Biosecurity

A Systems Perspective

Edited by Susan M. Hester, Lucie M. Bland, Edith Arndt, Sana Bau, James S. Camac, Evelyn Mannix, Raphaël Trouvé and Andrew P. Robinson

Chapter 14

Map

Creating Maps of Establishment Potential

14 Map
Creating Maps of Establishment Potential

James S. Camac, Estibaliz Palma, and John B. Baumgartner

ABSTRACT

A strong, efficient, and effective biosecurity system requires the determination of which areas are most susceptible to threat incursion, establishment, and spread. Maps of establishment potential are developed to inform where surveillance resources should be focused, to assess likely spread scenarios, estimate likelihoods of threat absence, create risk maps for prioritising threats and optimally minimise risk. This chapter outlines the three fundamental geographic barriers that influence establishment potential—propagule pressure, abiotic suitability, and biotic suitability. We describe how each barrier is commonly approximated using a range of methods such as species distribution models, pathway analyses, expert elicitation, border screening data, and distributional data of host or habitat. We provide particular emphasis on the various decisions and assumptions that must be made when constructing species distribution models for invasive species. Lastly, we provide a practical and pragmatic approach for integrating these various models and data to create maps of establishment potential. We illustrate this integrated approach using a hypothetical case study on oriental fruit fly (*Bactrocera dorsalis*) entering Australia.

GLOSSARY

Pathway analysis A model that attempts to approximate the propagule pressure or contamination rate of a threat entering a region or point of entry (e.g. ports or airports).

Species distribution model (SDM) A model that approximates a species' niche and maps this niche to the landscape to portray the geographic distribution of suitable conditions.

Establishment likelihood (or establishment potential) A probability that encompasses the three geographic barriers to threat establishment (i.e. the likelihood of arrival or propagule pressure, abiotic suitability, and biotic suitability).

Propagule pressure (also termed introduction rate or contamination rate) A measure of the likelihood or expected number of threat-specific contamination events entering a region.

Abiotic suitability The suitability of the non-living environment (e.g. climate, terrain, or disturbance regime) for a threat.

Biotic suitability The suitability of the living environment (e.g. habitat and interspecies relationships) for a threat. Commonly, this will be based on various habitat requirements (e.g. vegetation type, structure, food source) and sometimes may include interspecific interactions that are facilitative, competitive, or predatory.

Leakage rate The expected number of events (or the likelihood of events) bypassing border screening controls.

Viability rate The likelihood that a contamination event that bypasses border screening is viable for establishment (e.g. based on factors such as pathway survivability or contamination population size).

Model transferability The capacity for a model fit under one set of conditions to be applied to different conditions. In the context of SDMs, a model is said to be transferable if the species-environment relationships estimated under one context remains valid in other contexts.

INTRODUCTION

Threat risk maps are a fundamental decision-support tool in biosecurity, and many governments and industries invest heavily into some form of risk mapping (Elith 2017). Threat risk maps can be used to inform likelihoods of threat absence, where to position finite post-border surveillance resources, trade interventions, and cost-sharing between jurisdictions (see Box 14.1). Threat risk maps can also be used to determine where to prioritise public and stakeholder awareness campaigns for the purposes of early detection or delimitation (i.e. reporting from farmers; see Chapter 6. Detect).

BOX 14.1. USING ESTABLISHMENT LIKELIHOOD MAPS IN BIOSECURITY DECISION MAKING

Maps of establishment likelihoods are fundamental decision-support tools for biosecurity. Specifically, they can be used for:

Informing where to conduct surveillance for early detection

Surveillance for early detection should be prioritised in areas where the likelihood of establishment is highest (see Chapter 6. Detect). Such prioritisation can be done on a threat-by-threat basis using threat-specific establishment likelihood maps. Alternatively, establishment likelihood maps can be combined for multiple threats to systematically identify areas of shared high establishment potential, and thus, inform where to prioritise surveillance for multiple threats (Camac, Baumgartner, Hester, et al. 2021). Maps of establishment likelihood are also useful for assessing the proportion of establishment potential captured by existing or proposed surveillance designs, depending on budget constraints and the spatial arrangement of survey effort (Camac, Baumgartner, Hester, et al. 2021). This is especially important for early detection, which can be expensive to implement and maintain (Camac, Dodd, et al. 2020; see Chapter 6. Detect).

Informing likelihoods of threat absence

To gain and maintain market access, exporting countries may be required to prove that they are free of invasive species and diseases known to affect economic, environmental, and societal values (see Chapter 2. Biosecurity Systems and International Regulations). Freedom of invasive species or disease is commonly assumed if surveillance programs do not detect the threat, especially when they focus on areas of high establishment likelihood or high consequence. A fundamental problem with declaring freedom based solely on the absence of detections is that this approach assumes the surveillance program has perfect sensitivity (i.e. if the invasive species is present, it will always be detected). In practice, it is impossible to be

Map 233

certain a species is absent. Rather, absence can only be inferred in terms of likelihoods (see Chapter 6. Detect). Establishment likelihood maps can be used in combination with surveillance data (i.e. surveillance sensitivity and effort) to infer likelihoods of threat absence that explicitly account for the geographic barriers faced by threats in establishing at a location. Likelihoods of absence can be estimated for locations where surveillance occurs, but also for locations without surveillance to make inferences about absence at regional or national scales (see Chapter 6. Detect).

Initialising spread models

Maps of establishment likelihood can be used to add further realism to spread models for potential exotic threats. Spread models are initialised based on a known incursion point or the location of a known established population (Bradhurst et al. 2021). Initial incursions may be simulated at random across a region of interest, or be specified to occur at a particular location of interest perceived to be at high risk of establishment. Likelihood maps that account for the geographic barriers to establishment (namely, propagule pressure and suitability of biotic and abiotic environments) provide a coherent and transparent method for simulating initial establishment events that seed subsequent spread events. By simulating both the introduction and subsequent spread of a threat, the costs and benefits associated with different pre-border, border, and post-border management strategies can be assessed, ultimately allowing decision-makers to better anticipate and proactively manage future incursions (Bradhurst et al. 2021).

Developing threat risk maps

Establishment likelihood maps are not themselves measures of risk. However, when coupled with spatial estimates of the economic, environmental, and societal/cultural consequences of a threat (see Chapter 3. Anticipate), maps of establishment likelihood can be used to produce risk maps. Risk maps can inform where to conduct surveillance or control measures to minimise risk, for example in areas that exhibit both high likelihoods and consequences of establishment. Risk maps can also be used to prioritise locations for risk mitigation activities, especially those associated with proactive asset protection (see Chapter 7. Prepare, Respond, and Recover).

Depending on how they are built, threat risk maps can describe where an invasive species or disease may arrive, establish and spread, or cause significant economic, environmental, and societal harm. True risk maps describe the product of (1) the likelihood a species will successfully establish in an area and (2) the expected magnitude of harm if establishment occurs in that area. Such maps are particularly useful for identifying locations where both the establishment likelihood and potential consequences of an incursion are high, and thus, where risk mitigation controls should be prioritised.

In practice, however, most so-called risk maps do not incorporate both likelihoods and consequences. Rather, maps tend to focus on estimating likelihoods associated with a particular stage of the invasion process, such as the likelihood of a threat arriving at a port via imported goods (i.e. pathway analysis) or the likelihood an area is environmentally suitable for establishment to occur (Venette et al. 2010).

While many methods exist for constructing likelihood maps, two main approaches are used in biosecurity. The first—pathway analysis—involves estimating the arrival rate of a threat entering a country or arriving at a point in the landscape (e.g. ports). Commonly, pathway analyses are informed by border interception data coupled with movement data (or models) describing vectors

of spread (e.g. imported goods, people, wind) from areas of known establishment to locations of interest (Camac, Baumgartner, Garms, et al. 2021; Tingley et al. 2018).

The second commonly used approach is to construct species distribution models (SDMs) that approximate the spatial distribution of suitable environment needed for a threat to establish and persist. SDMs can be parameterised using data on a threat's known physiological tolerances (e.g. temperature tolerances; Kearney et al. 2008), occurrence records from its global or native range (Phillips, Anderson, and Schapire 2006), or a combination of the two (Kriticos et al. 2017). For invasive species, SDMs commonly predict climate suitability as a function of global climatic variables under the assumption that climate is the primary driver limiting a threat's potential geographic distribution.

Despite both pathway analyses and SDMs providing partial (and complementary) information on the barriers to establishment, there have been few attempts to integrate these two model types in biosecurity. Studies that have attempted to link pathway likelihoods with estimates of environmental suitability are often geographically restricted to points of entry (e.g. ports) and do not make inferences about relative likelihoods of establishment beyond these locations (e.g. Tingley et al. 2018).

In this chapter, we outline a general framework for estimating the geographic establishment potential of exotic threats using data types available in most countries. We provide practical guidance on how to (1) distribute pathway risk beyond points of entry using various "rules of thumb" and (2) estimate abiotic and biotic suitability for a threat. We illustrate this mapping framework using a case study of the oriental fruit fly (*Bactrocera dorsalis*) potentially establishing in Australia.

DEVELOPING MAPS OF ESTABLISHMENT LIKELIHOOD

Neither pathway analyses nor SDMs estimate the establishment potential (or establishment likelihood) of a threat. This is because the establishment potential of an exotic threat is governed by three spatial constraints (Figure 14.1), where each must be met for establishment to occur:

1. Can the threat reach the location of interest (i.e. propagule pressure)?
2. Are abiotic conditions suitable (e.g. climate suitability)?
3. Are biotic conditions suitable (e.g. presence of host or food)?

While pathway models inform the likelihood of a species arriving at a location (i.e. propagule pressure), they rarely account for abiotic suitability (e.g. climate) or biotic suitability (e.g. presence of hosts
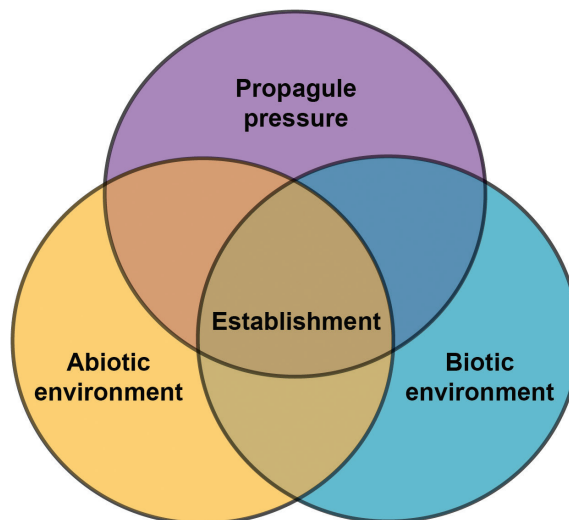

FIGURE 14.1   The three main elements governing the likelihood of establishment of exotic species in the introduced region. (Adapted from Catford, Jansson, and Nilsson 2009).

Map   235

or specific vegetation types). While SDMs can approximate climate suitability and to a lesser degree biotic suitability (if relevant predictors are available), they do not account for propagule pressure.

In the following sections, we provide overviews of each of the three barriers to threat establishment (Figure 14.1) and how they can be estimated using commonly available data and methods. Once estimated, the product (i.e. multiplication) of these three likelihoods give the probability of establishment conditioned on the three barriers.

ESTIMATING PROPAGULE PRESSURE

Most contemporary introductions of exotic threats have been facilitated by the movement of people and goods across continents and international borders (Hulme 2021). To determine the likelihood of an exotic threat arriving at a location (i.e. a pathway analysis), biosecurity practitioners not only require an understanding of how the threat may spread across jurisdictional boundaries, but also an understanding of the main vectors for post-border spread within these regions.

A pathway analysis estimating the expected likelihood of a post-border arrival requires information on: (1) volumes of pathway carriers (e.g. passengers, cargo, or mail) from infected locations to the region of interest; (2) the likelihood a carrier contains a threat viable for establishment; (3) the likelihood contaminated carriers are not detected at the border; and (4) how contaminated carriers disperse post-border.

Pre-border regulation and border screening of people and imported goods entering a country provide data useful for approximating the probability of exotic threat arrival. Data obtained from border screening can be used to fit statistical models that estimate the rate at which a threat arrives at a country's border or points of entry (see Chapter 5. Screen). If information is available regarding the viability of threats (e.g. estimated through transit survival likelihoods or average population size for detected contamination events) and the efficacy of border screening (e.g. by conducting endpoint surveys), statistical models can quantify the expected likelihood that viable introduction events bypass pre-border and border controls.

