Developing pragmatic maps of establishment likelihood for plant pests

Technical Report for CEBRA project 170607

James Camac¹, John Baumgartner¹, Andrew Robinson¹, Jane Elith¹, and

contributions from: Dr Ranjith Subasinghe² Dr Mark Stanaway² Dr Susie Collins² Dr Anaís Gibert¹ Dr Matthew Hill³ Dr Peter Caley³ Dr Simon Barry³

¹The Centre of Excellence for Biosecurity Risk Analysis, The University of Melbourne ²Department of Agriculture, Water and the Environment ³Data61, CSIRO

November 3, 2020





How to cite:

Camac, JS., Baumgartner, J., Robinson, A., Elith J (2020) Developing pragmatic maps of establishment likelihood for plant pests. Technical Report for CEBRA project 170607.

Contents

1	Executive summary 1			
2	Introduction			
3	Decision 3.1 Guio 3.1.1 3.1.2	tree for invasive species distribution modelling de for invasive species distribution modelling 1. Step 1: Data Sourcing 3.1.1.1 Pest biology 3.1.1.2 Occurrence data 3.1.1.3 Climatic data 3.1.1.5 Alternatives for data-poor species 3.1.2.1 Physiological data 3.1.2.2 Occurrence data	5 9 11 12 13 13 13 13 14 14 14	
	3.1.3 3.1.4 3.1.5	3Step 3: Model Development3.1.3.1Process-based models3.1.3.2Correlative models3.1.3.3Model settings & complexity3.1.3.4Sampling background & dealing with survey bias3.1.3.5Covariate uncertainty4Step 4: Model Validation5Step 5: Model Utility	17 18 20 22 24 24 27	
4	Framewo	ork for mapping establishment likelihoods for plant pests	32	
	4.1 Ider 4.2 Path 4.2.1 4.2.2 4.2.2 4.2.3 4.2.4 4.2.5 4.2.6 4.2.7 4.2.8 4.3 Exp	httifying high risk pathways iways iways iways iways Pathway: Air passengers 4.2.1.1 International tourists 4.2.1.2 Returning residents 4.2.1.3 Torres Strait visitors (Fruit fly specific) 2 Pathway: Mail 3 Pathway: Imported fertiliser 4 Pathway: International Containers 5 Pathway: Vessels 6 9	34 35 35 35 41 43 45 45 45 48 50 51 52 52 53	
	4.4 Abie 4.5 Biot	otic suitability	53 54	



	4.6	Estimating establishment likelihoods	56	
5	Cas 5.1 5.2 5.3 5.4 5.5	e study 1: Oriental fruit fly Background	57 58 59 65 69	
6	Cas 6.1 6.2 6.3 6.4 6.5	e study 2: Khapra beetle Background	 80 81 81 83 83 86 	
7	Cas 7.1 7.2 7.3 7.4 7.5	e study 3: Brown marmorated stink bug (BMSB) Background	96 97 97 99 106 109	
8	Cas 8.1 8.2 8.3 8.4 8.5	e study 4: Gypsy moth Background High risk pathways Abiotic suitability Biotic suitability Establishment likelihoods	120 121 121 123 128 133	
9	Info 9.1 9.2 9.3 9.4 9.5	rming biosecurity decisionsUtility 1: Informing where to conduct surveillance for early detectionUtility 2: Surveillance coverage relative to total establishment likelihoodUtility 3: Informing likelihoods of area freedom	145 146 147 149 152 153	
10	Futu 10.1 10.2 10.3 10.4 10.5 10.6	Jare research needs and extensionsMajor review of RRRACollecting empirical data for pathway post-border movementsAccounting for uncertaintyAggregating mapsPredicting future establishment likelihoodsVirtual analytics platform	154 154 156 156 157 157	
Bi	bliog	raphy	158	
A	A Spatial heterogeneity in pest risk assessments: a review of risk maps 1			
в	Wor	kshop	212	

С	Deal	ing wi	th uncertainty in predictor selection	228
D	edma	aps: al	ר R package for creating maps for early detection	259
	D.1	Softwa	are requirements	259
		D.1.1	Minimum system requirements	259
			D.1.1.1 Operating system specific requirements	260
		D.1.2	Required <i>R</i> packages	261
		D.1.3	Data directory	261
	D.2	Param	etrising edmaps	264
	D.3	Runni	ng edmaps	268
		D.3.1	Using edmaps with Docker (recommended)	269
		D.3.2	Using edmaps with renv	270
		D.3.3	Using edmaps by itself	270
Ε	Refe	rence	manual for edmaps R package	271

List of Figures

3.1	Guide for developing scientifically robust species distribution models for invasive pests. For further details please see text in sections 3.1.1 to	
3.2	3.1.5. Distribution and intensity of Global Biodiversity Information Facility	10
	(GBIF) occurrence records. Lighter colours equal higher intensity. Sourced	10
3.3	An output from the web interface Climate Matcher.	12 26
4.1	Polynomial relationship between the number of accommodation providers	
	and rooms contained in statistical area level 2 categories	37
4.2	Geographic distribution of tourist rooms across Australia.	38
4.3	Top) Distance from major Australian international airports; Bottom) Dis-	10
4 4	The approaching distribution of human provide tion a group Acatualia	40
4.4 4.5	Ten Distance from Cairne International Airport: Bottom) Distance weight	42
4.5	score	44
4.6	Log10 proportion of fertiliser tonnage deposited per grid cell	47
4.7	Geographic distribution of secondary ALUMC classifications.	55
5.1	A female oriental fruit fly, <i>Bactrocera dorsalis</i> , laying eggs by inserting her	
E 2	OVIPOSITOR IN the SKIN of a papaya.	57
5.2	of suitable climate for oriental fruit fly (<i>Bactrocera dorsalis</i>) or its synonym	
	Bactrocera invadens.	60
5.3	Global climatic suitability for oriental fruit fly estimated from range bag-	00
	ging	63
5.4	Australian projected climatic suitability for oriental fruit fly derived from	
	range bagging	64
5.5	Average Normalised Differential Vegetation Index (NDVI) observed be-	
	tween October 2018 and March 2019	67
5.6	Distribution of suitable biotic environment for oriental fruit fly.	68
5.7	Estimated environmental suitability for oriental fruit fly.	70
5.8	National establishment likelihood map for oriental fruit fly.	71
5.9 5.10	Oriental fruit fly establishment likelihood map for greater Cairns.	72
5.10	Oriental fruit fly establishment likelihood map for greater brisbane	73
5.11	Oriental fruit fly establishment likelihood map for greater Melbourne	74
5.13	Oriental fruit fly establishment likelihood map for greater Hobart	76
5.14	Oriental fruit fly establishment likelihood map for greater Adelaide.	77
5.15	Oriental fruit fly establishment likelihood map for greater Perth	78
5.16	Oriental fruit fly establishment likelihood map for greater Darwin	79



$\begin{array}{c} 6.1 \\ 6.2 \\ 6.3 \\ 6.4 \\ 6.5 \\ 6.6 \\ 6.7 \\ 6.8 \\ 6.9 \\ 6.10 \\ 6.11 \\ 6.12 \end{array}$	An adult khapra beetle (<i>Trogoderma granarium</i>) Log10 vessel pathway distance decay weights for khapra beetle Distribution of suitable biotic environment for khapra beetle National establishment likelihood map for khapra beetle	80 82 85 87 88 89 90 91 92 93 94 95
7.1 7.2 7.3	Adult brown marmorated stink bug (<i>Halyomorpha halys</i>) Published brown marmorated stink bug (<i>Halyomorpha halys</i>) CLIMEX (Kriticos <i>et al.</i> , 2017) and Maxent (Zhu <i>et al.</i> , 2012) distribution models Brown marmorated stink bug (<i>Halyomorpha halys</i>) Random forest, Maxent and Support Vector Machine distribution models published in Fraser	96 100
7.4	<i>et al.</i> (2017)	101
7.5	Australian climatic suitability for brown marmorated stink bug derived from range bagging	104
7.6	Distribution of suitable biotic environment for brown marmorated stink	103
7.7 7.8 7.9	Estimated environmental suitability for brown marmorated stink bug. National establishment likelihood map for brown marmorated stink bug. Brown marmorated stink bug establishment likelihood map for greater	110 111 112
7.10	Brown marmorated stink bug likelihood of establishment map for greater Brisbane.	112
7.11	Brown marmorated stink bug establishment likelihood map for greater Sydney.	114
7.12	Brown marmorated stink bug establishment likelihood map for greater Melbourne.	115
7.13	Brown marmorated stink bug establishment likelihood map for greater Hobart.	116
7.14	Brown marmorated stink bug establishment likelihood map for greater Adelaide.	117
7.15	Brown marmorated stink bug establishment likelihood map for greater Perth	118
7.16	Brown marmorated stink bug establishment likelihood map for greater Darwin.	119
8.1 8.2 8.3	Female Asian gypsy moth (<i>Lymantria dispar asiatica</i>) Log10 vessel pathway distance decay weights for gypsy moth Published gypsy moth CLIMEX (Paini <i>et al.</i> , 2018) and GARP ensemble	120 122
8.4	(Peterson <i>et al.</i> , 2007) distribution models	125 126



8.5 8.6 8.7	Australian climatic suitability for gypsy moth derived from range bagging. 127 Distribution of suitable biotic environment for gypsy moth
8.8	National establishment likelihood map for gypsy moth
8.9	Gypsy moth establishment likelihood map for greater Cairns 137
8.10	Gypsy moth establishment likelihood map for greater Brisbane 138
8.11	Gypsy moth establishment likelihood map for greater Sydney 139
8.12	Gypsy moth establishment likelihood map for greater Melbourne 140
8.13	Gypsy moth establishment likelihood map for greater Hobart 141
8.14	Gypsy moth establishment likelihood map for greater Adelaide 142
8.15	Gypsy moth establishment likelihood map for greater Perth
8.16	Gypsy moth establishment likelihood map for greater Darwin 144
9.1	The relationship between the proportion of national establishment like- lihood captured and the number of 5 km grid cells surveyed for each of
	the four case study pests
9.2	Scenario tree outlining the possible ways an absence can arise when
	surveillance sensitivity and specificity are imperfect
9.3	The Australian Plant Pest and Disease model (APPDIS) simulating the spread of tramp ant near Cairns
10.1	An example of an Australian Incoming Passenger Card

List of Tables

3.1 3.2	Common errors in biological databases and occurrence records Examples of presence-background SDM usage in the management of invasive species	16 28
 4.1 4.2 4.3 4.4 4.5 	Number of international arrivals between January 2018 and January 2019. Major Australian international airports	 35 39 46 49 51
5.1 5.2	Oriental fruit fly pathway likelihoods with associated expected number of post-border entries per annum based on estimated pathway volumes. ALUMC landuses likely to contain fruit fly host material.	59 66
6.1 6.2	Khapra beetle pathway likelihoods with associated expected number of post-border entries per annum based on estimated pathway volumes ALUMC landuses likely to contain khapra beetle host material	81 84
7.1 7.2	Brown marmorated stink bug pathway likelihoods with associated expected number of entries per annum based on estimated pathway volumes	98 107
8.1 8.2 8.3	Gypsy moth pathway likelihoods with associated expected number of entries per annum based on estimated pathway volumes	121 128 130

1. Executive summary

The federal Department of Agriculture, Water and the Environment (DAWE) and associated state biosecurity agencies play a fundamental role in surveying for the early detection of high impact invasive plant pests. This surveillance is inherently labour intensive and costly to maintain. As such, efficient allocation of increasingly finite surveillance resources across all risk areas presents a significant challenge for biosecurity decision-makers. Compounding the issue of prioritising which pest species to target in early detection surveillance, little to no information is available about where, when and how a new pest species is likely to arrive and establish in Australia. In order to determine where surveillance resources should be allocated for early detection, it is imperative we have an understanding of how the risk of pest establishment varies across space.

This project sought to develop a general and pragmatic framework for creating defensible maps of pest establishment likelihoods for the purposes of informing postborder surveillance for early detection of pests not currently present in Australia. Specifically, the project aimed to utilise border pathway leakage likelihoods derived from DAWE's pathway model, the Risk-Return Resource Allocation model (RRRA); combine these with various models describing pathway post-border spread; and weight expected arrival numbers by biotic (e.g. host abundance) and abiotic (e.g. climate) suitability for the pest/disease.

In this report, we outline the four primary outputs:

- 1. A guide designed to aid practitioners in developing robust and transparent species distribution models for use in biosecurity decision-making (See Chapter 3);
- A framework for creating maps of pest establishment likelihood for the purposes of informing early detection, incorporating 1) RRRA pathway leakage likelihoods, 2) methods for distributing risk associated with 10 high risk pathways, and 3) spatially-explicit measures of biotic and abiotic suitability for a pest (See Chapter 4);
- 3. An *R* package, edmaps, for implementing the above framework (Appendix D); and
- 4. Four case studies utilising the establishment likelihood mapping framework:
 - a) Oriental fruit fly (Bactrocera dorsalis; Chapter 5);
 - b) Khapra beetle (Trogoderma granarium; Chapter 6);
 - c) Brown marmorated stink bug (Halyomorpha halys; Chapter 7); and
 - d) Gypsy moth (*Lymantria dispar*; Chapter 8).

In delivering the above outputs, the project also:

- 1. outlines several applications in which these maps may be used (Chapter 9);
- 2. outlines several areas warranting future research (Chapter 10);
- 3. conducted a formal literature review examining how pest risk maps are commonly developed and the types of models and data that are used (Appendix A);
- 4. conducted a workshop (Appendix B) attended by 21 participants from various government agencies, academic institutions and private enterprises in order to elicit:
 - a) how biosecurity practitioners currently estimate the potential distribution of pests and diseases that are not currently present in a country; and
 - b) what were the major areas of uncertainty they faced when developing and using such models to inform biosecurity policy and management.
- 5. conducted an analysis examining the utility of model ensembles for dealing with predictor uncertainty in invasive species distribution models (Appendix C), and;
- 6. developed a web tool for estimating areas where correlative species distribution models are extrapolating, and are thus unlikely to be reliable (Climate Matcher).

Examination of the four case studies revealed that areas of highest establishment likelihood were generally in and around metropolitan areas. This is because these regions are expected to receive high propagule pressure from most pathways, and for the four case studies at least, are thought to contain adequate host material and suitable climate.

2. Introduction

Pest risk maps are a fundamental decision support tool used in biosecurity. Depending on how they are built, they can describe where an invasive species may arrive, establish, spread or cause significant economic, environmental and societal harm. They are used to inform strategic and tactical decisions about where to allocate finite surveillance resources, international and domestic trade interventions, and cost-sharing between jurisdictions (Venette *et al.*, 2010; Camac *et al.*, 2018). Pest risk maps can also be used to inform likelihoods of pest absence (Camac *et al.*, 2019) and are a critical communication tool used to raise public awareness, and thus, the efficacy of general surveillance (i.e. reporting from farmers and the general public). Pest risk maps are such a vital component of an effective and efficient biosecurity system, that now, many governments and industries are heavily investing into some form of risk mapping (Elith, 2017).

In the ideal case, pest risk maps should describe 1) the likelihood a species will successfully establish in an area, and 2) the associated magnitude of harm if an incursion occurred in that area (Venette *et al.*, 2010). However, in practice, pest risk maps rarely incorporate both likelihoods and consequences (Venette *et al.*, 2010), but rather focus on estimating likelihoods associated with a particular stage of the invasion process such as the likelihood that a pest may arrive at a port (i.e. pathway analyses), or the likelihood an area is environmentally suitable for pest establishment (Venette *et al.*, 2010; Elith, 2017). As such, most pest risk maps do not explicitly measure risk (i.e. likelihood maps still provide crucial information for informing where to conduct border screening and post-border surveillance for early detection.

Two approaches are commonly used to inform decisions about where to allocate early warning surveillance for pests that are believed to have not yet arrived in a country. The first involves conducting detailed pathway analyses that may use interception data coupled with pathway-specific trade volumes (e.g. number of imported containers) and information about which countries the pest is known to have established in (Yemshanov *et al.*, 2011; Douma *et al.*, 2016; Tingley *et al.*, 2018). The second, more common approach, is to construct models, typically referred to as species distribution models (SDMs) or niche models, using physiological data (Kearney *et al.*, 2008), global occurrence data (Zhu *et al.*, 2012; Phillips *et al.*, 2017), or a combination of the two (Kriticos *et al.*, 2017). These models generally use global climatic data (e.g. WorldClim, CliMond) as predictor inputs, and predict where climatic conditions are suitable for pest survival and establishment. These two approaches – pathways and SDMs – are characterising different things, though both can potentially contribute useful information for surveillance.

Interestingly, there have been few attempts to integrate pathway analyses with estimates of a pest's environmental suitability (See Appendix A for a detailed review). This is despite both pieces of information being critical in determining whether a species can successfully establish at a given location. Moreover, studies that do at-



tempt to link pathway likelihoods with estimates of environmental suitability are often geographically restricted to points of entry (e.g. ports), and as such, do not make inferences about relative risk of establishment beyond these points of entry (e.g. Tingley *et al.*, 2018). In part, this is not surprising, because using statistical approaches to estimate likelihoods of arrival across large geographic areas would require substantial computing power and detailed information about the likely secondary movements of pests once they arrive within a country – data that for a variety of reasons (e.g. privacy, limited resources) is often not recorded, or only recorded for a fraction of the area of interest.

Project purpose

DAWE has tasked the Centre of Excellence for Biosecurity Risk Analysis (CEBRA) to develop a transparent and pragmatic (given data limitations) approach that can be readily applied to priority plant pests for estimating national maps of plant pest establishment likelihood. Fundamentally, establishment likelihood maps for a given pest or disease must encompass:

- 1. its ability to arrive at a given location;
- 2. climatic suitability at mapped locations; and
- 3. the presence of hosts/vectors at mapped locations.

In developing this framework, the department requested that, in order to maximise 1) the model's applicability to most priority plant pests; and 2) its use the department, the proposed framework should use information that is readily available for most pests and incorporate data which is either collected or derived from existing departmental systems (e.g. Risk Return Resource Allocation model). Furthermore, given the large number of subjective decisions made throughout the modelling process, the department also requested general guidance in estimating climate suitability maps for invasive species.

In this final report of the two-year project, we outline a pragmatic framework for creating maps of pest establishment likelihood as a function of arrival rates, climate suitability and the presence of host material. In doing so, we provide general guidance in estimating an invasive species' climate suitability – from obtaining and cleaning data, to model development, model validation and model use in decision making. We then illustrate the model mapping framework by developing maps of establishment likelihood for four plant pests identified on the National Priority Plant Pest (NPPP) list: oriental fruit fly (*Bactrocera dorsalis*), khapra beetle (*Trogoderma granarium*), brown marmorated stink bug (*Halyomorpha halys*) and gypsy moth (*Lymantria dispar*). We selected these four pests as they all posed significant threat to Australian agriculture and/or environment and were exemplars of different pathways of entry, different host materials and differing potential climatic distributions.

What this report does not do

This report does not provide guidance on how to prioritise pests and diseases, how to optimally allocate surveillance, or how to account for climate change.

3. Decision tree for invasive species distribution modelling

Species Distribution Models (SDMs) for invasive species are often considered synonymous with pest risk maps (Venette *et al.*, 2010; Elith, 2017). The primary aim of such models is to relate environmental conditions to the persistence of a species. The estimated relationships from these models are then projected onto geographic space in order to produce maps of environmental suitability for a particular species or group of species. The degree of environmental suitability and the extent of suitable environment in a region of interest are then used to inform a variety of biosecurity decisions such as:

- 1. which pests/diseases should be prioritised;
- 2. where surveillance programs should be conducted;
- 3. the likelihood of pest absence; and
- 4. which government agencies should incur monitoring and eradication costs.

The utility of SDMs to inform such important biosecurity decisions has led to many government agencies and industries actively investing in some form of pest risk mapping. A consequence of such investment has been the spawning of initiatives such as the European Union's Pratique risk assessment program and government in-house development of risk mapping platforms such as the USA's NAPPFAST (Magarey *et al.*, 2007) and Australian's climate matching system (CLIMATE/CLIMATCH; Crombie *et al.* 2008). However, a wealth of additional tools also exists in scientific literature for estimating the potential distribution of a species. These methods range from those that model the underlying ecophysiological or biophysical processes governing a species' distribution (Kearney *et al.*, 2008; Higgins & Richardson, 2014; Kearney & Porter, 2017), to those that correlate environmental variables to patterns in species occurrences, and those that incorporate both mechanistic and correlative processes.

Despite the vast number of tools available, and the diversity of opinions on how to use them, there remains no clear evidence of a single, best approach for predicting an invasive species' potential distribution (Barry *et al.*, 2015). There are good arguments for thinking that there would not exist a best method for all species and all places: the ecology and biogeography of species' distributions is so diverse and so complex, that it is highly likely that different models will have different strengths. Even so, a fundamental problem is that to determine which approach performs consistently well requires an understanding of models' ability to predict the potential invasive distribution of a species, which for all practical purposes is unknowable (Elith, 2017). In order to get around this, modellers have mostly evaluated methods based on their predictive performance when predicting to current distribution data (Sequeira *et al.*, 2018; Yates *et al.*, 2018). However, these performance metrics do not explicitly examine the performance with respect to the true purpose of these models – their ability to predict the potential, currently unoccupied, distribution of a species. Accurate prediction outside this range (Elith, 2017; Fourcade *et al.*, 2017). In order to examine a model's transferability from one region to another, various cross-validation approaches have been used, such as parametrising a model based on native occurrences and then using it to predict to the observed invasive region (Yates *et al.*, 2018). However, this too is 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 *et al.*, 2010; Elith, 2017). As such, as with all modelling, whilst decisions about the correct method to use can be well informed by existing knowledge, there remains ambiguity in the final choice of model to predict an unobserved event. This chapter aims to provide the framework for making a well-informed decision about the method to use.

At the broadest level a distinction exists between methods that explicitly model process, and those that fit statistical (also known as correlative) models to the data – the patterns - that result from those processes. In theory, process-based models are expected to be better than pattern-based methods at predicting the potential distribution of a species (Dormann, 2007; Kearney & Porter, 2009; Elith et al., 2010). This is because these models attempt to explicitly model the dominant processes governing survival, growth and fecundity and how these relate to various environmental data. Relative to correlative approaches, if these models are parametrised correctly, they should be better able to estimate a species' fundamental niche - the distribution governed by abiotic factors such as climate - and thus, the species' potential distribution. While these process-based models are increasingly becoming an area of active research, they are still considerably less used to predict potential distributions of invasive species relative to correlative approaches (Elith, 2017). This is predominantly due to these models: 1) being considerably more complex to build and run; 2) requiring more time to collate appropriate physiological data; and 3) having to deal with difficulties of matching microclimate or laboratory conditions associated with physiological responses to broadscale climatic variables that are commonly available for mapping (Kearney & Porter, 2009; Yates et al., 2018).

By contrast, correlative models are generally faster and easier to fit relative to processbased models. Instead of explicitly modelling the processes governing a species' potential distribution (e.g. growth, mortality and fecundity), they work by deriving likelihood functions by correlating occurrence records (and absences if available) with environmental data using regression and machine learning techniques. These likelihood functions are then used to create suitability maps within the trained region as well as in novel space and/or time. If parametrised correctly, the predictive accuracy of correlative models is typically high in the region in which they were trained (Elith, 2017). In fact, comparisons between process-based and correlative models often show a high degree of congruence of predictions in the regions where correlative models are trained. However, a range of outcomes (from congruence to dissimilarity) are often observed when comparing predictions outside the trained region of correlative models (Morin & Lechowicz, 2008; Buckley *et al.*, 2010; Kearney *et al.*, 2010). This is because the ability of correlative models to accurately predict outside the trained region (termed transferability) is thought to be highly contingent on the degree to which predictions are made under novel environmental conditions that are not present within the trained region (termed extrapolation; Dormann 2007; Elith & Leathwick 2009; Sequeira *et al.* 2018; Yates *et al.* 2018).

Both process-based and correlative approaches come with a variety of technical and practical pros and cons - most of which have been exhaustively examined elsewhere (e.g. Kearney & Porter, 2009; Froese, 2012; Elith, 2017). However, irrespective of the approach taken, a wealth of subjective decisions and assumptions must be made at each step of the modelling process, from data sourcing and cleaning, to model development, model validation and model utility. These decisions impact predictions, so it is important to think them through logically and to document what decisions were made, and why. Predictions will never be perfect - indeed, given the paucity of data on some species, predictions may not be accurate at all. However a well thought out and documented process will produce predictions whose provenance is known, and will also provide good intuition about how to quantify uncertainties and test sensitivities. Ultimately, it will be these decisions and assumptions that govern whether scientists, stakeholders and government agencies can use the model and its predictions, especially when applied to invasive species where the true potential distribution is unknowable. This is well understood within the academic literature and has led to considerable discussion on how 'best' to source and clean data (Zizka et al., 2019), select background points (in the case of correlative presence-background models; Phillips et al. 2009; VanDerWal et al. 2009), choose parameters or predictors (Austin, 2002; Breiner et al., 2017; Fourcade et al., 2017), validate models (Elith et al., 2010; Sequeira et al., 2018; Yates et al., 2018), and deal with model uncertainty (Elith et al., 2010; Yates et al., 2018). Unfortunately, little consensus currently exists in the literature as to how these various issues should be handled. As a case in point, when developing correlative distribution models for invasive species, occurrence data used to fit the model are often restricted to those from the 'native' range (Thuiller et al., 2005; Elith, 2017), where the species is more likely to be in equilibrium with its environment. However, others have used or advocated for using the only the invaded range (Mau-Crimmins et al., 2006; Barbet-Massin et al., 2018), or in some cases, occurrence records from both the native and invaded range (Broennimann & Guisan, 2008; Jiménez-Valverde et al., 2011). This general lack of consensus within the academic literature, coupled with often poorly described methods and justifications, has led to inconsistent workflows being undertaken when developing these models. Consequently, a wide variety of possible potential distributions can be predicted for a single species, even when using the same modelling technique.

Species distribution models are an integral component of our risk map framework, which is outlined later in this report. As such, we first aim to improve the transparency, reproducibility and consistency of workflows used to create species distribution models for informing policy and management of invasive species. Specifically, we aimed to create a generic, method-agnostic (because there is no single 'best' method; Barry *et al.* 2015) decision tree that will guide users in developing models for invasive species, but also prompt users to explicitly consider and document the various decisions made at each step of the modelling process.

In the proceeding sections we describe a decision tree for invasive species distribution modelling. This decision tree was informed by practical and technical knowledge derived from a workshop attended by biosecurity practitioners and SDM experts, as well as analyses conducted by CEBRA and CSIRO using real and simulated species



cebra

Centr



3.1. Guide for invasive species distribution modelling

The guide presented in this report is designed to aid practitioners in developing robust and transparent species distribution models for use in biosecurity decision-making. It should be consulted irrespective of whether one is developing a new model or reproducing (or updating) an existing model. The decision tree is broken up into five sequential steps: data sourcing; data cleaning; model development; model validation; and model utility (Fig 3.1). Each major step contains a number of component steps that must be carefully considered, because the consequences ultimately affect model predictions, interpretation and how the model is used to inform policy and management. Since some steps then feed into later steps, the following descriptions gradually develop a suite of relevant information for a modeller.

Document what you have done!

In the following steps it is imperative that all gathered information and decisions made are fully documented. This ensures transparency in how data was sourced and used and allows experts to fully assess the merits of any analysis or model outputs.





Figure 3.1.: Guide for developing scientifically robust species distribution models for invasive pests. For further details please see text in sections 3.1.1 to 3.1.5.



3.1.1. Step 1: Data Sourcing

3.1.1.1. Pest biology

When considering whether to build or reproduce a species distribution model, the very first step should be to conduct a short literature review of grey (national and international governmental reports) and published literature that outlines what is known about the pest or disease. Specifically, the literature review should summarise details associated with: physiological climate tolerances, existing SDMs, required hosts, and any distributional data reported in the literature¹.

-`ģ́-Tip

A good place to start a literature review is to examine information available on CABI (https: //www.cabi.org/isc/). CABI contains a large collection of pest data sheets compiled by experts. These data sheets provide information on pest biology, distribution, spread pathways and many other factors. They also contain comprehensive reference lists indicating where information was sourced.

Lt is important to consult more than just CABI.

While CABI is generally kept up-to-date, it does not always contain the latest biological or model-derived information and also only contains coarse distributional data (typically country level). Moreover, it rarely contains specifics as to how these data were derived. It is therefore critical that the underlying data sources provided by CABI are consulted and supplemented by independent searches of the scientific literature (e.g. via Google Scholar or Web of Science).

¹Grey or published literature can sometimes provide detailed occurrence data that are not available in large online databases such as GBIF.



