is often missing and uncertainty is often very high. Meta-analysis could be used to identify the main drivers and species traits to be considered in parsimonious generic models.
- Most spread and outbreak models for native species are tailor-made for given species. Identifying the drivers, species traits and their interaction is also needed to develop generic models.
- Collecting appropriate data to parameterize risk models is a big issue. Models should be designed to be easily applicable, even when poor information is available on the PnP.
- Models should be interconnected, linked to economic impact models, and ideally consider multi-risks to fully answer the decision-maker's needs.

Competing Interests

The authors declare that there are no competing interests associated with the manuscript.

Author Contributions

C.R. and J.C.D. designed the structure of this short review, R.v.d.D. and J.C.D. mainly focused on the invasion part, C.R. and D.C. mainly focused on the emergence part, and C.R. lead the paper writing, all others contributed to later versions of the manuscript.

Funding

This project has received funding from the European Union's Horizon 2020 Programme for Research and Innovation under grant agreement no. 771271, project called HOMED «HOListic Management of Emerging forest pests and Diseases ». <http://homed-project.eu/>

Acknowledgements

We thank Monique Mourits (Wageningen University, NL) with whom we interacted on the economic impact, and the guest editors, Nicola Spence and Alan MacLeod (DEFRA, U.K.), who invited us to contribute to this special issue on biosecurity.

Abbreviations

AUC, area under the ROC curve; EKE, expert knowledge elicitation; PnPs, pests and pathogens; SDMs, species distribution models.