Estimates of the propagule pressure passing through various points of entry are then paired with information about how the carriers of the threat disperse post-border. For example, if a contaminated shipping container passes border screening at a port, it may be moved a considerable distance before being opened and unpacked, allowing the threat to escape into the wider landscape. Inadequate post-border monitoring of the movement of goods and people, alongside privacy concerns about collecting these types of data, are currently the main barriers to conducting pathway analyses that incorporate post-border movements.

Despite the limited availability of post-border movement data, opportunities exist to use available data to approximate the post-border dispersal of threats. Entry pathways are diverse, but in most cases are related to either human movement or human activities. Depending on the introduction pathway and the availability of national datasets, post-border movement of invasive species and diseases can be approximated in a variety of ways. For example, it can be assumed that international tourists entering a country will concentrate around areas of high tourist accommodation (at least initially) or that returning residents or international mail will distribute throughout a country as a function of human population density or where commerce activity is greatest. Most imported goods (e.g. food, nursery stock, or vehicles) can be assumed to be destined for areas of higher human population density. Post-border distribution of other goods (e.g. farm equipment and fertiliser) can be approximated by coupling land use type with measures of agricultural intensity (e.g. fertiliser usage statistics, farm numbers). There is evidence to support such assumptions, with studies commonly finding land use, road density, and human population density to be strongly correlated with first detections of exotic threats even when correcting for potential survey bias (e.g. Dodd et al. 2016).

ESTIMATING ABIOTIC SUITABILITY

Abiotic barriers limiting a species' potential distribution are diverse in nature and scale and can include climate, disturbance regimes, and terrain attributes. At large geographic scales (e.g. global or

continental), the potential distribution of a species is thought to be governed by climate (Araújo and Rozenfeld 2014). The last several decades have seen the development of global databases of climate and biological data as well as a vast range of statistical and mechanistic SDMs. In invasive species management, SDMs are commonly parameterised using global climatic variables (the most readily available global environmental data), and thus, are sometimes referred to as "climate suitability models".[1] These models have become a popular tool in biosecurity for quantifying the potential area of occupancy, which in turn can inform threat prioritization (McGeoch et al. 2016), post-border surveillance design (Camac, Baumgartner, Hester, et al. 2021), and estimates of expected economic impacts (Dodd, Baumgartner, and Kompas 2021; Dodd et al. 2020).

Many biosecurity agencies and associated research institutions actively invest in some form of invasive species climate suitability modelling, whether via initiatives such as the European Union's Pratique risk assessment program, the development of cloud computing systems such as Biosecurity Commons (Biosecurity Commons 2022), the development of proprietary software (e.g. CSIRO's CLIMEX; Kriticos et al. 2016), or governmental platforms such as the United States's NAPPFAST (Magarey et al. 2007) or Australia's CLIMATE/CLIMATCH (ABARES 2020; Climatch 2019).

Despite a vast number of SDM methods (and diverse opinions on how to use them), there remains no clear evidence of a single best approach for predicting invasive species' potential distribution (Barry et al. 2015). Irrespective of the approach chosen to predict climatic suitability for a species, a wealth of subjective decisions and assumptions are made along the model building process, from data sourcing and cleaning to model development and validation. Subjective or arbitrary decisions can lead to a wide variety of potential suitability predictions for a single species, even when the modelling algorithm is held constant (Camac, Baumgartner, Hester, et al. 2021).

In the following sections, we describe the main steps required for biosecurity practitioners to develop and use SDMs (Figure 14.2).

Data Sourcing

Literature Review

When considering whether to build or reproduce an SDM, the first step should be to conduct a short review of the grey literature (national and international governmental reports) and peer-reviewed literature. The literature review should summarise details associated with physiological climate tolerances, existing SDMs, required hosts, and any distributional data reported in the literature. A good place to start a literature review is to examine the Invasive Species Compendium maintained by the Centre for Agriculture and Bioscience International (CABI 2023). The compendium contains data summarising threats' biology, distribution, spread pathways, and associated information sources.

Occurrence Records

Ideally, SDMs should be based on occurrence datasets including both presence and absence localities (i.e. presence-absence methods) or abundance data. Since these are rarely available globally, the majority of SDMs are built using only presence localities (presence-only methods) or by comparing presence locations to a random sample from all available locations (presence-background methods).

In recent years, publicly available occurrence databases have been pivotal in driving the popularity and accessibility of presence-only/presence-background SDM methods. At the global scale, the most comprehensive dataset is the Global Biodiversity Information Facility (GBIF 2023). The data contained within GBIF can be used in different modelling frameworks for either parameterising models or for validating them. Species-specific databases or other expert-curated datasets are another valuable source of occurrence information. These databases are typically compiled and verified by experts and are thus considered highly reliable. In many cases, these data are already integrated into GBIF, however, this is not always the case. As such, it is important that searches are made for such speciality databases.
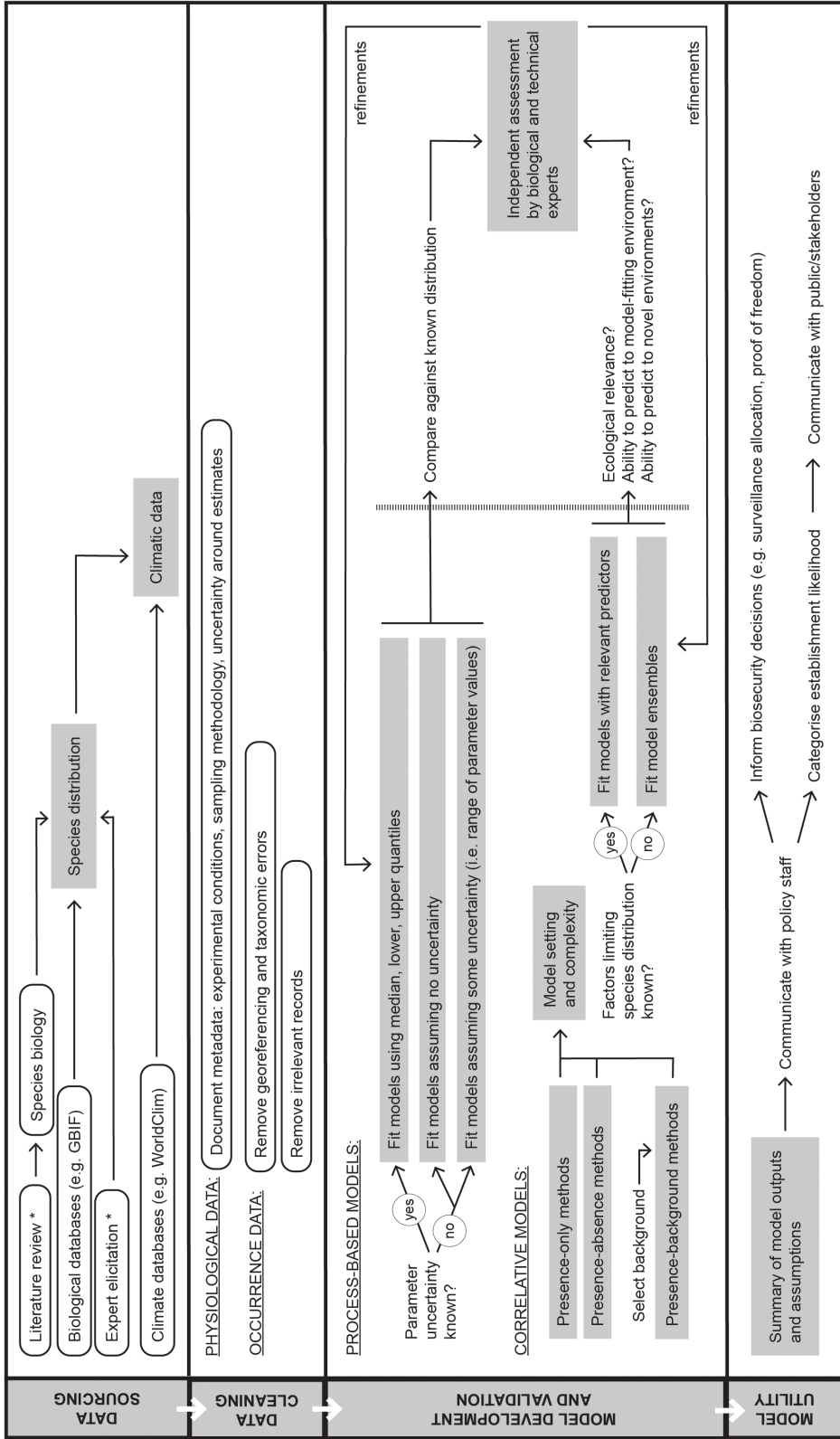
Map 237

FIGURE 14.2 Steps involved in creating a climatic suitability model, from data sourcing and cleaning to model development and validation, to model utility for decision making. (Adapted from Camac, Baumgartner, et al. 2020). *If the species' biology or distribution is unknown, expert elicitation can be used, or data can be collected from datasets for phylogenetically/morphologically similar species.

Physiological Data

Physiological data that may be incorporated into process-based SDMs are generally derived from controlled laboratory experiments published in the peer-reviewed literature and in academic theses. Several demographic and physiological databases have also been published to simplify parametrisation of process-based models, including:

- *GlobTherm*: A database that includes thermal tolerance metrics for 2,133 species of multicellular algae, plants, fungi, and animals, extracted from published studies (Bennett et al. 2018).
- *AmP*: A database of referenced data on animal energetics and parameter values of models based on Dynamic Energy Budget (DEB) theory for over 1,000 animal species (Marques et al. 2018).
- *COMPADRE*: A database of more than 5,500 population matrix models (and ancillary information) for more than 600 plant species worldwide (Salguero-Gómez et al. 2015).
- *COMADRE*: A database of more than 1,625 population matrix models (and ancillary information) for more than 340 animal species worldwide (Salguero-Gómez et al. 2016).

Climatic Data

Common global climate databases include:

- WorldClim (Fick and Hijmans 2017; WorldClim 2023),
- Climate Research Unit (CRU 2023; Harris et al. 2014),
- CliMond (CliMond 2014; Kriticos et al. 2012), and
- CHELSA (CHELSA 2023; Karger et al. 2017).

Each dataset contains different suites of long-term averaged climatic variables, and estimates long-term climatic averages across the earth using different data sources, interpolation algorithms, and temporal periods. It is therefore important for the user to carefully select the variables (and dataset) that are likely to be most relevant to the threat of interest.

Non-Climatic Abiotic Data

Non-climatic variables (e.g. land use, and disturbance patterns) also play an important role in governing a species' potential distribution. However, such information is generally not available at either the spatial extent or resolution required to parametrise SDMs. Where such information does exist (e.g. land use layers), it is often based on image classification algorithms from global satellite imagery. Consequently, such data are often coarse (either in terms of spatial resolution or categorisation), and thus, may have limited utility in SDM. Over time, the quantity, quality and accessibility of these datasets will undoubtedly improve.

Alternatives for Data-Poor Threats

When there is insufficient biological or distributional data for a given threat, there are three options to estimate relevant physiological tolerances or responses to climatic factors: (1) collate data on a threat with similar phylogeny or similar biological, behavioural, or physiological traits (Morales-Castilla et al. 2017); (2) use a trait-based SDM (e.g. Morin and Lechowicz 2008; Pollock, Morris, and Vesk 2012); or (3) use an expert elicitation framework (e.g. IDEA protocol; Hemming et al. 2018).

Data Cleaning

Occurrence Data

Species occurrence records derived from online databases (e.g. GBIF) are a valuable resource for parametrising and validating SDMs. However, issues with data quality such as errors in georeferencing, taxonomy, temporal periods, irrelevant records (e.g. fossil records, cultivated plant specimens, and zoo animals), and incomplete metadata can introduce significant biases in model fitting and validation (Maldonado et al. 2015). A summary of common errors found in occurrence databases can be found in Table 14.1.

Map   239

TABLE 14.1

Common Errors in Occurrence Databases.

Geo-referencing errors
- Incorrect habitat (marine or terrestrial)
- Coordinates and country mismatches
- Switched or equal latitude and longitude
- Records on country or province centroids
- Records based on rasterised collections
- Records with high spatial uncertainty or strong decimal rounding
- Coordinates entered as (0, 0)

Taxonomic errors
- Misidentifications
- Obsolete or incorrect taxonomy
- Incorrect spelling

Temporal errors
- Incorrect or missing dates
- Records not within climate temporal period

Irrelevant records
- Records from zoos, botanical gardens and museums
- Fossil records
- Migratory records
- Border interceptions
- Records from less reliable sources (e.g. iNaturalist)
- Duplicate records

Standardised tools for cleaning large-scale biodiversity databases include the R package CoordinateCleaner (Zizka et al. 2019), which identifies records from cultivated sources as well as geographic and temporal errors commonly found in biological databases (Table 14.1). Another useful tool is the R package taxize, which helps deal with taxonomic issues such as spelling mistakes, synonyms, or obsolete taxonomy (Boyle et al. 2022; Chamberlain and Szöcs 2013).