3.1.1.2. Occurrence data

Conducting a literature review is fundamental to a firm understanding of the biological constraints on a pest or disease, yet it is also important to source data on its current distribution. The best models are based on presence-absence or abundance data, but since these are rarely available globally, here we simply focus on "presence-only" data (i.e. records of species presences, with no explicit information on absences). One such source that should be consulted is the Global Biodiversity Information Facility (GBIF; https://www.gbif.org/). GBIF is the most comprehensive global biological database, with over 1 billion occurrence records for more than one million species (Fig 3.2). The data contained within GBIF is an invaluable source for both parametrising correlative SDMs, and validating process-based models.

-`ģ́-Tip

Data from GBIF can be directly sourced via its web interface, or imported directly into R using the R package **rgbif**. A nice tutorial of how to use rgbif can be found at: https://ropensci.org/tutorials/rgbif_tutorial/.



Figure 3.2.: Distribution and intensity of Global Biodiversity Information Facility (GBIF) occurrence records. Lighter colours equal higher intensity. Sourced from https://www.gbif.org on 09/01/2019.

Pest-specific databases (e.g. fruit fly database for the Afro-tropical region) are another invaluable source of occurrence information. These databases are typically compiled and verified by experts, and thus, are 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, and if available, they are consulted to determine whether distributional data is present.



3.1.1.3. Climatic data

Irrespective of whether one is reproducing a model or developing a new one, sourcing global climate data is essential. Common climate databases include:

- WorldClim (Fick & Hijmans 2017; http://www.worldclim.org)
- Climate Research Unit (Harris et al. 2014; http://www.cru.uea.ac.uk),
- CliMond (Kriticos et al. 2012; https://www.climond.org); and
- CHELSA (Karger et al. 2017; http://chelsa-climate.org).

Each dataset contains different suites of long-term averaged climatic variables. Each database 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 database) that are likely to be most useful for the pest/disease of interest. Sometimes the most relevant climatic variables might not be long-term averages but rather variables related to weather or extremes.

3.1.1.4. Non-climatic data

While climate is considered to be the predominant factor governing species distributions at large spatial scales (Araújo & Rozenfeld, 2014), other variables can also play an important role in governing a species' distribution (e.g. predator-prey interactions, landuse and disturbance patterns). However, such information is generally not available at either the spatial extent or resolution required to parametrise or predict invasive SDMs. Where such information does exist (e.g. landuse layers), it is often based on image classification algorithms from satellite imagery. As a consequence, such data are often coarse (either in terms of spatial resolution or categorisation), and thus, may have limited utility in species distribution modelling.

A summary of some freely available global landuse datasets can be found at https: //geography.com/free-global-land-cover-land-use-data/. Over time, the quantity and quality of these other datasets will undoubtedly improve and become more readily accessible. We therefore recommend users continue to be on the lookout for new large-scale spatial datasets that become available, and which could be used to estimate or refine SDMs.

3.1.1.5. Alternatives for data-poor species

In some situations there is limited biological or distributional data for a pest or disease (e.g. fewer than 30 records worldwide, or only records from a very restricted locality). In these cases, one has three options: 1) collate data on a species with similar phylogeny or similar biological, behavioural or physiological traits (Morales-Castilla *et al.*, 2017); 2) use a trait-based SDM (e.g. Morin & Lechowicz, 2008; Pollock *et al.*, 2012); or 3) use an expert elicitation framework (e.g. IDEA protocol; Hemming *et al.*, 2017) to estimate relevant physiological tolerances or responses to various climatic factors (e.g. Martin *et al.*, 2015).

3.1.2. Step 2: Data Cleaning

Irrespective of whether a process-based or correlative approach is used, data cleaning is a must when developing robust models of the potential distributions of pests or diseases.

3.1.2.1. Physiological data

Physiological data that are incorporated into process-based SDMs are generally derived directly from controlled laboratory experiments found in the peer-reviewed literature and academic theses. Commonly, these data are reported as means and standard errors and are presented in either tables or text. Such data are often considered definitive with minimal errors or data cleaning required. However, such experimental data suffer from two primary concerns. The first is whether the experimental conditions are transferable to the natural world, and the second is whether the individuals used are representative of the true variability within a species. Both questions are difficult to answer, especially the latter, where the true intra-specific variability is almost never known. However, 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 in regards to both the representativeness and transferability of experimental findings².

It is also important to record the uncertainty around any experiment-derived physiological responses. This is because such errors allow for the faithful propagation of uncertainty throughout the modelling process. Ultimately, this manifests itself in a range of possible distribution maps, ranging from best to worse case scenario.

Lastly, in light of the biological information that should be obtained from the literature review (stage 1), it is important to document physiological processes (e.g. factors governing reproductive maturity or diapause) that lack empirical estimates. Documenting these knowledge gaps highlights areas of future research. Moreover, it provides greater context to the decisions made within the model development stage in handling these unknown parameters.

²Ideally, this information should have been documented as part of stage 1 (Data sourcing).



-`ģ́-Tip

Several demographic and physiological databases have recently been published in an effort to simplify parametrisation of process-based models. These include:

GlobTherm; (Bennett et al., 2018): A database that includes thermal tolerance metrics for 2,133 species of multicellular algae, plants, fungi, and animals in 43 classes, 203 orders and 525 families from marine, intertidal, freshwater, and terrestrial realms, extracted from published studies.

AmP; (*Marques* et al., 2018): A database of referenced data on animal energetics, parameter values of models based on Dynamic Energy Budget (DEB) theory, and properties derived from these parameters. Contains records for over 1000 animal species.

COMPADRE; (*Salguero Gómez* et al., 2015): A database of more than 5,500 plant population matrix models derived from over 450 studies and more than 600 species worldwide. Also contains relevant ancillary information (e.g. ecoregion, growth form, taxonomy, phylogeny) that facilitates interpretation of the numerous demographic metrics that can be derived from population matrices.

COMADRE; (*Salguero Gómez* et al., 2016): A database of more than 1625 animal population matrix models derived from over 400 and more than 340 species worldwide. Also contains ancillary information (e.g. ecoregion, taxonomy, biogeography, etc.) that facilitates interpretation of the numerous demographic metrics that can be derived from population matrices.



3.1.2.2. Occurrence data

Species occurrence records derived from online databases (e.g. GBIF) are an invaluable resource for both parametrising and validating SDMs. However, issues with data quality such as errors in geo-referencing, taxonomy, temporal period, irrelevant records (e.g. cultivated specimens) and incomplete metadata can introduce significant biases in model fitting and validation procedures (Maldonado *et al.*, 2015). As such, careful scrutiny of these data is a must! Special care must also be taken for migratory pests, which may contain records in regions where persistence is not year-round.

Here, the decision tree prompts the user to methodically search their data for errors and irrelevant records commonly found within biological databases. A summary of common errors can be found in Table 3.1.

Geo-referencing errors	Taxonomic errors	Temporal errors	Irrelevant records
1. Incorrect habitat	1. Misidentifications;	1. Incorrect dates or	1. Records from zoos,
(Marine or Terrestrial)	2. Obsolete or incorrect	dates not supplied;	botanical gardens
2. Coordinates & country	y taxonomy;	2. Records not within	& museums;
mismatches;	3. Incorrect spelling.	climate temporal	2. Fossil records;
3. Switched or equal		period.	3. Migratory records;
latitude & longitude;			3. Biosecurity interceptions;
4. Records on country			4. Records from unreliable
or province centroids;			sources (e.g. iNaturalist).
5. Records based on			
rasterized collections;			
6. Records with high			
spatial uncertainty			
or strong decimal			
rounding;			
7. Duplicate records.			

Table 3.1.: Common errors in biological databases and occurrence records.

It is important to note that most GBIF records also contain details of the record's original source. This allows users to further scrutinize the reliability of the record.

Once data have been cleaned of the above errors, it is critical to remove records that are unlikely to have originated from established populations. This is because our interest is in delimiting the potential distribution of where a pest or disease could establish and subsequently spread. Unfortunately, such information is not always recorded in online metadata. As such, we recommend cross-referencing occurrence data against expert verified country statuses published in CABI (https://www.cabi.org/isc/). In the ideal case, were appropriate resources are available, all records should be scrutinized by experts in the relevant geographic area to verify whether they belong to an established population, and if not, removed prior to model development or validation.



-`ģ́-Tip

GBIF data are typically manually cleaned by poorly documented user-defined functions and procedures that ultimately affect the reproducibility of an analysis (*Zizka* et al., 2019).

Recently, there has been concerted effort to develop standardised tools for cleaning large-scale biodiversity databases, such as GBIF. One such tool is the R package CoordinateCleaner (Zizka et al., 2019). This package provides a standardized, reproducible and fast way to identify records from cultivated sources as well as identify many geographic and temporal errors commonly found in biological databases. A tutorial on how to use CoordinateCleaner can also be found at:

https://ropensci.github.io/CoordinateCleaner/articles/Cleaning_ GBIF_data_with_CoordinateCleaner.html.

Taxonomic errors such as spelling mistakes and use of synonyms or obsolete taxonomy can be identified and corrected using the R package taxize (Chamberlain et al., 2017). A tutorial for using taxize can be found at:

https://ropensci.org/tutorials/taxize_tutorial/.

3.1.3. Step 3: Model Development

While a vast number of tools exist, and a diversity of opinions on how to use them, there remains no strong evidence of a single *best* approach for predicting an invasive species' potential distribution (Barry *et al.*, 2015). Moreover, the choice in approach will ultimately depend on data availability and the model's intended application (see Section 3.1.5). It will also strongly depend on the available in-house technical expertise and various resource constraints (e.g. analyst time, access to commercial software). As a consequence, the decision tree presented in this report does not advocate for any particular method. Instead, it focuses on providing users with general guidance for both process-based and correlative methods.

-`<mark>`</mark>{ Tip

We highly recommend users consult Froese (2012). This report, prepared by the Queensland government, provides a comprehensive description of a range of SDM tools that can be used to inform biosecurity policy and management. Furthermore, it outlines the technical advantages and disadvantages of different approaches as well as their expected resource requirements (i.e. analyst time, skill and data requirements).

We also recommend users carefully consider how the model will be used and whether the chosen method is fit for purpose. *Guillera-Arroita* et al. (2015) provides a comprehensive summary of whether correlative models are appropriate for various biosecurity decision contexts. We summarise some of the main points of that paper in section 3.1.5.



3.1.3.1. Process-based models

Process-based models range in complexity from simple climatic threshold mapping to complicated biophysical models such as NicheMapper (Kearney & Porter, 2017). The commercial software CLIMEX (Kriticos et al., 2015) – a semi-mechanistic model – can also incorporate process when it is parametrised using experimental data. However, more commonly CLIMEX physiological parameters are inferred using global distributional data (Kriticos et al., 2015; Elith, 2017) and manual tweaking of stress parameters in the fitting process. 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 considered to be superior at predicting a species' fundamental distribution – the distribution governed by abiotic factors such as climate (Kearney & Porter, 2009; Elith, 2017). In practice, however, the reliability of these models is strongly dependent on whether it accounts for the fundamental processes delimiting a species' fundamental niche and whether the empirically derived parameters accurately reflect both the responses to climate as well as the intra-specific variability in these responses (Kearney & Porter, 2009). In most cases, the climatic factors governing a species' potential distribution are unknown or unknowable, and the resources required to 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.

If process-based models are built, we strongly recommend that irrespective of the method used, it is critical to account for parameter uncertainty (where possible) when using these models. This is because such uncertainty accounts not only for possible measurement errors in such experiments, but more importantly, will also provide an estimate of the intra-specific variability of a species, and thus, the range of possible climatic responses it may exhibit. This uncertainty can either be directly incorporated into the model or acknowledged via repeat model runs using parameter estimates at lower, upper and median quantiles. The outcome should be a range of climate suitability maps that encompass the entire plausible range of climatic responses.

3.1.3.2. Correlative models

There is a vast range of correlative methods that can be used to predict the potential distribution of a pest or disease. Broadly these methods fall into three distinct categories: 1) those based on presence-only data such as climatic envelopes (e.g. BIOCLIM, range bagging, convex hulls) and climate matching algorithms available in CLIMEX and CLIMATCH; 2) those that can be fitted to presence/background or presence/pseudoabsence data, such as machine learning algorithms and regression models (e.g. Maxent, Boosted Regression Trees, Generalised Linear Models, Generalised Additive Models); and 3) those based on presence/absence records such as binomial regression techniques. Determining which model type to use is ultimately dependent on data availability, assumptions the modeller is willing to make, and what objective or decision context the model is designed to solve (Guillera-Arroita *et al.*, 2015; Elith, 2017). However, irrespective of the approach, all three work by attempting to relate occurrences to spatially explicit covariates (Elith *et al.*, 2010).



Presence-only methods:

Presence-only methods commonly estimate a species' potential distribution in one of three ways: 1) by 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 indices (CLIMATCH, CLIMEX; Crombie *et al.* (2008); Kriticos *et al.* (2015)); or 3) by estimating convex hulls or hyper volumes; Blonder *et al.* (2017). These methods use only presence locations and make no comparison with the range of environments in the broader landscape or at absence sites. That is, they ignore the set of environmental conditions potentially available to a species within the region of interest (Elith, 2017). Climatic variables can be equally or unequally weighted or the most limiting variable can be used for prediction (e.g. BIOCLIM). Distance matrices such as the Gower metric or Mahalanobis distance can also be used to predict climatic similarity between records of occurrence and all unvisited sites. Range bagging (Drake, 2015) is a recent development in presence-only distribution modelling that appears to show great promise to invasive species distribution models (For more information see Appendix (C).

The primary appeal of these models is that they deal directly with the most common type of data available – presence-only records – without having to make additional assumptions and decisions associated with the choice of background/pseudo-absences found in other methods below. However, the primary disadvantage of these methods is that they can be highly sensitive to outliers (e.g. when delimiting climatic ranges) and sampling biases (i.e. some sites are more likely to be sampled than others) present within the occurrence records (Elith, 2017). The latter is particularly problematic because sample biases are commonplace in invasive species occurrences and these methods provide no direct means by which one can account for such biases. These methods also do not estimate probabilities of occurrence, since the prevalence is unknown (Elith, 2017).

Presence-background methods:

Presence-background methods derive suitability scores from relative frequencies of surveyed occurrences (positive cases) relative to another non-positive case. The meaning of this alternative case can vary in subtle but important ways. For some methods, such as Maxent (Phillips *et al.*, 2006), the non-positive case is commonly referred to as the background – a set of points that randomly sample the landscape or locations available, irrespective of whether a species is present or not. Increasing evidence shows that this approach is best for modelling presence-only data with logistic regression (Merow *et al.*, 2013; Elith, 2017).

Other methods or interpretations treat the non-positive cases as implied absences or pseudo-absences (e.g. GARP). Typically these implied absences are either placed anywhere except where presences occur, or 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 *et al.*, 2008; Lobo *et al.*, 2010). However, relative to the presence-background approach, this method requires additional ad hoc decisions about both the position and number of pseudo-absences, and is generally perceived as a less rigorous statistical framework (Renner *et al.*, 2015).



Irrespective of which approach is used, both assume a species is in equilibrium with their environment – that is, the species has had the opportunity to establish in all suitable areas within the model-fitting region. These methods also do not estimate probabilities of occurrence as this requires knowing the true prevalence of the modelled organism (Guillera-Arroita *et al.*, 2015). Rather, these models estimate relative probabilities that are assumed to be proportional to the true probabilities. The implication of this is that predictions within a map for one species are sensibly ranked, but predictions across species are not comparable – all you know for each species is that high is high and low is low, not what the exact numbers mean. Lastly, these approaches require subjective decisions about the extent, distribution and number of non-positive cases to be sampled. The impacts of these decisions are often underappreciated by species distribution modellers, despite evidence highlighting how different decisions can result in substantial effects on model predictions and interpretation (Phillips *et al.*, 2009; Warton & Shepherd, 2010; Syfert *et al.*, 2013).

Presence-absence methods:

Presence-absence methods derive suitability scores from relative frequencies of surveyed occurrences and absences in different environmental conditions. Presence-absence models are the ideal case, because they can more readily be used to account for imperfect detection and survey effort while also removing the need to select background points (Lahoz-Monfort *et al.*, 2014; Guillera-Arroita *et al.*, 2015). More importantly, unlike other methods they can be used to estimate probabilities as opposed to relative probabilities (Guillera-Arroita *et al.*, 2015). As such, if the required data are available, this approach should always be preferenced over either of the two alternatives mentioned above. However, in practice such data are rarely available at the geographic grain and extent commonly required to estimate the potential distribution of an invasive species (Elith & Leathwick, 2009; Elith, 2017).

3.1.3.3. Model settings & complexity

It is critical to carefully consider and document the settings used when fitting a model. All methods, irrespective of whether they are presence-absence, presence-background or presence-only, contain settings that can be modified by the user that ultimately affect how a model is fit and interpreted. For example, the commonly used tool Maxent contains a variety of settings which control the selection and number of background points, the complexity of covariate response curves (via feature selection and regularization) and types of model outputs that can be extracted (e.g. raw, cumulative, logistic, or logit; Merow *et al.*, 2013).

Too often, users underappreciate the importance of these settings and instead opt to use defaults (Merow *et al.*, 2013). However, depending on the species and decision context, such defaults may be highly inappropriate. For example, by default Maxent randomly selects background points across all locations represented by predictor data supplied by the user (Merow *et al.*, 2013). As a consequence, if one modelled an invasive species using global maps of climatic variables, Maxent would sample background points from across the world. Such a background implies that the species has an unlimited dispersal capacity and has had the opportunity to establish anywhere in the world – a highly unrealistic assumption. Instead, a non-default background should

have been used, with the extent restricted to a region that could reasonably be assumed to have been available to the species (VanDerWal *et al.*, 2009; Merow *et al.*, 2013).

It is also important to carefully consider model complexity when fitting correlative models (Merow *et al.*, 2014). Highly complex models (i.e. those with many predictors and interactions) are more susceptible to over-fitting, especially when the number of occurrence records is low (Merow *et al.*, 2014; Breiner *et al.*, 2015). Over-fitting is an error that occurs when a model too closely fits a limited set of data such that it describes idiosyncrasies of that dataset (i.e. fits to the noise rather than the signal). A consequence of over-fitted models is that because they model idiosyncratic patterns in the data used to parametrise them, they tend to be less transferable when predicting outcomes outside the fitted dataset.

Modellers have several options available to them to determine the right amount of complexity, and thus, exposure to over-fitting species distribution models (Elith & Leathwick, 2009; Merow *et al.*, 2014). A common approach is to use cross-validation (e.g. Valavi *et al.*, 2018), a statistical procedure that attempts to measure a model's transferability (i.e. ability to predict outside the fitted data). It achieves this by splitting a dataset into subsets, fitting the model to some of the subsets (training sets) and predicting to the others (test sets), and typically repeating this procedure until all subsets have been used for testing. This process can then be repeated for each of a set of competing models (i.e. models using different predictors and varying complexity). The models' ability to accurately predict the testing data can be compared, with higher predictive performance inferring higher model transferability and by implication lower over-fitting error.

-`ģ́-Tip

When conducting cross-validation, data are typically assigned randomly to training and testing subsets. However, in a spatial context, such random assignments will likely exhibit high spatial autocorrelation (i.e. observations close to each other have similar characteristics); potential consequences include over-estimation of predictive power and model transferability. To minimise this issue, some (Valavi et al., 2018) have advocated for dividing the study area into blocks and assigning the records within these block to cross-validation folds. While in theory this seems straightforward, in practice it can be a difficult exercise.

The R package **blockCV**(*Valavi* et al., 2018) *has vastly simplified this process of partitioning spatial data into blocks. A plain English description of what this package can do can be found at:*

https://methodsblog.com/2018/11/29/blockcv-english/

When there are very few occurrence data, cross-validation procedures will have limited capacity to infer model transferability (i.e. due to high uncertainty in performance measures). In these situations, it is best to use simple models, such as those with few predictors (Breiner *et al.*, 2015), few or no interactive (product) terms, and those that utilise simple non-linear responses or features. Simple models minimise over-fitting and have greater tractability, which can help in understanding the primary drivers of species' occurrence patterns, and are likely to be more easily generalised to new datasets (Merow *et al.*, 2014). In some models, the complexity of response curves can be configured in user-defined settings. In Maxent, users can set the maximum complexity of responses by adjusting the allowable feature types and the extent of regu-



larisation (Merow *et al.*, 2013). For Generalised Additive Models, the complexity of response curves can be modified by specifying the degrees of freedom available in various smoothing terms (here higher degrees of freedom allow for more complex response curves).

3.1.3.4. Sampling background & dealing with survey bias

Background points or pseudo-absences are required when estimating the potential distribution of a species using two-class methods such as Maxent, GARP, Boosted Regression Trees, binomial GLMs and GAMs, or when using spatial point process models (PPMs; Renner & Warton 2013). Recent evidence has shown that how these samples are chosen, and the number used, can have a substantial impact on both model interpretation and predictive capacity (Syfert *et al.*, 2013; Elith, 2017). However, consensus is currently lacking as to how background (or pseudo-absences) should be sampled for invasive species, and what number should be used (Fourcade *et al.*, 2014).

Ideally, background points should: 1) characterise the range of environments the species of interest could have dispersed to; and 2) exhibit the same patterns in survey bias as observed in the occurrence records. However, in practice, accounting for survey bias or defining the geographic extent in which a species may have dispersed is no easy task (VanDerWal *et al.*, 2009; Syfert *et al.*, 2013), especially for invasive species. Invasive species by definition are undergoing active range shifts (Elith, 2017). These range shifts make it difficult to define the extent across which a species could have dispersed, and thus, the extent across which background or pseudo-absences should be sampled. An additional complexity is that often these range shifts will exhibit strong geographic biases in survey effort, with invaded regions more likely to be intensively sampled relative to native regions (Elith, 2017).

When two-class methods are used to estimate the potential distribution of an invasive species, occurrence records are commonly restricted to the native range (Thuiller *et al.*, 2005; Elith, 2017). This is done because a species is more likely to be in equilibrium with its environment in its native range – that is, it is assumed a species has had greater opportunity to disperse across its native range and occupy sites with suitable environments. Not only does this minimise violations of the assumption of equilibrium, but it also simplifies the process of background sampling by allowing the modeller to justifiably sample points across the native range extent.

However, problems can arise when discarding occurrence records from outside the native range, including: 1) it can dramatically reduce what is already a small set of occurrence records; 2) it does not solve issues of survey bias within the native range; and 3) it can reduce the model's ability to predict the suitability of environments that are outside the native range. These issues can ultimately manifest in ill-specified models, and/or models that predict a restricted potential distribution relative to what may actually be the case (Broennimann & Guisan, 2008; Jiménez-Valverde *et al.*, 2011). This is of particular concern among biosecurity practitioners, who regularly use SDMs to inform where a pest could establish, and consequently, where to allocate finite surveillance resources. As such, biosecurity practitioners are typically hesitant to discard occurrence records (see synthesis of workshop discussion in Appendix B), and possibly as a consequence, tend to use presence-only methods such as climate matching algorithms (e.g. CLIMATCH, CLIMEX), convex hulls and hyper volumes (Drake, 2015; Blonder *et al.*, 2017), which don't require background points (Elith, 2017).



Nevertheless, a variety of methods exist for handling survey bias and/or constraining the geographic region from which background is sampled. A brief description of some of these methods is provided below:

Defining the extent from which to sample background

If only the native occurrence records are used, modellers commonly sample background points across the entire extent under the assumption the species has had the opportunity to disperse across the region. However, if occurrences from the invaded range are included, modellers use several different approaches. One approach is to construct convex hulls (often with buffers) around the set of occurrence records and then sample background from within that hull. Another approach is to sample background from within a radius around each occurrence record – the size of the radius typically being defined by maximum dispersal distances (VanDerWal *et al.*, 2009). A third option is to sample background in a pre-defined set of bioclimatic zones (e.g. Köppen-Geiger zones; Hill & Terblanche, 2014), which might represent occupied zones or those believed to provide suitable environmental conditions for the modelled organism³.

Dealing with survey bias

Occurrence data are often biased toward human population centres and roads (Reddy & Dávalos, 2003). If not corrected, this bias may result in a model that describes patterns in survey effort as oppose to the true distribution of suitable environmental conditions for the species (Phillips et al., 2009). A common way to account for survey bias is to sample the background such that it reflects the same sample selection bias as the occurrence data. For example, if presence data are only taken from near roads or cities, then background data should be taken from the same areas (Ferrier *et al.*, 2002). Many methods exist for achieving this (e.g. Fourcade et al., 2014), but we briefly detail the three most commonly used. The simplest is called target-group sampling, which uses occurrence records from a broad set of species that are likely have similar observation activities of collectors as the species of interest (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. The second method is a continuation of the first, whereby the target-group occurrence records are used to construct a interpolated bias grid that can then be used to inform the distribution and intensity of background samples. The primary advantage of this latter approach is that it provides greater flexibility in defining the total number of background points one wishes to use. The third option is to create a bias grid as a function of regression techniques that correlate occurrence records against factors likely to explain geographic survey bias (e.g. distance from roads and cities). Outside the bias grid approach, other modellers use bias covariates (Warton et al., 2013; Fithian et al., 2015).

Whilst these methods are the best available for dealing with bias, they are not commonly applied in invasive species modelling. This is partly through lack of technical expertise, but also because at the scale of modelling (across countries) and with the patchiness of records across countries, modellers often cannot drill down to the grain required to consider relevant factors such as distance to roads. However, bias is a substantial problem that needs to be considered when modelling with presence-only data, and it is worth assessing how best to characterise variation in survey effort. Some mod-

³Though this still requires defining the extent in which background should be sampled.



ellers simply reduce the density of records (i.e. thinning) in relatively over-sampled locations.

3.1.3.5. Covariate uncertainty

Accurately predicting a species' distribution into novel environmental space using a correlative approach requires that the model includes the causal predictors that govern a species' niche (Austin, 2002; Elith, 2017). However, rarely do we know what these variables are, makes sensible selection of model predictors a significant challenge (Barry *et al.*, 2015).

A range of studies have attempted to determine which predictor variables are best to include, and in lieu of truly causal predictors, how proximal these variables may be (for a review, see: Barry *et al.*, 2015). However, despite this work, there has been little attempt to review or develop methods to simply identify "good predictive variables" from those typically available to modellers, and how to best use them to reduce errors associated with model transferability.

Based on work conducted by CSIRO in collaboration with CEBRA, we recommend that modellers account for covariate uncertainty and improve model transferability by constructing ensembles of simple bivariate models. Specifically, when the choice of which predictors to use is unknown or unguided, constructing an ensemble of small (e.g. two-variable) models allows for model construction to be focussed on selection of "good" performing variables (ones that perform well in both native and non-native ranges), rather than relying on model fit as the indication of the "best" model. This allows for putatively more proximal variables to be identified and to apply them in a simple manner that not only increases transferability, but also provides a more tractable approach to developing maps of pests' potential distributions. For full details of these methods please see Appendix C.

3.1.4. Step 4: Model Validation

Irrespective of whether a model is process-based or correlative, it should be validated against independent datasets (i.e. occurrence data not used in the model parametrisation process). For process-based models this is a relatively straightforward process of overlaying localities of known establishments onto maps of predicted suitability and then examining whether patterns in both broadly match (e.g. Kearney *et al.*, 2008, 2010; Kriticos *et al.*, 2017). However, due to a variety of difficulties, correlative invasive species SDMs are rarely validated using independent data (West *et al.*, 2016).

Correlative SDMs have been evaluated using a variety of approaches. For example, some have examined patterns in model residuals and whether predictor importance and fitted functions are consistent with what is known about the species' ecology and physiological tolerances (Elith & Leathwick, 2009; Elith, 2017). More commonly however, emphasis has been on using formal statistical summaries such as area under the receiver operating curve (AUC), kappa, and explained deviance to examine a model's ability to accurately predict the fitted data, or ideally, a set of independent data (Elith, 2017).

While these statistical summaries provide invaluable information on how well a model can predict a species' known range, their application to invasive species distribution 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 effectively use such statistical methods in this context would require knowing the potential distribution of a species – something that is unknowable.

To overcome these issues, some have fitted invasive species SDMs using only native range occurrence data and then evaluating the models' ability to predict occurrences in the known invaded range (i.e. a measure of its potential range). However, the reliability of this test is dependent on whether the processes limiting a species' distribution in its native range are the same in the invaded range or the potential invaded range (i.e. regions where the pest has not yet arrived). It is also based on the strong assumption that the occurrences in the invaded range are representative of persistence⁴ (Elith, 2017).

Others have evaluated invasive species SDMs according to their ability to predict performances in the model-fitting range. The assumption here is that if the model accurately predicts the known distribution of a species, then it is more likely to accurately predict its potential distribution. Simulation analyses conducted in a previous CEBRA report shows some support for this assumption (Barry *et al.*, 2015).

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

- 1. models are first assessed for their ecological relevance, by: using expert knowledge; examining functional response curves; sourcing additional data including physiological information (e.g. temperature tolerances); or comparing models with process-based models that do not use occurrence records, as well as other published models.
- 2. one formally examines a model's ability to predict in the model-fitting region (e.g. by using AUC) if the method involves estimating response curves⁵.
- 3. regions of novel environmental space (i.e. environments not found in the modelfitting region) be highlighted on any suitability map derived from a correlative model estimating climatic response curves.