References

- 1 Canadell, J.G. and Raupach, M.R. (2008) Managing forests for climate change mitigation. *Science* **320**, 1456–1457 <https://doi.org/10.1126/science.1155458>
- 2 Brockerhoff, E.G. and Liebhold, A.M. (2017) Ecology of forest insect invasions. *Biol. Invasions* **19**, 3141–3159 <https://doi.org/10.1007/s10530-017-1514-1>
- 3 Ghelardini, L., Luchi, N., Pecori, F., Pepori, A.L., Danti, R., Della Rocca, G. et al. (2017) Ecology of invasive forest pathogens. *Biol. Invasions* **19**, 3183–3200 <https://doi.org/10.1007/s10530-017-1487-0>
- 4 Rosenzweig, C., Casassa, G., Karoly, D.J., Imeson, A., Liu, C., Menzel, A. et al. (2007) Assessment of observed changes and responses in natural and managed systems. *Climate Change 2007: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, M.L. Parry, O.F. Canziani, J.P. Palutikof, P.J. van der Linden and C.E. Hanson, Eds., Cambridge University Press, Cambridge, U.K., 79–131
- 5 Robinet, C. and Roques, A. (2010) Direct impacts of recent climate warming on insect populations. *Integr. Zool.* **5**, 132–142 <https://doi.org/10.1111/j.1749-4877.2010.00196.x>
- 6 Jactel, H., Petit, J., Desprez-Loustau, M.-L., Delzon, S., Piou, D., Battisti, A. et al. (2012) Drought effects on damage by forest insects and pathogens: a meta analysis. *Glob. Change Biol.* **18**, 267–276 <https://doi.org/10.1111/j.1365-2486.2011.02512.x>
- 7 Seidl, R., Thom, D., Kautz, M., Martin-Benito, D., Peltoniemi, M., Vacchiano, G. et al. (2017) Forest disturbances under climate change. *Nat. Clim. Change* **7**, 395–402 <https://doi.org/10.1038/nclimate3303>
- 8 Venette, R.C., Kriticos, D.J., Magarey, R.D., Hoch, F.H., Baker, R.H.A., Worner, S.P. et al. (2010) Pest risk maps for invasive alien species: a roadmap for improvement. *BioScience* **60**, 349–362 <https://doi.org/10.1525/bio.2010.60.5.5>
- 9 Venette, R.C. (2015) *Pest Risk Modelling and Mapping for Invasive Alien Species*, p. 252, CAB International, Boston
- 10 EFSA PLH Panel, Jeger, M., Bragard, C., Caffier, D., Candresse, T., Chatzivassiliou, E. et al. (2018) Guidance on quantitative pest risk assessment. *EFSA J.* **16**, 5350 <https://doi.org/10.2903/j.efsa.2018.5350>
- 11 Blackburn, T.M., Pyšek, P., Bacher, S., Carlton, J.T., Ducan, R.P., Jarošík, V. et al. (2011) A proposed unified framework for biological invasions. *Trends Ecol. Evol.* **26**, 333–339 <https://doi.org/10.1016/j.tree.2011.03.023>
- 12 McGeoch, M.A., Genovesi, P., Bellingham, P.J., Costello, M.J., McGrannachan and, C. and Sheppard, A. (2016) Prioritizing species, pathways, and sites to achieve conservation targets for biological invasion. *Biol. Invasions* **18**, 299–314 <https://doi.org/10.1007/s10530-015-1013-1>
- 13 Douma, J.C., Pautasso, M., Venette, R., Robinet, C., Hemerik, L., Mourits, M. et al. (2016) Pathway models for analyzing and managing the introduction of alien plant pests – an overview and categorization. *Ecol. Model.* **339**, 58–67 <https://doi.org/10.1016/j.ecolmodel.2016.08.009>
- 14 Douma, J.C., Robinet, C., Hemerik, L., Mourits, M.M., Roques, A. and van der Werf, W. (2015) Development of probabilistic models for quantitative pathway analysis of plant pests introduction for the EU territory. *EFSA Support. Publ.* **12**, EN-809, 435pp <https://doi.org/10.2903/sp.efsa.2015.EN-809>
- 15 Douma, J.C., Hemerik, L., van der Werf, W., Magnusson, C. and Robinet, C. (2017) Development of a pathway model to assess the exposure of European pine trees to pine wood nematode via the trade of wood. *Ecol. Appl.* **27**, 769–785 <https://doi.org/10.1002/eap.1480>