Once data have been cleaned, it is critical to remove records that are likely to be transient in nature (i.e. unlikely to be associated with a persistent population). This is because biosecurity applications typically focus on delimiting the area where a threat could establish and spread. CABI (https://www.cabidigitallibrary.org/product/qi) is an example of an expert-verified database that contains the known establishment status of threats at the country or provenance scale. Where appropriate resources are available, all records should be scrutinised by experts in the relevant geographic area to verify whether they are likely to be records of established populations, and if not, they should be removed prior to model development or validation.

Physiological Data

Experimental physiological data suffer from two primary concerns: whether the experimental conditions are transferable to the natural world and whether the individuals used are representative of variability within a species (Briscoe et al. 2019). Both questions are difficult to answer, especially the latter, where intra-specific variability is almost never known. By documenting the experimental approach, the number of individuals used, where they were sourced, and the number of populations sampled, the modeller and consulted experts have the required information to make informed decisions regarding both the representativeness and transferability of experimental findings.

It is also important to record the uncertainty around any experiment-derived physiological responses to allow for the propagation of uncertainty throughout the modelling process. Ultimately, this manifests itself in a range of possible distribution maps, ranging from best to worse case scenarios. It is also important to document physiological processes (e.g. factors governing reproductive

maturity or diapause) that lack empirical estimates. Documenting these knowledge gaps highlights areas for future research and provides greater context for the decisions made during model development.

Model Development

There remains no clear evidence of a single, best approach for predicting an invasive species' potential distribution (Barry et al. 2015). To determine which approach performs consistently well requires an understanding of models' ability to predict the potential invasive distribution of a species, which for practical purposes is unknowable (Elith 2017). As such, modellers have mostly evaluated methods based on their predictive performance when predicting to current distribution data (Sequeira et al. 2018, Yates et al. 2018), but these performance metrics do not explicitly examine performance with respect to the true purpose of these models—their ability to predict the potential, currently unoccupied distribution of a species.

Accurate prediction in the region where the model was fitted does not guarantee accurate prediction outside this range (Elith 2017; Fourcade, Besnard, and Secondi 2018). To examine a model's transferability from one region to another, various cross-validation approaches have been used. One commonly used approach is to parametrise a model based on native-range occurrences and use it to predict to the observed invasive region (Yates et al. 2018). This approach is also fraught with problems because the distributional data in the native or invaded range may not be a true indication of a species' ability to persist under environmental conditions not currently occupied (Elith 2017; Elith, Kearney, and Phillips 2010). As with all modelling, while decisions about the most useful method can be informed by existing knowledge, there remains ambiguity in the final choice of model to predict an unobserved event.

Broadly, SDMs can be classified as either process-based models or correlative models. Determining which model to use is ultimately dependent on data availability, tolerance to the assumptions made, and the decision context (Elith 2017; Guillera-Arroita et al. 2015).

Correlative Models

Correlative SDMs exploit the statistical association between spatial environmental data and occurrence records to implicitly capture biological processes limiting the distribution of the species. Irrespective of which correlative approach is used, an "equilibrium assumption" applies—that is, that a species is in equilibrium with their environment (i.e. the species has had the opportunity to establish in all suitable areas within the model-fitting region). A vast range of correlative methods can be used to predict the potential distribution of a species or disease. These methods fall into four categories:

- **Presence-only models** commonly estimate a species' potential distribution in two ways: (1) using occurrence records to delimit the climatic range in which a species may persist (e.g. envelope methods such as BIOCLIM, Booth et al. 2014; and range bagging, Drake 2015); (2) by calculating climatic similarity relative to the climate of known occurrence locations (CLIMATCH; ABARES 2020, CLIMEX; Kriticos et al. 2016). The primary appeal of presence-only models is that they use the most commonly available type of data (presence records) without additional assumptions and decisions associated with other methods (see below). However, presence-only models can be highly sensitive to outliers (e.g. when delimiting climatic ranges) and, like all correlative methods, to sampling biases (i.e. some sites are more likely to be sampled than others; Elith 2017).
- **Presence-background models** derive suitability scores from the frequencies of surveyed occurrences (i.e. positive cases) relative to non-positive cases. The meaning of non-positive cases can vary in important ways (see Box 14.2). For some popular methods, such as Maxent (Phillips et al. 2006), the non-positive case is commonly referred to as the

Map 241

background (i.e. a set of points that randomly sample the landscape or locations available, irrespective of whether a species is present or not).

- **Presence-pseudo-absence models** treat the non-positive cases as implied absences or pseudo-absences (e.g. Genetic Algorithm for Rule-Set Prediction; GARP). These implied absences are either placed (1) anywhere except where presences occur or (2) in geographic or environmental zones considered to be unlikely to contain the species of interest (Elith 2017). The peer-reviewed literature contains several suggestions on how to choose locations for pseudo-absences or define reliable absences for presence-only data (e.g. Le Maitre, Thuiller, and Schonegevel 2008; Lobo, Jiménez-Valverde, and Hortal 2010). Relative to the presence-background approach, this method requires additional subjective decisions about both the position and number of pseudo-absences and is generally perceived as a less rigorous statistical framework (Renner et al. 2015).
- **Presence-absence models** derive suitability scores from relative frequencies of surveyed occurrences and absences in different environmental conditions. Presence-absence models are the ideal correlative model because they can more readily account for imperfect detection and survey effort while also removing the need to arbitrarily select background points (Guillera-Arroita et al. 2015; Lahoz-Monfort, Guillera-Arroita, and Wintle 2014). Unlike other methods, they can be used to estimate true probabilities of presence as opposed to relative probabilities (Guillera-Arroita et al. 2015). If the required data are available, presence-absence models should always be preferred over other correlative approaches. In practice, such data are rarely available at the geographic resolution and extent required to estimate the potential distribution of an invasive species and their use in invasive species management is limited (Elith and Leathwick 2009).

BOX 14.2. CHALLENGES AND ISSUES IN USING CORRELATIVE MODELS FOR INVASIVE SPECIES

Correlative methods that do not utilise data on absences (i.e. presence-only, presence-background, and presence-pseudoabsence models) do not estimate true probabilities of occurrence, as this requires knowing the true prevalence of the modelled organism (Guillera-Arroita et al. 2015). Rather, these correlative models estimate relative probabilities that are commonly assumed to be proportional to the true probabilities. When models are fitted for multiple species, direct comparisons can be made between their general patterns of predictions (e.g. which areas are predicted to have high vs. low suitability) but not between their predicted suitability scores. In other words, a suitability score of 0.9 for one species, may not be the same as a 0.9 for another species.

Presence-background and presence-pseudoabsence models rely on additional subjective decisions about the extent, distribution, and number of background locations or pseudo-absences. The impacts of these decisions are often underappreciated, despite evidence highlighting how different arbitrary modelling decisions and assumptions can lead to substantial differences in model predictions and interpretation (Phillips et al. 2009; Syfert, Smith, and Coomes 2013; Warton and Shepherd 2010).

Ideally, for presence-background models, background points should characterise the range of environments to which the species of interest could have dispersed and should exhibit the same pattern of survey bias as exists in the occurrence records. In practice, defining the extent of potential dispersal is challenging, especially for invasive species with actively expanding ranges (VanDerWal et al. 2009). To reduce the impact of such complications associated with modelling species with expanding invasive ranges, presence-background models often restrict the species' occurrence records to the native range. This decision also simplifies the

process of background sampling by limiting the choice of sample points to the extent of the species' native range. Yet, discarding records beyond the species' native range will typically reduce the sample size and may result in potentially unreliable extrapolation when predicting climatic suitability across the species' exotic range (Broennimann and Guisan 2008). For this reason, modellers may be hesitant to discard occurrence records in the invaded range.

When SDMs use records from both the native and the exotic range, different approaches can be used to select background points. The first approach is based on constructing convex hulls around the occurrence records and sampling the background points from within the hulls. The second approach is to sample background points from within a radius from each occurrence record; the size of the radius being defined by the user, usually based on the species' dispersal ability (VanDerWal et al. 2009). The third approach involves sampling background points from a pre-defined set of bioclimatic zones believed to provide suitable environmental conditions for the modelled species.

Occurrence data are often biased towards areas with high population and infrastructure density. To avoid confusion between patterns of survey effort and patterns of climatic suitability, survey bias needs to be accounted for in the SDMs. A simple method to do this is called target-group sampling. This method uses occurrence records from a broad set of species for which sampling bias is likely to be similar to that for the species of interest, due to similar behaviour by those reporting occurrences (Phillips et al. 2009). The sites for all records from all species in the target group then make up the full set of available information on survey effort and can be directly used as background data within the region of interest. A second method extends the target-group approach, using the target-group occurrences to construct an interpolated bias grid that informs the appropriate distribution and intensity of background samples. Alternatively, data describing factors likely to explain geographic survey bias (e.g. distance from roads and cities) can be used directly to calculate such bias grids (Warton, Renner, and Ramp 2013) or can be combined in regression models to estimate them.

While the above methods can efficiently deal with sampling bias, they are not commonly applied in invasive species modelling. This is due partly to lack of technical expertise, and partly to the broad spatial scale of modelling (across countries) that prevents the inclusion of relevant factors at smaller scales (e.g. distance to roads). However, sampling bias can substantially challenge predictions of climatic suitability for invasive species based on presence-only data, and spatial variation in survey effort should be addressed in these models where feasible.

Process-Based Models

Process-based (mechanistic) SDMs are models that explicitly incorporate biological processes that limit distributions. These models range in complexity from simple climatic threshold mapping to complicated biophysical models such as NicheMapR (Kearney and Porter 2017). The commercial software CLIMEX (Kriticos et al. 2016) can also incorporate processes when it is parametrised using experimental data.

The primary advantage of process-based models is that they attempt to model the causal relationships between climate and various vital attributes such as survival, growth, and fecundity. They are therefore assumed to be superior at predicting a species' fundamental distribution (Elith 2017; Kearney and Porter 2009), though limited evidence supports this assumption (Uribe-Rivera et al. 2022). The reliability of process-based models is strongly dependent on whether they account for the fundamental processes delimiting a species' fundamental niche and whether the empirically derived parameters accurately reflect both the responses to climate and intra-specific variability in these responses (Kearney and Porter 2009). In most cases, the climatic factors governing a species' potential distribution are unknown or unknowable, and the resources required to

Map 243

obtain such data (i.e. lab experiments and field validation), coupled with time and in-house technical skills required to build these models, currently make them infeasible for most biosecurity agencies (Briscoe et al. 2019).

Model Settings and Complexity

Correlative methods used to fit SDMs allow the user to fine-tune a variety of settings to achieve the desired model fit and interpretation. Such settings include, among others, the selection and number of background points, the complexity of covariate response curves (i.e. the relationship between the value of an environmental variable, e.g. annual mean temperature, and suitability for the modelled species) and the type of model outputs that can be extracted.

Default settings for SDMs methods are not always appropriate. For example, MAXENT background points are, by default, randomly sampled across the entire extent of the climatic data supplied by the user. Background points should typically be restricted to regions that the species has had an opportunity to reach and occupy (VanDerWal et al. 2009). Thus, when determining appropriate model settings, the species and the problem at hand should always be considered.

Model predictions can be affected by model complexity, which can strongly influence the transferability of the model to conditions outside the range of environmental conditions used to fit the model. Highly complex models (i.e. those with many predictors and interactions) are susceptible to over-fitting to noise and peculiarities of the model-fitting data, especially when the number of occurrence records is low (Breiner et al. 2015). Such models may not generalise well, limiting their ability to accurately predict climatic suitability in regions outside the model-fitting range. This issue can be minimised by optimising a model such that it more accurately predicts data not included in model-fitting.

This can be achieved through the process of cross-validation, whereby the occurrence dataset is split into model-fitting and model-testing subsets, and average performance over the replicate models is evaluated.[2] This process can be repeated for each of a set of competing models that use different predictors and varying complexity, to identify models that predict well and exhibit high transferability. In cases when cross-validation is not feasible (e.g. due to small sample size), opting for simpler models is recommended—with few predictor variables, and few or no interactive terms. Simple models retain greater tractability and are likely to be more easily generalised to new datasets (Merow, Smith, and Silander Jr 2013).