The mapping of novel environments (Elith, 2017) is a crucially important step when using models that attempt to estimate covariate response curves (e.g. Maxent, GAMs, GLMs etc.) because it helps interpretation of model output and also guides users as to where predictions may be highly unreliable (i.e. where the model is extrapolating).

Lastly, in order to maximise the robustness of these models, we recommend that 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 pest biology are incorporated and that the models are appropriately implemented.

⁴This also affects process-based models.

⁵Statistical performance measures in the fitted region will be less applicable for climatic envelope models (e.g. range bagging, BIOCLIM), which define suitability as a function of the extent of occurrence records in environmental space.



-`<mark>`</mark>{ Tip

Two popular methods used to identify and map novel environments:

MESS (Multivariate Environmental Similarity Surface; Elith et al. 2010): can identify, distinguish and quantify novel univariate ranges. It does not account for changes in correlation structure among these predictors (i.e. new combinations of covariates). MESS is currently implemented as a tool in Maxent and can also be calculated in R using: dismo::mess() or ecospat::ecospat.mess().

ExDet (Extrapolation detection; Mesgaran et al. 2014): can detect, distinguish and quantify two types of novelties: novel univariate range and novel combinations of covariates. ExDet is not implemented in Maxent but can be readily estimated within R using: ecospat::ecospat.climan()

Web interface: In order to improve the accessibility to these extrapolation algorithms, CE-BRA, as part of this project, has developed a web interface called Climate Matcher (Fig. 3.3) that allows users to examine novel climatic conditions relative to selected countries or user uploaded points/polygons. Users can then download the resultant output and use it to highlight areas where SDM predictions may be dubious. This web interface can be accessed at:



https://apps.cebra.unimelb.edu.au/climate_matcher/

Figure 3.3.: An output from the web interface Climate Matcher – An easy to use tool for examining novel climatic conditions. In this example, ExDet is used to examine areas of novel climate relative to Afghanistan (Yellow). Red = Novel climatic range, White = Analogous climate, Blue = Novel correlations.



3.1.5. Step 5: Model Utility

Clear model documentation

Invasive species SDMs are regularly used to inform a variety of biosecurity management decisions (see Table 3.2). However, to ensure the appropriate use of a model, the first step is to ensure fundamental details are clearly articulated and summarised for biosecurity decision-makers. In particular, it is critically important to summarise what data were used, how the model was developed, the assumptions it made and how one should interpret and use the output. In essence, the documentation that happens at each step of the decision tree (Fig 3.1) should be collated and summarised into a non-technical summary readily interpretable by policy staff.

Versatility of different model methods

The type of model will have important implications for how it is used to inform biosecurity management. Process-based models tend to directly measure quantities of interest that govern whether species can persist in a given environment (e.g. fecundity). As a consequence, the outputs are often easier to interpret and can be readily used to inform priority pest lists, optimal surveillance and post-border cost-sharing arrangements.

Correlative models are also commonly used to inform biosecurity decisions. However, while these models may be appropriate for some decision contexts (e.g. spatial prioritisation of surveillance) they are often inappropriate for many biosecurity applications (Table 3.2). For example, Guillera-Arroita *et al.* (2015) highlighted that using presence-background models such as Maxent to inform pest prioritisation or surveillance optimisation routines 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 please see Table 3.2.

Thresholding model outputs

It is common practice to convert maps which provide 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 possible risk to the broader public, we strongly recommend against discretising model outputs for other biosecurity decision contexts. This is because the decision of where to threshold is often highly subjective and dependent on both the expected consequences of a pest outbreak and the tolerance for incorrectly assigning a suitable location as unsuitable. More important, however, is that discretising model outputs rarely improves decision-making because it degrades the information supplied by the model (Table 3.2).
Table 3.2.: Examples of presence-background SDM usage in the management of invasive species. Adapted from Guillera-Arroita *et al.* (2015). AOO = Area of Occupancy; S/M = Single or Multi-species. For a comprehensive review of SDM applications to invasive species management please see Guillera-Arroita *et al.* (2015).

Application	SDM usage (input into)	S/M	Consequences of biased occupancy estimation	Is relative likelihood or ranking a sufficient input?	Has binary conversion been used?	Consequences of using binary conversion
Target management actions (e.g. survey or control)	Rank sites according to habitat suitability	S	Efforts to survey or control invasion may be spatially misguided.	Both suitable if only interested in identifying best sites. Ranking not sufficient if aim is to identify sites within a certain suitability compared to best sites.	Yes	A 'threshold' may be used to delimit best sites but binary conversion leads to loss of information.
Optimal surveillance	Input to optimisation: find surveillance level that minimises overall costs (monitoring + management) given site occupancy probability.	S	Overall costs will increase, as the level of surveillance guiding early management is not optimal.	Computing expected costs requires probabilities; can result in suboptimal solution otherwise.	No	Not valid. Method based on acknowledging uncertainty in species presence to evaluate trade-offs in expected costs.
Invasive species monitoring or status assessment	Calculate species' AOO or compare temporal changes in AOO (by fitting models at different points in time).	S	Invasion status may be incorrectly assessed due to biased estimation of AOO or its temporal trends (real trend missed or spurious trend detected).	Relative likelihoods & rankings not able to estimate AOO, and not sufficient input to track temporal trends (as not comparable across time; except for PA data when detectability constant in space and time).	Yes	Degrades AOO estimation, even when probabilities are estimated; It does not solve the inadequacy of relative likelihoods or rankings.

Application	SDM usage (input into)	S/M	Consequences of biased occupancy estimation	Is relative likelihood or ranking a sufficient input?	Has binary conversion been used?	Consequences of using binary conversation
Importation decisions (pre-border risk assessment)	SDMs fitted with data from other areas used to calculate potential AOO, for categorising risk (e.g. low, medium, high), sometimes within a scoring system that also considers non-spatial attributes (e.g. weed risk assessments).	S	May result in beneficial species being denied entry, or harmful species being imported.	Relative likelihoods and rankings are not suitable for estimating AOO.	Yes	Degrades AOO estimation, even when probabilities are estimated; It does not solve the inadequacy of relative likelihoods or rankings.
Post-border cost-sharing arrangements	SDMs fitted with data from other areas used to calculate the proportion of potential AOO within each jurisdiction.	S	Could lead to inappropriate allocation of management costs across jurisdictions.	Relative likelihood is sufficient, as the proportion of AOO is not affected by a constant scaling. Correct site ranking does not guarantee appropriate cost sharing.	Yes	Degrades the estimation of proportions of AOO.
Post-border prioritisation lists for control or eradication	Calculate potential AOO for several species to prioritise management resources across them.	М	Management effort may be allocated to species that actually have low potential for spread and vice versa.	Relative likelihoods and rankings are not sufficient because they are not comparable across species.	Yes	Degrades AOO estimation, even when probabilities are estimated; It does not solve the inadequacy of relative likelihoods or rankings

Application	SDM usage (input into)	S/M	Consequences of biased occupancy estimation	Is relative likelihood or ranking a sufficient input?	Has binary conversion been used?	Consequences of using binary conversation
Evaluate management options or scenarios, in terms of the probability of eradicating a species over some defined time horizon	Input to spatially-explicit meta-population modelling methods (e.g. RAMAS GIS) to: 1. Define the spatial extent of suitable habitat (or 'patches'), by choosing a threshold below which environments are considered 'unsuitable'. 2. Set the carrying capacity of 'suitable' patches as a function of suitability. 3. Set vital rates at 'suitable' patches as a function of suitability.	S	Poor estimates of probability of eradication and therefore choice of management options.	 (1) Using relative likelihood allows delineating patches of distinct suitability, but can be problematic because the choice of threshold is arbitrary (i.e. not based on observed prevalence). Occupancy probabilities provide a more natural way to determine the spatial structure of the meta-population. (2-3) Can be problematic (as can be the use of occupancy probabilities) unless a good model can be developed to link relative likelihood to carrying capacity / vital rates (e.g. using other data). Ideally one would directly model 'maximum abundance' or vital rates as a function of environmental covariates. 	(1) Yes. Needed by definition. (2-3) No	 (1) Needed to define patches (but note that, if suitability does not indicate sharp boundaries, a patch-based meta-population model may not be appropriate). (2-3) Not valid, non- thresholded values are needed to distinguish the suitability of patches.

Table 3.2 continued from previous page

Cebra I trans to Balling for

Application	SDM usage (input into)	S/M	Consequences of biased occupancy estimation	Is relative likelihood or ranking a sufficient input?	Has binary conversion been used?	Consequences of using binary conversation
Evaluate management options or scenarios, in terms of the probability of eradicating a species over some defined time horizon	Input to spatially explicit individual- based, cellular automata or multi-agent model, to determine the performance of individuals in each cell in each time step.	S	Poor estimates of probability of eradication and therefore choice of management options.	Can be problematic (as can be the use of occupancy probabilities) unless a good model can be developed to link them to whichever rules govern the performance of individuals within cells.	No	Not valid, non-thresholded values are needed.
Assessment of the impact of climate/land use change on invasive species	SDM fitted with current data used to make predictions under changing environmental conditions (e.g. based on projected scenarios).	S	Incorrect evaluation of impacts of climate/land use change, as both current and projected species range are incorrectly estimated.	Relative likelihood suitable to evaluate relative change but not to compute AOO and derived quantities such as species persistence. A ranking has the risk of missing critical isoclines/thresholds and lead to incorrect impact predictions.	Yes	Degrades AOO estimation, even when probabilities are estimated; It does not solve the inadequacy of relative likelihoods or rankings

Table 3.2 continued from previous page

4. Framework for mapping establishment likelihoods for plant pests

Federal and state biosecurity agencies play a vital role in surveying for the early detection of high impact invasive plant pests. Such surveillance is labour intensive and costly to maintain, and as such, is rarely implemented uniformly across geographic space but rather positioned in regions considered to have high likelihoods and/or consequences of pest establishment (De Lima *et al.*, 2011; Whittle *et al.*, 2013; Renton *et al.*, 2017; National Fruit Fly Strategy Action Plan, 2017). However, determining the geographic distribution of establishment likelihoods, especially when a pest may enter via multiple pathways, poses a significant challenge to biosecurity agencies worldwide. This is because accurate estimation of establishment likelihoods requires information about: 1) the major pathways that allow pests to enter a country; 2) the probability that a carrier along a pathway (e.g. passenger) contains a pest; 3) how pathway vectors will distribute post-border; and 4) whether the abiotic and biotic environment at a given location is suitable for pest establishment. However, such detailed knowledge is rarely available for large geographic extents (e.g. all of Australia) or is simply not known for many plant pests.

In this chapter we outline a pragmatic (given data constraints) and transparent framework for developing Australian national maps of establishment likelihood that we believe should be applicable to the majority of terrestrial plant pests that are likely to enter Australia. At its foundations, this mapping framework is based on the principle that for a pest to successfully establish it must first overcome three geographic barriers (Catford *et al.*, 2009), namely:

- 1. can it reach the location of interest (i.e. arrival rate)?
- 2. are the abiotic conditions suitable (e.g. climate suitability)?
- 3. are the biotic conditions suitable (e.g. presence of host/food)?

In the proceeding sections, we outline how each of these three geographic barriers can be estimated using readily available data sources and methods. We also outline an approach in which these three geographic barriers can be integrated in order to derive a national map of establishment likelihood. We illustrate the implementation of this framework using four priority plant pest case studies: oriental fruit fly (*Bactrocera dorsalis*), khapra beetle (*Trogoderma granarium*), brown marmorated stink bug (BMSB; *Halyomorpha halys*) and gypsy moth (*Lymantria dispar*).

cebra

ARisk vs. establishment likelihoods

It is important to note that the maps developed in this project are technically not maps of risk but rather establishment likelihoods. Risk maps, by definition, are the product of spatially explicit establishment likelihoods multiplied by spatially explicit consequences (likelihoods \times consequences). In this project, the focus was on the first component of risk, establishment likelihoods. We therefore avoid the term "risk maps" when referring to the mapping framework or associated outputs developed as part of this project.

Framework not suitable for marine pests

The mapping framework described in the following sections is explicitly designed for terrestrial plant pests. The framework is not suited to marine pests which utilise different pathways of entry and are exposed to different environmental and biotic variables.



4.1. Identifying high risk pathways

In the ideal case, post-border arrival rates should be estimated using a comprehensive pathway analysis that incorporates: 1) volumes of multiple pathway carriers entering the country from infected countries; 2) border surveillance effort and sensitivity (i.e. in order to estimate leakage rate); and 3) post-border movements of pathway vectors (i.e. carriers). However, in practice, pathway analyses rarely incorporate post-border movements, but instead focus on estimating arrivals at various points of entry (e.g. ports; Paini & Yemshanov, 2012; Tingley *et al.*, 2018). This lack of post-border movement in pathway analyses is possibly due to two factors. The first is that governments worldwide have a strong focus on border surveillance in the effort to mitigate risk and associated post-border costs. Second, obtaining high quality post-border movement data across large geographic extents from multiple pathways will likely come at significantly greater cost and may have social and political implications (e.g. privacy concerns in passenger movements).

The Department of Agriculture (DAWE's predecessor) invested significant effort into quantifying the likelihood that high priority pests or functional groups (e.g. fruit fly) enter Australia from up to 54 pathways. Likelihoods were estimated using a formal expert elicitation process involving more than 120 experts with familiarity with the Australian biosecurity system. These national-level entry likelihoods were then integrated into a probabilistic framework (i.e. a Netica Bayesian network) called the Risk-Return Resource Allocation (RRRA) model. Fundamentally, RRRA is defined by a set of entry pathways, a set of organisms of concern and a set of biosecurity interventions that occur both pre-border and at the border (Craik *et al.*, 2017). These are combined with estimates of outbreak consequences in order to evaluate the effective-ness of biosecurity controls designed to stop pests, diseases and weeds from entering Australia and provide insights as to where finite biosecurity resources should be allocated to maximise risk mitigation (Craik *et al.*, 2017).

As RRRA currently contains Australia's most comprehensive database on pest by pathway arrival likelihoods, the mapping framework developed as part of this project was designed to directly utilise national pathway probabilities derived from RRRA – specifically probabilities of arrival after accounting for pre-border and border controls. However, in practice, the risk map framework can utilise any national pathway probability, irrespective of how it is estimated. Moreover, while RRRA does not contain likelihoods of entry by country of origin, the risk map framework developed here could be readily applied to such data should it become available.

While national pathway probabilities are useful for identifying high risk pathways, they do not provide information about the subsequent distribution of arrivals that evade pre-border and border checks, or whether they are likely to result in establishments. In the following sections we outline methods, based on data pragmatic first principles, for distributing vectors/carriers post-border for ten high risk pathways, as identified from RRRA or interception data, for four national priority plant pests: oriental fruit fly (*Bactrocera dorsalis*), khapra beetle (*Trogoderma granarium*), brown marmorated stink bug (BMSB; *Halyomorpha halys*) and gypsy moth (*Lymantria dispar*).

4.2. Pathways

Approved Arrangement Sites

The method outlined below does not explicitly account for Approved Arrangement (AA) sites. This is due to several reasons. First, pathway likelihoods used are leakage rates after accounting for pre-border and border controls (of which AA sites form part of the latter). Second, we were unable to obtain a comprehensive spatial layer of all AA sites or the volume of pathway goods sent to each. Consequently, leakage from AA sites could not be incorporated into this analysis as we were unable to appropriately account for the differential risk profiles among AA sites.

cebra

4.2.1. Pathway: Air passengers

Between January 2018 and January 2019 there were a total of 23,356,510 international passengers entering Australia (Table 4.1). These numbers encompassed permanent settlers, long-term returning residents and long-term and short-term visitors. As tourists are likely to disperse across the country differently to returning residents, and because they may differ in their probability of carrying a pest, we treated them independently. For the purposes of our analyses we classified short and long term visitors as tourists and the remainder as returning residents. This meant that between January 2018 and January 2019 there were a total of 10,633,690 tourists and 12,722,820 returning residents entering Australia. Unfortunately, due to limited reliable information on passengers' final destinations, we were unable to ascertain the number of passengers arriving at each international airport. This was predominately due to passengers entering Australia via a major international airport (e.g. Sydney) and then proceeding to catch domestic flights to their final destinations (e.g. Hobart), for which we are unable to differentiate tourist from returning resident. As a consequence, country-level arrival numbers were distributed across Australia using methods described below.

Passenger type	Number
Long-term returning resident	153,060
Short-term returning resident	12,454,100
Permanent settlers	115,660
Long-term arriving visitors	657,790
Short-term arriving visitors	9,975,900
TOTAL	23,356,510

Table 4.1.: Number of international arrivals between January 2018 and January 2019.Data source: 3401.0 – Overseas Arrivals and Departures, Australia, Feb 2019.

4.2.1.1. International tourists

In the days immediately following arrival into the country — when the risk of a passenger carrying a pest is likely greatest — international tourists are not expected to



travel significant distances and are likely to congregate in areas with high densities of tourist accommodation. For this reason, we distributed national numbers of international tourists across Australia as a function of the number of tourist rooms and distance from major international airports:

$$Tourists_i = \frac{tourist rooms_i \times f(airport distance_i)}{\sum_{i=1}^{n} tourist rooms_i \times f(airport distance_i)} \times Tourist passengers.$$
(4.1)

Where *Tourist passengers* is the total number of tourists entering Australia in a given year and tourist rooms_{*i*} is the number of tourist rooms in a grid cell¹.

Tourism room data were downloaded from the ABS (Cat. 8635.0 - Tourist Accommodation, Australia, 2015-16) for each state in Excel format. These data cover 4,453 establishments with 203,766 rooms and 552,877 beds². Small area statistics for 2015-16 are classified to the Australian Statistical Geography Standard (ASGS): Volume 3 – Non ABS Structures, 2015 Edition (cat. no. 1270.0.55.003) effective from September quarter 2015. Small area data are available for States and Territories, Tourism Regions (TR) and Statistical Area Level 2 (SA2). A correspondence of SA2 to TR and detailed TR maps for each state and territory can be found in Tourism Region Maps and Allocation File, Australia (ABS Cat. 9503.0.55.001).

A total of 914 of 988 accommodation records matched those of the SA2 data set. Discrepancies were mainly to do with the amalgamation of SA2 areas. All accommodation areas containing more than 3 accommodation providers were allocated to the appropriate SA2 by splitting the quantities equally between the two SA2 areas so that 951 SA2 areas were reconciled with the data. There were 37 SA2 areas with three or fewer accommodation providers that were not reconciled as the names could not be easily resolved and the impact for this analysis was considered negligible. The final totals from the reallocated data contain 4,413 accommodation providers, 203,682 rooms and 552,557 beds.

The number of rooms is expected to be the most representative of the risks associated with tourist pathways as rooms represent the family/friend groups that travel together. By using rooms, the large international tourist establishments are more reasonably included. As some of the SA2 records had the number of establishments listed, but not number of rooms. Number of rooms has been regenerated based on a polynomial approximation (Fig 4.1) to the number of accommodation providers in the SA2, where SA2s with many accommodation providers usually have larger establishments (Fig. 4.1). The final geographic distribution of tourist room density can be seen in Fig. 4.2.

¹A grid cell could be any size of interest (here we use 1 km²).

²Noting that some areas were missing room or bed numbers.





Figure 4.1.: Polynomial relationship between the number of accommodation providers and rooms contained in Statistical Area level 2 categories. Figure courtesy of Dr Mark Stanaway.



Private accommodation

The accommodation data obtained from the ABS does not include private providers (e.g. Airbnb). As such, these data do not fully encapsulate all accommodation providers that tourists may use. However, the use of such accommodation by tourists is expected to occur in similar areas where more formal accommodation providers occur (e.g. regions with tourist attractions).



Figure 4.2.: Geographic distribution of tourist rooms across Australia. For presentation purposes data has been aggregated to 5 km grid cells.



The airport distance weight for each grid cell, $f(\text{airport distance}_i)$ (eqn 4.1) is an attempt to account for the situation that international tourists are unlikely to travel significant distances from airports when they first arrive in Australia. Ideally, this weight would be informed by empirical data on where passengers travel in the first few days from arrival. However, in the absence of such information, a weight could be informed by any function thought to approximate the likely distances travelled by tourists in the first few days since arrival. For example, a continuous decay function such as a negative exponential (eqn: 4.2) could be used.

$f(airport distance_i) = e^{-\beta \times airport distance_i}$.	(4.2)
--	-------

After consultation with departmental staff, a decision was made to assume that 50% of tourists are likely to be distributed within 200 km of a major international airport (Table 4.2). Effectively, this meant that β was assigned as $\log(0.5)/200 = -0.0035$. The distance from international airports and the derived weight score can be seen in Fig (4.3).

International airport	Longitude	Latitude
Darwin	130.8774	-12.40809
Cairns	145.7557	-16.87301
Brisbane	153.1097	-27.40341
Sydney	151.1658	-33.93649
Canberra	149.1911	-35.30748
Melbourne	144.8511	-37.66876
Adelaide	138.5375	-34.93862
Perth	115.9741	-31.94080
Adelaide Perth	$138.5375 \\115.9741$	-34.93862 -31.94080

Table 4.2.: Major Australian international airports.







Figure 4.3.: Top) Distance from major Australian international airports; Bottom) Distance weight score derived from negative exponential under the assumption that 50% of tourists will be distributed within 200 km of an international airport. For presentation purposes data has been resampled to 5 km grid cells.



4.2.1.2. Returning residents

Again we did not have sufficient information as to where returning residents go upon entering Australia. However, these passengers are likely to be distributed post-border as a function of population density (Fig. 4.4), whereby a greater proportion of returning residents will comes from (and will return to) high population areas. As such, we estimated the expected number of returning residents occurring in each grid cell, *i*, as:

Returning residents_i =
$$\frac{\text{population density}_i}{\sum\limits_{i=1}^{n} \text{population density}_i} \times \text{Returning residents.}$$
 (4.3)

Where *Returning residents* is the total number of returning residents entering Australia in a given year.

Population data were derived from the 2016 Australian Census of Population and Housing produced by the Australian Bureau of Statistics (Cat. 2074.02016). Specifically, we first joined Mesh Block-scale counts of the Usual Resident Population (URP) with Mesh Block polygons (Cat. 1270.0.55.001), and then transferred counts onto a 1 km² raster with Australian Albers coordinate system (i.e. equal area projection). This entailed calculating, for each grid cell, the proportion of the cell overlapped by each intersecting Mesh Block polygon, and distributing population counts accordingly assuming uniform distribution of the population within Mesh Blocks. Population counts were summed across the set of intersecting Mesh Blocks (or parts thereof) to achieve an estimate of the count at the 1 km² scale.





Figure 4.4.: The geographic distribution of human population across Australia. Data based on 2016 census. For presentation purposes data has been aggregated to a 5 km grid.

4.2.1.3. Torres Strait visitors (Fruit fly specific)

Between January 2018 and January 2019 there were approximately 51,0000 air passengers arriving at Cairns international airport from the Torres Strait. This pathway has been identified by the Risk Return Resource Allocation (RRRA) as a significant pathway for the arrival of exotic fruit fly into mainland Australia. This is because parts of Torres Strait exhibit regular incursions of *Bactrocera dorsalis* that likely originate from established populations on Papua New Guinea³.We distributed air passengers arriving via this pathway as a function of distance from Cairns international airport and the proportion of the national population present each grid cell *i*:

$$TS passengers_{i} = \frac{\text{population density}_{i} \times f(\text{Cairns airport distance}_{i})}{\sum_{i=1}^{n} \text{population density}_{i} \times f(\text{Cairns airport distance}_{i})} \times TS \text{ passengers.}$$

$$(4.4)$$

Where *TS passengers* is the total number of Torres Strait passengers entering Australia in a given year.

Like the tourist passenger pathway, we assumed that distribution of Torres Strait passengers would exhibit a negative exponential decay with distance (eqn: 4.2). After consultation with departmental staff, a decision was made to assume that 50% of Torres Strait air passengers were likely to be distributed within 10 km of Cairns international airport in areas of higher population density. Effectively, this meant that β was assigned as $-\log(0.5)/10 = -0.07$. This much stronger decay function with distance combined with population density was made under the assumption that passengers from the Torres Strait were most likely to take up accommodation with friends and relatives in and around Cairns. The distance from Cairns international airports and the derived weight score can be seen in Fig. 4.5.

³Although this is fruit fly specific, there are other pests that occur in the Torres Strait Islands and northern Cape York for which this pathway may be applicable.







Figure 4.5.: Top) Distance from Cairns international airport; Bottom) Distance weight score derived from negative exponential under the assumption that 50% of Torres Strait air passengers will be distributed within 10 km of Cairns International Airport. For presentation purposes data has been aggregated to a 5 km grid.



4.2.2. Pathway: Mail

In 2017–18 Australia received 152 million international mail articles (DAWR, 2018). As a wide range of goods can readily be sent by mail, this pathway is a significant source of entry for many pests and diseases (e.g. RRRA has identified that both khapra beetle and Xylella may enter via this pathway). As we were unable to obtain high resolution empirical data as to how this volume of international mail was distributed across Australia, we assumed, as for returning residents, that mail entering Australia was distributed as a function of population density (see Fig. 4.4). Specifically, the mail estimated to arrive in a given grid cell, *i*, was estimated as:

$$Mail_{i} = \frac{\text{population density}_{i}}{\sum_{i=1}^{n} \text{population density}_{i}} \times Mail.$$
(4.5)

Where Mail is the total number of mail units entering Australia.

4.2.3. Pathway: Imported fertiliser

RRRA has identified that some pests, such as khapra beetle, may enter Australia via imported fertiliser or its associated packaging. Between 2015 and 2016 approximately 6907 units of fertiliser were imported into Australia (data supplied by the DAWE). In order to distribute these units across Australia, we utilised the latest statistics on fertiliser use summarised by the Australian Bureau of Statistics (4627.0 - Land Management and Farming in Australia, 2016-17)). These data were derived from the 2016-17 Rural Environmental and Agricultural Commodities Survey (REACS) and provide estimates of total fertiliser tonnage applied to each Natural Resource Management region across Australia. While both the volume of fertiliser imports and the ABS distribution data are not for the 2017–2018 year, we assume that the spatial distribution of risk remains the same. Specifically, we rasterized NRM vector data (again using equal area projection – Australian Albers) and then multiplied the fertiliser tonnage in each raster cell, Fertiliser tonnage, by the presence (1) or absence (0) of ALUMC landuse categories (see Section 4.5 for more details) likely to utilise imported fertiliser, Fertiliser landuse_i (Table 4.3). This allowed us to account for some of the spatial heterogeneity in fertiliser use within an NRM region (Fig. 4.6). We then can distribute national numbers of imported fertiliser using:

$$Fertiliser units_{i} = \frac{Fertiliser tonnage_{i} \times Fertiliser landuse_{i}}{\sum_{i=1}^{n} Fertiliser tonnage_{i} \times Fertiliser landuse_{i}} \times Fertiliser units.$$
(4.6)

Where *Fertiliser units* is the total number of imported fertiliser units entering Australia in a given year.



Landuse	Code	Landuse (cont)	Code (cont)
Cropping	330	Irrigated beverage and spice crops	432
Cereals	331	Irrigated hay and silage	433
Beverage and spice crops	332	Irrigated oilseeds	434
Hay and silage	333	Irrigated sugar	435
Oilseeds	334	Irrigated cotton	436
Sugar	335	Irrigated alkaloid poppies	437
Cotton	336	Irrigated pulses	438
Alkaloid poppies	337	Irrigated rice	439
Pulses	338	Irrigated perennial horticulture	440
Perennial horticulture	340	Irrigated tree fruits	441
Tree fruits	341	Irrigated olives	442
Olives	342	Irrigated tree nuts	443
Tree nuts	343	Irrigated vine fruits	444
Vine fruits	344	Irrigated shrub berries and fruits	445
Shrub berries and fruits	345	Irrigated perennial flowers and bulbs	446
Perennial flowers and bulbs	346	Irrigated perennial vegetables and herbs	447
Perennial vegetables and herbs	347	Irrigated citrus	448
Citrus	348	Irrigated grapes	449
Grapes	349	Irrigated seasonal horticulture	450
Seasonal horticulture	350	Irrigated seasonal fruits	451
Seasonal fruits	351	Irrigated seasonal flowers and bulbs	452
Seasonal flowers and bulbs	352	Irrigated seasonal vegetables and herbs	453
Seasonal vegetables and herbs	353	Irrigated turf farming	454
Grazing irrigated modified pastures	420	Intensive horticulture	510
Irrigated pasture legumes	421	Production nurseries	511
Irrigated legume/grass mixtures	423	Shadehouses	512
Irrigated sown grasses	424	Glasshouses	513
Irrigated cropping	430	Glasshouses - hydroponics	514
Irrigated cereals	431	·	

Table 4.3.: Landuses likely to use imported fertiliser.

♣ Domestic fertiliser use

The ABS data used to distribute fertiliser across the country is based on farm usage. While farm usage is likely to account for a substantial quantity of imported fertiliser, its use in private gardens may also be high. Unfortunately, we do not have adequate data on domestic use in order to partition its usage and distribute farm vs domestic use differently.