- 16 EFSA. (2014) Guidance on expert knowledge elicitation in food and feed safety risk assessment. *EFSA J.* **12**, 3734 <https://doi.org/10.2903/j.efsa.2014.3734>
- 17 Franklin, J. (2010) *Mapping Species Distributions: Spatial Inference and Prediction*, Cambridge University Press, Cambridge <https://doi.org/10.1017/CBO9780511810602>
- 18 Phillips, S.J., Anderson, R.P. and Schapire, R.E. (2006) Maximum entropy modeling of species geographic distributions. *Ecol. Model.* **190**, 231–259 <https://doi.org/10.1016/j.ecolmodel.2005.03.026>
- 19 Kriticos, D.J., Maywald, G.F., Yonow, T., Zurcher, E.J., Herrmann, N.I. and Sutherst, R.W. (2015) *CLIMEX Version 4: Exploring the Effects of Climate on Plants, Animals and Diseases*, 184 pp, CSIRO, Canberra
- 20 Liebhold, A. and Bascompte, J. (2003) The Allee effect, stochastic dynamics and the eradication of alien species. *Ecol. Lett.* **6**, 133–140 <https://doi.org/10.1046/j.1461-0248.2003.00405.x>
- 21 Lustig, A., Worner, S.P., Pitt, J.P., Doscher, C., Stouffer, D.B. and Senay, S.D. (2017) A modeling framework for the establishment and spread of invasive species in heterogeneous environments. *Ecol. Evol.* **7**, 8338–8348 <https://doi.org/10.1002/ece3.2915>
- 22 Yemshanov, D., Koch, F.H., McKenney, D.W., Downing, M.C. and Sapio, F. (2009) Mapping invasive species risks with stochastic models: a cross-border United States-Canada application for *Sirex noctilio* Fabricius. *Risk Anal. Int. J.* **29**, 868–884 <https://doi.org/10.1111/j.1539-6924.2009.01203.x>
- 23 Robinet, C., Roques, A., Pan, H., Fang, G., Ye, J., Zhang, Y. et al. (2009) Role of human-mediated dispersal in the spread of the pinewood nematode in China. *PLoS One* **4**, e4646 <https://doi.org/10.1371/journal.pone.0004646>
- 24 Chapman, D.S., White, S.M., Hoofman, D.A.P. and Bullock, J.M. (2015) Inventory and review of quantitative models for spread of plant pests for use in pest risk assessment for the EU territory. *EFSA Support. Publ.* **12**, EN-795 190 pp <https://doi.org/10.2903/sp.efsa.2015.EN-795>
- 25 Leung, B., Roura-Pascual, N., Bacher, S., Heikkilä, J., Brotons, L., Burgman, M.A. et al. (2012) TEASIng apart alien species risk assessments: a framework for best practices. *Ecol. Lett.* **15**, 1475–1493 <https://doi.org/10.1111/ele.12003>
- 26 Parry, H., Sadler, R. and Kriticos, D. (2013) Practical guidelines for modelling post-entry spread in invasion ecology. *NeoBiota* **18**, 41–66 <https://doi.org/10.3897/neobiota.18.4305>
- 27 Robinet, C., Kehlenbeck, H., Kriticos, D.J., Baker, R.H.A., Battisti, A., Brunel, S. et al. (2012) A suite of models to support quantitative assessment of spread in pest risk analysis. *PLoS One* **7**, e43366 <https://doi.org/10.1371/journal.pone.0043366>
- 28 Hudgins, E.J., Liebhold, A.M. and Leung, B. (2017) Predicting the spread of all invasive forest pests in the United States. *Ecol. Lett.* **20**, 426–435 <https://doi.org/10.1111/ele.12741>
- 29 Ward, S.F., Fei, S. and Liebhold, A.M. (2019) Spatial patterns of discovery points and invasion hotspots of non-native forest pests. *Global Ecol. Biogeogr.* **28**, 1749–1762 <https://doi.org/10.1111/geb.12988>
- 30 Branco, M., Nunes, P., Roques, A., Fernandes, M.R., Orazio, C. and Jactel, H. (2019) Urban trees facilitate the establishment of non-native forest insects. *NeoBiota* **52**, 25–46 <https://doi.org/10.3897/neobiota.52.36358>
- 31 Fahrner, S. and Aukema, B.H. (2018) Correlates of spread rates for introduced insects. *Glob. Ecol. Biogeogr.* **27**, 734–743 <https://doi.org/10.1111/geb.12737>