Covariate Uncertainty

Accurately predicting abiotic suitability for species through correlative models requires knowledge of the causal predictors (covariates) that govern the species' niche. However, selection of model predictors remains a significant challenge (Barry et al. 2015). There are no obvious "best" predictive variables, and users tend to rely on variables provided in readily available datasets (e.g. WorldClim). When the choice of predictors is unguided, we recommend that a practitioner ensemble (i.e. model average) a range of simple (2–3 parameter) competing models such that overfitting is less likely and model/covariate uncertainty is encapsulated in the final suitability output. Ensembles of such simple models have shown great promise in predicting distributions of rare (Breiner et al. 2015) and non-native species (Hill et al. 2022).

Model Validation

Irrespective of whether a model is process-based or correlative, the outputs should be validated against independent datasets (i.e. occurrence data not used to parametrise the model). For process-based models, validation is straightforward and involves overlaying localities of known establishments onto maps of predicted suitability and then examining whether patterns broadly match (Kearney et al. 2008; Kriticos et al. 2017). However, validation of correlative SDMs for invasive species faces a variety of difficulties. Validation tools for correlative models range from examination of the model residuals to formal statistical summaries (e.g. area under the receiver

operating curve AUC, kappa, explained deviance) that examine the model's ability to accurately predict the fitted, or ideally independent, data (Elith 2017). Unfortunately, the application of these statistical evaluations to invasive species modelling is not particularly appropriate (Elith 2017). This is because the objective of invasive species SDMs is to predict a species' potential (i.e. unobserved) distribution, not just its known range, which is what these statistical measures inform.

To overcome these issues, some authors have fitted the invasive species SDMs using occurrence records from the native range only, and then validated the predictions by either using data in the same region or by using data from the known invaded range. While there is some support for the former validation approach (Barry et al. 2015), using the known invaded region for validation is dependent on: (1) whether limiting factors for the species' distribution in the native and invaded ranges are comparable (i.e. whether the model is indeed transferrable), and (2) whether occurrences in the invaded range are representative of persistence.

Due to the difficulties in evaluating and validating correlative invasive distribution models, we suggest that:

- Models are first assessed for their ecological relevance through: expert knowledge; examining functional response curves; sourcing additional data including physiological information (e.g. temperature tolerances).
- Model predictive ability in the model-fitting region is assessed (e.g. by using AUC) if the method involves estimating response curves (e.g. non-ensemble models).
- Regions of novel environmental space (i.e. environmental conditions not found in the model-fitting region) are examined, especially for suitability maps derived from a correlative model that estimates response curves.
- Model outputs and associated model documentation (i.e. how the model was parametrised, from data sourcing through to validation) be independently assessed by both technical and biological experts. Doing so will ensure key aspects of the species' biology are incorporated and that models are appropriately implemented.

Model Utility for Decision Making

To ensure the appropriate use of a model, fundamental details must be clearly articulated and summarised for biosecurity decision-makers. It is critically important to summarise what data were used, how the model was developed, assumptions that were made, and how one should interpret and can use the output. In essence, the documentation should account for all decisions made in Figure 14.2.

The type of model will have important implications for how it can be used to inform biosecurity decision-making. Correlative models are commonly used to inform a variety of biosecurity decisions. However, while these models may be appropriate in some decision contexts (e.g. spatial prioritisation of surveillance) they may be inappropriate in others. Guillera-Arroita et al. (2015) highlighted that using presence-background models such as Maxent to inform multi-threat prioritisation is often inappropriate because outputs are relative probabilities that are not comparable across species. For a comprehensive guide as to whether relative likelihoods or ranked scores can be used for various biosecurity decision contexts, we recommend readers consult Guillera-Arroita et al. (2015).

It is also common practice to convert maps of continuous predictions (e.g. true and relative probabilities) into a binary (suitable/not suitable) or categorical (high/moderate/low) heat map (Guillera-Arroita et al. 2015). While this is a simple way to communicate to the broader public, we strongly advise against discretising model outputs in biosecurity decision contexts. This is because the decision of where a particular threshold should be is often highly subjective and dependent on both the expected consequences of an invasive species outbreak and the tolerance for incorrectly labelling a suitable location as unsuitable. Discretising model outputs rarely improves decision-making because it degrades the information supplied by the model (Guillera-Arroita et al. 2015).

Map   245

ESTIMATING BIOTIC SUITABILITY

Biotic barriers such as various habitat requirements (e.g. vegetation type, structure, food source) and inter-specific relationships (e.g. facilitative, competitive, or predatory) can have significant impacts on species' persistence, and consequently, their ability to establish in new locations. While such information is sometimes incorporated in SDMs, it is rarely done so for invasive species. This is because: (1) methods best suited to modelling the potential distribution of an invasive species are likely those that estimate the fundamental niche; (2) biotic influences are considered to act on much finer spatial scales than those at which models are commonly applied; (3) biotic predictors are often confounded with climatic model predictors; and (4) biotic factors vary substantially across species and rarely are there relevant standardised global datasets that can be used to approximate such effects (Elith 2017). While data describing global distributions of major biomes and vegetation types are available, their classification systems are typically too broad to meaningfully estimate biotic effects for most species (Brown et al. 2022). Biotic relationships are incredibly complex, and these relationships are mostly unknown in novel environments where new habitats, predators, facilitators, and food sources may play a significant role in species' establishment and long-term persistence.

Despite these challenges in statistically estimating biotic impacts on species' distributions, biosecurity practitioners and governments often use a variety of datasets to inform the availability of suitable habitat or food sources required by potential threats in their regions of interest. These datasets are often derived from direct investment by government departments, and consequently, are country- or region-specific. As these spatial data are not standardised for international use, they are typically developed on finer scales and can include classification systems containing hundreds of classes. Frequently used datasets include highly detailed distributions of land use, vegetation types, known vectors or host material, or estimates of vegetation cover or greenness. For example, the Australian government has invested heavily in the development of a national 50m land use raster containing over 100 land use types that span conservation, agriculture (often identifying the production type), residential/urban, and water land use types (ABARES 2019), with similar datasets available in many other countries.

DEVELOPING A PRAGMATIC ESTABLISHMENT LIKELIHOOD MAP FOR ORIENTAL FRUIT FLY IN AUSTRALIA

Here we provide a pragmatic example of developing an establishment likelihood map for the oriental fruit fly in Australia that explicitly encapsulates the three barriers to establishment (i.e. propagule pressure, abiotic suitability, and biotic suitability).

The oriental fruit fly, *Bactrocera dorsalis*, is a fruit fly native to Asia that has spread and established in over 65 countries, including parts of America, Oceania, and most of sub-Saharan Africa (Figure 14.3). The species poses a significant threat to many horticultural industries worldwide with over 300 different types of fruit and vegetables susceptible (Vargas, Piñero, and Leblanc 2015).

Where oriental fruit fly has yet to establish, countries have implemented networks of pheromone traps with the dual purpose of early detection of incursions and providing evidence to trading partners that the threat is absent (see Chapter 6. Detect). However, the effectiveness of such surveillance networks at detecting new incursions is ultimately dependent on whether traps are positioned in locations where the likelihood of incursion and establishment is greatest, underscoring the relevance and value of maps of establishment likelihood.

ESTIMATING PROPAGULE PRESSURE

The first step in developing an oriental fruit fly establishment likelihood map is to determine the main pathways through which the species may enter the country, and the likelihoods of it (1) evading border screening procedures and (2) remaining viable for an establishment event to occur. Many

FIGURE 14.3 Adult female oriental fruit fly, *Bactrocera dorsalis* (Hendel), laying eggs by inserting her ovipositor in a papaya. (Photograph by Scott Bauer, USDA.)

approaches can be used to estimate these likelihoods, including empirical methods that utilise border interception data, expert elicitation, or a combination of the two.

Here, we use a hypothetical example. Let us assume that oriental fruit fly can enter Australia via four main pathways: international tourists; residents returning from overseas; imported fruit and vegetables; and travellers from the Torres Strait Islands (located between Australia and Papua New Guinea) carrying contaminated fruit or vegetables into Cairns Airport. Let us also assume that estimates exist for annual likelihoods of one or more viable contamination events (i.e. a surviving population size capable of establishing) for each pathway (see Table 14.2).

TABLE 14.2

Hypothetical Likelihoods of Viable Contamination Entering Australia for Four Pathways of Entry of the Oriental Fruit Fly.

Pathway	Probability (Viable contamination | passes border inspection)
International tourists	0.01
Returning residents	0.005
Imported fruit and vegetables	0.001
Passengers from Torres Strait Islands to Cairns	0.05

Note: Here, probability (viable contamination | passes border inspection) refers to the conditional probability that (1) a leakage event occurs and (2) it is viable such that a post-border establishment event could occur.

Map 247

People travelling from the Torres Strait Islands to mainland Australia via Cairns is the pathway with the highest probability (0.05), translating to a 5% chance of a viable contamination event bypassing border screening each year. This pathway has the highest likelihood, due to parts of the Torres Strait regularly experiencing incursions of oriental fruit fly that likely originate from Papua New Guinea, where the species has a persistent established population (Camac, Baumgartner, et al. 2020). The second highest pathway likelihood is associated with international tourists, who may be less aware of Australia's quarantine rules and are therefore more likely to bring in contaminated fruit relative to returning residents (0.01 vs. 0.005, respectively). Imported fruit and vegetables have a low likelihood of a viable contamination (0.001), because they are cold treated prior to export. Assuming pathways are independent, these likelihoods translate to a combined annual probability of approximately 6.5%[3] that one or more viable leakage events occur in a year.

The next step is to determine how pathway contaminations are likely to disperse beyond points of entry into the wider landscape. Ideally, these post-border movements would be informed by empirical data; however, such data are rarely available at the required scale and resolution. It is often necessary to make informed assumptions about how pathway contaminations will disperse through geographic space using pre-existing data and probabilistic functions. Mathematically, this can be depicted as:

$$\Pr(Arrival)_{i,k} = \Pr(Viable\ contamination\,|\,passes\ border\ inspection)_k \times f(i,k) \qquad \text{Eq. (14.1)}$$

Here, $\Pr(Arrival)_{i,k}$ is the expected probability a contaminant from pathway k arrives at location i. It is estimated as a function of the likelihood a contaminant passes border inspection multiplied by $f(i,k)$, a weighting function that describes the proportion of pathway units (e.g. people, containers, etc.) expected to arrive at location i. In the following sections we briefly describe some simple ways for formulating $f(i,k)$, for the four pathways used in this hypothetical example.

Torres Strait Pathway

We assume that most passengers entering Cairns from the Torres Strait Islands are likely to be distributed as a function of population density within the city. We assume that 50% of passengers will remain within 10 km of the airport, which is positioned in close proximity to the city. We account for this by using a negative exponential distance-decay function, where the likelihood of a passenger dispersing to a location decreases with distance from airport, such that 50% of incoming passengers will be distributed within 10 km of the airport. As these individuals are likely to be staying with friends and relatives, we also weight this by the resident human population in each location based on national census data.

The weight (i.e. $f(i,k)$ in Eq. 14.1) for this pathway can be specified as:

$$\Pr(Torres\ Strait\ arrival)_i = e^{\left(-\frac{log(0.5)}{10} \times Distance_i\right)} \times Proportion\ of\ population_i \qquad \text{Eq. (14.2)}$$

where the 0.5 within the log specifies the proportion of passengers dispersed within a particular distance from the airport, specified as the denominator. When mapped onto a raster grid, this weighted function results in the heat map shown in Figure 14.4.