Figure 4.6.: Log10 proportion of fertiliser tonnage deposited per grid cell. For presentation purposes data has been aggregated to a 5 km grid.



4.2.4. Pathway: International Containers

In 2017 a total of 5,125,422 containers were discharged at Australian ports (Table 4.4). In order to estimate the post-border risk associated with international shipping containers, we require detailed information as to the ports at which they are discharged and their post-border destinations. Unfortunately, information on the final destinations of containers after discharge at ports is not readily available for 2017–2018. However, the Australian Bureau of Statistics contains an experimental statistical analysis of the post-border movements of international shipping containers entering Australia in 2009–2010 (ABS Cat No. 5368.0.55.018). This experimental analysis explicitly estimated the number of international shipping containers arriving at each postal area from each of five of Australia's busiest marine ports (i.e. Sydney, Melbourne, Brisbane, Fremantle and Adelaide). Collectively, these ports account for 98.29% of containers being discharged in Australia based on 2017–2018 data (Table 4.4). The analysis also revealed that approximately 10% of containers being discharged in a given state are transported interstate (See ABS Cat No. 5368.0.55.018) for details). However, the majority of containers are generally distributed in postcodes surrounding the port of entry.

Under the assumption that the distribution of containers has not substantially changed since 2009–2010, we used this ABS experimental dataset to inform the post-border distribution on 2017–2018 containers. To do this we first combined 2009–2010 ABS container count data with a postal area shape file (Version 2011; ABS Cat. 1270.0.55.003). Container numbers arriving in each postal area polygon, p, from each of the five major ports, j, can then be converted to a proportion using:

Proportion of containers_{*p,j*} =
$$\frac{\text{N containers}_{p,j}}{\sum\limits_{p=1}^{n} \text{N containers}_{p,j}}$$
. (4.7)

This vector data is then rasterised to the appropriate raster projection (i.e. Australian Albers) and cell resolution, such that it produces a stack of five rasters – one stack for each port. These rasters are then further normalised and multiplied by the current number of containers arriving at each port (e.g. using 2017–2018 container numbers). The total number of containers arriving in a given raster cell *i*, is then just the sum across raster layers:

N containers_i =
$$\sum_{j=1}^{n} \left(\frac{\text{Proportion of containers}_{i,j}}{\sum_{i=1}^{n} \text{Proportion of containers}_{i,j}} \times \text{Total containers}_{j} \right)$$
, (4.8)

where *Total containers*_j is the total number of containers discharged at a port j in a given year.



Port	Port code	Longitude	Latitude	Number of containers
Sydney	AUSYD	151.1906880	-33.86068000	1762876
Melbourne	AUMEL	144.9158500	-37.81745100	1759221
Brisbane	AUBNE	153.1715600	-27.37148000	819838
Fremantle	AUFRE	115.7480600	-32.04720000	466318
Adelaide	AUADL	138.5079505	-34.79916667	229529
Townsville	AUTSV	146.8307000	-19.25312000	36020
Darwin	AUDRW	130.8452700	-12.47116000	15501
Port Kembla	AUPKL	150.8904500	-34.45455000	12751
Newcastle	AUNTL	151.7718500	-32.91538000	5472
Geelong	AUGEX	144.3646500	-38.10478000	3917
Launceston	AULST	147.1239500	-41.42815000	3534
Esperance	AUEPR	121.8983600	-33.87233000	3429
Gladstone	AUGLT	151.2415300	-23.82563000	1658
Cairns	AUCNS	145.7803800	-16.92603000	1293
Hobart	AUHBA	147.3337700	-42.88306700	1271
Port Alma	AUPTL	150.8628300	-23.58269000	941
Port Hedland	AUPHE	118.5738700	-20.31500500	454
Dampier	AUDAM	116.7010900	-20.66269000	255
Mackay	AUMKY	149.2226300	-21.10678800	222
Broome	AUBME	122.2125300	-18.00359000	196
Port Pirie	AUPPI	138.0139000	-33.18032000	101
Devonport	AUDPO	146.3663100	-41.18003000	94
Portland	AUPTJ	141.6150800	-38.35179000	85
Bunbury	AUBUY	115.6665500	-33.32127000	81
Perth	AUPER	115.8554770	-31.96527778	61
Weipa	AUWEI	141.8697740	-12.67042600	57
Wyndham	AUWYN	128.1011700	-15.45192700	51
Burnie	AUBWT	145.9111800	-41.05032000	46
Geraldton	AUGET	114.5930900	-28.77461000	42
Albany	AUALH	117.8892400	-35.03078000	33
Port Lincoln	AUPLO	135.8695100	-34.72168000	27
Whyalla	AUWYA	137.5944290	-33.01487000	26
Gove	AUGOV	136.6783200	-12.20134000	10
Bundaberg	AUBDB	152.3884570	-24.76722222	4
Port Bonython	AUPBY	137.7641800	-32.99250000	3
Onslow	AUONS	115.1012200	-21.63765000	3
Botany Bay	AUBTB	151.2166000	-33.97067000	2

Table 4.4.: Number of international containers discharging at major Australian ports in2017–2018. Data obtained from the Department of Agriculture.



4.2.5. Pathway: Vessels

Unlike the container pathway (see below), where pests may be transported great distances from the point of entry, stowaway pests are more likely to disperse in the immediate vicinity of the discharge port – either by natural means or via the initial unloading or short-distance movement of vessel items. To that end, we assumed the spatial distribution of the likelihood of a pest arriving via this pathway would best be described by the number of containers discharged at each port (supplied by DAWE; Table 4.4), Container_{*j*} – a variable likely to be highly correlated with vessel numbers ⁴ – and then weighting this by a negative exponential distance-decay function for each port, Port distance_{*i*,*j*}. Specifically, we first constructed a weighted grid for each port using:

Port weight_{*i*, *j*} = Containers_{*j*} × exp(
$$\beta$$
 × Port distance_{*i*,*j*}) (4.9)

where Port weight_{*i*,*j*} produces the number of containers for each raster cell, *i*, associated with port, *j*. Like the tourist and Torres Strait pathways, β describes how steep the distance decay should be (i.e. how severely should dispersal be limited) and ultimately will depend on the pest's initial dispersal capabilities. Effectively, Port weight_{*i*,*j*} is our bias grid for each port, where more risk is allocated to regions close to ports that receive larger quantities of containers (and by implication vessels).

In order to distribute national numbers of international vessel visits across all ports, we then sum these risk scores across ports for each *i*, and then normalise these such that we receive a proportion of risk attributed to each raster cell *i*, that can then be multiplied by the national number of vessels arriving into Australia:

$$\text{Vessel arrivals}_{i} = \frac{\sum_{j=1}^{n} \text{Port weight}_{i,j}}{\sum_{i=1}^{n} \sum_{j=1}^{n} \text{Port weight}_{i,j}} \times \text{Vessels}$$
(4.10)

⁴We were unable to obtain accurate numbers of vessels (and the type of vessel) docked at each port. If this data becomes available, it can be readily substituted into the model.



4.2.6. Pathway: Imported Machinery

Based on 2015–2016 data supplied by DAWE, Australia imported approximately 1,216,726 units of machinery from overseas, with the vast majority being new vehicles (Table 4.5).

As we did not have spatially explicit data describing how these imported machinery were distributed post-border, coupled with the fact that new vehicles were the dominant good imported via this pathway, we assumed that machinery was distributed as a function of population density (see Fig. 4.4). Specifically, the number of imported machinery units arriving in a given grid cell, *i*, was estimated using:

Imported machinery_i = $\frac{\text{population density}_i}{\sum_{i=1}^{n} \text{population density}_i} \times \text{Machinery imports},$ (4.11)

where *Machinery imports* is the total number of machinery items imported into Australia.

Table 4.5.: Breakdown of goods imported via the machinery pathway. Data based on2015–2016 data and obtained from the Department of Agriculture, Water andthe Environment.

Machinery type	Percentage
Conveyances	0.38
Parts	3.48
Used machinery	0.14
New machinery	0.69
New vechicles	92.79
Used vechicles	1.43
Tools	1.10



4.2.7. Pathway: Imported nursery stock

Based on 2015–2016 data supplied by the DAWE, Australia imported approximately 4,475 units of nursery stock from overseas. These goods are imported by a variety of small, medium and large businesses and are likely to be distributed according to population density – the higher the population, the greater the demand (see Fig. 4.4). Thus, we estimate, the number of imported nursery items arriving in a given grid cell, i, using:

Imported nursery stock_i = $\frac{\text{population density}_i}{\sum_{i=1}^{n} \text{population density}_i} \times \text{Imported nursery stock}, (4.12)$

where *Imported nursery stock* is the total number of nursery items imported into Australia.

4.2.8. Pathway: Imported plant-based food

Imported food (e.g. fruit and vegetables), while often cold-treated, can still pose a significant entry risk for some plant pests. Based on 2015–2016 data supplied by DAWE, Australia imported approximately 15,832 units of plant-based food from overseas. While detailed information on the final destination of these goods remains unknown, it is highly likely that because they are for human consumption they will be distributed according to human population density (see Fig. 4.4). Thus, we estimate the number of imported food items arriving in a given grid cell, *i*, using:

Imported food_i =
$$\frac{\text{population density}_i}{\sum\limits_{i=1}^{n} \text{population density}_i} \times \text{Imported food},$$
 (4.13)

where *Imported food* is the total number of plant-based food items imported into Australia.



4.3. Expected number of pest arrivals

Assuming the number of units associated with each pathway have been distributed across the country, we can calculate the expected number of pest arrivals at a location. This is done by multiplying the expected number of pathway units in each raster cell, Pathway units_{*i*,*k*}, by the respective pathway per unit probability, $Pr(contamination_k)$, derived from RRRA or other pathway methods. For example, if a pest could enter the country via the four pathways described above (i.e. tourists, returning residents, mail and Torres Strait), one would first estimate arrivals by pathway:

Pest arrivals_{*i*,*k*} = Pathway units_{*i*,*k*} ×
$$Pr(contamination_k)$$
 (4.14)

The total number of pest arrivals can then be calculated by simply summing the expected number of arrivals in each grid cell *i* across entry pathways, *k*:

Total pest
$$\operatorname{arrivals}_{i} = \sum_{k=1}^{4} \operatorname{Pest} \operatorname{arrivals}_{i,k}$$
 (4.15)

4.4. Abiotic suitability

Climate is likely to be the major abiotic barrier to most pest establishment, especially at large spatial scales (Araújo & Rozenfeld, 2014). The geographic distribution of suitable climate can be estimated using a wide variety of approaches including climatmatching algorithms (e.g. CLIMATCH, CLIMEX; Crombie et al., 2008; Kriticos et al., 2015), environmental convex hulls and Range Bagging (e.g. Drake, 2015), two-class correlative species distribution models (e.g. Maxent Phillips et al., 2006), physiological models (e.g. NicheMapper; Kearney & Porter, 2017), semi-mechanistic models (e.g. CLIMEX; Kriticos et al., 2015), or when data is poor, expert-derived suitability maps (e.g. Martin et al., 2015). Many tools exist, and there are diverse opinions on how to use them, but there remains no strong evidence of a single *best* approach for predicting an invasive species' potential distribution (Barry et al., 2015). As a consequence, our establishment likelihood mapping framework is agnostic as to how abiotic suitability is estimated. However, we strongly recommend users consult Chapter 3 to obtain practical guidance on how to robustly estimate climatic suitability for a species. Regardless of the method used, it is important that prior to input into this mapping framework the spatial layer is in the same projection and has the same extent and resolution as the layers used to distribute pest arrival numbers in the previous subsection (e.g. Australian Albers and 1 km resolution).



4.5. Biotic suitability

The presence or absence of host material or suitable habitat will be another major factor affecting a pest's establishment success. The geographic distribution of host material can be informed using a variety of methods and datasets. For example, for a plant disease that disperses via an insect carrier, biotic suitability could be estimated by modelling the climatic distribution of the insect, the host, or a combination of the two. Landuse layers may also be used to inform the distribution of host material – particularly for agricultural plant species.

One such spatial layer that is likely to be highly applicable to most high priority plant pests is the Australian Land Use and Management Classification (ALUMC version 8; ABARES, 2019), provided by the Australian Collaborative Land Use and Management Program. ALUMC is the most comprehensive and up-to-date Australian landuse layer (current layer was compiled in December 2018) that classifies the dominant landuse of every 50×50 m grid cell across Australia in a three-tiered hierarchical structure (Fig. 4.7). The deeper the layer, the more detail is provided about the land use. Other spatial layers that may also be useful for informing habitat suitability include those that attempt to quantify the amount of vegetation cover (fractional vegetation cover) or vegetation greenness (Normalised Difference Vegetation Index; NDVI) or simply the geographic distribution of major vegetation types (National Vegetation Information System; NVIS). Like abiotic suitability, the establishment likelihood mapping framework described in this report is agnostic as to how one estimates biotic suitability, mostly because the types of models and data used will be highly species dependent.



Figure 4.7.: Geographic distribution of secondary ALUMC classifications. ABARES December 2018.



4.6. Estimating establishment likelihoods

Given spatially explicit estimates of the number of arrivals per grid cell, *i*, coupled with the abiotic and biotic suitability of these grid cells, one can then estimate a relative establishment likelihood (i.e. the expected number of arrivals weighted by biotic and abiotic suitability) as:

$$\label{eq:stablishment} \begin{split} \text{Establishment likelihood}_i = \text{Total pest arrivals}_i \times \text{Abiotic suitability}_i \times \text{Biotic suitability}_i, \\ (4.16) \end{split}$$

where abiotic and biotic suitability components are assumed to be normalised (if required), such that their scores are scaled between 0 and 1. This ensures that both biotic and abiotic factors are equally weighted (i.e. are equally important). If there is evidence that these factors should not be equally weighted, an adjustment factor can be added to either suitability score.

5. Case study 1: Oriental fruit fly



Figure 5.1.: A female oriental fruit fly, *Bactrocera dorsalis*, laying eggs by inserting her ovipositor in the skin of a papaya. Photo by Scott Bauer.



5.1. Background

Oriental fruit fly, *Bactrocera dorsalis* (Hendel), poses a substantial threat to horticultural industries in locations where the species could become established, with over 300 different types of fruit and vegetables potentially infested by the species (Vargas *et al.*, 2015). Recently, *B. dorsalis* has undergone significant taxonomic changes such that it now also encompasses *B. philippinensis* Drew and Hancock, *B. papayae* Drew and Hancock and *B. invadens* (Vargas *et al.*, 2015). This expanded classification has substantially increased the geographic range of *B. dorsalis*, which includes the Pacific region, Africa, South America, and parts of the USA (Vargas *et al.*, 2015).

Fruit flies such as oriental fruit fly present a significant phytosanitary threat to the international competitiveness of Australia's horticultural industry. Over the past decade many countries have joined the World Trade Organisation (WTO), which requires members to manage imports according to the Sanitary and Phytosanitary Measures Agreement (SPS Agreement) and the International Standards for Phytosanitary Measures (ISPMs). In many cases, this has led to trading partners imposing stringent quarantine measures that require increasing evidence of pest freedom in order to justify phytosanitary standards. This, in turn, has put strain on Australia's competitive ability to gain and maintain market access in the international horticultural market.

In the effort to minimise additional phytosanitary procedures and maximise international market access, each Australian state has established networks of fruit fly lure traps. These networks of traps have been set up according to the draft national code of practice (Anon, 2008) and international guidelines (International Atomic Engergy Agency, 2003; International Plant Protection Convention, 2006) for fruit fly surveillance (International Plant Protection Convention, 2006) and are a primary tool used to provide evidence of fruit fly freedom and early detection. In the following sections we develop a map of establishment likelihood for *Bactrocera dorsalis* to facilitate the placement of finite surveillance resources for early detection.

5.2. High risk pathways

The Risk Return Resource Allocation Model (RRRA) identified three high risk pathways that could allow *Bactrocera dorsalis* to enter mainland Australia: tourists, returning residents and passengers travelling from the Torres Strait into Cairns (Table 5.1). For details on how units associated with each pathway were distributed post-border see Chapter 4.

Tourist contamination likelihoods

The Department of Agriculture, Water and Environment (DAWE) considered tourists entering Australia to be less biosecurity aware, and thus, more likely to be carrying fruit fly contaminated goods. In order to incorporate this belief, we assumed the that two-thirds of the RRRA air passenger risk was assigned to tourists with the remaining third assigned to returning residents. It is expected that this weighting will be revised once empirical data become available. **Table 5.1.:** Oriental fruit fly pathway likelihoods with associated expected number of post-border entries per annum based on estimated pathway volumes. Data derived from RRRA, with two-thirds of the air passenger risk assigned to tourists and the remaining third assigned to returning residents.

Pathway	Pr(Contamination)	Units	Expected entries
Tourists	1.57e - 06	13,941,270	21.9
Returning residents	7.86e - 07	15,486,050	12.2
Torres Strait visitors	4.12e - 04	51,000	21.0
TOTAL			55.1

5.3. Abiotic suitability

The potential distribution of oriental fruit fly (*Bactrocera dorsalis*) has been approximated using a variety of techniques such as CLIMEX (Stephens *et al.*, 2007), Maxent and GARP (De Meyer *et al.*, 20010) as well as several other techniques tested as part of this project (Figure 5.2 & Appendix C). While some broad similarities exist (e.g. Central Africa and northern parts of South America are highly suitable) substantial differences also exist between models. This is particularly obvious when examining the distribution of suitable climate in Australia with CLIMEX indicating much of the central and northern east coast of Australia being highly suitable, while GARP suggests the northern coastline is most suitable and Maxent suggests minimal suitable climate in Australia. The reasons for such discrepancies are many, each approach relies on different assumptions, the models were parametrised and validated using different sources of data and information, and the choice of covariates or constraining factors differed among the models.





Figure 5.2.: CLIMEX, GARP and Maxent model estimates of the global distribution of suitable climate for oriental fruit fly (*Bactrocera dorsalis*) or its synonym *Bactrocera invadens*. Dark colours signify greater suitability. CLIMEX (Stephens *et al.*, 2007); GARP and Maxent (De Meyer *et al.*, 20010)

While climatic influences on some biological processes have been estimated and incorporated into published CLIMEX models for this pest (Stephens *et al.*, 2007), we decided against using this model. We did this because while CLIMEX models do use some biological information, they are not truly mechanistic because they still require parameter fitting via inference from the pest's known distribution – typically via manual tweaking of parameters (Kriticos *et al.*, 2015). This manual tweaking of parameters introduces a high subjectivity to the model fitting, which is exacerbated by the reliance on a single parametrisation of the model.

Instead, based on work conducted as part of this project (see Appendix C), we opted to use a novel method known as range bagging. Range bagging was recently proposed by Drake (2015). This algorithm uses presence-only data to estimate the environmental limits of species' habitat by subsetting the multidimensional environments (to user-defined levels of dimensionality), and then using convex hulls to estimate boundaries in each subset of environmental dimensions. Range bagging repeatedly fits models to a random assortment of occurrence samples (and covariate choices) and averages the outcome by using votes (how often a given environment occurs inside niche boundaries) on the ranges of convex hulls obtained from bootstrap samples across all the environmental dimensions. Effectively, its output is the proportion of models that estimate a given location has suitable climate. For example, a suitability score of 0.1 would indicate that only 10% of the ensembled convex hulls deemed that location climatically suitable (i.e. contained the environmental conditions at that location). By contrast, a score of 0.9 would indicate that 90% of estimated convex hulls deemed that location climatically suitable.

The approach has seen recent applications to invasion biology, and appears promising for biosecurity applications. Part of the appeal of this approach is that absences or background data are not required – only presence data are needed (see Chapter 3 & Appendix C). This in turn, removes a number of subjective decisions required in the modelling process and instead focuses solely on the data which we do have – presence locations. Another major advantage of range bagging is that it can readily be used to deal with uncertainty in covariate selection. This is done by specifying low dimensionality (e.g. 2-dimensions) and allowing the algorithm to randomly select from among a suite of possible covariates – effectively resulting in an ensemble of hundreds of competing models. Finally, range bagging also has the advantage of generating outputs that are directly comparable across species, since the predicted quantity (i.e. the proportion of replicate hulls that contain given environmental conditions) is constant across models. In contrast, presence-background methods can be poorly calibrated, with varying estimated relationships between predicted suitability and true probability of occurrence (e.g. Guillera-Arroita *et al.* 2014), complicating such comparison.

Here, we fit range bagging models using code supplied by Drake 2015, with dimensionality set to 2 (meaning only two covariates are fitted at a time), the number of bootstrapped models set to 100 and the proportion of occurrence records used per model set at 0.5. We allowed the algorithm to sample from all 19 WorldClim version 2 (Fick & Hijmans, 2017) bioclimatics parameters (i.e. BIO01 to BIO19) derived from the published 10 minute (approximately 20 km resolution) raster layers. We used "simple" two-dimensional combinations in order to minimise biases associated with model over-fitting (see Chapter 3 and Appendix C).

We sourced occurrence records of *Bactrocera dorsalis*, including its synonyms *B. philippinensis*, *B. papayae* and *B. invadens* (Vargas *et al.*, 2015), directly from GBIF and supple-



mented these data with those collated by Hill *et al.* 2017. Prior to running the range bagging algorithms we first cleaned these occurrence records by using cleaning routines in the recently published CoordinateCleaner R package (Zizka *et al.*, 2019). Specifically, we removed records that:

- 1. had equal latitude and longitudes or were within 0.5 degrees radius of coordinates 0,0;
- 2. were within a 5 km radius of a capital city centroid;¹
- 3. were within a 10 km radius of either a country or province centroid;
- 4. were within 1 degree radius around the GBIF headquarters in Copenhagen, Denmark;
- 5. were within 100 m radius around the centroids of known biodiversity institutions;
- 6. were located in the ocean; or
- 7. contained mismatches between the country specified in the record and the record's coordinates.

We also removed duplicate records and thinned occurrence records to one point per 20 km (the resolution of the WorldClim version 2 climate data). Following this, we removed all occurrence records that were present in countries with no known establishment records based on CABI country-level distributional data. This ensured that the remaining occurrences were most likely from established populations, and thus suitable for inclusion in the range bagging analysis. Models were then fitted subsequently projected and resampled to a 1 km grid – our desired map resolution. The mapped distributional data and model outputs can be seen in Figures 5.3 & 5.4.

¹In large databases such as GBIF, if a record does not have accurate GIS coordinates, users sometimes provide country or capital city coordinates. While such coordinates are useful for providing information on the country or province of the record, they are likely to be too coarse when estimating climatic suitability for a species, which involves mapping occurrence records onto fine-resolution gridded climatic data.



Figure 5.3.: Global climatic suitability for oriental fruit fly estimated from range bagging. Blue points indicate cleaned distributional data collated from GBIF and Hill *et al.* 2017. Raster resolution: 10 minutes (approx 20 km). Suitability is the proportion of ensembled convex hulls that identify a location as climatically suitable across bootstrapped combinations of environmental variables.




Figure 5.4.: Australian projected climatic suitability for oriental fruit fly derived from range bagging. Raster aggregated to a 5 km grid for presentation purposes. Suitability is the proportion of ensembled convex hulls that identify a location as climatically suitable across bootstrapped combinations of environmental variables.



5.4. Biotic suitability

We utilised the tertiary raster layer of Australian Land Use and Management Classification (ALUMC version 8; ABARES, 2019), provided by the Australian Collaborative Land Use and Management Program, to define the presence and absence of fruit fly host material. Specifically, we created a binary raster whereby any 50 m² grid cell containing commodities vulnerable to fruit fly, such as tree fruits, vine fruits, olives, citrus, vegetables, herbs, and shrub berries, as well as urban areas, was assumed to contain host material and assigned a value of 1 (Table 5.2). All other grid cells containing other dominant land uses were classified as 0 and assumed to contain no fruit fly host material. This binary raster was then converted to a 1 km grid resolution by finding the maximum value of aggregated grid cells. Note that this layer is an approximation of the geographic distribution of host material, and does not account for possible native and weed hosts, for which we currently have incomplete distributional data.

While the ALUMC binary layer provides insights into the presence or absence of fruit fly host material, it does not provide an indication of the abundance of host material in any given season. We approximated the host abundance (conditional on the host being present) by multiplying the binary host presence raster with average Normalised Difference Vegetation Index (NDVI; a measure of raster cell greenness), scaled between 0 and 1, observed between October 2018 to March 2019 (The period when the flies would be most active; Fig. 5.5). We obtained this data from the Bureau of Meteorology at a raster resolution of 0.05 by 0.05 degrees (approximately 5 km scale), which was then resampled to a 1 km grid. The resulting biotic suitability map can be seen in Fig. 5.6).

Biotic suitability_i = Host presence_i × $\frac{\text{NDVI}_i - min(\text{NDVI}_i)}{max(\text{NDVI}_i) - min(\text{NDVI}_i)}$ (5.1)

Landuse	Code
Perennial horticulture	340
Perennial tree fruits	341
Perennial oleaginous fruits	342
Perennial tree nuts	343
Perennial vine fruits	344
Perennial shrub nuts, fruits and berries	345
Perennial vegetables and herbs	347
Perennial citrus	348
Perennial grapes	349
Seasonal horticulture	350
Seasonal fruits	351
Seasonal vegetables and herbs	353
Abandoned perennial horticulture	365
Irrigated perennial horticulture	440
Irrigated perennial tree fruits	441
Irrigated perennial oleaginous fruits	442
Irrigated perennial tree nuts	443
Irrigated perennial vine fruits	444
Irrigated perennial shrub nuts, fruits and berries	445
Irrigated perennial vegetables and herbs	447
Irrigated citrus	448
Irrigated grapes	449
Irrigated seasonal horticulture	450
Irrigated seasonal fruits	451
Irrigated seasonal vegetables and herbs	453
Abandoned irrigated perennial horticulture	465
Residential and farm infrastructure	540
Urban residential	541
Rural residential with agriculture	542
Rural residential without agriculture	543
Remote communities	544

 Table 5.2.: ALUMC landuses likely to contain fruit fly host material.





Figure 5.5.: Average Normalised Differential Vegetation Index (NDVI) observed between October 2018 and March 2019. Data obtained from the Australian Bureau of Meteorology. Raster aggregated to a 5 km grid for presentation purposes.





Figure 5.6.: Distribution of suitable biotic environment for oriental fruit fly. Estimated as a function of ALUMC and NDVI data. Raster aggregated to a 5 km grid for presentation purposes.



5.5. Establishment likelihoods

According to RRRA, *Bactrocera dorsalis* is most likely to enter via three pathways: international tourists, returning residents and air passengers from the Torres Strait (Table 5.1). Tourists are distributed as a function of the density of tourist accommodation and distance from major airport. By contrast returning residents are distributed as a function of population density, while the Torres Strait arrivals are distributed as a function of both population density and distance from Cairns airport. The implication of these assumptions is that in the days following arrival the majority of arrivals will likely occur in major metropolitan areas that are close to airports and contain high densities of tourist accommodation.

However, whether these arrivals are likely to result in establishment depends on both the abiotic (e.g. climate suitability) and biotic (e.g. presence of food source) environment (Fig. 5.7). In terms of food availability (i.e. biotic suitability), the expectation is that all Australian cities are likely to be suitable because residents grow host fruit and vegetables in backyards. However, what differentiates establishment likelihoods between cities is the climate suitability. For example, we estimated a relatively high arrival rate for Bactrocera dorsalis - a tropical fruit fly - in and around Melbourne, but the climate suitability in this region is much lower compared to more tropical areas such as Cairns (which also has a high arrival rate) (see Fig. 5.4). Thus, while Melbourne contained one of the highest arrival rates, the highest establishment likelihoods for this species were found in more tropical metropolitan cities (e.g. Brisbane and Cairns; Fig. 5.8). Fig. 5.9–5.16 provide more detailed maps of the geographic distribution of *Bac*trocera dorsalis establishment likelihoods around major cities and tourist locations. For finer scale (i.e. 1 km resolution) maps please see associated interactive maps attached to this report. Interactive maps on expected number of arrivals across all pathways (not shown here) are also included as additional files attached to this report.





Figure 5.7.: Estimated environmental suitability for oriental fruit fly. Suitability is defined as the product of abiotic (i.e. climate suitability) and biotic suitability (i.e. host presence × NDVI). Raster aggregated to a 5 km grid for presentation purposes





Figure 5.8.: National establishment likelihood map for oriental fruit fly. Raster aggregated to a 5 km grid for presentation purposes. Only establishment scores above 10^{-8} plotted.





Figure 5.9.: Oriental fruit fly establishment likelihood map for greater Cairns. Only establishment scores above 10^{-8} plotted.





Figure 5.10.: Oriental fruit fly establishment likelihood map for greater Brisbane. Only establishment scores above 10^{-8} plotted.





Figure 5.11.: Oriental fruit fly establishment likelihood map for greater Sydney. Only establishment scores above 10^{-8} plotted.