- 32 Mech, A.M., Thomas, K.A., Marsico, T.D., Herms, D.A., Allen, C.R., Ayres, M.P. et al. (2019) Evolutionary history predicts high-impact invasions by herbivorous insects. *Ecol. Evol.* **9**, 12216–12230 <https://doi.org/10.1002/ece3.5709>
- 33 Burnham, K.P. and Anderson, D.R. (2002) *A Practical Information-Theoretic Approach. Model Selection and Multimodel Inference*, 2nd ed, Springer, New York
- 34 Battisti, A., Stastny, M., Netherer, S., Robinet, C., Schopf, A., Roques, A. et al. (2005) Expansion of geographic range in the pine processionary moth caused by increased winter temperatures. *Ecol. Appl.* **15**, 2084–2096 <https://doi.org/10.1890/04-1903>
- 35 Fält-Nardmann, J.J.J., Tikkanen, O.P., Ruohomäki, K., Otto, L.F., Leinonen, R., Pöyry, J. et al. (2018) The recent northward expansion of *Lymantria monacha* in relation to realised changes in temperatures of different seasons. *Forest Ecol. Manag.* **427**, 96–105 <https://doi.org/10.1016/j.foreco.2018.05.053>
- 36 Jepsen, J.U., Hagen, S.B., Ims, R.A. and Yoccoz, N.G. (2008) Climate change and outbreaks of the geometrids *Operophtera brumata* and *Epirrita autumnata* in subarctic birch forest: evidence of a recent outbreak range expansion. *J. Animal Ecol.* **77**, 281–291 <https://doi.org/10.1111/j.1365-2656.2007.01339.x>
- 37 Henriques, J., Inacio, M.L., Lima, A. and Sousa, E. (2012) New outbreaks of charcoal canker on young cork oak trees in Portugal. *Integr. Prot. Oak For.* **76**, 85–88 ISBN 978-92-9067-253-1
- 38 Matek, M. and Pernek, M. (2018) First record of *Dendrolimus pini* outbreak on aleppo pine in Croatia and severe case of population collapse caused by entomopathogen *Beauveria bassiana*. *South-East Eur. For.* **9**, 91–96 <https://doi.org/10.15177/seefor.18-17>
- 39 Elith, J., Graham, C.H., Anderson, R.P., Dudik, M., Ferrier, S., Guisan, A. et al. (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography* **29**, 129–151 <https://doi.org/10.1111/j.2006.0906-7590.04596.x>
- 40 Thuiller, W., Lavorel, S. and Araúja, M.B. (2005) Niche properties and geographical extent as predictors of species sensitivity to climate change. *Global Ecol. Biogeogr.* **14**, 347–357 <https://doi.org/10.1111/j.1466-822X.2005.00162.x>
- 41 Roques, L., Roques, A., Berestycki, H. and Kretzschmar, A. (2008) A population facing climate change: joint influences of Allee effects and environmental boundary geometry. *Popul. Ecol.* **50**, 215–225 <https://doi.org/10.1007/s10144-007-0073-1>
- 42 Robinet, C., Rousset, J. and Roques, A. (2014) Potential spread of the pine processionary moth in France: preliminary results from a simulation model and future challenges. *Ann. Forest Sci.* **71**, 149–160 <https://doi.org/10.1007/s13595-013-0287-7>
- 43 Dajoz, R. (1998) *Les insectes et la forêt*, p. 594, Editions Tec & Doc, Lavoisier, Paris,
- 44 Berryman, A.A. (1987) The theory and classification of outbreaks. In *Insect Outbreaks* (Barbosa, P. and Schultz, J.C.), pp. 3–30, Academic Press <https://doi.org/10.1016/B978-0-12-078148-5.50005-1>
- 45 Aparicio, J.P., Corley, J.C. and Rabinovich, J.E. (2013) Life history traits of *Sirex noctilio* F. (Hymenoptera: Siricidae) can explain outbreaks independently of environmental factors. *Math. Biosci. Eng.* **10**, 1265–1279 <https://doi.org/10.3934/mbe.2013.10.1265>
- 46 Wilder, J.W., Voorhis, N., Colbert, J.J. and Sharov, A. (1994) A three variable differential equation model for gypsy moth population dynamics. *Ecol. Model.* **72**, 229–250 [https://doi.org/10.1016/0304-3800\(94\)90085-X](https://doi.org/10.1016/0304-3800(94)90085-X)
- 47 Van Der Putten, W.H., Macel, M. and Visser, M.E. (2010) Predicting species distribution and abundance responses to climate change: why it is essential to include biotic interactions across trophic levels. *Philos. Tran. R. Soc. B* **365**, 2025–2034 <https://doi.org/10.1098/rstb.2010.0037>