Returning Residents and Imported Fruit and Vegetables

We assume that Australian citizens returning from overseas and imported fruit and vegetables will both be dispersed as a function of human population density derived from Australian census data. This acknowledges that (1) most returning residents will quickly return to their place of residence
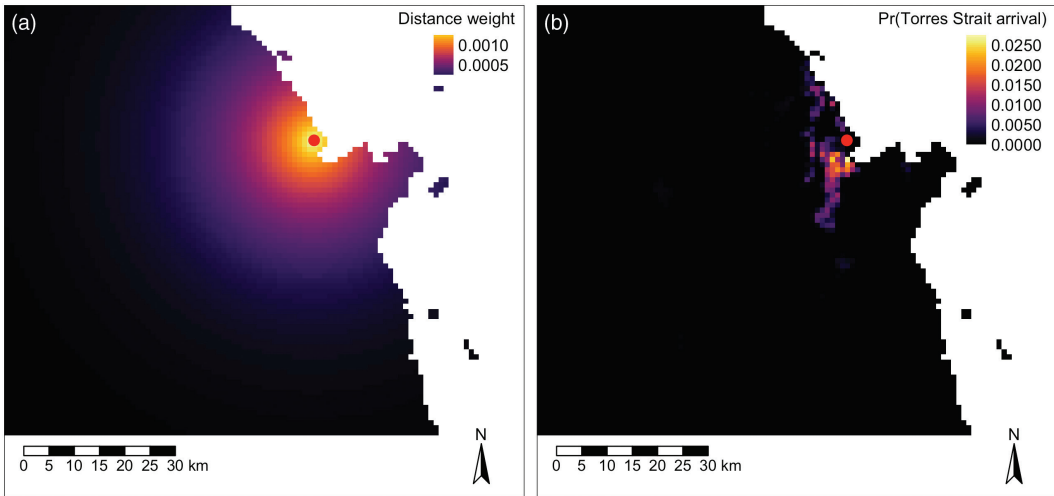
FIGURE 14.4 Weights defining the post-entry distribution of contamination likelihood around Cairns (red point), on the hypothetical Torres Strait Island pathway. Shown are (a) the impact of distance on relative likelihood of oriental fruit fly arrival; and (b) the relative likelihood of arrival after accounting for human population density. Warmer colours signify higher arrival likelihoods/distance weights. Raster data were projected to the Australian Albers (EPSG: 3577) coordinate system and then aggregated from a 1 km to 5 km raster grid to aid visualisation.

irrespective of the distance from airport, and (2) imported goods will go where demand is greatest. As such we specify the weight, $f(i,k)$, for this pathway as:

$$\Pr\left(Returning\ resident\ arrival\right)_i, \Pr\left(Imported\ food\ arrival\right)_i = Proportion\ of\ population_i \quad \text{Eq. (14.3)}$$

When mapped onto a raster grid, this weighted function results in the heat map shown in Figure 14.5.

International Tourists

International tourists are not expected to travel significant distances in the days immediately following arrival (i.e. when the risk of a passenger carrying contaminated fruit or other goods is likely greatest). We will assume that 50% of passengers will disperse within 250 km from an international airport, and we assume they will congregate in areas with high densities of tourist accommodation. Here, we will use a similar approach as for the Torres Strait Islands pathway, whereby we assume a negative exponential decline with distance from international airports, weighted by the proportion of tourist rooms in a location (data derived from Camac, Dodd, et al. 2020):

$$\Pr\left(Tourist\ arrival\right)_i = e^{\left(-\frac{log(0.5)}{250} \times Airport\ distance_i\right)} \times Proportion\ of\ tourist\ beds_i \quad \text{Eq. (14.4)}$$

Estimating Likelihoods of Arrival Across Pathways

Once the pathway-specific likelihoods have been distributed across the region of interest, we can then determine the probability of one or more viable contaminations (i.e. propagule pressure) arriving at location i across k pathways of interest:

$$\Pr\left(Arrival\right)_i = 1 - \prod_{k=1}^{n}\left(1 - \Pr\left(Arrival\right)_{i,k}\right) \quad \text{Eq. (14.5)}$$

Where, $\Pr\left(Arrival\right)_{i,k}$ is the likelihood a viable contamination from pathway k arrives at location i (Figure 14.6).

Map   **249**
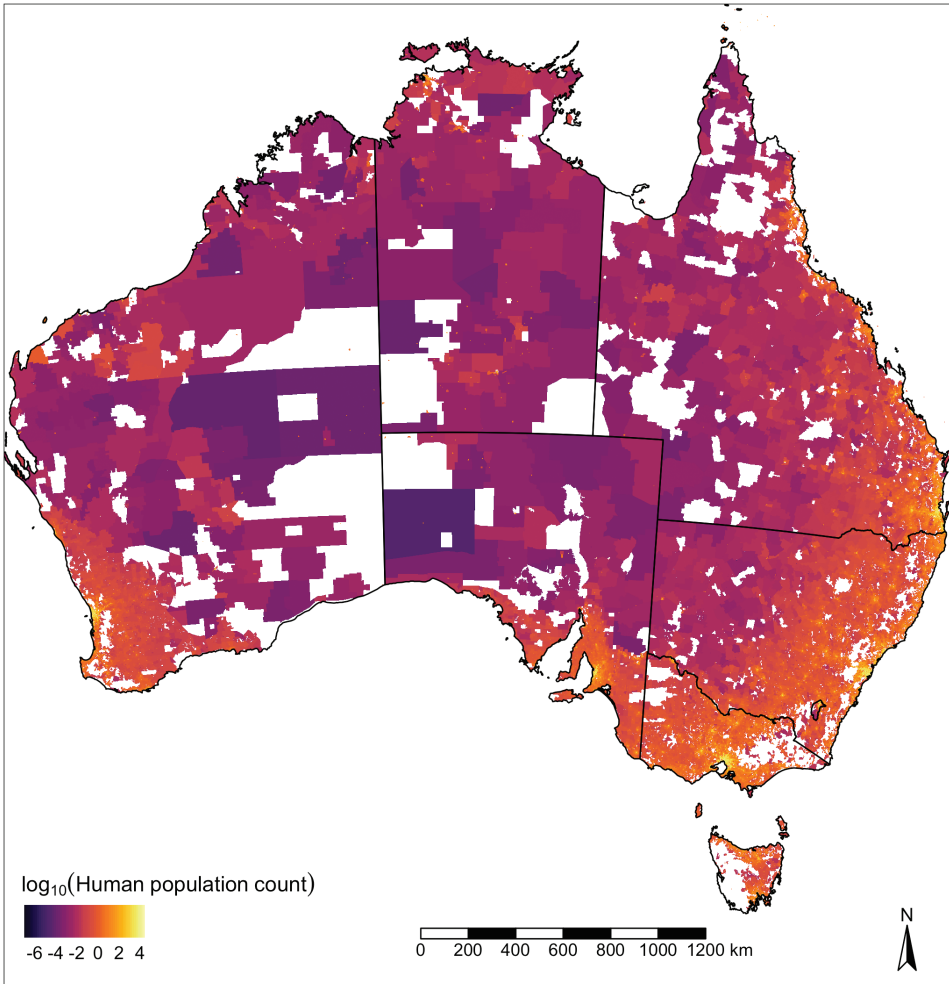

FIGURE 14.5 Human population count ($\log_{10}$ scale). (Map derived from 2016 Australian Census data (Australian Bureau of Statistics 2017)). Warmer colours signify higher human population counts. Raster data were projected to the Australian Albers (EPSG: 3577) coordinate system and aggregated from a 1 km to 5 km raster grid to aid visualisation.

ESTIMATING ABIOTIC SUITABILITY

The potential distribution of the oriental fruit fly has been approximated using a variety of techniques, including CLIMEX (Stephens, Kriticos, and Leriche 2007), Maxent, and GARP (De Meyer et al. 2010). While some broad similarities exist (e.g., Central Africa and northern parts of South America are highly suitable) substantial differences also exist among model predictions. Acknowledging the inherent uncertainty in both model and covariate selection, we opted to approximate the geographic distribution of suitable climate for the oriental fruit fly using a method known as range bagging (Drake 2015). Range bagging estimates the environmental limits of a species' habitat by calculating convex hulls in environmental space, around environmental conditions at known occurrence locations. This process is then repeated using random subsets, of user-defined size, of both occurrence records as well as available environmental covariates (e.g. mean annual rainfall, mean annual temperature, etc.). Suitability is then defined for each raster cell as the proportion of replicates for which
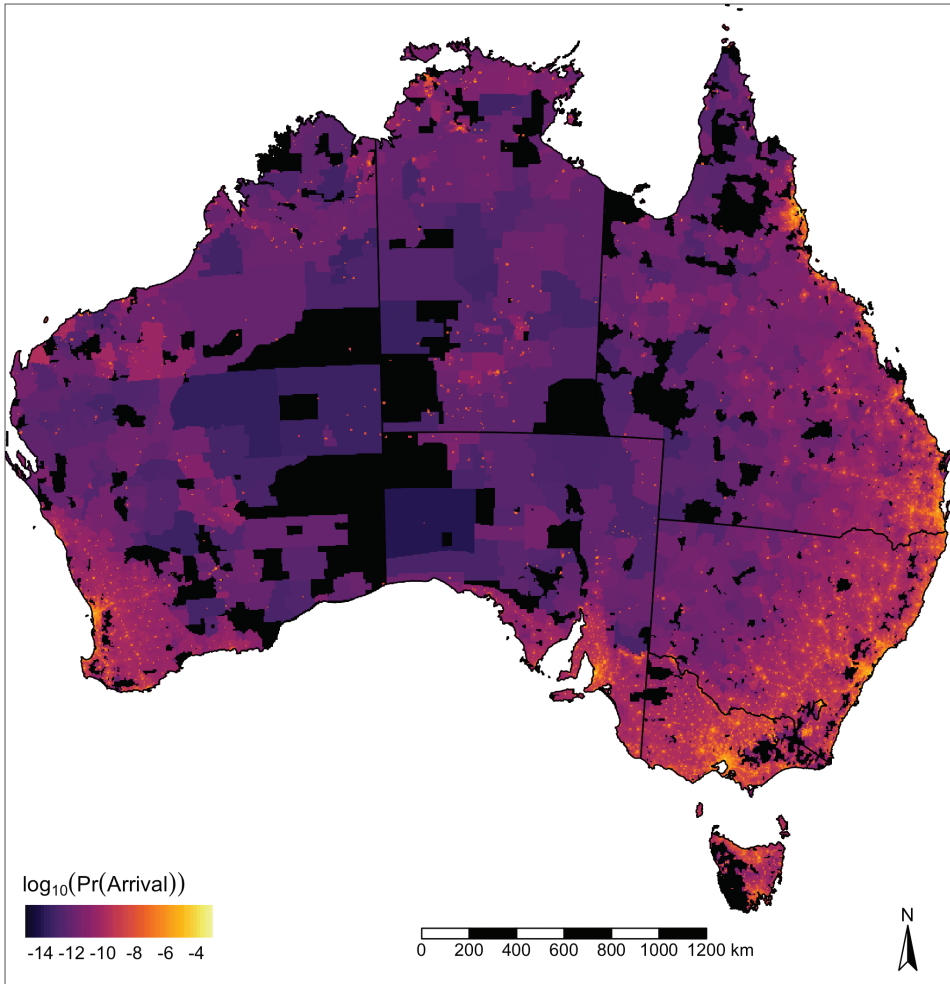
FIGURE 14.6   The ($\log_{10}$) likelihood that one or more viable oriental fruit fly contaminations arrive across the four hypothetical pathways. Warmer colours signify higher arrival likelihoods. Raster data were projected to the Australian Albers (EPSG: 3577) coordinate system and aggregated from a 1 km to a 5 km raster grid to aid visualisation.

the location is labelled as suitable (i.e. falls within the hull). For example, a suitability score of 0.1 would indicate that 10% of the estimated convex hulls deemed that particular location climatically suitable.

The approach has seen recent application to invasion biology and appears promising in the context of biosecurity because no absences or background data are required (Camac, Baumgartner, Garms, et al. 2021, Camac, Baumgartner, Hester, et al. 2021, Camac, Baumgartner, et al. 2020, Camac, Dodd, et al. 2020). This in turn removes several subjective decisions required in the modelling process. The method may also reduce inaccuracies that can arise from projecting to novel environmental conditions (i.e. environmental conditions outside the model-fitting data). This is because, unlike some other methods (e.g. Maxent, GLM, GAM), the method does not attempt to estimate response curves, but rather defines convex hull boundaries in environmental space based on known occurrences. Another major advantage of range bagging is that climatic suitability scores are comparable among species. Finally, range bagging also allows one to explicitly account for uncertainty in covariate selection by specifying low dimensionality (e.g. two dimensions) and allowing the algorithm to randomly select from among a suite of possible covariates—effectively resulting in an ensemble of potentially hundreds or thousands of competing models' parameterisations.

Map   251

Here, we used the range bagging algorithm with dimensionality set to two covariates used in each replicate, the number of bootstrapped models set to 100, and the proportion of occurrence records used per model set at 0.5. We used ensembles of "simple" two-dimensional models to minimise biases associated with model over-fitting and collinearity, and thus, maximise the model's transferability into novel environments (Camac, Baumgartner, Hester, et al. 2021, Camac, Baumgartner, et al. 2020). Ensembles of small models, each with only two variables, have been shown to frequently outperform standard SDM methods (Breiner et al. 2018). We allowed the algorithm to sample from all 19 WorldClim 2 (Fick and Hijmans 2017) bioclimatic parameters (i.e. BIO01 to BIO19) derived from the published 10 arc-minute (approximately 15 km resolution) raster layers.