Figure 5.12.: Oriental fruit fly establishment likelihood map for greater Melbourne. Only establishment scores above 10^{-8} plotted.





Figure 5.13.: Oriental fruit fly establishment likelihood map for greater Hobart. Only establishment scores above 10^{-8} plotted.





Figure 5.14.: Oriental fruit fly establishment likelihood map for greater Adelaide. Only establishment scores above 10^{-8} plotted.





Figure 5.15.: Oriental fruit fly establishment likelihood map for greater Perth. Only establishment scores above 10^{-8} plotted.





Figure 5.16.: Oriental fruit fly establishment likelihood map for greater Darwin. Only establishment scores above 10^{-8} plotted.

6. Case study 2: Khapra beetle



Figure 6.1.: An adult khapra beetle (*Trogoderma granarium*). Photo by Udo Schmidt.

6.1. Background

Khapra beetle (*Trogoderma granarium*) is a highly destructive pest of stored grain and dry foodstuff that has caused losses of up to 74% (Botha *et al.*, 2005). It is highly resistant to many contact insecticides or fumigants and can undergo diapause under adverse conditions for significant lengths of time, proving difficult to eradicate once established (CABI).

Currently, khapra beetle occurs in many middle-eastern countries (e.g. Iran & Iraq), Africa and South Asia, but remains absent from Australia, New Zealand, South-East Asia and South America (CABI). Countries without this pest maintain strong quarantine controls and restrictions to mitigate risk of importing the pest and to prevent market access for countries with known infestations. Consequently, khapra beetle poses a significant threat to Australia's export-orientated grain industry.

To facilitate decisions associated with the deployment of finite surveillance resources for early detection, we developed a map of establishment likelihoods for khapra beetle (*Trogoderma granarium*). The details of this map, including how it was constructed, can be found in the following sections.

6.2. High risk pathways

The Risk Return Resource Allocation Model (RRRA) identified five high risk pathways that could allow *Trogoderma granarium* to enter Australia: tourists, returning residents, international mail, imported fertiliser and international vessels (Table 6.1). See Chapter 4 for details on how units associated with each pathway were distributed postborder. For the purpose of this analysis, we assumed that arrivals associated with vessels would decay exponentially with distance, such that 50% of the arrivals would occur within 1 km of the port of arrival (i.e. β was assigned as $-\log(0.5)/1 = -0.693$; Section 4.2.5; Fig. 6.2). We deemed this appropriate, because khapra beetle cannot fly and is therefore highly unlikely to travel large distances. Rather, it is likely to disperse in close proximity to ports via natural movement, or more likely, short distance hitch-hiking events where goods are taken off vessels.

Table 6.1.: Khapra beetle pathway likelihoods with associated expected number of postborder entries per annum based on estimated pathway volumes. Data derived from RRRA, with two-thirds of the air passenger risk assigned to tourists and the remaining third assigned to returning residents.

Pathway	Pr(Contamination)	Units	Expected entries
Tourists	1.07e - 06	13,941,270	15.0
Returning residents	5.37e - 07	15,486,050	8.3
Mail	8.01e - 08	152,000,000	12.2
Fertiliser	3.92e - 04	6,907	2.7
Vessels	1.44e - 04	18,612	2.7
TOTAL			40.9



Tourist contamination likelihoods

The Department of Agriculture, Water and Environment (DAWE) considered tourists entering Australia to be less biosecurity aware, and thus, more likely to be carrying goods contaminated by khapra beetle. In order to incorporate this belief, we assumed that two-thirds of the RRRA air passenger risk was assigned to tourists, with the remaining third assigned to returning residents. It is expected that this weighting will be revised once empirical data become available.



Figure 6.2.: Log10 vessel pathway distance decay weights for khapra beetle. A negative exponential decay function was used where 50% of arrivals from vessels would be distributed within 1 km of a major marine port.



6.3. Abiotic suitability

Khapra beetle is typically a pest of hot, dry climates or of commodities stored elsewhere in hot dry conditions (Banks, 1977). While previous research (e.g. Banks, 1977; Botha *et al.*, 2005) has indicated that much of Australia's interior, including some grain growing regions, are likely to be climatically suitable, the pest is generally believed to be capable of establishing in most locations. This is because: 1) the pest commonly occurs in stored grain and food products ¹, and thus, is unlikely to be subjected to ambient climatic conditions reflected in the parametrisation of climate suitability models; and 2) under adverse conditions, this species can undergo significant periods of diapause, allowing larvae to persist for approximately 9 months without food and up to 6 years with food (Burges, 1963). As such, for the purposes of developing a map of establishment likelihoods, we assumed that the environmental conditions within storage facilities containing grain and other vulnerable food commodities would be equally suitable for khapra beetle to establish. Thus, no climatic suitability model was constructed for this pest.

6.4. Biotic suitability

We utilised the tertiary raster layer of the Australian Land Use and Management Classification (ALUMC version 8; ABARES, 2019) to define the presence and absence of khapra beetle host material. Specifically, we created a binary raster whereby any 50 m² grid cell likely to contain storage of commodities vulnerable to khapra beetle, such as oilseeds, cereals or pulses (CABI), were assumed to contain host material and assigned a value of 1 (Table 6.2). Urban areas, feedlots and other intensive animal farming (e.g. dairy, piggeries, poultry) were also assumed to contain host material, as they are likely to contain various quantities of local and imported stored grain and other dry foodstuff. Crop fields of vulnerable commodities were not included as suitable locations as there has been no documented cases of infestations on crops (CABI). All other grid cells containing other dominant land uses were classified as 0 and assumed to contain no khapra beetle host material. This binary raster was then converted to a 1 km grid resolution by finding the maximum value of aggregated grid cells (Fig. 6.3).

¹There have been no documented infestations of unharvested grain in and around farms (CABI).



Landuse	Code
Intenstive animal production	520
Dairy sheds and yards	521
Feedlots	522
Poultry farms	523
Piggeries	524
Horse studs	526
Saleyards/stockyards	527
Abandoned intenstive animal production	528
Manufacturing and industrial	530
General purpose factory	531
Food processing factory	532
Major industrial complex	533
Bulk grain storage	534
Abandoned manufacturing and industrial	538
Residential and farm infrastructure	540
Urban residential	541
Rural residential with agriculture	542
Rural residential without agriculture	543
Remote communities	544
Farm buildings/infrastructure	545
Ports and water transport	574

Table 6.2.: ALUMC landuses likely to contain khapra beetle host material.





Figure 6.3.: Distribution of suitable biotic environment for khapra beetle. Estimated from ALUMC data. Raster aggregated to a 5 km grid for presentation purposes.



6.5. Establishment likelihoods

According to RRRA, khapra beetle is most likely to enter via five pathways: air passengers (tourists & returning residents), international mail, imported fertiliser and international vessels (Table 6.1). In our establishment likelihood mapping framework, both international mail and returning residents are distributed as a function of human population density. By contrast, international tourists were distributed as a function of the density of tourist accommodation and distance from international airport, fertiliser imports were distributed according to NRM usage statistics, and arrivals by vessels were distributed as a function of the proportion of containers arriving at a port (assumed to be correlated with number of vessels arriving) and distance from port. The implication of these assumptions is that in the days following entry the majority of arrivals will likely occur in major metropolitan areas that are close to ports (both air and marine), and contain high densities of tourist accommodation. Consequently, Sydney and Melbourne are expected to receive the most arrivals due to their high population density and tourist accommodation coupled with each containing ports that receive a substantial proportion of the total number of containers (and presumably international vessels) entering Australia. In these regions it is likely that the majority of khapra beetles will end up in residential pantries, and therefore, pose little risk of further spread and establishment. In agricultural regions, khapra beetle establishment, at least according to RRRA, is most likely to occur in regions that use imported fertiliser. However, the likelihood of khapra beetle arrival and establishment in these regions is much lower than in urban centres due to the majority of arrivals entering via international air passengers and mail (Table 6.1).

As the locations most likely to receive khapra beetle are also those which are most likely to contain appropriate food sources (Fig. 6.3), the establishment map (Fig. 6.4) is largely a function of where khapra beetle is most likely to arrive. Figures 6.5–6.12 provide more detailed maps of the geographic distribution of *Trogoderma granarium* establishment likelihoods around major cities and tourist locations. For finer scale (i.e. 1 km resolution) maps please see associated interactive maps attached to this report. Interactive maps on expected number of arrivals across all pathways (not shown here) are also included as additional files attached to this report.





Figure 6.4.: National establishment likelihood map for khapra beetle. Raster aggregated to a 5 km grid for presentation purposes. Only establishment scores above 10^{-8} plotted.





Figure 6.5.: Khapra beetle establishment likelihood map for greater Cairns. Only establishment scores above 10^{-8} plotted.





Figure 6.6.: Khapra beetle establishment likelihood map for greater Brisbane. Only establishment scores above 10^{-8} plotted.





Figure 6.7.: Khapra beetle establishment likelihood map for greater Sydney. Only establishment scores above 10^{-8} plotted.





Figure 6.8.: Khapra beetle establishment likelihood map for greater Melbourne. Only establishment scores above 10^{-8} plotted.





Figure 6.9.: Khapra beetle establishment likelihood map for greater Hobart. Only establishment scores above 10^{-8} plotted.





Figure 6.10.: Khapra beetle establishment likelihood map for greater Adelaide. Only establishment scores above 10^{-8} plotted.





Figure 6.11.: Khapra beetle establishment likelihood map for greater Perth. Only establishment scores above 10^{-8} plotted.





Figure 6.12.: Khapra beetle establishment likelihood map for greater Darwin. Only establishment scores above 10^{-8} plotted.

7. Case study 3: Brown marmorated stink bug (BMSB)



Figure 7.1.: Adult brown marmorated stink bug (Halyomorpha halys). Photo by Alpsdake.



7.1. Background

Brown marmorated stink bug (BMSB; *Halyomorpha halys*), also known as yellow-brown stink bug, is a highly polyphagous (>100 hosts) plant pest that poses a substantial threat to horticulture worldwide (CABI 2019). Over the last two decades, BMSB has rapidly spread from its native range in East Asia (China, Japan, Korea and Taiwan) into Europe, North America and Canada, and in doing so, has cause significant agricultural losses (Rice *et al.*, 2014; Valentin *et al.*, 2017). Australia and New Zealand are currently free of BMSB, despite large regions in both countries being deemed climatically suitable (Zhu *et al.*, 2012; Fraser *et al.*, 2017; Kriticos *et al.*, 2017). However, maintaining this pest-free status is becoming increasingly difficult as Australians import larger quantities of potentially contaminated goods from an expanding list of countries, ultimately increasing the risk to Australia's billion dollar horticultural industry.

Interception data coupled with expert opinion indicate that for both Australia and New Zealand the number of BMSB arrivals and the likelihood of establishment will be greatest when the pest over winters in the Northern Hemisphere. This is because when BMSB over winters it aggregates in large numbers in both residential and industrial buildings. As a consequence, BMSB is a common stowaway in passenger luggage and imported bulk freight, cargo and vehicles (DAWR, 2017; Ormsby, 2018). Moreover, because large numbers of this pest can be found in stowaway, successful establishment of this pest will be more likely as it can more easily overcome possible founder effects that inhibit the successful establishment of many other pests.

In an effort to facilitate decisions associated with allocation of finite surveillance resources for early detection, we developed a map of establishment likelihoods for brown marmorated stink bug (*Halyomorpha halys*). The details of this map, including how it was constructed, can be found in the following sections.

7.2. High risk pathways

The Risk Return Resource Allocation Model (RRRA) identified five high risk pathways that could allow brown marmorated stink bug (*Halyomorpha halys*) to enter Australia: tourists, returning residents, imported nursery stock, machinery, and plant-based food (Table 7.1). Based on RRRA estimates, international tourists and returning residents are expected to be the highest risk pathways in terms of expected number of post-border arrivals, followed by imported nursery stock and plant-based food. These estimates appear to be in contradiction to recent interception data from both Australia and New Zealand, which indicates that in addition to air passengers, imported machinery – particularly from Asia and Europe – is likely to be a significant pathway of entry for this pest (DAWR, 2017; Ormsby, 2018). This contradiction between RRRA and interception data is most likely because RRRA likelihoods for this pest are derived from an aggregate group of pests that encompasses horticultural bugs, thrips and mites.

For the purpose of this project, we used RRRA pathway likelihoods to create a map of establishment likelihood for this pest. However, we also included the machinery pathway into the model, despite RRRA indicating it is responsible for a very low expected number of post-border arrivals.

Caution!

We recommend that this establishment likelihood map be interpreted with caution due to the issues associated with pathway leakage estimates (see above), and ultimately this map should be revised when revised pathway leakage rates become available.

Table 7.1.: Brown marmorated stink bug pathway likelihoods with associated expected number of entries per annum based on estimated pathway volumes. Data derived from RRRA, with two-thirds of the air passenger risk assigned to toursits and the remaining third assigned to returning residents.

Pathway	Pr(Contamination)	Units	Expected entries
Tourists	1.57e - 06	13,941,270	21.900
Returning residents	7.86e - 07	15,486,050	12.200
Nursery stock	5.34e - 04	4,475	2.400
Machinery	1.40e - 09	1,216,726	0.002
Plant-based food	1.49e - 04	15,832	2.400
TOTAL			38.800

Nourist contamination likelihoods

The Department of Agriculture, Water and Environment (DAWE) considered tourists entering Australia to be less biosecurity aware, and thus, more likely to be carrying BMSB contaminated goods. In order to incorporate this belief, we assumed the that two-thirds of the RRRA air passenger risk was assigned to tourists with the remaining third assigned to returning residents. It is expected that this weighting will be revised once empirical data become available.



7.3. Abiotic suitability

The potential distribution of brown marmorated stink bug (*Halyomorpha halys*) has been approximated using a variety of methods such as CLIMEX (Kriticos *et al.*, 2017), Maxent (Zhu *et al.*, 2012; Fraser *et al.*, 2017), Random forests (Fraser *et al.*, 2017), Support Vector Machines (Fraser *et al.*, 2017) and ensembles of multiple of these methods (Fraser *et al.*, 2017). Much like in the oriental fruit fly case study (Chapter 5), while some models exhibit some similarities (e.g. eastern USA is almost always classified as suitable) substantial differences also exist between models (Fig. 7.2 & 7.3). This is particularly obvious when examining the distribution of suitable climate in Australia between the models. In the CLIMEX model, much of the central and north-east coast of Australia is estimated to be highly suitable. By contrast, the Maxent models compiled by Zhu *et al.* (2012) and to a lesser degree Fraser *et al.* (2017) indicate that the south-east and south-west of Australia are most optimal. Again, the reasons for such discrepancies are many, each approach relies on different assumptions, the models were parametrised and validated using different sources of data and information, and the choice of covariates or constraining factors differed among the models.




Figure 7.2.: Published brown marmorated stink bug (*Halyomorpha halys*) CLIMEX (Kriticos *et al.*, 2017) and Maxent (Zhu *et al.*, 2012) distribution models. Darker shades equal more suitable climate.





Figure 7.3.: Brown marmorated stink bug (*Halyomorpha halys*) Random forest, Maxent and Support Vector Machine distribution models published in Fraser *et al.* (2017).



While climatic influences on some biological processes have been estimated and incorporated into published CLIMEX models for this pest (Kriticos *et al.*, 2017), we opted to use a range bagging model. We did this because while CLIMEX models do use some biological information, they are not truly mechanistic because they still require parameter fitting via inference from the pest's known distribution – typically via manual tweaking of parameters. This manual tweaking of parameters introduces a high subjectivity to the model fitting, which is exacerbated by the reliance of a single parametrisation of the model. By contrast, range bagging, while a correlative model, is less prone to manual tweaking of parameters and most importantly allows one to ensemble a range of plausible models (Appendix C). Furthermore, its output is easily interpretable with its suitability score indicating the proportion of ensembled models that deemed a given location as climatically suitable. Lastly, the other reason we opted for range bagging was that the estimated potential distribution from this model, relative to other published models, better matched the distribution thought probable by DAWE in-house BMSB experts.

We used the range bagging algorithm with dimensionality set to 2 (meaning only two covariates are fitted at a time), the number of bootstrapped models set to 100 and the proportion of occurrence records used per model set at 0.5. We allowed the algorithm to sample from all 19 WorldClim version 2 (Fick & Hijmans, 2017) bioclimatic parameters (i.e. BIO01 to BIO19) derived from the published 10 minute (approximately 20 km resolution) raster layers. We used "simple" two-dimensional combinations in order to minimise biases associated with model over-fitting (see Chapter 3 and Appendix C).

We sourced occurrence records of *Halyomorpha halys* from Kriticos *et al.* 2017. Prior to running the range bagging algorithms we first cleaned these occurrence records by using cleaning routines in the recently published CoordinateCleaner R package (Zizka *et al.*, 2019). Specifically, we removed records that:

- 1. had equal latitude and longitudes or were within 0.5 degrees radius of coordinates 0,0;
- 2. were within a 5 km radius of a capital city centroid;¹
- 3. were within a 10 km radius of either a country or province centroid;
- 4. were within a 1 degree radius around the GBIF headquarters in Copenhagen, Denmark;
- 5. were within a 100 m radius around the centroids of known biodiversity institutions;
- 6. were located in the ocean; or
- 7. contained mismatches between the country specified in the record and the record's coordinates.

¹In large databases such as GBIF, if a record does not have accurate GIS coordinates, users sometimes provide either country or capital city coordinates. While such coordinates are useful for providing information on the country or province of the record, they are likely to be too coarse when estimating the climatic suitability of a species, which involves mapping occurrence records onto fine-resolution gridded climatic data.



We also removed duplicate records and thinned them to one point per 20 km (the resolution of the WorldClim version 2 climate data). Following this, we removed all occurrence records that were present in countries with no known establishment records based on CABI country-level distributional data. In addition to this, we manually removed occurrences in central and western provinces of Canada (Alberta and Saskatchewan) which have been associated with human-mediated transport from the USA and do not represent established populations (Kriticos *et al.*, 2017). The distribution record near Kenai (Alaska) was also removed as it is considered a vagrant observation, most likely associated with aerial transport (Kriticos *et al.*, 2017). This cleaning ensured that the remaining occurrences were most likely from established populations, and thus suitable for inclusion in the range bagging analysis. Models were then fitted, and subsequently projected and resampled to a 1 km grid – our desired map resolution. The mapped distributional data and model outputs can be seen in Figures 7.4 & 7.5. The outcome of this approach was a suitability map that appeared to be in between the CLIMEX model and the Maxent model produced by Zhu *et al.* (2012).



Figure 7.4.: Global climatic suitability of brown marmorated stink bug derived from range bagging. Blue points refer to cleaned distributional data collated from Kriticos *et al.* (2017). Raster resolution: 10 min (approx 20 km). Suitability is the proportion of ensembled convex hulls that identify a location as climatically suitable across bootstrapped combinations of environmental variables.





Figure 7.5.: Australian climatic suitability for brown marmorated stink bug derived from range bagging. Raster aggregated to a 5 km grid for presentation purposes. Suitability is the proportion of ensembled convex hulls that identify a location as climatically suitable across bootstrapped combinations of environmental variables.



7.4. Biotic suitability

We utilised the tertiary raster layer of Australian Land Use and Management Classification (ALUMC version 8; ABARES, 2019), provided by the Australian Collaborative Land Use and Management Program, to define the presence and absence of brown marmorated stink bug host material. Specifically, we created a binary raster whereby any 50 m² grid cell containing commodities vulnerable to BMSB, such as tree fruits and nuts, vine fruits, olives, citrus and shrub berries (Table 7.2). Given the species also feeds on many ornamental plants present in gardens and street-side vegetation (e.g. maple and oak trees, and many flowering plants), we also included urban and rural residential areas. All other grid cells containing other dominant land uses were classified as 0 and assumed to contain no BMSB host material. This binary raster was then converted to a 1 km grid resolution by finding the maximum value of aggregated grid cells. Note that this layer is an approximation of the geographic distribution of host material, and does not account for possible native hosts, for which we currently have incomplete distributional data.

While the ALUMC binary layer provides insights into the presence or absence of BMSB host material, it does not provide an indication of the abundance of host material in any given season. We approximated the host abundance (conditional on the host being present) by multiplying the binary host presence raster with average Normalised Difference Vegetation Index (NDVI; a measure of raster cell greenness), scaled between 0 and 1, observed between October 2018 to March 2019 (the period when the pest would be most active; Fig. 7.6) using the same approach used in the oriental fruit fly case study (see Section 5.4). We obtained this data from the Bureau of Meteorology at a raster resolution of 0.05 by 0.05 degrees (approximately 5 km scale), which was then resampled to a 1 km grid (Fig. 5.5).

Table 7.2.: ALUMC landuses	likely to contain brown marmorated	stink bug host mate-
rial.		

Landuse	Code	Landuse (cont)	Code (cont)
Cereals	331	Irrigated vine fruits	444
Beverage and spice crops	332	Irrigated shrub berries and fruits	445
Pulses	338	Irrigated perennial flowers and bulbs	446
Perennial horticulture	340	Irrigated perennial vegetables and herbs	447
Perennial tree fruits	341	Irrigated citrus	448
Perennial oleaginous fruits	342	Irrigated grapes	449
Perennial tree nuts	343	Irrigated seasonal horticulture	450
Perennial vine fruits	344	Irrigated seasonal fruits	451
Perennial shrub berries and fruits	345	Irrigated seasonal flowers and bulbs	452
Perennial flowers and bulbs	346	Irrigated seasonal vegetables and herbs	453
Perennial Citrus	348	Abandoned irrigated perennial horticulture	465
Perennial Grapes	349	Intensive horticulture	510
Seasonal horticulture	350	Production nurseries	511
Seasonal fruits	351	Shadehouses	512
Seasonal flowers and bulbs	352	Glasshouses	513
Seasonal vegetables and herbs	353	Glasshouses - hydroponic	514
Irrigated cereals	431	Abandoned intensive horticulture	515
Irrigated beverage and spice crops	432	Residential and farm infrastructure	540
Irrigated pulses	438	Urban residential	541
Irrigated perennial horticulture	440	Rural residential with agriculture	542
Irrigated tree fruits	441	Rural residential without agriculture	543
Irrigated olives	442	Remote communities	544
Irrigated tree nuts	443		





Figure 7.6.: Distribution of suitable biotic environment for brown marmorated stink bug. Estimated as a function of ALUMC and NDVI data. Raster aggregated to a 5 km grid for presentation purposes.



7.5. Establishment likelihoods

According to RRRA, brown marmorated stink bug is most likely to enter via five pathways: tourists, returning residents, imported nursery stock, imported plant-based and to a lesser extent imported machinery² (Table 7.1). In our establishment likelihood mapping framework, returning residents and imported nursery stock, plant-based food and machinery are distributed as a function of human population density. By contrast international tourists were distributed as a function of the density of tourist accommodation and distance from international airports. The implication of these assumptions is that in the days following entry the majority of arrivals will likely occur in major metropolitan areas that are close to airports, and contain high densities of tourist accommodation. Consequently, cities such as Sydney and Melbourne are expected to receive the most arrivals due to their high population density and tourist accommodation.

However, whether these arrivals are likely to result in establishment depends on both the abiotic (e.g. climate suitability) and biotic (e.g. presence of food source) environment (Fig. 7.7). In terms of food availability (i.e. biotic suitability), all Australian cities appear to be suitable because residents are likely to grow host plants in backyards (e.g. citrus, olives, maple trees) and, at a broad scale, have fairly similar spring/summer NDVI. However, what differentiates establishment likelihoods between cities is the climate suitability (Fig. 7.5). Based on the range bagging model, northern Australia contains mostly unsuitable climate for BMSB, meaning that arrivals in places such as Darwin have a lower likelihood of establishment. By contrast, cities such as Brisbane, Sydney, Perth and to a lesser degree Melbourne (i.e. the cities expected to receive the greatest number of arrivals based on population and tourist accommodation) appear to contain optimal climate, and thus, arrivals in these cities are unlikely to be impeded by climate (Fig. 7.8). Figures 7.9–7.16 provide more detailed maps of the geographic distribution of brown marmorated stink bug establishment likelihoods around major cities and tourist locations. For finer scale (i.e. 1 km resolution) maps please see associated interactive maps attached to this report. Interactive maps on expected number of arrivals across all pathways (not shown here) are also included as additional files attached to this report.

²Although there is contradictory evidence to suggest the expected number of arrivals via the imported machinery pathway is much higher than estimated in RRRA.





Figure 7.7.: Estimated environmental suitability for brown marmorated stink bug. Suitability is defined as the product of abiotic (i.e. climate suitability) and biotic (i.e. host presence × NDVI) suitability. Raster aggregated to a 5 km grid for presentation purposes.





Figure 7.8.: National establishment likelihood map for brown marmorated stink bug. Raster aggregated to a 5 km grid for presentation purposes. Only establishment scores above 10^{-8} plotted.





Figure 7.9.: Brown marmorated stink bug establishment likelihood map for greater Cairns. Only establishment scores above 10^{-8} plotted.





Figure 7.10.: Brown marmorated stink bug likelihood of establishment map for greater Brisbane. Only establishment scores above 10^{-8} plotted.





Figure 7.11.: Brown marmorated stink bug establishment likelihood map for greater Sydney. Only establishment scores above 10^{-8} plotted.





Figure 7.12.: Brown marmorated stink bug establishment likelihood map for greater Melbourne. Only establishment scores above 10^{-8} plotted.





Figure 7.13.: Brown marmorated stink bug establishment likelihood map for greater Hobart. Only establishment scores above 10^{-8} plotted.





Figure 7.14.: Brown marmorated stink bug establishment likelihood map for greater Adelaide. Only establishment scores above 10^{-8} plotted.





Figure 7.15.: Brown marmorated stink bug establishment likelihood map for greater Perth. Only establishment scores above 10^{-8} plotted.





Figure 7.16.: Brown marmorated stink bug establishment likelihood map for greater Darwin. Only establishment scores above 10^{-8} plotted.

8. Case study 4: Gypsy moth



Figure 8.1.: Female Asian gypsy moth (*Lymantria dispar asiatica*). Photo by Alexander Schintlmeister.



8.1. Background

Gypsy moth (*Lymantria dispar*) is a highly polyphagous insect defoliator consisting of three sub-species: Asian gypsy moth (*Lymantria dispar asiatica*), Japanese gypsy moth (*Lymantria dispar japonica*), and European gypsy moth (*Lymantria dispar dispar)*. Field and laboratory tests conducted in North America (Liebhold *et al.*, 1995) and Australia (Matsuki *et al.*, 2001) show that more than 650 species of trees and shrubs across as many as 53 families are susceptible to gypsy moth defoliation, including many Australian endemic Acacias and Eucalyptus. Native to Eurasia, the species complex has successfully established in eastern USA, and in doing so has caused significant damage to many important North American tree species with more than 34×10^6 ha of forest being defoliated since 1924 (Tobin *et al.*, 2007).

Gypsy moths are currently absent from both Australia and New Zealand. However, the species poses a significant threat to much of Australia's native vegetation and commericial plantation industry (Matsuki *et al.*, 2001). Thus, in an effort to facilitate decisions associated with where finite surveillance resources should be undertaken for early detection, we developed a map of establishment likelihoods for gypsy moth (*Lymantria dispar*). The details of this map, including how it was constructed, can be found in the following sections.

8.2. High risk pathways

The Risk Return Resource Allocation Model (RRRA) identified three high risk pathways that could allow gypsy moth to enter Australia: imported machinery, containers and vessels (Table 8.1). For details on how units associated with each pathway were distributed post-border see Chapter 4. For the purpose of this analysis, we assumed that arrivals associated with vessels would decay exponentially with distance, such that 50% of the arrivals would occur within 1 km of the port of arrival (i.e. β was assigned as $-\log(0.5)/1 = -0.693$; See section 4.2.5; Figure 8.2). We deemed this as appropriate, because the two primary modes of dispersal for this species from vessels is short distance: 1) wind-borne passive movement of first instars (Liebhold *et al.*, 1992); and 2) hitch-hiking events from goods/luggage taken from vessels.

Pathway	Pr(Contamination)	Units	Expected entries
Machinery	1.11e - 05	1,216,726	13.5
Containers	6.26e - 07	5,125,422	3.2
Vessels	3.45e - 05	18,612	0.6

Table 8.1.: Gypsy moth pathway likelihoods with associated expected number of entries per annum based on estimated pathway volumes. Data derived from RRRA.





Figure 8.2.: Log10 vessel pathway distance decay weights for gypsy moth. A negative exponential decay function was used where 50% of arrivals from vessels would be distributed within 1 km of a major marine port.



8.3. Abiotic suitability

The potential distribution of gypsy moth has been approximated using a variety of CLIMEX models, however, the majority of these have focused explicitly on particular areas of interest (e.g. USA and Europe) with few predicting the global distribution of suitable climate for this species. The two exceptions to this were the CLIMEX model recently published by Paini *et al.* (2018) and the GARP ensemble model published by Peterson *et al.* (2007). Much like in the other case studies these different models exhibited some similarities (e.g. eastern USA and Europe exhibiting highly suitable climate) but also some differences (e.g. central Africa and South America; Fig. 8.3). In terms of Australia, these models exhibited somewhat similar results, with both indicating that the south-west and south-east contained suitable climate.