- 48 Nenzén, H.K., Filotas, E., Peres-Neto, P. and Gravel, D. (2017) Epidemiological landscape models reproduce cyclic insect outbreaks. *Ecol. Complex.* **31**, 78–87 <https://doi.org/10.1016/j.ecocom.2017.04.004>
- 49 Turchin, P., Ellner, S.P., Kendall, B.E., Murdoch, W.W., Fischlin, A., Casas, J. et al. (2003) Dynamical effects of plant quality and parasitism on population cycles of larch budmoth. *Ecology* **84**, 1207–1214 [https://doi.org/10.1890/0012-9658\(2003\)084\[1207:DEOPQAJ2.0.CO;2](https://doi.org/10.1890/0012-9658(2003)084[1207:DEOPQAJ2.0.CO;2)
- 50 Baier, P., Pennerstorfer, J. and Schopf, A. (2007) PHENIPS-A comprehensive phenology model of ips typographus (L.) (Col., Scolytinae) as a tool for hazard rating of bark beetle infestation. *Forest Ecol. Manag.* **249**, 171–186 <https://doi.org/10.1016/j.foreco.2007.05.020>
- 51 Buse, J., Schröder, B. and Assmann, T. (2007) Modelling habitat and spatial distribution of an endangered longhorn beetle - a case study for saproxylic insect conservation. *Biol. Conserv.* **137**, 372–381 <https://doi.org/10.1016/j.biocon.2007.02.025>
- 52 Ismail, R., Mutanga, O. and Kumar, L. (2010) Modeling the potential distribution of pine forests susceptible to *Sirex noctilio* infestations in Mpumalanga, South Africa. *Trans. GIS* **14**, 709–726 <https://doi.org/10.1111/j.1467-9671.2010.01229.x>
- 53 Taylor, R.A., Bauer, L.S., Poland, T.M. and Windell, K.N. (2010) Flight performance of *Agrilus planipennis* (Coleoptera: Buprestidae) on a flight mill and in free flight. *J. Insect Behav.* **23**, 128–148 <https://doi.org/10.1007/s10905-010-9202-3>
- 54 Robinet, C., David, G. and Jactel, H. (2019) Modelling the distances travelled by flying insects based on the combination of flight mill and mark-release-recapture experiments. *Ecol. Model.* **402**, 85–92 <https://doi.org/10.1016/j.ecolmodel.2019.04.006>
- 55 Robinet, C., Van Opstal, N., Baker, R. and Roques, A. (2011) Applying a spread model to identify the entry points from which the pine wood nematode, the vector of pine wilt disease, would spread most rapidly across Europe. *Biol. Invasions* **13**, 2981–2995 <https://doi.org/10.1007/s10530-011-9983-0>
- 56 Holt, J., Leach, A.W., Knight, J.D., Griessinger, D., MacLeod, A., van der Gaag, D.J. et al. (2012) Tools for visualizing and integrating pest risk assessment ratings and uncertainties. *EPP0 Bull.* **42**, 35–41 <https://doi.org/10.1111/j.1365-2338.2012.02548.x>
- 57 Marmion, M., Parviainen, M., Luoto, M., Heikkinen, R.K. and Thuiller, W. (2009) Evaluation of consensus methods in predictive species distribution modelling. *Diver. Distrib.* **15**, 59–69 <https://doi.org/10.1111/j.1472-4642.2008.00491.x>
- 58 Boulanger, Y., Gray, D.R., Cooke, B.J. and de Grandpré, L. (2016) Model-specification uncertainty in future forest pest outbreak. *Glob. Change Biol.* **22**, 1595–1607 <https://doi.org/10.1111/gcb.13142>
- 59 Fournier, A., Barbet-Massin, M., Rome, Q. and Courchamp, F. (2017) Predicting species distribution combining multi-scale drivers. *Glob. Ecol. Conserv.* **12**, 215–226 <https://doi.org/10.1016/j.gecco.2017.11.002>
- 60 Caswell, H. (2019) *Sensitivity Analysis: Matrix Methods in Demography and Ecology*, Springer International Publishing, pp. 299, Cham, Switzerland <https://doi.org/10.1007/978-3-030-10534-1>
- 61 Sobol, I.M. (2001) Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Math. Comput. Simulat.* **55**, 271–280 [https://doi.org/10.1016/S0378-4754\(00\)00270-6](https://doi.org/10.1016/S0378-4754(00)00270-6)
- 62 Koch, F.H., Yemshanov, D., McKenney, D.W. and Smith, W.D. (2009) Evaluating critical uncertainty thresholds in a spatial model of forest pest invasion risk. *Risk Anal.* **29**, 1227–1241 <https://doi.org/10.1111/j.1539-6924.2009.01251.x>
- 63 Blower, S.M. and Dowlatabadi, H. (1994) Sensitivity and uncertainty analysis of complex models of disease transmission: an HIV model, as an example. *Int. Stat. Rev.* **62**, 229–243 <https://doi.org/10.2307/1403510>