To parameterise the range bagging models, we obtained global occurrence records for oriental fruit fly from the Global Biodiversity Information Facility (GBIF 2023) and supplemented these data with those collated by Hill, Gallardo, and Terblanche (2017). Acknowledging that such biological databases are susceptible to data quality issues (see Table 14.1), we cleaned these data using routines in the recently published CoordinateCleaner R package (Zizka et al. 2019).

We also removed duplicate records and thinned occurrence records to one point per 15 km (the resolution of the WorldClim version 2 climate data). Following this, we removed all records that occurred in countries lacking known established and persistent populations, based on CABI expert-curated country-level distributional data.

Once all cleaning and cross-referencing was complete, the data were used with the range bagging method to approximate the geographic distribution of suitable average climate (Figures 14.7 and 14.8).

ESTIMATING BIOTIC SUITABILITY

To approximate the geographic distribution of suitable biotic environment within Australia, we used the Australian Land Use and Management Classification raster (ALUM version 8; ABARES 2019) to define the presence and absence of fruit fly host material. Specifically, we created a binary raster whereby any 50 m^2 grid cell containing commodities vulnerable to fruit fly (e.g. tree fruits, vine fruits, olives, citrus, vegetables, herbs, and shrub berries), as well as urban
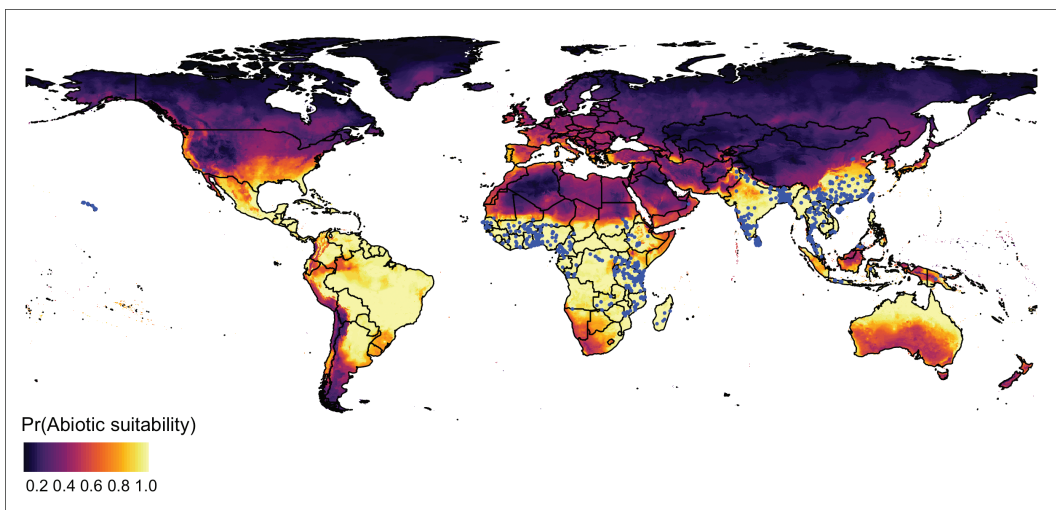

FIGURE 14.7 Estimated global climate suitability for the oriental fruit fly at the 10 arc-minute (approx. 15 km) scale. Suitability is the proportion of ensembled convex hulls derived from range bagging that identify a location as climatically suitable across bootstrapped combinations of environmental variables. Blue dots signify occurrence records used in range bagging parameterisation. Warmer colours signify higher likelihoods of climatic suitability.
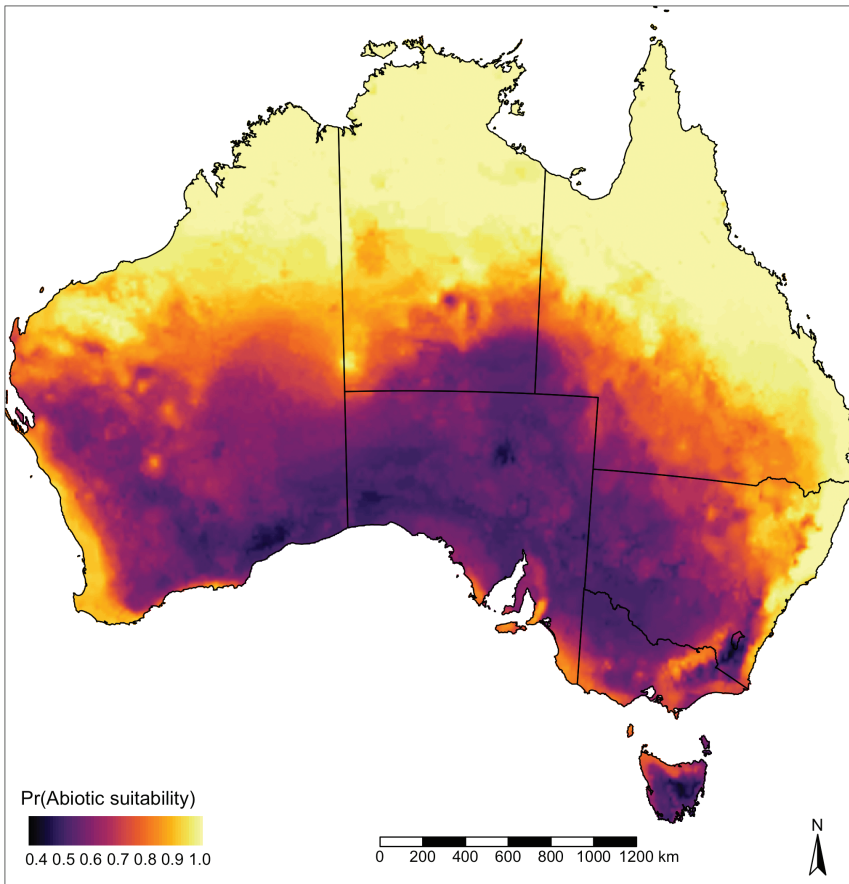
FIGURE 14.8   Estimated climate suitability in Australia for the oriental fruit fly. Suitability is the proportion of ensembled convex hulls derived from range bagging that identify a location as climatically suitable across bootstrapped combinations of environmental variables. Warmer colours signify higher climatic suitability. Raster data were projected to the Australian Albers (EPSG: 3577) coordinate system and resampled to a 5 km raster grid to aid visualisation.

areas, was assumed to contain host material and assigned a value of 1. All grid cells lacking such land uses assumed to contain no fruit fly host material and were classified as 0. This binary raster was converted to a 1 km grid resolution by determining whether any of the aggregated grid cells contained host material.

The Normalised Difference Vegetation Index (NDVI) describes the amount of live green vegetation and therefore can be a good indicator of the amount of suitable habitat or food source. We combined our binary host presence layer with normalised NDVI (scaled to the 0–1 range) by calculating the product of the two layers, thereby scaling host presence according to cell vegetation greenness (Figure 14.9). In effect, this means locations with no host material have a biotic suitability score of zero, and those with host material have a score conditioned on the amount of greenness at that location (our crude measure of host abundance).

DERIVING MAPS OF ESTABLISHMENT POTENTIAL

Now that we have produced three maps that encompass our belief about the probabilities of viable post-border arrivals and the suitability of the biotic and abiotic environments, we can estimate
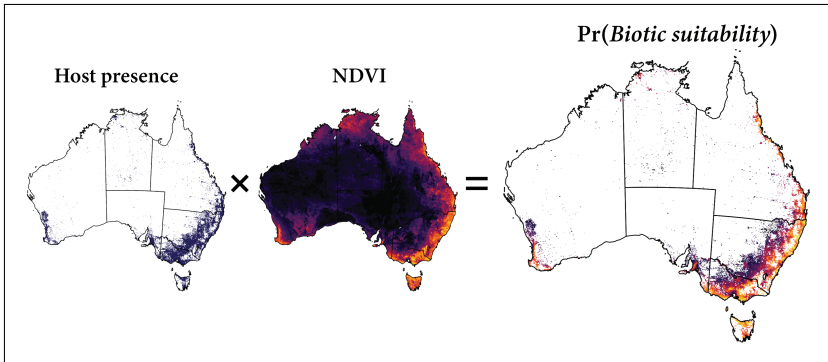
Map 253

FIGURE 14.9 Biotic suitability was calculated as the product of binary host presence (present vs. absent; blue indicates presence) and NDVI (an indicator of vegetation greenness; warm colours denote higher NDVI).

establishment potential (see Figure 14.1) by taking the product of the three maps, as shown in Figure 14.10.

The result of this product (of the three geographic barriers) is our pragmatic estimate of the likelihood one or more establishment events could occur at a given location in Australia (Figure 14.10). This map can then be used to inform biosecurity decision making.
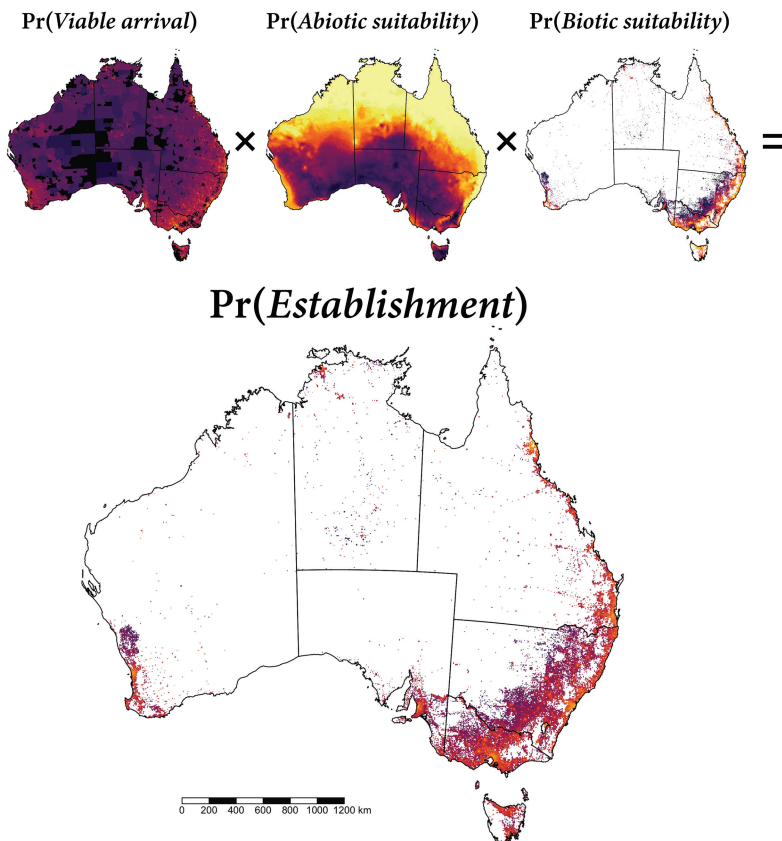

FIGURE 14.10 The product of the three geographic barriers to oriental fruit fly establishment can be used to approximate establishment potential (likelihood of establishment).

IN A NUTSHELL

- Invasive species' establishment potential is a function of three fundamental factors: (1) how likely it is to arrive at a location; (2) whether the climatic environment is suitable; and (3) whether suitable habitat or hosts exist. If there is no chance of arrival at a location, or if condition 2 and/or 3 are unmet, establishment cannot occur.
- Estimating establishment potential is critical for informing post-border surveillance, estimating likelihoods of threat absence, simulating incursion, and spread scenarios, and threat prioritisation.
- Pathway analyses estimate likelihoods of arrival at different points of entry (e.g. ports) and SDMs are often used to predict an invasive species' potential distribution.
- Practical and pragmatic approaches can be used to integrate outputs of SDMs and pathway models by linking them with commonly available spatial datasets and assumptions about post-border movements.
- Despite the limited availability of post-border movement data, opportunities exist to use existing data to approximate the post-border dispersal of threats.

NOTES

1 SDMs are also commonly known as habitat suitability models, bioclimatic envelope models, and ecological niche models, among other names.
2 It is important to note that how well a model predicts the known distribution of a species may not be representative of the model's ability to predict a species' potential (i.e. unobserved) distribution.
3 The conditional probability that one or more viable leakage events occur across pathways is estimated as 1 minus the product of the probability that no viable contaminations occur in each pathway. Mathematically this is written as: $1 - ((1 - 0.01) \times (1 - 0.005) \times (1 - 0.001) \times (1 - 0.05)) = 0.0651$.

REFERENCES

ABARES. 2019. "Catchment scale land use of Australia - Update December 2018." accessed 9 November 2022. https://doi.org/10.25814/5c7728700fd2a.