While climatic influences on some biological processes have been estimated and incorporated into published CLIMEX models for this pest (Paini *et al.*, 2018), we opted to use a range bagging model. We did this because while CLIMEX models do use some biological information, they are not truly mechanistic because they still require parameter fitting via inference from the pest's known distribution – typically via manual tweaking of parameters. This manual tweaking of parameters introduces a high subjectivity to the model fitting, which is exacerbated by the reliance on a single parametrisation of the model. By contrast, range bagging, while a correlative model, is less prone to manual tweaking of parameters, allows one to ensemble across a range of plausible models (Appendix C), and has an output that is readily interpretable and comparable across species – the proportion of ensembled models that deem a location climatically suitable.

Specifically, we used the range bagging algorithm with dimensionality set to 2 (meaning only two covariates are fitted at a time), the number of bootstrapped models set to 100 and the proportion of occurrence records used per model set at 0.5. We allowed the algorithm to sample from all 19 WorldClim version 2 (Fick & Hijmans, 2017) bioclimatic parameters (i.e. BIO01 to BIO19) derived from the published 10 minute (approximately 20 km resolution) raster layers. We used "simple" two-dimensional combinations in order to minimise biases associated with model over-fitting (Chapter 3 & Appendix C).

We sourced occurrence records for *Lymantria dispar* directly from GBIF. Prior to running the range bagging algorithms we first cleaned these occurrence records by using cleaning routines in the recently published CoordinateCleaner R package (Zizka *et al.*, 2019). Specifically, we removed records that:

- 1. had equal latitude and longitudes or were within 0.5-degrees radius of coordinates 0,0;
- 2. were within a 5 km radius of a capital city centroid,¹
- 3. were within a 10 km radius of either a country or province centroid;
- 4. were within a 1 degree radius around the GBIF headquarters in Copenhagen, Denmark;

¹In large databases such as GBIF, if a record does not have accurate GIS coordinates, users sometimes provide country and capital city coordinates. While such coordinates are useful for providing information on the country or province of the record, they are likely to be too course when estimating the climatic suitability of a species, which involves mapping occurrence records onto fine-resolution gridded climatic data.



- 5. were within a 100 m radius around the centroids of known biodiversity institutions;
- 6. were located in the ocean; or
- 7. contained mismatches between the country specified in the record and the record's coordinates.

We also removed duplicate records and thinned them to one point per 20 km (the resolution of the WorldClim version 2 climate data). Following this, we removed all occurrence records that were present in countries with no known establishment records based on CABI country-level distributional data. This ensured that the remaining occurrences were most likely from established populations, and thus suitable for inclusion in the range bagging analysis. Models were then fitted, and subsequently projected and resampled to 1 km grid – our desired map resolution. The mapped distributional data and model outputs can be seen in Figures 8.4 & 8.5. Broadly, the outcome of this modelling approach was a suitability map that was similar to the CLIMEX model published by Paini *et al.* (2018), whereby both south-west and south-east coastlines of Australia were deemed climatically suitable.





Figure 8.3.: Published gypsy moth CLIMEX (Paini *et al.*, 2018) and GARP ensemble (Peterson *et al.*, 2007) distribution models. Darker shades indicate more suitable climate.



Figure 8.4.: Global climatic suitability for gypsy moth, derived from range bagging. Blue points refer to cleaned distributional data collated from GBIF. Raster resolution: 10 minutes (approx 20 km). Suitability is the proportion of ensembled convex hulls that identify a location as climatically suitable across bootstrapped combinations of environmental variables.





Figure 8.5.: Australian climatic suitability for gypsy moth derived from range bagging. Raster aggregated to a 5 km grid for presentation purposes. Suitability is the proportion of ensembled convex hulls that identify a location as climatically suitable across bootstrapped combinations of environmental variables.



8.4. Biotic suitability

We utilised the tertiary raster layer of Australian Land Use and Management Classification (ALUMC version 8; ABARES, 2019), provided by the Australian Collaborative Land Use and Management Program, to define the presence and absence of gypsy moth host material. Specifically, we created a binary raster whereby any 50 m² grid cell likely to contain a vulnerable landuse (e.g. those containing susceptible endemic, ornamental and wood/food production tree species) such as nature conservation, national parks, wood production forestry, plantation forests, tree fruits and tree nuts (Table 8.2). Given the species also feeds on many ornamental plants present in gardens and streetside vegetation (e.g. maple and oak trees), we also included urban and rural residential areas. All other grid cells containing other dominant land uses were classified as 0 and assumed to contain no BMSB host material. This binary raster was then converted to a 1 km grid resolution by finding the maximum value of aggregated grid cells.

Landuse	Code	Landuse (cont)	Code (cont)
Nature conservation	110	Plantation forests	310
Strict nature reserves	111	Hardwood plantation forestry	311
Wilderness area	112	Softwood plantation forestry	312
National park	113	Other forest plantation	313
Natural feature protection	114	Environmental forest plantation	314
Habitat/species management area	115	Beverage and spice crops	332
Protected landscape	116	Perennial tree fruits	341
Other conserved area	117	Perennial tree nuts	343
Managed resource protection	120	Perennial shrub berries and fruits	345
Managed biodiversity	121	Perennial flowers and bulbs	346
Managed surface water supply	122	Irrigated plantation forests	410
Managed groundwater	123	Irrigated hardwood plantation forestry	411
Managed landscape	124	rrigated softwood plantation forestry	412
Traditional indigenous uses	125	Irrigated other forest plantation	413
Other minimal use	130	Irrigated environmental forest plantation	414
Defence land - natural areas	131	Irrigated beverage and spice crops	432
Stock route	132	Irrigated tree fruits	441
Residual native cover	133	Irrigated tree nuts	443
Rehabilitation	134	Irrigated shrub berries and fruits	445
Grazing native vegetation	210	Irrigated perennial flowers and bulbs	446
Production native forests	220	Urban residential	541
Wood production forestry	221	Rural residential with agriculture	542
Other forest production	222	Rural residential without agriculture	543

Table 8.2.: ALUMC landuses likely to contain Gyspy moth host material.

While this raster provides an indication of the extent of host material across Australia, it is likely to be an over-estimate because not all natural and conservation landuses will contain vulnerable tree or shrub species (e.g. Grasslands, sedgelands etc.). As such, we further masked areas of natural vegetation that are unlikely to contain Acacia or Eucalyptus species (two genera shown to contain susceptible species Liebhold et al. 1995). We did this by using the National Vegetation Information System; (NVIS) - A 100 m raster that attributes each cell to one of up to 89 vegetation classes defined by vegetation structure and dominant species. To mask the 1 km landuse layer. We first created a binary NVIS layer by identifying all vegetation types likely to contain vulnerable species (Table 8.3). Specifically, any vegetation type that explicitly included Acacia or Eucalyptus as dominant species, or were likely to contain these species as sub-dominants (e.g. Open woodlands) were classified as 1. We also included cleared, non-native vegetation and buildings to ensure susceptible agricultural landuses were not masked. We then aggregated the raster to a 1 km grid by finding the maximum value of aggregated grid cells. The landuse and NVIS rasters were then multiplied to provide a host distribution raster

However, while the above raster provided details about the presence and absence of host material it does not provide an indication of its abundance in any given season. We approximated the host abundance (conditional on the host being present) by further multiplying the binary host presence raster with the Min-Maxed² normalised average Normalised Difference Vegetation Index (NDVI; a measure of raster cell greenness) observed between October 2018 to March 2019 (The period when the pest would be most active; Fig. 8.6). We obtained this data from the Bureau of Meteorology at a raster resolution of 0.05 by 0.05 degrees (approximately 5 km scale), which was then resampled to a 1 km grid (Fig. 5.5).

Biotic suitability _i =Vulnerable landuse (ACLUM) _i ×	
Vulnerable vegetation (NVIS) $_i \times$	(8 1)
$NDVI_i - min(NDVI_i)$	(0.1)
$\overline{max(\text{NDVI}_i) - min(\text{NDVI}_i)}$	

²To ensure index was appropriately scaled between 0 and 1

Table 8.3.: NVIS vegetation classes likely to contain Gyspy moth host material.	Table 8.3.: NVIS vegetation classes likely to contain Gyspy moth host material.			
Vegetation	Code			
Cool temperate rainforest	1			
Tropical or sub-tropical rainforest	2			
Eucalyptus (+/- tall) open forest with a dense broad-leaved and/or tree-fern understorey (wet sclerophyll)	3			
Eucalyptus open forests with a shrubby understorey	4			
Eucalyptus open forests with a grassy understorey	5			
Warm temperate rainforest	6			
Tropical Eucalyptus open forests and woodlands with a tall annual grassy understorey	7			
Eucalyptus woodlands with a shrubby understorey	8			
Eucalyptus woodlands with a tussock grass understorey	9			
Eucalyptus woodlands with a hummock grass understorey	10			
Tropical mixed spp forests and woodlands	11			
Brigalow (Acacia harpophylla) forests and woodlands	13			
Other Acacia forests and woodlands	14			
Other forests and woodlands	16			
Eucalyptus low open woodlands with hummock grass	18			
Eucalyptus low open woodlands with tussock grass	19			
Mulga (Acacia aneura) woodlands +/- tussock grass +/- forbs	20			
Other Acacia tall open shrublands and [tall] shrublands	21			
Acacia (+/- low) open woodlands and shrublands with chenopods	22			
Acacia (+/- low) open woodlands and shrublands with hummock grass	23			
Acacia (+/- low) open woodlands and shrublands +/- tussock grass	24			
Acacia (+/- low) open woodlands and sparse shrublands with a shrubby understorey	25			
Mallee with hummock grass	27			
Low closed forest or tall closed shrublands (including Acacia, Melaleuca and Banksia)	28			
Mallee with a dense shrubby understorey	29			

Table 8.3.: (continued) Vegetation Code Mulga (Acacia aneura) open woodlands and sparse shrublands +/- tussock grass 45Eucalyptus open woodlands with shrubby understorey 47Eucalyptus open woodlands with a grassy understorey 48Mulga (Acacia aneura) woodlands and shrublands with hummock grass 51Mulga (Acacia aneura) open woodlands and sparse shrublands with hummock grass 52Eucalyptus low open woodlands with a shrubby understorey 53Eucalyptus tall open forest with a fine-leaved shrubby understorey 54Mallee with an open shrubby understorey 55Eucalyptus (+/- low) open woodlands with a chenopod or samphire understorey 56Eucalyptus woodlands with ferns, herbs, sedges, rushes or wet tussock grassland 59Eucalyptus tall open forests and open forests with ferns, herbs, sedges, rushes or wet tussock grasses 60 Mallee with a tussock grass understorey 61Dry rainforest or vine thickets 62 Eucalyptus woodlands with a chenopod or samphire understorey 65Open mallee woodlands and sparse mallee shrublands with a hummock grass understorey 66

Vegetation	Code
Open mallee woodlands and sparse mallee shrublands with a tussock grass understorey	67
Open mallee woodlands and sparse mallee shrublands with an open shrubby understorey	68
Open mallee woodlands and sparse mallee shrublands with a dense shrubby understorey	69
Other open Woodlands	79
Other sparse shrublands and sparse heathlands	80
Regrowth or modified forests and woodlands	90
Regrowth or modified shrublands	91
Unclassified forest	96
Unclassified native vegetation	97
Cleared, non-native vegetation, buildings	98
Unknown/No data	99





Figure 8.6.: Distribution of suitable biotic environment for gypsy moth. Estimated as a function of ALUMC, NVIS and NDVI data. Raster aggregated to a 5 km grid for presentation purposes.

8.5. Establishment likelihoods

According to RRRA, gypsy moth is most likely to enter via three pathways: imported machinery, containers and vessels (Table 8.1). In our establishment likelihood mapping framework, imported machinery (i.e. largely new and used cars and parts) is distributed as a function of human population density. By contrast, shipping containers are distributed based on ABS final postcode destination data, and arrivals via vessels



is distributed as a function of distance from marine port and the quantity of containers arriving at that port. The implication of these assumptions is that in the days following entry, the majority of arrivals will occur in major metropolitan areas that are close to marine ports and that receive large quantities of shipping containers.



Figure 8.7.: Estimated environmental suitability for gypsy moth. Suitability is defined as the product of abiotic (i.e. climate suitability) and biotic suitability (i.e. host presence × NDVI × NVIS). Raster aggregated to a 5 km grid for presentation purposes.

However, whether these arrivals are likely to result in establishment depends on both the abiotic (e.g. climate suitability) and biotic (e.g. presence of food source) environment (Fig. 8.7). In terms of food availability (i.e. biotic suitability), all Australian



cities appear to be suitable (i.e. contain appropriate hosts) and, at a broad scale, have fairly similar spring/summer NDVI. However, based on the range bagging climate suitability model (Fig. 8.5), gypsy moth appears to have a narrow climatic niche in parts of the east coast and south-west of Australia. Thus, the success of arrivals coming into Australia is likely to be restricted to these regions that contain high population density and receive shipping containers, such as Sydney, Perth and Melbourne (Fig. 8.8). Figures 8.9–8.16 provide more detailed maps of the geographic distribution of gypsy moth establishment likelihoods around major cities and tourist locations. For finer scale (i.e. 1 km resolution) maps please see associated interactive maps attached to this report. Interactive maps showing the expected number of arrivals across all pathways (not shown here) are also included as additional files attached to this report.




Figure 8.8.: National establishment likelihood map for gypsy moth. Raster aggregated to a 5 km grid for presentation purposes. Only establishment scores above 10^{-8} plotted.





Figure 8.9.: Gypsy moth establishment likelihood map for greater Cairns. Only establishment scores above 10^{-8} plotted.





Figure 8.10.: Gypsy moth establishment likelihood map for greater Brisbane. Only establishment scores above 10^{-8} plotted.





Figure 8.11.: Gypsy moth establishment likelihood map for greater Sydney. Only establishment scores above 10^{-8} plotted.





Figure 8.12.: Gypsy moth establishment likelihood map for greater Melbourne. Only establishment scores above 10^{-8} plotted.





Figure 8.13.: Gypsy moth establishment likelihood map for greater Hobart. Only establishment scores above 10^{-8} plotted.





Figure 8.14.: Gypsy moth establishment likelihood map for greater Adelaide. Only establishment scores above 10^{-8} plotted.





Figure 8.15.: Gypsy moth establishment likelihood map for greater Perth. Only establishment scores above 10^{-8} plotted.





Figure 8.16.: Gypsy moth establishment likelihood map for greater Darwin. Only establishment scores above 10^{-8} plotted.

9. Informing biosecurity decisions

Maps of pest establishment likelihoods are a fundamental decision support tool used to inform many biosecurity decisions. Commonly, such maps are based on pathway analyses that determine the likelihood of entry at various points locations (e.g. ports and airports) or are based on the geographic distribution of suitable abiotic and/or biotic environmental conditions. However, these methods are rarely integrated to provide a more holistic measure of establishment potential. That is they rarely account for the three fundamental geographic barriers to establishment: 1) the ability of a pest/disease to arrive at a given location; 2) the abiotic suitability of the location; and 3) the biotic suitability of the location.

The establishment likelihood mapping framework outlined in Chapter 4 and implemented in chapters 5–8, encapsulates these three fundamental geographic barriers to establishment. In the following sections we briefly outline how these maps can be used to:

- 1. inform where to conduct surveillance for early detection;
- 2. estimate surveillance coverage relative to total regional establishment likelihoods;
- 3. inform likelihoods of area freedom;
- 4. inform incursion seeds in pest/disease spread modelling; and
- 5. develop true risk maps.



9.1. Utility 1: Informing where to conduct surveillance for early detection

A primary purpose for maps of pest/disease establishment likelihoods is to inform where to conduct surveillance for early detection, and thus minimise the spread, and economic, societal and environmental consequences that might otherwise arise from an incursion. In this context, the surveillance for early detection should be prioritised in areas where the likelihood of establishment is highest.

The outputs of the mapping framework outlined in Chapter 4 allow this to be readily done by using the following steps:

- Step 1: Determine the management scale at which a surveillance program will be implemented. For example, surveillance may be coordinated in terms of a national grid where each grid cell has a specific area (e.g. 5 km²) or it may be in terms of geographic management units such as Natural Resource Management (NRM) regions.
- Step 2: Aggregate establishment scores to the management unit scale. This can be done by calculating the summed (if management units have equal area) or median (when management units have unequal area) establishment score across 1 km² grid cells (the default resolution of outputs) that fall within each management unit.
- **Step 3:** Allocate finite surveillance resources proportional to the establishment scores estimated in management units.
- **Step 4:** Use the raw 1 km² resolution establishment likelihood map to distribute surveillance within a management unit proportional to the establishment likelihood estimated at grid cells.

Caution!

This approach is a first approximation for determining where to conduct surveillance for early detection. It does not, however, inform the optimal allocation of surveillance. To do so would require minimising the net expected costs of conducting surveillance. That is, the optimal surveillance effort is that which minimises the sum of the costs of surveillance and the expected costs (e.g. eradication costs and market losses) of failing to detect an incursion.



9.2. Utility 2: Surveillance coverage relative to total establishment likelihood

The maps of establishment likelihoods can also be used to assess how much of the national (or jurisdictional) establishment likelihood is captured by surveillance. This can be achieved by:

- Step 1: Determining the spatial scale at which a surveillance unit (e.g. trap, survey or surveillance) is deemed to be representative. For example, an area of 5 km² would be set if this was the scale at which a network of pheromone traps or surveys were used to inform the status of a pest/disease.
- **Step 2:** Aggregate (i.e. by summing), the establishment likelihood map (default resolution is 1 km²) to the desired spatial scale defined above.
- **Step 3:** Calculate the proportion of establishment potential captured in grid cells with surveillance relative to the summed total establishment likelihood estimated in the region of interest.

A similar approach can also be used to determine the maximum proportion of geographic risk that could be captured, given a particular amount of area surveyed. Here, the maximum proportion of risk captured is estimated by assuming that finite surveillance resources are always allocated to the highest risk locations¹.

To illustrate this, let's assume that the surveillance for each of the four case study pests, oriental fruit fly (Bactrocera dorsalis), khapra beetle (Trogoderma granarium), brown marmorated stink bug (BMSB; Halyomorpha halys) and gypsy moth (Lymantria dispar)) were designed to inform pest status at a 5 km scale. Furthermore, lets assume there was a national budget to survey up to 50,000 km² (i.e. $2,000 \times 5$ km²) for each pest. We could examine how changes in surveyed area affect the maximum proportion of establishment potential sampled, by simply taking the raw 1 km² map of establishment likelihoods for each pest and aggregating (summing) the establishment scores to a 5 km² grid. Surveillance resources could then be allocated to highest establishment score cells in descending order until the surveillance budget is exhausted (i.e. when 2,000 cells have been sampled). One can then plot how the proportion of establishment potential captured varies with area sampled (e.g. see Fig. 9.1). The steepness of the curves will be determined by the extent to which risk is spatially aggregated. For example, based on these conditions, oriental fruit fly has the most aggregated risk, and thus, requires the lowest area surveyed (233 grid cells) to reach 80% establishment potential coverage. By contrast, gypsy moth has the most dispersed establishment likelihoods, predominately because of its wide host distribution coupled with widely dispersed pathway carriers (e.g. containers), which as a result requires the most area (1451 grid cells) surveyed to reach 80% establishment potential coverage.

¹In practice, if high risk locations are not aggregated in geographic space, then achieving the maximum proportion of geographic risk for a given amount of surveillance area may be difficult due to logistical and time constraints associated with travel.





Figure 9.1.: The relationship between the proportion of national establishment likelihood captured and the number of 5 km grid cells surveyed for each of the four case study pests. A) Oriental fruit fly; B) Khapra beetle; C) Brown marmorated stink bug; D) Gypsy moth. Dotted line represents the point at which 80% of the national establishment potential is captured under the assumptions that surveillance is prioritised to cells with high establishment likelihoods and that surveillance effort within each cell is sufficiently high to reliably detect the pest. Further details on how establishment likelihood was estimated for each pest can be found in Chapters 5–8.



9.3. Utility 3: Informing likelihoods of area freedom

General and targeted surveillance programs are an integral component of any biosecurity system. Typically focussing on areas of high establishment likelihood or high consequence (Hauser & McCarthy, 2009; Stanaway *et al.*, 2011), they are used to inform the status (presence or absence) of pests and diseases that affect economic, environmental and societal values in both pre- and post-outbreak conditions. In countries where a plant pest or disease is capable of establishing, internationally recognised surveillance programs (e.g., ISPM26; International Plant Protection Convention 2006) are used to determine pest freedom in high-value agricultural and urban centres (International Atomic Engergy Agency, 2003). In these landscapes, if surveillance programs do not detect a plant pest, then pest freedom is often assumed, allowing the export country to gain or maintain market access with other World Trade Organisation (WTO) countries without the need for additional phytosanitary procedures (e.g., International Plant Protection Convention 1995, 1998).

A fundamental problem with declaring pest freedom based solely on the absence of any detections is that it assumes the surveillance program has perfect sensitivity. That is, if the pest is present, then the surveillance program will always detect it. In reality, however, surveillance programs almost never have perfect sensitivity (Kery, 2002; Royle & Link, 2006; Hauser & McCarthy, 2009). The complication of imperfect sensitivity means that the lack of detections from a surveillance program can arise from one of two processes: 1) the pest is absent; or 2) the pest is present but the surveillance program failed to detect it (i.e. a false negative; Fig. 9.2). This is problematic, as unlike false positives, which can be double-checked and verified, it is impossible to be absolutely certain a pest is absent (Wintle *et al.*, 2004, 2012; Martin, 2017). Rather, absence can only be inferred in terms of likelihoods.

The traditional approach to estimating likelihoods of absence is based on the confidence of rejecting the null hypothesis that a pest or disease is present. Here, confidence, a, is defined by the probability a surveillance program detects a pest or disease. Typically, this is estimated as a function of the sensitivity of a surveillance component (e.g. diagnostic test, trap or survey), S, to detect the object of interest, at some predetermined prevalence level, p, and some amount of surveillance effort (e.g. number of tests, traps or repeat surveys), N_{effort} :

$$a = 1 - (1 - p \times S)^{N_{\text{effort}}} \tag{9.1}$$

Absence is then inferred if no detection was recorded and the confidence of detection, *a*, is at or above some pre-defined tolerance of being wrong (typically 0.05; Cannon 2002; Ormsby 2016). For example, if our tolerance for being wrong is 0.05 (i.e. wrong 5 times in every 100), then we'd need our confidence to be 0.95 or higher in order to reject the null hypothesis that the pest is present, and declare a pest absent. This approach has been used in ecology to determine the absence of rare species (McArdle, 1990; Kery, 2002) as well as other disciplines such as demonstrating absence in livestock disease (Cannon, 2002), plant pests (Ormsby, 2016) and inspecting consignments (International Plant Protection Convention, 2008).

While the null hypothesis approach described above provides a measure of confidence in pest absence, it does so solely as a function of surveillance sensitivity at some pre-defined prevalence and some amount of surveillance effort. It therefore ignores



Figure 9.2.: Scenario tree outlining the possible ways an absence can arise when surveillance sensitivity and specificity are imperfect.

possible differential likelihoods of pest establishment across geographic space. For example, it does not include information about the likely barriers faced by a pest in arriving and successfully establishing — Can it get there? Is the climate suitable? Is there a food source? This is particularly problematic if trying to estimate likelihoods of absence over a large geographic extent because if these barriers are not accounted for, then confidence is likely to be underestimated in locations where these barriers are formidable, and overestimated in locations where barriers are weak or non-existent.

Recently, Camac *et al.* (2019) proposed using Bayes' theorem as an alternative approach to quantifying the likelihood a pest or disease is absent (eqn 9.2). There are several advantages of using Bayes' theorem over null hypothesis testing. First, the likelihood of pest absence is measured on a more logical scale — probabilities of absence. This is in contrast to the null hypothesis approach (described above) which does not estimate probabilities of absence but rather infers absence based on the conditional probability of detection assuming the object of interest is present (Cannon, 2002; Barrett *et al.*, 2010). While this might sound like a subtle difference, the two approaches estimate quite different quantities, and calculating them with the same data can give different results (Rout, 2017).

$$Pr(Absence_{i}) = 1 - \frac{(1 - Pr(Detection_{i})) Prior_{i}}{(1 - Prior_{i}) + (1 - Pr(Detection_{i})) Prior_{i}}$$
(9.2)

$$Pr(Detection_{i}) = 1 - (1 - Pr(sensitivity))^{N_{effort_{i}}}$$

$$Prior = Pr(Present_{i})$$

However, the most notable advantage of the Bayesian approach is that it can directly incorporate additional sources of information that are otherwise ignored by the null hypothesis approach. It achieves this using a model parameter known as the *prior* (McCarthy, 2007). In the context of estimating likelihoods of pest absence, the prior describes the belief that a pest or disease is present at a location. This is where maps of establishment likelihoods come into play.

The establishment likelihood maps derived from this framework encompass the three major geographic barriers faced by pests when arriving and establishing — Can it get there? Is the climate suitable? Is there a food source? As such they can be used to account for these geographic barriers by acting as a spatially explicit prior probability of pest/disease presence. A major practical advantage of this is that the amount of surveillance required to declare freedom at a given level (e.g. 0.95) will now be informed by the establishment likelihood at that location. This means that locations deemed to have low establishment likelihoods will require less surveillance effort to achieve a desired probability of absence compared to regions of high establishment potential – thus focusing finite surveillance resources where they are needed most. To use these maps in this framework the following steps should be taken:

- **Step 1:** Determine the spatial scale at which surveillance infrastructure is used to inform pest/disease absence. For example, an area of 5 km² would be set if this was the scale at which a network of pheromone traps or surveys were used to informs the status of a pest/disease.
- **Step 2:** Aggregate (i.e. by summing), the establishment likelihood map (default resolution is 1 km²) to the desired spatial scale defined above.
- **Step 3:** Convert grid cell establishment likelihoods to relative probabilities by dividing scores by the summed total establishment likelihood of the region of interest².
- **Step 4:** Estimate the probability of detection (i.e. Pr(Detection_i)) for each grid cell, *i*, by specifying the known sensitivity of a surveillance unit as well as the amount of surveillance effort conducted in each grid cell (i.e. N_{effort_i}).
- **Step 5:** Calculate probabilities of absence for each grid cell, *i*, by using estimated establishment probabilities as the prior (Prior_{*i*}) in equation 9.2.

Caution!

Before applying this approach it is strongly recommended that users consult Camac et al. (2019) which goes into the assumptions and limitations of the approach more thoroughly.

²This step is needed because the establishment likelihoods are not probabilities but rather weighted expected number of arrivals per year.



9.4. Utility 4: Informing seeds in pest/disease spread modelling

The Australian Animal Disease Model (AADIS) has recently been extended to simulate the spread and control of plant pests, on both a regional and national scale (Bradhurst *et al.*, 2019). The new model is called APPDIS – the Australian Plant Pest and Disease model (Figure 9.3). APPDIS is an agent-based spread model that can represent a plant pest population as a point incursion, an established population at specified locations, or as a mechanistic estimation based on configurable environmental criteria. It is particularly useful for simulating possible incursions and how different management strategies may affect subsequent spread and eradication success.

Currently, APPDIS simulates initial incursions and subsequent spread based on either a user-defined point locality of incursion, or a random seed which may occur in any cell with host material. Greater biological realism can be incorporated into APPDIS by sampling these seeds probabilistically, directly from the maps of establishment likelihoods derived from the framework outlined in Chapter 4. The consequence of this will be that APPDIS will simulate spread and management strategies in regions where incursions are more likely to occur, and thus, allow decision-makers to better anticipate and manage future incursions.



Figure 9.3.: The Australian Plant Pest and Disease model (APPDIS) simulating the spread of tramp ant near Cairns.



9.5. Utility 5: Developing true risk maps

Currently the mapping framework developed in this project estimates establishment likelihoods (i.e expected number of arrivals weighted by abiotic and biotic suitability). While this is extremely useful for informing where to conduct surveillance for early detection, it does not inform where to conduct surveillance in order to minimise risk (i.e. likelihood \times consequence). For this to occur we also need estimates of how economic, environmental and societal values vary across geographic space as well as an understanding of how a pest or disease may impact such values.

Recently, as part of estimating the *Value of Australia's biosecurity system* (Dodd *et al.*, 2019), the Centre of Excellence for Biosecurity Risk Analysis (CEBRA) has developed spatially explicit value estimates for a range of asset classes. These include³:

- **Provisioning assets** (agriculture [cropping/grazing/horticulture]; forestry; water provisioning and filtration; indigenous subsistence);
- **Regulatory services** (erosion control; flood mitigation; toxin mediation; carbon sequestration; genepool/pollination);
- **Cultural services** (domestic recreation/amenity; international tourism; existence/bequest; indigenous cultural value);
- Infrastructure assets (residential dwellings);
- **Companion animal assets** (domestic animals [dogs/cats/fish/birds/etc.]; horses [recreational use])

These layers could be combined with estimates of the proportional reduction in value caused by a pest or disease outbreak to derive spatially explicit consequence layers. These consequence layers can be multiplied by the establishment likelihood maps⁴ to construct true risk maps. These maps can then be used to identify areas that exhibit both high establishment likelihoods and high consequences, and thus, where to allocate surveillance for risk mitigation.