- 64 Yemshanov, D., Koch, F.H., Lyons, D.B., Ducey, M. and Koehler, K. (2012) A dominance-based approach to map risks of ecological invasions in the presence of severe uncertainty. *Divers. Distrib.* **18**, 33–46 <https://doi.org/10.1111/j.1472-4642.2011.00848.x>
- 65 Scientific Committee, E.F.S.A., Benford, D., Halldorsson, T., Jeger, M.J., Knutsen, H.K., More, S. et al. (2018) Scientific opinion on the principles and methods behind EFSA's guidance on uncertainty analysis in scientific assessment. *EFSA J.* **16**, 5122, pp. 235 <https://doi.org/10.2903/j.efsa.2018.5122>
- 66 Molnar, C. (2020) Interpretable Machine learning. A Guide for Making Black Box Models Explainable. Ebook, pp. 312
- 67 Gobeyn, S., Mouton, A.M., Cord, A.F., Kaim, A., Volk, M. and Goethals, P.L.M. (2019) Evolutionary algorithms for species distribution modelling: a review in the context of machine learning. *Ecol. Model.* **392**, 179–195 <https://doi.org/10.1016/j.ecolmodel.2018.11.013>
- 68 Guisan, A., Thuiller, W. and Zimmermann, N.E. (2017) *Habitat Suitability and Distribution Models with Applications in R (Ecology, Biodiversity and Conservation)*, Cambridge University Press, Cambridge
- 69 Guisan, A. and Zimmermann, N.E. (2000) Predictive habitat distribution models in ecology. *Ecol. Model.* **135**, 147–186 [https://doi.org/10.1016/S0304-3800\(00\)00354-9](https://doi.org/10.1016/S0304-3800(00)00354-9)
- 70 Lorena, A.C., Jacintho, L.F.O., Siqueira, M.F., De Giovanni, R., Lohmann, L.G., De Carvalho, A.C.P.L.F. et al. (2011) Comparing machine learning classifiers in potential distribution modelling. *Expert Syst. Appl.* **38**, 5268–5275 <https://doi.org/10.1016/j.eswa.2010.10.031>
- 71 Maher, S.P., Randin, C.F., Guisan, A. and Drake, J.M. (2014) Pattern-recognition ecological niche models fit to presence-only and presence-absence data. *Meth. Ecol. Evol.* **5**, 761–770 <https://doi.org/10.1111/2041-210X.12222>
- 72 Boogar, A.R., Salehi, H., Pourghasemi, H.R. and Blaschke, T. (2019) Predicting habitat suitability and conserving *Juniperus* spp. habitat using SVM and maximum entropy machine learning techniques. *Water* **11**, 2049 <https://doi.org/10.3390/w11102049>
- 73 Watts, M.J. and Worner, S.P. (2011) Modelling insect habitat suitability with artificial neural networks. Chapter 8, In *"Insect Habitats: Characteristics, Diversity and Management"* (Harris, E.L. and Davies, N.E.), pp. 163–195, Novascience Press New York, USA ISBN: 978-1-60876-396-2
- 74 Mouton, A.M., De Baets, B. and Goethals, P.L.M. (2010) Ecological relevance of performance criteria for species distribution models. *Ecol. Model.* **221**, 1995–2002 <https://doi.org/10.1016/j.ecolmodel.2010.04.017>
- 75 Breiman, L. (2001) Random forests. *Mach. Learn.* **45**, 5–32 <https://doi.org/10.1023/A:1010933404324>
- 76 Murdoch, W.J., Singh, C., Kumbier, K., Abbasi-Asl, R. and Yu, B. (2019) Definitions, methods, and applications in interpretable machine learning. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 22071–22080 <https://doi.org/10.1073/pnas.1900654116>
- 77 Soliman, T., Mourits, M.C.M., van der Werf, W., Hengeveld, G.M., Robinet, C. and Oude Lansink, A.G.J.M. (2012) Framework for modelling economic impacts of invasive species, applied to pine wood nematode in Europe. *PLoS One* **7**, e45505 <https://doi.org/10.1371/journal.pone.0045505>
- 78 Petucco, C., Lobianco, A. and Cauria, S. (2020) Economic evaluation of an invasive forest pathogen at a large scale: the case of ash dieback in France. *Environ. Model. Assess.* **25**, 1–21 <https://doi.org/10.1007/s10666-019-09661-1>
- 79 Bradshaw, C.J.A., Leroy, B., Bellard, C., Roiz, D., Albert, C., Fournier, A. et al. (2016) Massive yet grossly underestimated global costs of invasive insects. *Nat. Commun.* **7**, 12986 <https://doi.org/10.1038/ncomms12986>
- 80 Dickinson, J.L., Zuckerman, B. and Bonter, D.N. (2010) Citizen science as an ecological research tool : challenges and benefits. *Annu. Rev. Ecol. Evol. Syst.* **41**, 149–172 <https://doi.org/10.1146/annurev-ecolsys-102209-144636>
- 81 César de Sá, N., Marchante, H., Marchante, E., Cabral, J.A., Honrado, J.P. and Vicente, J.R. (2019) Can citizen science data guide the surveillance of invasive plants? A model-based test with Acacia trees in Portugal. *Biol. Invasions* **21**, 2127–2141 <https://doi.org/10.1007/s10530-019-01962-6>
- 82 Eschen, R., Rigaux, L., Sukovata, L., Vettraino, A.M., Marzano, M. and Grégoire, J.-C. (2015) Phytosanitary inspection of woody plants for planting at European Union entry points: a practical enquiry. *Biol. Invasions* **17**, 2403–2413 <https://doi.org/10.1007/s10530-015-0883-6>