ABARES. 2020. Climatch v2.0 User Manual, edited by Australian Bureau of Agricultural and Resource Economics and Sciences. Canberra.

Araújo, M. B., and A. Rozenfeld. 2014. "The Geographic Scaling of Biotic Interactions." *Ecography* 37 (5):406–415. https://doi.org/10.1111/j.1600-0587.2013.00643.x.

Australian Bureau of Statistics. 2017. "2074.0 - census of population and housing: Mesh block counts, Australia, 2016." https://www.abs.gov.au/AUSSTATS/abs@.nsf/Lookup/2074.0Main+Features12016.

Barry, S., J. Elith, D. Heersink, P. Caley, M. Kearney, P. Tenant, and A. Arthur. 2015. Final report: CEBRA 1402B Tools and approaches for invasive species distribution modelling for surveillance. Tech. rep., CSIRO: Centre of Excellence for Biosecurity Risk Analysis, The University of Melbourne.

Bennett, J. M., P. Calosi, S. Clusella-Trullas, B. Martínez, J. Sunday, A. C. Algar, M. B. Araújo, B. A. Hawkins, S. Keith, I. Kühn, C. Rahbek, L. Rodríguez, A. Singer, F. Villalobos, M. Ángel Olalla-Tárraga, and I. Morales-Castilla. 2018. "GlobTherm, a Global Database on Thermal Tolerances for Aquatic and Terrestrial Organisms." *Scientific Data* 5 (1):180022. https://doi.org/10.1038/sdata.2018.22.

Biosecurity Commons. 2022. "Biosecurity Commons." https://www.biosecuritycommons.org.au/.

Booth, T. H., H. A. Nix, J. R. Busby, and M. F. Hutchinson. 2014. "Bioclim: The First Species Distribution Modelling Package, Its Early Applications and Relevance to Most Current MaxEnt Studies." *Diversity and Distributions* 20 (1):1–9. https://doi.org/10.1111/ddi.12144.

Boyle, B. L., B. S. Maitner, G. G. C. Barbosa, R. K. Sajja, X. Feng, C. Merow, E. A. Newman, D. S. Park, P. R. Roehrdanz, and B. J. Enquist. 2022. "Geographic Name Resolution Service: A Tool for the Standardization and Indexing of World Political Division Names, With Applications to Species Distribution Modeling." *bioRxiv*:2022.04.25.489424. https://doi.org/10.1101/2022.04.25.489424.

Bradhurst, R., D. Spring, M. Stanaway, J. Milner, and T. Kompas. 2021. "A Generalised and Scalable Framework for Modelling Incursions, Surveillance and Control of Plant and Environmental Pests." *Environmental Modelling & Software* 139:105004. https://doi.org/10.1016/j.envsoft.2021.105004.

Map   255

Breiner, F. T., A. Guisan, A. Bergamini, and M. P. Nobis. 2015. "Overcoming Limitations of Modelling Rare Species by Using Ensembles of Small Models." *Methods in Ecology and Evolution* 6 (10):1210–1218. https://doi.org/10.1111/2041-210X.12403.

Breiner, F. T., M. P. Nobis, A. Bergamini, and A. Guisan. 2018. "Optimizing Ensembles of Small Models for Predicting the Distribution of Species With Few Occurrences." *Methods in Ecology and Evolution* 9 (4):802–808. https://doi.org/10.1111/2041-210X.12957.

Briscoe, N. J., J. Elith, R. Salguero-Gómez, J. J. Lahoz-Monfort, J. S. Camac, K. M. Giljohann, M. H. Holden, B. A. Hradsky, M. R. Kearney, S. M. McMahon, B. L. Phillips, T. J. Regan, J. R. Rhodes, P. A. Vesk, B. A. Wintle, J. D. L. Yen, and G. Guillera-Arroita. 2019. "Forecasting Species Range Dynamics With Process-Explicit Models: Matching Methods to Applications." *Ecology Letters* 22 (11):1940–1956. https://doi.org/10.1111/ele.13348.

Broennimann, O., and A. Guisan. 2008. "Predicting Current and Future Biological Invasions: Both Native and Invaded Ranges Matter." *Biology Letters* 4 (5):585–589.: https://doi.org/10.1098/rsbl.2008.0254.

Brown, C. F., S. P. Brumby, B. Guzder-Williams, T. Birch, S. B. Hyde, J. Mazzariello, W. Czerwinski, V. J. Pasquarella, R. Haertel, S. Ilyushchenko, K. Schwehr, M. Weisse, F. Stolle, C. Hanson, O. Guinan, R. Moore, and A. M. Tait. 2022. "Dynamic World, Near Real-Time Global 10 m Land Use Land Cover Mapping." *Scientific Data* 9 (1):251.https://doi.org/10.1038/s41597-022-01307-4.

CABI. 2023. "CABI compendium invasive species." accessed 2 June 2023. https://www.cabidigitallibrary. org/product/qi

Camac, J. S., J. B. Baumgartner, B. Garms, A. Robinson, and T. Kompas. 2021. Estimating trading partner exposure risk to new pests or diseases. Technical Report for CEBRA project 190606. Centre of Excellence for Biosecurity Risk Analysis, The University of Melbourne.

Camac, J. S., J. B. Baumgartner, S. Hester, R. Subasinghe, and S. Collins. 2021. Using edmaps & Zonation to inform multi-pest early-detection surveillance designs. Technical Report for CEBRA project 20121001: Centre of Excellence for Biosecurity Risk Analysis, The University of Melbourne.

Camac, J. S., J. B. Baumgartner, A. Robinson, and J. Elith. 2020. *Developing Pragmatic Maps of Establishment Likelihood for Plant Pests.* Centre of Excellence for Biosecurity Risk Analysis, The University of Melbourne.

Camac, J. S., A. Dodd, N. Bloomfield, and A. Robinson. 2020. *Sampling to Support Claims of Area Freedom: Technical Report for the Department of Agriculture.* Centre of Excellence for Biosecurity Risk Analysis, The University of Melbourne.

Catford, J. A., R. Jansson, and C. Nilsson. 2009. "Reducing Redundancy in Invasion Ecology by Integrating Hypotheses into a Single Theoretical Framework." *Diversity and Distributions* 15 (1):22–40.https://doi. org/10.1111/j.1472-4642.2008.00521.x.

Chamberlain, S., and E. Szöcs. 2013. "Taxize: Taxonomic Search and Retrieval in R [version 2; Peer Review: 3 Approved]." *F1000Research* 2 (191). https://doi.org/10.12688/f1000research.2-191.v2.

CHELSA. 2023. "Climatologies at high resolution for the earth's land surface areas." accessed 2 June 2023. http://chelsa-climate.org.

Climatch. 2019. "Climatch." accessed 2 June 2023. https://climatch.cp1.agriculture.gov.au/.

CliMond. 2014. "CliMond: global climatologies for bioclimatic modelling." accessed 2 June 2023. https:// www.climond.org.

CRU. 2023. "Climatic Research Unit." accessed 2 June 2023. https://www.uea.ac.uk/groups-and-centres/ climatic-research-unit.

De Meyer, M., M. P. Robertson, M. W. Mansell, S. Ekesi, K. Tsuruta, W. Mwaiko, J. F. Vayssières, and A. T. Peterson. 2010. "Ecological Niche and Potential Geographic Distribution of the Invasive Fruit Fly *Bactrocera Invadens* (Diptera, Tephritidae)." *Bulletin of Entomological Research* 100 (1):35–48. https:// doi.org/10.1017/S0007485309006713.

Dodd, A., J. B. Baumgartner, and T. Kompas. 2021. *Final Report: Stakeholder Perceptions of and Options for Refining the CEBRA 'Value' Model for Use Within DAWE, CEBRA Project 20100401.* Centre of Excellence for Biosecurity Risk Analysis, The University of Melbourne.

Dodd, A. J., M. A. McCarthy, N. Ainsworth, and M. A. Burgman. 2016. "Identifying Hotspots of Alien Plant Naturalisation in Australia: Approaches and Predictions." *Biological Invasions* 18 (3):631–645. https:// doi.org/10.1007/s10530-015-1035-8.

Dodd, A., N. Stoeckl, J. B. Baumgartner, and T. Kompas. 2020. *Key Result Summary: Valuing Australia's Biosecurity System, CEBRA Project 170713.* Centre of Excellence for Biosecurity Risk Analysis, The University of Melbourne.

Drake, J. M. 2015. "Range Bagging: A New Method for Ecological Niche Modelling from Presence-Only Data." *Journal of the Royal Society Interface* 12 (107):20150086. https://doi.org/10.1098/rsif.2015.0086.

Elith, J. 2017. "Predicting Distributions of Invasive Species." In *Invasive Species: Risk Assessment and Management*, edited by Andrew P. Robinson, Mark A. Burgman, Mike Nunn and Terry Walshe, 93–129. Cambridge: Cambridge University Press.

Elith, J., M. Kearney, and S. Phillips. 2010. "The Art of Modelling Range-Shifting Species." *Methods in Ecology and Evolution* 1 (4):330–342. https://doi.org/10.1111/j.2041-210X.2010.00036.x.

Elith, J., and J. R. Leathwick. 2009. "Species Distribution Models: Ecological Explanation and Prediction Across Space and Time." *Annual Review of Ecology, Evolution, and Systematics* 40 (1):677–697. https://doi.org/10.1146/annurev.ecolsys.110308.120159.

Fick, S. E., and R. J. Hijmans. 2017. "WorldClim 2: New 1-Km Spatial Resolution Climate Surfaces for Global Land Areas." *International Journal of Climatology* 37 (12):4302–4315. https://doi.org/10.1002/joc.5086.

Fourcade, Y., A. G. Besnard, and J. Secondi. 2018. "Paintings Predict the Distribution of Species, or the Challenge of Selecting Environmental Predictors and Evaluation Statistics." *Global Ecology and Biogeography* 27 (2):245–256. https://doi.org/10.1111/geb.12684.

GBIF. 2023. "Global Biodiversity Information Facility." accessed 2 June 2023. https://www.gbif.org.

Guillera-Arroita, G., J. J. Lahoz-Monfort, J. Elith, A. Gordon, H. Kujala, P. E. Lentini, M. A. McCarthy, R. Tingley, and B. A. Wintle. 2015. "Is My Species Distribution Model Fit for Purpose? Matching Data and Models to Applications." *Global Ecology and Biogeography* 24 (3):276–292. https://doi.org/10.1111/geb.12268.

Harris, I., P. D. Jones, T. J. Osborn, and D. H. Lister. 2014. "Updated High-Resolution Grids of Monthly Climatic Observations – the CRU TS3.10 Dataset." *International Journal of Climatology* 34 (3):623–642. https://doi.org/10.1002/joc.3711.

Hemming, V., M. A. Burgman, A. M. Hanea, M. F. McBride, and B. C. Wintle. 2018. "A Practical Guide to Structured Expert Elicitation Using the IDEA Protocol." *Methods in Ecology and Evolution* 9 (1):169–180. https://doi.org/10.1111/2041-210X.12857.

Hill, M., P. Caley, J. Camac, J. Elith, and S. Barry. 2022. "A novel method accounting for predictor uncertainty and model transferability of invasive species distribution models." *bioRxiv*:2022.03.14.483865. https://doi.org/10.1101/2022.03.14.483865.

Hill, M. P., B. Gallardo, and J. S. Terblanche. 2017. "A Global Assessment of Climatic Niche Shifts and Human Influence in Insect Invasions." *Global Ecology and Biogeography* 26 (6):679–689. https://doi.org/10.1111/geb.12578.

Hulme, P. E. 2021. "Unwelcome Exchange: International Trade as a Direct and Indirect Driver of Biological Invasions Worldwide." *One Earth* 4 (5):666–679. https://doi.org/10.1016/j.oneear.2021.04.015.

Karger, D. N., O. Conrad, J. Böhner, T. Kawohl, H. Kreft, R. W. Soria-Auza, N. E. Zimmermann, H. P. Linder, and M. Kessler. 2017. "Climatologies at High Resolution for the earth's Land Surface Areas." *Scientific Data* 4 (1):170122. https://doi.org/10.1038/sdata.2017.122.

Kearney, M., B. L. Phillips, C. R. Tracy, K. A. Christian, G. Betts, and W. P. Porter. 2008. "Modelling Species Distributions Without Using Species Distributions: the Cane Toad in Australia Under Current and Future Climates." *Ecography* 31 (4):423–434. https://doi.org/10.1111/j.0906-7590.2008.05457.x.