³Each of the items in parentheses are a separate value layer.

⁴Assuming expected weighted arrivals are converted to proportions

10. Future research needs and extensions

Several future research opportunities and extensions became apparent during the development of the framework for estimating the spatial distribution of establishment likelihoods. We briefly document some of these in the following sections.

10.1. Major review of RRRA

RRRA is currently Australia's most comprehensive database on pest by pathway arrival likelihoods. However, during the development of this project it became apparent that in some cases RRRA pathway likelihoods were inconsistent with both contemporary knowledge and interception data on what were the high risk pathways of entry for specific pests. Perhaps the clearest example of this was for brown marmorated stink bug. According to RRRA estimates, BMSB had a very low likelihood of entry via imported machinery (i.e. 0.002 entries per year), vessels and shipping containers. However, contemporary interception data collated in Australia (DAWR, 2017) and New Zealand (Ormsby, 2018) contradicts this, with both highlighting several BMSB interceptions in imported machinery (e.g. new and used cars), vessels and other imported goods. This contradiction between RRRA and interception data is most likely because RRRA likelihoods for this pest are derived from an aggregate group of pests that encompass horticultural bugs, thrips and mites. Another possible explanation is that pathway entry likelihoods may have changed since 2015 when the majority of RRRA likelihoods were initially quantified.

We therefore recommend a major review of RRRA pathway leakage rates (and the methods used to estimate them) for all Australian priority plant pests. We also recommend that interception databases be integrated with RRRA such that contemporary pathway likelihoods can be estimated and possibly allow likelihoods to vary by country of origin.

10.2. Collecting empirical data for pathway post-border movements

Further research is also required to better understand the post-border movements of high risk pathway carriers.

For example, an assumption used in our mapping framework was that tourists are distributed as a function of tourist accommodation and distance from international airports. While we had spatially explicit empirical data to approximate the density of tourist accommodation, the distance from airport decay function was not informed by empirical data. Information obtained from Incoming Passenger Cards (IPC), could be



particularly useful for determining the proportion of passengers intending to reside in each suburb or postcode. It can also be used to estimate these proportions by country of residence or to estimate distance decay functions describing tourist distribution with respect to international airports.

PLEASE COMPLETE IN ENGLISH WITH A BI	UE OR BLACK		Are you bringing into Australia:
Family/surname			 Goods that may be prohibited or subject to restrictions, such as medicines, steroids, illegal pornography, firearms, weapons or illicit drugs? Yes No
Given names			 More than 2250mL of alcoholic beverages or 25 cigarettes or 25g of tobacco products? Yes No
Passport number			 Goods obtained overseas or purchased duty and/or tax free in Australia with a combined total price of more than AUD\$900, including gifts? Yes
			4. Goods/samples for business/commercial use? Yes No
Flight number or name of ship			 AUD\$10,000 or more in Australian or foreign currency equivalent? Yes No Note: If a customs or police officer asks, you must report travelers cheques, chemical and a contract of the area negative bla instruments of one amount
Intended address in Australia		$ \rightarrow $	6 Meat poultry fish seafood eggs dairy fruit vegetables? Yes No
	State		7. Grains, seeds, bulbs, straw, nuts, plants, parts of plants, traditional Yes No medicines or herbs, wooden articles?
Do you intend to live in Australia for the next 12 months?	Yes	No	 Animals, parts of animals, animal products including equipment, pet food, eggs, biologicals, specimens, birds, fish, insects, shells, bee products? Yes No
If you are NOT an Australian citizen:			 Soil, items with soil attached or used in freshwater areas e.g. sports/recreational equipment, shoes? Yes No
Do you have tuberculosis?	Yes	No	10. Have you been in contact with farms, farm animals, wilderness areas Yes No reshwater streams/lakes etc in the past 30 days?
Do you have any criminal conviction/s?	Yes	No 🕨	▶ 11. Were you in Africa. South/Central America or the Caribbean in the last 6 days? Yes No
 complete. I understand failure to answe questions may have serious consequent 	r any ces.		THE CARD
Complete. I understand tailure to answe questions may have serious consequen Your CONTACT DETAILS IN AUSTRALIA Phone () E-mail OR	r any ces.		EMERGENCY CONTACT DETAILS (FAMILY OR FRIEND) Name E-mail, Phone OR
Complete. I understand tailure to answe questions may have serious consequen YOUR CONTACT DETAILS IN AUSTRALIA Phone () E-mail OR Address	r any ces.	Sta	EMERGENCY CONTACT DETAILS (FAMILY OR FRIEND) Name E-mail, Phone OR Mail address
Complete. I understand tailure to answe questions may have serious consequen YOUR CONTACT DETAILS IN AUSTRALIA Phone () E-mail OR Address PLEASE COMPLETE IN ENGLISH	▶ PLEASE ×	Star	tate Mail address
Complete. I understand tailure to answe questions may have serious consequen YOUR CONTACT DETAILS IN AUSTRALIA Phone () E-mail OR Address PLEASE COMPLETE IN ENGLISH In which country did you board this flight or ship? What is your usual occupation?	 PLEASE × Migra perm to Au 	Star AND ANSW ating manently istralia	Edgy Horkity Feas THE CARD English Emergency contact betails (FAMILY or Friend) Name E-mail, Phone OR Phone OR Wears Months Days Your intended length of stay in Australia OR
Complete. I understand tailure to answe questions may have serious consequent YOUR CONTACT DETAILS IN AUSTRALIA Phone () E-mail OR Address PLEASE COMPLETE IN ENGLISH In which country did you board this flight or ship? What is your usual occupation? What is your usual occupation? Nationality as shown on passport Date Day Month Year	PLEASE × PLEASE × Migra perm to Au MAKE SUR HAVE COM BOTH SIDI THIS CA PRESENT THIS CA	Star AND ANSW ating nanently istralia RE YOU PLETED VES OF ARD. HIS CARD	EMARY THE CARD English EMERGENCY CONTACT DETAILS (FAMILY OR FRIEND) Name E-mail, Phone OR Mail address Were A or B or C B Visitor or temporary entrant Years Months Your intended length of stay in Australia Your country of residence Your main reason for travel (× one only) Convention/conference Your main reason for travel (× one only) Convention/conference

Figure 10.1.: An example of an Australian Incoming Passenger Card.

Updated analyses of shipping container post-border movements are also required. Currently, the latest analysis of shipping container movements comes from an experimental ABS dataset conducted on 2009–2010 data (ABS Cat. 1270.0.55.003). More contemporary data will likely provide a better reflection of current movements of shipping containers, and thus, how hitch hiker pests and diseases on this pathway are currently being distributed across Australia.

Lastly, research is required into the time spent by pathway carriers in areas of low and high establishment likelihoods. This information will aid in determining the a pest's exposure to suitable conditions and potentially whether it is likely to survive the journey.



10.3. Accounting for uncertainty

Currently, the mapping framework does not incorporate uncertainties in either pathway leakage rates or post-border movements. This is partly because such uncertainties were either not known, or were computationally difficult to propagate at large spatial scales (i.e. national level). We therefore recommend that future work should be directed at estimating these uncertainties and adding computationally efficient functionality for propagating and mapping multiple sources of uncertainty within the framework. Adding such functionality will improve biosecurity decisions by allowing decisionmakers to assess the range of plausible establishment likelihoods that could exist given current knowledge and uncertainties.

10.4. Aggregating maps

There is some interest in designing early detection surveillance programs for groups of pests as opposed to species-specific programs. One advantage of group-level surveillance is that, relative to species-level surveillance, it may reduce costs by focussing more on regions where the establishment likelihoods are high for multiple pests or diseases.

In the course of this project, some have asked whether the establishment likelihood maps can be combined across multiple pests in order to inform group-level decisions about where to allocate early detection surveillance. In practice, aggregating establishment likelihood maps across pests can be done by simply summing the grid-cell establishment scores across species. However, whether aggregating across a suite of pests is appropriate will likely depend on whether:

- 1. they share the same pathways of entry and whether the ranking of pathway likelihoods is the same;
- 2. they utilise the same host material; and
- 3. they have similar distributions of climatic suitability.

If the group of pests do not meet the above conditions then the derived aggregated map is unlikely to be useful in informing where to conduct surveillance. For example, we would advise against aggregating oriental fruit fly and BMSB because while they may share many hosts, areas of climatic suitability differ greatly (one being tropical the other being more temperate), and they can enter via different pathways. As such, future research should be done to determine which combinations of pests are appropriate to aggregate.



10.5. Predicting future establishment likelihoods

It is also important to examine how establishment likelihoods may change in the future, especially in order to design adaptive and robust early detection surveillance programs. This would require building models that can forecast:

- changes in both passenger volumes and commercial trade among countries under different climate and trade scenarios;
- changes in pest/disease climate suitability under climate change;
- changes in host material distribution caused by both climate change and other human activities;
- changes in which countries are infected with the pest/disease; and
- changes in pathway leakage likelihoods.

10.6. Virtual analytics platform

We have endeavoured to design a generalisable and transparent framework for mapping establishment likelihoods for plant pests that can be readily implemented in *R*. In order to minimise user error we have packaged the code into an *R* package called edmaps (i.e. Early Detection Maps). We have also provided functions that allow users to automate the workflow by pointing the mapping framework to a structured Excel spreadsheet and running a single command in *R*. However, implementing and using the outputs of this software will likely still be limited by the user's: 1) ability to install software dependencies; 2) computational resources; and 3) ability to interact with outputs. In order to mitigate these issues and improve user uptake, more work should be done to improve accessibility and users' ability to tap into other data sources both within and outside the Department.

One solution is to develop a biosecurity-focused online virtual analytics platform (i.e. a cloud-based workspace) that provides practitioners with a "one stop modelling shop" that will not only greatly simplify the access and integration of multiple sources of commonly used information (available both within and outside Departmental systems) but also provide users with a central repository of state-of-the-art analytical workflows that can inform scientifically robust policy- and decision-making. The Bio-diversity and Climate Change Virtual Laboratory (BCCVL) is one such online environment where researchers can access and integrate multiple sources of data and use a range of analytical methods to inform conservation-based decision-making.

Another fundamental advantage of such an online platform is that it can provide a coherent environment for integrating modelled outputs commonly used along the invasion continuum from pre-border to post-border. For example, real-time interception data could feed directly into the RRRA module and thereby provide contemporary national pathway leakage likelihoods required in mapping post-border pest establishment likelihoods. These maps of establishment likelihood, in turn, could then be combined with other data sources and analytical tools to inform probabilities of area freedom (Camac *et al.*, 2019), spread modelling (APPDIS; Bradhurst *et al.*, 2019), optimal surveillance (Kompas & Che, 2009; Kompas *et al.*, 2017) and constructing spatially explicit risk maps (i.e. maps of likelihoods × consequences).

Acknowledgements

This report is a product of the Centre of Excellence for Biosecurity Risk Analysis (CE-BRA). In preparing this report, the authors acknowledge the financial and other support provided by the Australian Department of Agriculture, Water and the Environment and its predecessors, the New Zealand Ministry for Primary Industries and the University of Melbourne. In particular, we thank Dr Ranjith Subasinghe, Dr Susie Collins, Dr Ajay Niranjane and Dr Mark Stanaway for their advice. We also thank Dr Anaís Gibert for contributing to Appendix A and Dr Matt Hill, Dr Peter Caley and Dr Simon Barry for their work on Appendix C of this report. Lastly, we sincerely thank all the participants that attended our workshop in October 2018.

Bibliography

- ABARES (2019) Catchment Scale Land Use of Australia Update December 2018. URL https://doi.org/10.25814/5c7728700fd2a.
- Anon (2008) Draft national code of practice for the management of mediterranean fruit fly. Tech. rep.
- Araújo MB, Rozenfeld A (2014) The geographic scaling of biotic interactions. *Ecography*, **37**, 406–415.
- Austin MP (2002) Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. *Ecological Modelling*, **157**, 101–118.
- Banks HJ (1977) Distribution and establishment of *Trogoderma granarium* Everts (Coleoptera: Dermestidae): Climatic and other influences. *Journal of Stored Products Research*, **13**, 183–202.
- Barbet-Massin M, Rome Q, Villemant C, Courchamp F (2018) Can species distribution models really predict the expansion of invasive species? *PLoS ONE*, **13**, e0193085–14.
- Barrett S, Whittle P, Mengersen K, Stoklosa R (2010) Biosecurity threats: the design of surveillance systems, based on power and risk. *Environmental and Ecological Statistics*, 17, 503–519.
- Barry S, Elith J, Heersink D, Caley P, Kearney M, Tenant P, Arthur A (2015) Final report: CEBRA 1402B Tools and approaches for invasive species distribution modelling for surveillance. Tech. rep., CSIRO.
- Bennett JM, Calosi P, Clusella-Trullas S, *et al.* (2018) Data Descriptor: GlobTherm, a global database on thermal tolerances for aquatic and terrestrial organisms. *Nature Publishing Group*, **5**, 1–7.
- Blonder B, Morrow CB, Maitner B, et al. (2017) New approaches for delineating ndimensional hypervolumes. *Methods in Ecology and Evolution*, **9**, 305–319.
- Booth TH, Nix HA, Busby JR, Hutchinson MF (2014) bioclim: the first species distribution modelling package, its early applications and relevance to most current MaxEnt studies. *Diversity and Distributions*, **20**, 1–9.
- Botha J, Szito A, Emery R, Grimm M, Hardie D (2005) Industry biosecurity plan for the grains industry. Tech. rep., Plant Health Australia.
- Bradhurst R, J M, M S (2019) Developing models for the spread and management of national priority plant pests. Tech. Rep. 170606, Centre of Excellence for Biosecurity Risk Analysis.
- Breiner FT, Guisan A, Bergamini A, Nobis MP (2015) Overcoming limitations of modelling rare species by using ensembles of small models. *Methods in Ecology and Evolution*, **6**, 1210–1218.
- Breiner FT, Nobis MP, Bergamini A, Guisan A (2017) Optimizing ensembles of small models for predicting the distribution of species with few occurrences. *Methods in Ecology and Evolution*, **9**, 802–808.
- Broennimann O, Guisan A (2008) Predicting current and future biological invasions: both native and invaded ranges matter. *Biology Letters*, **4**, 585–589.
- Buckley LB, Urban MC, Angilletta MJ, Crozier LG, Rissler LJ, Sears MW (2010) Can

mechanism inform species' distribution models? *Ecology Letters*, **13**, 1041–1054.

- Burges HD (1963) Studies on the Dermestid beetle Trogoderma granarium Everts. VI.—Factors inducing diapause. *Bulletin of Entomological Research*, **54**, 571–587.
- Camac JS, Clarke S, Niranjane A, de Majnik J (2019) Estimating probabilities of pest absence across geographic space. Tech. Rep. 1606D, Centre of Excellence for Biosecurity Risk Analysis.
- Camac JS, Spring D, Stanaway M, Kompas T (2018) National exotic fruit fly surveillance program. Tech. rep., Centre for Environmental and Economic Research.
- Cannon RM (2002) Demonstrating disease freedom—combining confidence levels. *Preventive Veterinary Medicine*, **52**, 227–249.
- Catford JA, Jansson R, Nilsson C (2009) Reducing redundancy in invasion ecology by integrating hypotheses into a single theoretical framework. *Diversity and Distributions*, **15**, 22–40.
- Chamberlain S, Szocs E, Boettiger C, *et al.* (2017) Taxonomic information from around the web. R package version 0.9.0. URL https://github.com/ropensci/taxize.
- Craik W, Palmer D, Sheldrake R (2017) Priorities for Australia's biosecurity system: An independent review of the capacity of the national biosecurity system and its underpinning Intergovernmental Agreement.
- Crombie J, Brown L, Lizzio J, Hood G (2008) Climatch v1.0. URL http://data. daff.gov.au:8080/Climatch/climatch.jsp.
- DAWR (2017) Draft pest risk analysis for brown marmorated stink bug (Halyomorpha halys). Tech. rep., Department of Agriculture and Water Resources, Commonwealth of Australia.
- DAWR (2018) Department of Agriculture and Water Resources: Annual Report 2017-18.
- De Lima F, Poogoda S, Smallridge C, *et al.* (2011) Trapping to better predict and prove fruit fly presence. Tech. rep., Cooperative Research Centre for National Plant Biosecurity.
- De Meyer M, Robertson MP, Mansell MW, *et al.* (20010) Ecological niche and potential geographic distribution of the invasive fruit fly Bactrocera invadens(Diptera, Tephritidae). *Bulletin of Entomological Research*, **100**, 35–48.
- Dodd A, Stoeckl N, Kompas T (2019) The value of Australia's biosecurity system. Tech. Rep. 1607A, Centre of Excellence for Biosecurity Risk Analysis.
- Dormann CF (2007) Promising the future? Global change projections of species distributions. *Basic and Applied Ecology*, **8**, 387–397.
- Douma JC, Pautasso M, Venette RC, *et al.* (2016) Pathway models for analysing and managing the introduction of alien plant pests—an overview and categorization. *Ecological Modelling*, **339**, 58–67.
- Drake JM (2015) Range bagging: a new method for ecological niche modelling from presence-only data. *Journal of The Royal Society Interface*, **12**, 20150086–9.
- Elith J (2017) Predicting distributions of invasive species. In: *Invasive species risk assessment and management* (eds. Robinson AP, Walshe T, Burgman MA, Nunn M), pp. 93–129. Cambridge University Press.
- Elith J, Kearney M, Phillips S (2010) The art of modelling range-shifting species. *Methods in Ecology and Evolution*, **1**, 330–342.
- Elith J, Leathwick JR (2009) Species distribution models: Ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics,*



40, 677–697.

- Ferrier S, Watson G, Pearce J, Drielsma M (2002) Extended statistical approaches to modelling spatial pattern in biodiversity in northeast New South Wales. I. Specieslevel modelling. *Biodiversity and Conservation*, **11**, 2275–2307.
- Fick SE, Hijmans RJ (2017) WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, **37**, 4302–4315.
- Fithian W, Elith J, Hastie T, Keith DA (2015) Bias correction in species distribution models: pooling survey and collection data for multiple species. *Methods in Ecology and Evolution*, **6**, 424–438.
- Fourcade Y, Besnard AG, Secondi J (2017) Paintings predict the distribution of species, or the challenge of selecting environmental predictors and evaluation statistics. *Global Ecology and Biogeography*, **27**, 245–256.
- Fourcade Y, Engler JO, Rödder D, Secondi J (2014) Mapping Species Distributions with MAXENT Using a Geographically Biased Sample of Presence Data: A Performance Assessment of Methods for Correcting Sampling Bias. *PLoS ONE*, **9**, e97122–13.
- Fraser D, Kumar S, Aguilar G (2017) Mapping the Potential Global Range of the Brown Marmorated Stink Bug, Halyomorpha halys, with Particular Reference to New Zealand. *Climate*, **5**, 75–14.
- Froese J (2012) A guide to selecting species distribution models to support biosecurity decision-making. Tech. rep.
- Guillera-Arroita G, Lahoz-Monfort JJ, Elith J (2014) Maxent is not a presence–absence method: a comment on thibaud et al. *Methods in Ecology and Evolution*, 5, 1192–1197. doi:10.1111/2041-210X.12252. URL https://besjournals.onlinelibrary. wiley.com/doi/abs/10.1111/2041-210X.12252.
- Guillera-Arroita G, Lahoz-Monfort JJ, Elith J, *et al.* (2015) Is my species distribution model fit for purpose? Matching data and models to applications. *Global Ecology and Biogeography*, **24**, 276–292.
- Harris I, Jones PD, Osborn TJ, Lister DH (2014) Updated high-resolution grids of monthly climatic observations - the CRU TS3.10 Dataset. *International Journal of Climatology*, 34, 623–642.
- Hauser CE, McCarthy MA (2009) Streamlining 'search and destroy': cost-effective surveillance for invasive species management. *Ecology Letters*, **12**, 683–692.
- Hemming V, Burgman MA, Hanea AM, McBride MF, Wintle BC (2017) A practical guide to structured expert elicitation using the IDEA protocol. *Methods in Ecology and Evolution*, **9**, 169–180.
- Higgins SI, Richardson DM (2014) Invasive plants have broader physiological niches. *Proceedings of the National Academy of Sciences*, **111**, 10610–10614.
- Hill MP, Gallardo B, Terblanche JS (2017) A global assessment of climatic niche shifts and human influence in insect invasions. *Global Ecology and Biogeography*, **26**, 679–689.
- Hill MP, Terblanche JS (2014) Niche Overlap of Congeneric Invaders Supports a Single-Species Hypothesis and Provides Insight into Future Invasion Risk: Implications for Global Management of the Bactrocera dorsalis Complex. *PLoS ONE*, **9**, e90121–12.
- International Atomic Engergy Agency (2003) *Trapping guidelines for area-wide fruit fly programmes*. Austria.
- International Plant Protection Convention (1995) *Requirements for the establishment of pest free areas: ISPM 4.* {Food and Agriculture Organization of the United Nations, Rome, Italy}.

- International Plant Protection Convention (1998) *Determination of pest status in an area: ISPM 8.* {Food and Agriculture Organization of the United Nations, Rome, Italy}.
- International Plant Protection Convention (2006) *Establishment of pest free areas for fruit flies (Tephritidae): ISPM 26.* Food and Agriculture Organization of the United Nations, Rome, Italy.
- International Plant Protection Convention (2008) *Methodologies for sampling of consignments: ISPM 31.* {Food and Agriculture Organization of the United Nations, Rome, Italy}.
- Jiménez-Valverde A, Peterson AT, Soberón J, Overton JM, Aragón P, Lobo JM (2011) Use of niche models in invasive species risk assessments. *Biological Invasions*, **13**, 2785–2797.
- Karger DN, Conrad O, Böhner J, *et al.* (2017) Data Descriptor: Climatologies at high resolution for the earth's land surface areas. *Nature Publishing Group*, **4**, 1–20.
- Kearney M, Phillips BL, Tracy CR, Christian KA, Betts G, Porter WP (2008) Modelling species distributions without using species distributions: the cane toad in Australia under current and future climates. *Ecography*, **31**, 423–434.
- Kearney M, Porter W (2009) Mechanistic niche modelling: combining physiological and spatial data to predict species' ranges. *Ecology Letters*, **12**, 334–350.
- Kearney MR, Porter WP (2017) NicheMapR an R package for biophysical modelling: the microclimate model. *Ecography*, **40**, 664–674.
- Kearney MR, Wintle BA, Porter WP (2010) Correlative and mechanistic models of species distribution provide congruent forecasts under climate change. *Conservation Letters*, **3**, 203–213.
- Kery M (2002) Inferring the Absence of a Species: A Case Study of Snakes. *The Journal of Wildlife Management*, **66**, 330–338.
- Kompas T, Che N (2009) A practical optimal surveillance measure: the case of papaya fruit fly in Australia.
- Kompas T, Ha P, Spring D (2017) Optimal Trapping and Review of Victoria's Current Mediterranean Fruit Fly Surveillance Program. Report prepared for the Victorian Department of Economic Development, Jobs, Transport and Resources.
- Kriticos D, Maywald GF, Yonow T, Zurcher EJ, Herrmann NI, Sutherst RW (2015) *Climex Version 4: Exploring the effects of climate on plants, animals and diseases.* CSIRO.
- Kriticos DJ, Kean JM, Phillips CB, Senay SD, Acosta H, Haye T (2017) The potential global distribution of the brown marmorated stink bug, Halyomorpha halys, a critical threat to plant biosecurity. *Journal of Pest Science*, **90**, 1–11.
- Kriticos DJ, Webber BL, Leriche A, Ota N, Macadam I, Bathols J, Scott JK (2012) CliMond: global high-resolution historical and future scenario climate surfaces for bioclimatic modelling. *Methods in Ecology and Evolution*, **3**, 53–64.
- Lahoz-Monfort JJ, Guillera-Arroita G, Wintle BA (2014) Imperfect detection impacts the performance of species distribution models. *Global Ecology and Biogeography*, **23**, 504–515.
- Landau WM (2018) The drake r package: a pipeline toolkit for reproducibility and high-performance computing. *Journal of Open Source Software*, **3**. URL https://doi.org/10.21105/joss.00550.
- Le Maitre DC, Thuiller W, Schonegevel L (2008) Developing an approach to defining the potential distributions of invasive plant species: a case study of Hakeaspecies in South Africa. *Global Ecology and Biogeography*, **17**, 569–584.
- Liebhold A, Gottschalk KW, Muzika RM, Montgomery ME, ODay K, Kelley B (1995)



Suitability of North American tree species to the Gypsy moth: A summary of field and laboratory test. Tech. Rep. No. NE-211.

- Liebhold AM, Halverson JA, Elmes GA (1992) Gypsy moth invasion in north america: A quantitative analysis. *Journal of Biogeography*, **19**, 513–520.
- Lobo JM, Jiménez-Valverde A, Hortal J (2010) The uncertain nature of absences and their importance in species distribution modelling. *Ecography*, **33**, 103–114.
- Magarey RD, Fowler GA, Borchert DM, Sutton TB, Colunga-Garcia M (2007) NAPP-FAST: An internet system for the weather-based mapping of plant pathogens. *Plant Disease*, **91**, 336–345.
- Maldonado C, Molina CI, Zizka A, *et al.* (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**, 973–984.
- Marques GM, Augustine S, Lika K, Pecquerie L, Domingos T, Kooijman SALM (2018) The AmP project: Comparing species on the basis of dynamic energy budget parameters. *PLoS Computational Biology*, **14**, e1006100–23.
- Martin T (2017) Surveillance for detection of pests and diseases: how sure can we be of their absence? In: *Invasive species risk assessment and management* (eds. Robinson AP, Walshe T, Burgman MA, Nunn M), pp. 348–384. Cambridge.
- Martin TG, Murphy H, Liedloff A, *et al.* (2015) Buffel grass and climate change: a framework for projecting invasive species distributions when data are scarce. *Biological Invasions*, **17**, 3197–3210.
- Matsuki M, Kay M, Serin J, Floyd R, Scott JK (2001) Potential risk of accidental introduction of Asian gypsy moth (Lymantria dispar) to Australasia: effects of climatic conditions and suitability of native plants. *Agricultural and Forest Entomology*, **3**, 305– 320.
- Mau-Crimmins TM, Schussman HR, Geiger EL (2006) Can the invaded range of a species be predicted sufficiently using only native-range data? *Ecological Modelling*, **193**, 736–746.
- McArdle BH (1990) When are rare species not there? Oikos, 57, 276–277.
- McCarthy MA (2007) *Bayesian methods for ecology*. Cambridge University Press, Cambridge.
- Merow C, Smith MJ, Edwards TCJ, *et al.* (2014) What do we gain from simplicity versus complexity in species distribution models? *Ecography*, **37**, 1267–1281.
- Merow C, Smith MJ, Silander Jr JA (2013) A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. *Ecography*, **36**, 1058–1069.
- Mesgaran MB, Cousens RD, Webber BL (2014) Here be dragons: a tool for quantifying novelty due to covariate range and correlation change when projecting species distribution models. *Diversity and Distributions*, **20**, 1147–1159.
- Morales-Castilla I, Davies TJ, Pearse WD, Peres-Neto P (2017) Combining phylogeny and co-occurrence to improve single species distribution models. *Global Ecology and Biogeography*, **26**, 740–752.
- Morin X, Lechowicz MJ (2008) Contemporary perspectives on the niche that can improve models of species range shifts under climate change. *Biology Letters*, **4**, 573–576.
- National Fruit Fly Strategy Action Plan (2017) National Fruit Fly Strategy Action Plan. Tech. rep.

Ormsby M (2016) Evaluation of Import and Export Parameters for Fruit Fly Export Restriction Zones. Tech. rep., New Zealand Ministry for Primary Industries.