Kearney, M., and W. Porter. 2009. "Mechanistic Niche Modelling: Combining Physiological and Spatial Data to Predict species' Ranges." *Ecology Letters* 12 (4):334–350. https://doi.org/10.1111/j.1461-0248.2008.01277.x.

Kearney, M. R., and W. P. Porter. 2017. "NicheMapR – an R Package for Biophysical Modelling: The Microclimate Model." *Ecography* 40 (5):664–674. https://doi.org/10.1111/ecog.02360.

Kriticos, D. J., J. M. Kean, C. B. Phillips, S. D. Senay, H. Acosta, and T. Haye. 2017. "The Potential Global Distribution of the Brown Marmorated Stink Bug, *Halyomorpha Halys*, a Critical Threat to Plant Biosecurity." *Journal of Pest Science* 90 (4):1033–1043. https://doi.org/10.1007/s10340-017-0869-5.

Kriticos, D. J., G. Maywald, T. Yonow, E. Zurcher, N. Herrmann, and B. Sutherst. 2016. *CLIMEX Version 4: Exploring the Effects of Climate on Plants, Animals and Diseases*. Canberra: CSIRO.

Kriticos, D. J., B. L. Webber, A. Leriche, N. Ota, I. Macadam, J. Bathols, and J. K. Scott. 2012. "CliMond: Global High-Resolution Historical and Future Scenario Climate Surfaces for Bioclimatic Modelling." *Methods in Ecology and Evolution* 3 (1):53–64. https://doi.org/10.1111/j.2041-210X.2011.00134.x.

Lahoz-Monfort, J. J., G. Guillera-Arroita, and B. A. Wintle. 2014. "Imperfect Detection Impacts the Performance of Species Distribution Models." *Global Ecology and Biogeography* 23 (4):504–515. https://doi.org/10.1111/geb.12138.

Le Maitre, D. C., W. Thuiller, and L. Schonegevel. 2008. "Developing an Approach to Defining the Potential Distributions of Invasive Plant Species: A Case Study of Hakea Species in South Africa." *Global Ecology and Biogeography* 17 (5):569–584. https://doi.org/10.1111/j.1466-8238.2008.00407.x.

Map   257

Lobo, J. M., A. Jiménez-Valverde, and J. Hortal. 2010. "The Uncertain Nature of Absences and Their Importance in Species Distribution Modelling." *Ecography* 33 (1):103–114. https://doi.org/10.1111/j.1600-0587.2009.06039.x.

Magarey, R. D., G. A. Fowler, D. M. Borchert, T. B. Sutton, M. Colunga-Garcia, and J. A. Simpson. 2007. "NAPPFAST: An Internet System for the Weather-Based Mapping of Plant Pathogens." *Plant Disease* 91 (4):336–345. https://doi.org/10.1094/PDIS-91-4-0336.

Maldonado, C., C. I. Molina, A. Zizka, C. Persson, C. M. Taylor, J. Albán, E. Chilquillo, N. Rønsted, and A. Antonelli. 2015. "Estimating Species Diversity and Distribution in the Era of Big Data: To What Extent Can We Trust Public Databases?" *Global Ecology and Biogeography* 24 (8):973–984. https://doi.org/10.1111/geb.12326.

Marques, G. M., S. Augustine, K. Lika, L. Pecquerie, T. Domingos, and A. L. M. Kooijman. 2018. "The AmP Project: Comparing Species on the Basis of Dynamic Energy Budget Parameters." *PLOS Computational Biology* 14 (5):e1006100. https://doi.org/10.1371/journal.pcbi.1006100.

McGeoch, M. A., P. Genovesi, P. J. Bellingham, M. J. Costello, C. McGrannachan, and A. Sheppard. 2016. "Prioritizing Species, Pathways, and Sites to Achieve Conservation Targets for Biological Invasion." *Biological Invasions* 18 (2):299–314. https://doi.org/10.1007/s10530-015-1013-1.

Merow, C., M. J. Smith, and J. A. Silander Jr. 2013. "A Practical Guide to MaxEnt for Modeling species' Distributions: What It Does, and Why Inputs and Settings Matter." *Ecography* 36 (10):1058–1069. https://doi.org/10.1111/j.1600-0587.2013.07872.x.

Morales-Castilla, I., T. J. Davies, W. D. Pearse, and P. Peres-Neto. 2017. "Combining Phylogeny and Co-Occurrence to Improve Single Species Distribution Models." *Global Ecology and Biogeography* 26 (6):740–752. https://doi.org/10.1111/geb.12580.

Morin, X., and M. J. Lechowicz. 2008. "Contemporary Perspectives on the Niche That can Improve Models of Species Range Shifts Under Climate Change." *Biology Letters* 4 (5):573–576. https://doi.org/10.1098/rsbl.2008.0181.

Phillips, S. J., R. P. Anderson, and R. E. Schapire. 2006. "Maximum Entropy Modeling of Species Geographic Distributions." *Ecological Modelling* 190 (3):231–259. https://doi.org/10.1016/j.ecolmodel.2005.03.026.

Phillips, S. J., M. Dudík, J. Elith, C. H. Graham, A. Lehmann, J. Leathwick, and S. Ferrier. 2009. "Sample Selection Bias and Presence-Only Distribution Models: Implications for Background and pseudo-Absence Data." *Ecological Applications* 19 (1):181–197. https://doi.org/10.1890/07-2153.1.

Pollock, L. J., W. K. Morris, and P. A. Vesk. 2012. "The Role of Functional Traits in Species Distributions Revealed Through a Hierarchical Model." *Ecography* 35 (8):716–725. https://doi.org/10.1111/j.1600-0587.2011.07085.x.

Renner, I. W., J. Elith, A. Baddeley, W. Fithian, T. Hastie, S. J. Phillips, G. Popovic, and D. I. Warton. 2015. "Point Process Models for Presence-Only Analysis." *Methods in Ecology and Evolution* 6 (4):366–379. https://doi.org/10.1111/2041-210X.12352.

Salguero-Gómez, R., O. R. Jones, C. R. Archer, Y. M. Buckley, J. Che-Castaldo, H. Caswell, D. Hodgson, A. Scheuerlein, D. A. Conde, E. Brinks, H. de Buhr, C. Farack, F. Gottschalk, A. Hartmann, A. Henning, G. Hoppe, G. Römer, J. Runge, T. Ruoff, J. Wille, S. Zeh, R. Davison, D. Vieregg, A. Baudisch, R. Altwegg, F. Colchero, M. Dong, H. de Kroon, J.-D. Lebreton, C. J. E. Metcalf, M. M. Neel, I. M. Parker, T. Takada, T. Valverde, L. A. Vélez-Espino, G. M. Wardle, M. Franco, and J. W. Vaupel. 2015. "The COMPADRE Plant Matrix Database: An Open Online Repository for Plant Demography." *Journal of Ecology* 103 (1):202–218. https://doi.org/10.1111/1365-2745.12334.

Salguero-Gómez, R., O. R. Jones, C. R. Archer, C. Bein, H. de Buhr, C. Farack, F. Gottschalk, A. Hartmann, A. Henning, G. Hoppe, G. Römer, T. Ruoff, V. Sommer, J. Wille, J. Voigt, S. Zeh, D. Vieregg, Y. M. Buckley, J. Che-Castaldo, D. Hodgson, A. Scheuerlein, H. Caswell, and J. W. Vaupel. 2016. "COMADRE: A Global Data Base of Animal Demography." *Journal of Animal Ecology* 85 (2):371–384. https://doi.org/10.1111/1365-2656.12482.

Sequeira, A. M. M., P. J. Bouchet, K. L. Yates, K. Mengersen, and M. J. Caley. 2018. "Transferring Biodiversity Models for Conservation: Opportunities and Challenges." *Methods in Ecology and Evolution* 9 (5):1250–1264. https://doi.org/10.1111/2041-210X.12998.

Stephens, A. E. A., D. J. Kriticos, and A. Leriche. 2007. "The Current and Future Potential Geographical Distribution of the Oriental Fruit Fly, *Bactrocera dorsalis* (Diptera: Tephritidae)." *Bulletin of Entomological Research* 97 (4):369–378. https://doi.org/10.1017/S0007485307005044.

Syfert, M. M., M. J. Smith, and D. A. Coomes. 2013. "The Effects of Sampling Bias and Model Complexity on the Predictive Performance of MaxEnt Species Distribution Models." *PLOS ONE* 8 (2):e55158. https://doi.org/10.1371/journal.pone.0055158.

Tingley, R., P. García-Díaz, C. R. R. Arantes, and P. Cassey. 2018. "Integrating Transport Pressure Data and Species Distribution Models to Estimate Invasion Risk for Alien Stowaways." *Ecography* 41 (4):635–646. https://doi.org/10.1111/ecog.02841.

Uribe-Rivera, D. E., G. Guillera-Arroita, S. M. Windecker, P. Pliscoff, and B. A. Wintle. 2022. "The Predictive Performance of Process-Explicit Range Change Models Remains Largely Untested." *Ecography*:e06048. https://doi.org/10.1111/ecog.06048.

VanDerWal, J., L. P. Shoo, C. Graham, and S. E. Williams. 2009. "Selecting pseudo-Absence Data for Presence-Only Distribution Modeling: How Far Should You Stray from What You Know?" *Ecological Modelling* 220 (4):589–594. https://doi.org/10.1016/j.ecolmodel.2008.11.010.

Vargas, R. I., J. C. Piñero, and L. Leblanc. 2015. "An Overview of Pest Species of *Bactrocera* Fruit Flies (*Diptera: Tephritidae*) and the Integration of Biopesticides With Other Biological Approaches for Their Management With a Focus on the Pacific Region." *Insects* 6 (2):297–318. https://doi.org/10.3390/insects6020297.

Venette, R. C., D. J. Kriticos, R. D. Magarey, F. H. Koch, R. H. A. Baker, S. P. Worner, N. N. Gómez Raboteaux, D. W. McKenney, E. J. Dobesberger, D. Yemshanov, P. J. De Barro, W. D. Hutchison, G. Fowler, T. M. Kalaris, and J. Pedlar. 2010. "Pest Risk Maps for Invasive Alien Species: A Roadmap for Improvement." *BioScience* 60 (5):349–362. https://doi.org/10.1525/bio.2010.60.5.5.

Warton, D. I., I. W. Renner, and D. Ramp. 2013. "Model-Based Control of Observer Bias for the Analysis of Presence-Only Data in Ecology." *PLOS ONE* 8 (11):e79168. https://doi.org/10.1371/journal.pone.0079168.

Warton, D. I., and L. C. Shepherd. 2010. "Poisson Point Process Models Solve the "pseudo-Absence Problem" for Presence-Only Data in Ecology." *The Annals of Applied Statistics* 4 (3):1383–1402. https://doi.org/10.1214/10-AOAS331.

WorldClim. 2023. "WorldClim." accessed 2 June 2023. http://www.worldclim.org.

Yates, K. L., P. J. Bouchet, M. J. Caley, K. Mengersen, C. F. Randin, S. Parnell, A. H. Fielding, A. J. Bamford, S. Ban, A. M. Barbosa, C. F. Dormann, J. Elith, C. B. Embling, G. N. Ervin, R. Fisher, S. Gould, R. F. Graf, E. J. Gregr, P. N. Halpin, R. K. Heikkinen, S. Heinänen, A. R. Jones, P. K. Krishnakumar, V. Lauria, H. Lozano-Montes, L. Mannocci, C. Mellin, M. B. Mesgaran, E. Moreno-Amat, S. Mormede, E. Novaczek, S. Oppel, G. Ortuño Crespo, A. T. Peterson, G. Rapacciuolo, J. J. Roberts, R. E. Ross, K. L. Scales, D. Schoeman, P. Snelgrove, G. Sundblad, W. Thuiller, L. G. Torres, H. Verbruggen, L. Wang, S. Wenger, M. J. Whittingham, Y. Zharikov, D. Zurell, and A. M. M. Sequeira. 2018. "Outstanding Challenges in the Transferability of Ecological Models." *Trends in Ecology & Evolution* 33 (10):790–802. https://doi.org/10.1016/j.tree.2018.08.001.

Zizka, A., D. Silvestro, T. Andermann, J. Azevedo, C. Duarte Ritter, D. Edler, H. Farooq, A. Herdean, M. Ariza, R. Scharn, S. Svantesson, N. Wengström, V. Zizka, and A. Antonelli. 2019. "CoordinateCleaner: Standardized Cleaning of Occurrence Records from Biological Collection Databases." *Methods in Ecology and Evolution* 10 (5):744–751. https://doi.org/10.1111/2041-210X.13152.