- Ormsby M (2018) Technical Review Proposed Treatments for BMSB (Halyomorpha halys (Stål); Pentatomidae). Tech. rep., Ministry for Primary Industries, New Zealand Government.
- Paini DR, Mwebaze P, Kuhnert PM, Kriticos DJ (2018) Global establishment threat from a major forest pest via international shipping: Lymantria dispar. *Scientific Reports*, **8**, 13723.
- Paini DR, Yemshanov D (2012) Modelling the arrival of invasive organisms via the international marine shipping network: A Khapra beetle study. *PLoS ONE*, **7**, e44589– 9.
- Peterson AT, Williams R, Chen G (2007) Modeled global invasive potential of Asian gypsy moths, Lymantria dispar. *Entomologia Experimentalis et Applicata*, **125**, 39–44.
- Phillips CB, Kean JM, Vink CJ, Berry JA (2017) Utility of the CLIMEX 'match climates regional' algorithm for pest risk analysis: an evaluation with non-native ants in New Zealand. *Biological Invasions*, **20**, 777–791.
- Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, **190**, 231–259.
- Phillips SJ, Dudik M, Elith J, Graham CH, Lehmann A, Leathwick J, Ferrier S (2009) Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecological Applications*, **19**, 181–197.
- Pollock LJ, Morris WK, Vesk PA (2012) The role of functional traits in species distributions revealed through a hierarchical model. *Ecography*, **35**, 716–725.
- Reddy S, Dávalos LM (2003) Geographical sampling bias and its implications for conservation priorities in Africa. *Journal of Biogeography*, **30**, 1719–1727.
- Renner IW, Elith J, Baddeley A, *et al.* (2015) Point process models for presence-only analysis. *Methods in Ecology and Evolution*, **6**, 366–379.
- Renner IW, Warton DI (2013) Equivalence of MAXENT and poisson point process models for species distribution modeling in ecology. *Biometrics*, **69**, 274–281.
- Renton M, Triska M, Broughton S, *et al.* (2017) Design and Evaluation of Targeted Biosecurity Surveillance Systems. Tech. rep., Plant Biosecurity Cooperative Research Centre.
- Rice KB, Bergh CJ, Bergmann EJ, *et al.* (2014) Biology, Ecology, and Management of Brown Marmorated Stink Bug (Hemiptera: Pentatomidae). *Journal of Integrated Pest Management*, **5**, 1–13.
- Rout TM (2017) Declaring eradication of an invasive species. In: *Invasive species risk assessment and management* (eds. Robinson AP, Walshe T, Burgman MA, Nunn M), pp. 334–347. Cambridge.
- Royle JA, Link WA (2006) Generalized site occupancy models allowing for false positive and false negative errors. *Ecology*, **87**, 835–841.
- Salguero Gómez R, Jones OR, Archer CR, *et al.* (2015) The compadre plant matrix database: an open online repository for plant demography. *Journal of Ecology*, **103**, 202–218.
- Salguero Gómez R, Jones OR, Archer CR, *et al.* (2016) COMADRE: a global data base of animal demography. *Journal of Animal Ecology*, **85**, 371–384.
- Sequeira AMM, Bouchet PJ, Yates KL, Mengersen K, Caley MJ (2018) Transferring biodiversity models for conservation: Opportunities and challenges. *Methods in Ecology and Evolution*, **9**, 1250–1264.
- Stanaway MA, Mengersen KL, Reeves R (2011) Hierarchical Bayesian modelling of early detection surveillance for plant pest invasions. *Environmental and Ecological*



Statistics, **18**, 569–591.

- Stephens A, Kriticos D, Leriche A (2007) The current and future potential geographical distribution of the oriental fruit fly, bactrocera dorsalis (diptera: Tephritidae). *Bulletin of Entomological Research*, **97**, 369–378.
- Syfert MM, Smith MJ, Coomes DA (2013) The Effects of Sampling Bias and Model Complexity on the Predictive Performance of MaxEnt Species Distribution Models. *PLoS ONE*, **8**, e55158–10.
- Thuiller W, Richardson DM, Pyšek P, Midgley GF, Hughes GO, Rouget M (2005) Nichebased modelling as a tool for predicting the risk of alien plant invasions at a global scale. *Global Change Biology*, **11**, 2234–2250.
- Tingley R, García-Díaz P, Arantes CRR, Cassey P (2018) Integrating transport pressure data and species distribution models to estimate invasion risk for alien stowaways. *Ecography*, **41**, 635–646.
- Tobin PC, Liebhold AM, Anderson Roberts E (2007) Comparison of methods for estimating the spread of a non-indigenous species. *Journal of Biogeography*, **34**, 305–312.
- Ushey K (2019) renv: Project Environments. URL https://CRAN.R-project.org/ package=renv. R package version 0.8.2.
- Valavi R, Elith J, Lahoz-Monfort JJ, Guillera-Arroita G (2018) blockCV: An rpackage for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models. *Methods in Ecology and Evolution*, **10**, 225–232.
- Valentin RE, Nielsen AL, Wiman NG, Lee DH, Fonseca DM (2017) Global invasion network of the brown marmorated stink bug, Halyomorpha halys. *Scientific Reports*, 7, 1–12.
- VanDerWal J, Shoo LP, Graham C, Williams SE (2009) Selecting pseudo-absence data for presence-only distribution modeling: How far should you stray from what you know? *Ecological Modelling*, **220**, 589–594.
- Vargas RI, Piñero JC, Leblanc L (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**, 297–318.
- Venette RC (2017) Climate Analyses to Assess Risks from Invasive Forest Insects: Simple Matching to Advanced Models. *Current Forestry Reports*, **3**, 255–268.
- Venette RC, Kriticos DJ, Magarey RD, *et al.* (2010) Pest risk maps for invasive alien species: A roadmap for improvement. *Bioscience*, **60**, 349–362.
- Warton DI, Renner IW, Ramp D (2013) Model-based control of observer bias for the analysis of presence-only data in ecology. *PLoS ONE*, **8**, e79168–9.
- Warton DI, Shepherd LC (2010) Poisson point process models solve the "pseudoabsence problem" for presence-only data in ecology. *The Annals of Applied Statistics*, 4, 1383–1402.
- West AM, Kumar S, Brown CS, Stohlgren TJ, Bromberg J (2016) Field validation of an invasive species Maxent model. *Ecological Informatics*, **36**, 126–134.
- Whittle PJL, Stoklosa R, Barrett S, Jarrad FC, Majer JD, Martin PAJ, Mengersen K (2013) A method for designing complex biosecurity surveillance systems: detecting nonindigenous species of invertebrates on Barrow Island. *Diversity and Distributions*, **19**, 629–639.
- Wintle BA, McCarthy MA, Parris KM, Burgman MA (2004) Precision and bias of methods for estimating point survey detection probabilities. *Ecological Applications*, **14**, 703–712.



- Wintle BA, Walshe TV, Parris KM, McCarthy MA (2012) Designing occupancy surveys and interpreting non-detection when observations are imperfect. *Diversity and Distributions*, **18**, 417–424.
- Yates KL, Bouchet PJ, Caley MJ, et al. (2018) Outstanding Challenges in the Transferability of Ecological Models. *Trends in Ecology & Evolution*, **33**, 790–802.
- Yemshanov D, Koch FH, Ducey M, Koehler K (2011) Trade-associated pathways of alien forest insect entries in Canada. *Biological Invasions*, **14**, 797–812.
- Zhu G, Bu W, Gao Y, Liu G (2012) Potential geographic distribution of brown marmorated stink bug invasion (Halyomorpha halys). *PLoS ONE*, **7**, e31246–10.
- Zizka A, Silvestro D, Andermann T, *et al.* (2019) CoordinateCleaner: Standardized cleaning of occurrence records from biological collection databases. *Methods in Ecology and Evolution*, **289**, 110–8.

A. Spatial heterogeneity in pest risk assessments: a review of risk maps

Spatial heterogeneity in pest risk assessments: a review of risk maps

Technical Review for CEBRA project 170607

Anias Gibert¹, Andrew Robinson¹, and James Camac¹

¹The Centre of Excellence for Biosecurity Risk Analysis, The University of Melbourne

November 3, 2020



Contents

1	Execu	itive summary	1
2	Intro	duction	3
3	Mate	rial and Methods	5
	3.1		6
4	Resul	lts	13
	4.1	Description of the studies reported	13
	4.2	Description of risk maps	14
	4	4.2.1 Entry risk	17
	4	4.2.2 Establishment risk	18
	4	4.2.3 Spread risk and impact	20
5	Discu	assion	21
	5.1	Accounting for conditional dependencies and impacts	21
	5.2	Accounting for uncertainty	21
	5.3	Estimating entry risk	22
	5.4	Estimating establishment risk	24
	5.5	Estimating spread risk and impact	24
	5.6 0	Conclusion	25

List of Figures

3.1	Geographical distribution of risk assessments collated in this report	6
4.1	Type of documents reported in this review. Guideline refers to RA frame- work provided by governmental or intergovernmental agencies.	13
4.2	Risk assessments by taxa. <i>Unspecified</i> refers to cases where no examples – and thus no taxa – were reported	14
4.3	The use of quantitative and qualitative methods in plant pest risk assessment over the last 28 years. Lines show the average number of studies	
	by year (± 95% CI).	15
4.4	Stages of invasion process encapsulated as either a single risk map or a	
	map that integrates multiple stages.	15
4.5	Methods used to create risk maps: a) Common methods used to integrate multiple risk components or attributes; b) Methods used to account for uncertainty; c) Methods used to estimate risk at each step of the invasion	
	process	16
4.6	Risk attributes commonly defined in pest risk assessments for each stage of the invasion process (including impact). Grey = Non-spatial risk attributes; shades of blue = spatial data. Different blue colours signify	
	the formats of spatial data commonly used	17
4.7	Type of data used to estimate risk in pest risk maps	18
4.8	Points of entry commonly used in pest risk maps.	19

List of Tables

3.1	Definition of the risk of entry, establishment, spread and impact, based	0
~ ~	on IPCC terminology [32]. \ldots	8
3.2	Commonly used methods to estimate components of risk	9
3.3	Methods used to integrate multiple risk components or attributes when	
	forming risk maps.	10
3.4	Methods used to account for risk uncertainty	11
3.5	Definition of the format of spatial data reported to address spatial het-	
	erogeneity in risk.	12
5.1	Database	26
Executive summary

Over the last two decades, pest *risk maps* have been used as a decision support tool for informing strategic and tactical decisions about where to allocate finite surveillance resources, international and domestic trade restrictions, and cost sharing between jurisdictions. However, no standardised approach currently exists, with pest risk maps varying in the stage/barriers of the invasion continuum they incorporate (i.e. entry, establishment, spread), which risk attributes they use (e.g. climate suitability, host availability, trade), what scale they are estimated at (e.g. region, cities, ports), and whether outputs are qualitative, quantitative, or include impacts.

Here we conducted a systematic review of grey and peer-reviewed literature that examined what data sources, methods and barriers to establishment and spread were commonly incorporated into pest risk maps. In total, we reviewed 74 peer-reviewed and grey literature documents published between 1991 and 2018.

We found that plant pest risk maps varied substantially in the methods, data sources and establishment barriers incorporated. The vast majority focused on estimating establishment potential (96%) via climatic suitability modelling (75%) and/or host distribution (69%), typically under the implicit assumption that likelihoods of arrival were uniform across the geographic space of interest.

By contrast, when entry or arrival likelihoods were estimated (44% of risk maps) – typically using trade patterns in pathway models – focus was predominately on major points of entry (e.g. ports) with most models not incorporating spread beyond these points or attempts to infer establishment potential.

Lastly, we found that only a small percentage of plant pest risk maps incorporated spatially explicit measures of impacts/consequences of an incursion (4%). This meant that the majority of reviewed pest risk maps did not explicitly measure risk (conventially defined as: *likelihood* × *consequence*).

We strongly advocate for more integration of methods used to estimate entry and spread with tools commonly used to inform establishment potential. Specifically, we call for pest risk maps to explicitly incorporate estimates of the three fundamental factors governing pest establishment and spread; namely: arrival rates, abiotic suitability and biotic suitability. Furthermore, economic tools must be integrated with the above in order to estimate spatially explicit consequences of pest outbreaks, and thus, true measures of risk in which to prioritise pests and allocate finite biosecurity resources.

Introduction

Globalisation of human movement and trade has dramatically increased the exposure of countries to new pests and diseases that may have devastating economic, environmental and social impacts [1]. As a consequence, governments must now contend with allocating finite risk mitigation resources to an ever increasing number of potential threats that may enter a country.

Traditionally, pests and diseases have been prioritised based on methods that account for life history traits, species geography and some measure of potential economic, environmental and/or social harm. An exemplar of this type of prioritisation has been the "Weed Risk Assessment" [2]. While these methods have been shown to be effective in prioritising existing and potential threats, they rarely identify which regions are most exposed to emerging threats, and therefore, do not inform where to allocate finite post-border surveillance, especially for maximising early detection.

Over the last two decades, *risk maps* have been developed for many plant pests and diseases [3–6]. These maps have been used in plant biosecurity as a decision support tool for informing strategic and tactical decisions about where to allocate finite surveil-lance resources, international and domestic trade restrictions, and cost sharing between jurisdictions [3, 6]. They have also been used to inform likelihoods of pest absence [7] and are a critical communication tool for raising public awareness in regions of high exposure, and thus, increasing the coverage and effectiveness of public reporting of incursions. Pest risk maps are becoming such a vital component of an effective and efficient biosecurity system that many governments and industries are now heavily investing into some form of risk mapping [8].

This strong interest in pest and disease risk mapping has resulted in a vast array of methods and tools being developed. Several books and reviews have been published to guide users through the development of such maps. However, no standardised approach currently exists, with many pest risk maps varying in the stage/barriers of the invasion continuum they incorporate (i.e. entry, establishment, spread), which risk attributes they use (e.g. climate suitability, host availability, trade), what scale they are estimated at (e.g. region, cities, ports), and whether outputs are qualitative, quantitative, or include impacts.

In our view, pest risk maps used to inform tactical biosecurity decisions such as where to conduct early detection surveillance or risk mitigation resources must, at their foun-

dation, encapsulate the three fundamental barriers to pest establishment and spread [9], namely:

- 1. can it reach the location of interest (i.e. arrival rate)?
- 2. are the abiotic conditions suitable (e.g. climate suitability)?
- 3. are the biotic conditions suitable (e.g. presence of host/food)?

If maps do not account for these three barriers, then estimated risks of establishment are likely to be overstated in some regions and understated in others. For example, if a pest risk map is based on a combination of climatic suitability and host availability, but does not account for propagule pressure or arrival rate, it will almost certainly overstate the expsoure risk in regions where arrival rates are low, and understate it in regions where they are high.

In order to assess whether these three funamental barriers are routinely incorporated into plant pest risk maps and/or risk assessments, we conducted a systematic review of grey (i.e publicly available government reports) and peer-reviewed literature published between 1991 and 2018. Specifically, we examined which barriers were routinely incorporated, which data sources and tools were used to estimate them, and which, if any, potential impacts were included.

Material and Methods

Literature search

We conducted systematic searches via Google Scholar using the search term ("risk assessment" OR "risk map") AND (pathway OR entry OR arrival) AND (map OR mapping OR spatial OR geographic) AND (pest OR invasive). We also used references listed in two recent meta-analyses on risk assessment [10] and pathway-models [11]. To ensure we also obtained relevant grey literature, we also conducted searches on government websites using the search terms "risk maps" and "plant" AND "risk assessment". All searches were conducted between October and December 2018. Non-English literature was not examined in this review.

In total, we compiled 74 articles and reports published between 1991 and 2018. Geographic zones spanned from Antarctica [12] to the tropics (Fig. 3.1), with the majority of reports/studies originating from Europe and the USA. A majority of risk analyses collated in this review followed various guidelines [e.g. 13–22]. A synthesis of these reviews can be found in [23]. Broadly, these guidelines advised on: 1) what were main components of risk assessment; 2) how to source relevant data and information; and 3) how to deal with uncertainty.

Data inclusion and exclusion

In total, we inspected the full text of 136 articles/reports/maps. Individual studies had to satisfy the following criteria to be included in our review: 1) report a quantitative or qualitative estimate of at least one of pest entry, establishment, spread, and/or impacts; and 2) focus on plant pests or their specific pathways of entry. We excluded studies with an emphasis on pest ranking for priority lists as our focus was on risk assessments used for informing post-border early detection and control measures [e.g. 24]. In each document, we recorded which components (i.e. stages) of the invasion process were included (i.e. entry, establishment and spread), what risk attributes were used and whether impacts or consequences were estimated. We also recorded the location of the risk assessment, the type of spatial data used (see Section 3.1), and what methods and data were used.



Figure 3.1: Geographical distribution of risk assessments collated in this report.

3.1 Definitions

Risk analysis and risk assessment

In this report the terms "risk assessment" or "risk analysis" (RA) refer to any documents (maps, reports, articles) that attempt to measure risk associated with entry, establishment, spread or impact of plant pests.

Spatially implicit versus spatially explicit

We found three types of pest risk assessments: 1) geographically neutral, when spatial considerations are not taken into account for any risk attribute; 2) spatially implicit, where regional differences are taken into account, but actual location is not important; and 3) spatially explicit, where data inputs or processes are spatially explicit and where the geographic distribution of risk is important.

Entry, establishment, spread and impact

Pest risk assessments were generally decomposed into quantitative or qualitative scores of pest entry, establishment, spread and impact. However, some authors used different demarcations. For example, some used the terminology "introduction" to refer to both "entry" and "establishment" [25–30], while others used it to refer to "establishment" and "spread" [31]. Here, we used the terminology entry, establishment, spread and impact, and reassigned others terminologies following IPCC definitions [32] (Table 3.1).

Methods used to develop risk maps

We identified three categories of methods: 1) methods used to estimate geographic risk components; 2) methods used to integrate multiple risk components (i.e. entry, establishment, spread and impact) or attributes (e.g. habitat suitability, host distribution) on a map, and 3) methods used to combine both risk and uncertainty on a map.

We reported nine types of methods used to estimate the geographic risk component; these ranged from weighted rules applied to risk sub-components to complex processbased species distribution models (Table 3.2). We also reported four methods used to integrate multiple risk components on a map (Table 3.3), and an additional three methods used to account for risk uncertainty (i.e. integrating both risk and incertainty on a map, Table 3.4).

Raster data, points/lines/nodes, polygons/centroids

We reported which format of spatial data was used to address the spatial heterogeneity in risk. Data have been attributed to three main categories (Table 3.5): 1) vector data as points, lines or nodes; 2) vector data as polygons or centroids; and 3) raster data. Vector data represent features as discrete points, lines (or arcs, i.e. connection between two points), and polygons (i.e. areas with closed boundary) or centroids (i.e. areas around a point, see Table 3.5). Raster data represent the landscape as a rectangular matrix of square cells (grid, see Table 3.5). We also reported cases when the risk was spatially implicit.

Table 3.1: Definition of the risk of entry, establishment, spread and impact, based on IPCC terminology [32].

Components of	Definition
LISK	
Entry	Movement of a pest into an area where it is not yet present, or present but
5	not widely distributed and being officially controlled.
Establishment	Perpetuation, for the foreseeable future, of a pest within an area after entry.
Spread	Expansion of the geographical distribution of a pest within an area.
Impact	Adverse consequences of pests, including those concerning uncultivated/unmanaged plants, wild flora, habitats and ecosystems.

Table 3.2: Commonly used methods to estimate components of risk.

Method	Definition	References
Correlative species distribution models	Models (e.g. MaxEnt, CLIMEX, CLIMATCH, BIOCLIM, GLMs, BRTs, GARP) that focus on correlating occurrence records with spatially explicit environ- mental data. These models implicitly capture processes delimiting the distri- bution of a species.	[33–35]
Process-based species distribution models	Models (e.g. degree-day models, NAPPFAST, CLIMEX, NicheMapper) that explicitly model the physiological processes governining a species' distribution.	[4, 27, 33, 35]
Weighted-rules	The weighted multiplication of standardised risk attributes. Here weights represent the relative importance of each risk attribute. Weights are commonly derived from expert judgement.	[33, 36, 37]
Pathway models	Models that predict the movement of propagules or vectors from source lo- cations to points of entry (e.g. ports). Commonly, these models are based on trade networks and can be either spatially implicit or explicit.	[11, 38, 39]
Distance-decay functions	Functions that assume risk or likelihoods decay with distances from high risk locations (e.g. points of entry & established populations). These functions are frequently used by USDA forest service in developing pest risk maps.	[37]
Statistical methods	Various statistical methods such as linear regression, or probability distribu- tions (Poisson, binomial function) have been used to estimate components of risk.	[40, 41]
Maximum overlay	A process in which multiple risk sub-components are overlayed and the max- imum score used.	[42]
Population models	Models that predict changes in population dynamics over time, including how their size and composition changes (e.g. Leslie matrices, mortality models, Ricker models). While these can be made spatially explicit, the vast majority are spatially implicit.	[40, 43–47]
Dispersal models	Models that predict the dispersal dynamics of a species (e.g. kernels, gravity models, linear models, jump diffusion). These model can be stochastic or deterministic.	[46, 48]

Table 3.3: Methods used to integrate multiple risk components or attributes when forming risk maps.

Method	Definition	References
Map overlays & masking	Two approaches are encapsulated here. The first works by map- ping a risk attribute (e.g. entry likelihoods) and then masking ar- eas considered low risk or unsuitable (e.g. via information about climatic suitability or host distribution). An alternative is to not combine risk attributes but rather plot them independently and allow users to examine the geographic distribution of each risk measure	[42, 49, 50]
Weighted-rules	The weighted multiplication (or summation) of standardised risk attributes in geographic space.	[33, 36, 37]
Spatial models	All key risk attributes and processes are modelled in a single spatially explicit model. Commonly, these models incorporate stochastic processes such that uncertainties can also be explicitly accounted for.	[51–55].
Pareto dominance	A game theory method that uses multiple criteria to inform deci- sion making. The method works by iteratively ranking locations based on their multi-criteria risk scores.	[4, 56]

Table 3.4: Methods used to account for risk uncertainty

Method	Definition	References
Thresholding	Areas of high uncertainty are masked in the mapping procedure. The threshold of what constitutes high uncertainty should align with the level of uncertainty that is acceptable to the risk manager.	[33]
Categorisation	Risk classification systems that partition risk scores based on both their expectation and associated uncertainty. For example, a clas- sification such as "high risk; high certainty" and "high risk; low uncertainty".	[57]
Optimisation	Methods that rank locations according to both risk and its uncer- tainty.	mean variance approach: [53] ; second-degree stochastic dominance [52, 53]; hyper-volume based approach [54, 55]

 Table 3.5: Definition of the format of spatial data reported to address spatial heterogeneity in risk.

Spatial data	Definition	Used to represent features such as
Vector data as points or lines	Points are XY coordinates, lines (or arcs) are pairs of coordinates, nodes are the intersec-	Points: ports ([38]), major cities ([58]), distribution centres ([59]).
	tion points where two or more lines meet. They represent a common XY coordinate	Lines: road segments ([38, 48]), rivers ([60]).
	pairs between intersecting lines. Generally, XY coordinates are a latitude and longitude.	Nodes: retailers, nurseries ([44]), major cities ([38]) in a transport network.
Vector data as poly- gons or centroids	Polygons are areas with a closed boundary created by multiple lines. A centroid is gen- erally the centre of a polygon or area.	Polygons: regions ([33]), shires ([40]), provinces ([39]). Centroids: ZIP Code area centroids
Raster data (grid)	The landscape is represented as a rectangu- lar matrix of square cells (grid), where each cell has a value (e.g. satellite images, aerial photographs, digitised maps).	([61]). Grids: climatic maps, but also vege- tation cover (satellite images in [33]), metropolitan area (nightime view by satellites images in [62, 63]), edaphic conditions (digitised maps in [64]).

Results

4.1 Description of the studies reported

In total, we identified 74 relevant documents published between 1991 and 2018. These documents came from both grey literature (i.e. government and international organisation reports, 47%), and peer-reviewed literature (53%). These documents came in a variety of formats such as scientific journal articles, technical reports, maps, books, and guidelines (Fig. 4.1). Guidelines and theoretical papers with no working examples were excluded from subsequent analyses (i.e. 16% of studies).



Figure 4.1: Type of documents reported in this review. Guideline refers to RA framework provided by governmental or intergovernmental agencies.

We found that while the majority of pest RAs focused on a particular species or disease (90%), others focused on modelling pathways irrespective of the species. Risk assessments on arthropods were the most common, followed by fungi and protists, plants, nematodes and molluscs, and bacteria (Fig. 4.2). Most risk assessments were single species (73%), however, others included as many as 800 species [65].

Both quantitative and qualitative methods appear to be equally represented within the risk assessment literature. However, over the last decade, the average trend suggests

that quantitative approaches are becoming increasingly used (though this trend is not significant; Fig. 4.3).

Pest risk assessments commonly reported on one (31%) or two (35%) stages of the invasion process (i.e. entry, establishment, spread and impact). Fewer reported on three (15%) or all (24%) components. Establishment component is the most commonly estimated (in 89% of RAs), followed by entry (66%), spread (45%), and impact (32%).

4.2 Description of risk maps

Most RAs (82%) were spatially explicit, 16% were spatially implicit, and 3% geographically neutral. Some studies contained both spatially explicit and implicit examples. Almost all spatially explicit RAs provided a risk map; 2% reported results in another media [e.g. 58].

Risk maps were commonly in the form of semi-quantitative rank scores (54%), but quantitative (26%) and exclusively qualitative (i.e. low vs high risk, 20%) approaches were also commonly used.

Risk maps either integrated one (56%) or multiple stages of invasion risk (44%, Fig. 4.4). Depending on the stage of invasion considered, maps used a variety of metrics such as the likelihood of pest entry, the potential impact of a pest establishment, or projected resource at risk (i.e. the distribution of timber volumes of potential tree hosts at risk of infestation). Methods used to integrate several components on a risk map varied from simple overlays of maps, sometimes with weighting rules, to highly complex spatial models or optimisation procedures (Fig. 4.5 a, Table 3.4).

Only 2% of risk maps accounted for uncertainty in measured risk. The three main methods for accounting for uncertainty were: threshold mapping, uncertainty classification and optimisation procedures (Fig. 4.5 b, Table 3.5).



Figure 4.2: Risk assessments by taxa. *Unspecified* refers to cases where no examples – and thus no taxa – were reported.



Figure 4.3: The use of quantitative and qualitative methods in plant pest risk assessment over the last 28 years. Lines show the average number of studies by year (\pm 95% CI).



Figure 4.4: Stages of invasion process encapsulated as either a single risk map or a map that integrates multiple stages.



Figure 4.5: Methods used to create risk maps: a) Common methods used to integrate multiple risk components or attributes; b) Methods used to account for uncertainty; c) Methods used to estimate risk at each step of the invasion process.



Figure 4.6: Risk attributes commonly defined in pest risk assessments for each stage of the invasion process (including impact). Grey = Non-spatial risk attributes; shades of blue = spatial data. Different blue colours signify the formats of spatial data commonly used.

4.2.1 Entry risk

Entry risk was incorporated in 44% of risk maps. Entry risk was commonly inferred directly from trade data (86%, Fig. 4.6), followed by natural spread (14%), international travellers (i.e. airlines passengers, 9%), and pests escaping from human facilities (e.g. urban areas, landfills, quarantine zones), 9%).

Entry risk was typically reported as either likelihoods of entry or as the number of entering propagules. Commonly, these methods assumed entry likelihoods were strongly correlated with trade volumes [e.g. 33, 55], or the density of importers [e.g. the number of nurseries within a territory 34].

Commonly, entry risk models contained a large number of attributes describing likelihoods of survival, dispersal and/or transportation. However, the vast majority of these likelihoods were spatially neutral (Fig. 4.6). For example, a species' dispersal limit was commonly assumed to not vary across geographic or environmental space.

Several methods have been used to estimate entry risk (Fig. 4.5 c): weighted-rules (45%), pathway models (36%) and distance-decay functions (41%) were the most used.



Figure 4.7: Type of data used to estimate risk in pest risk maps.

Most RAs used databases on trade to estimate entry risk (77%, Fig. 4.7). However, expert elicitation and data derived from the literature were also commonly used 86%. In general most data was readily available online. In total we found 107 databases were used to estimate risks associated with entry, establishment, spread and impact. A complete list can be found in Table 5.1.

Entry risk was mostly estimated at points of entry (e.g. ports) but also at high risk areas such as markets, distribution centres and urban area (Fig. 4.8). However, in some cases, some analyses estimated territory or national arrival rates, and then distributed them within the region of interest using risk proxies such as human population density, housing density (see city lights data [66]), employees numbers, or traffic survey data.

4.2.2 Establishment risk

Most risk maps incorporated some measure of establishment likelihood (96%, Fig. 4.4). Typically, this was done as a function of climate (75%, Fig. 4.6), but also included the distribution of hosts (69%), land cover (33%) and other abiotic factors such as elevation and disturbance (29%). Plant species richness and phenology were also sometimes used (in less than 6%). In the vast majority of cases, these factors were used in raster format (i.e. mapped as a grid, Fig. 4.5). While establishment risk was mostly considered a spatially explicit problem, spatially implicit examples did exist. For example, some



Figure 4.8: Points of entry commonly used in pest risk maps.

risk assessments either estimated establishment risk as a proportion of suitable area for establishment [e.g. 30], or more simply, assumed a country or region was at high risk if it contained any suitable environment [e.g. 67].

Climate and habitat suitability was mostly inferred via niche modelling (54%); with correlative and process-based models being equally used (Fig. 4.5 c). Thus, establishment risk was often provided as a suitability index, rather than a likelihood of establishment – however, users typically inferred that high suitability equalled high establishment risk. Data used to parameterise these models mostly originated from online climate (e.g. WorldClim: http://www.worldclim.org, CRU: http://www.cru.uea.ac.uk and CliMond: https://www.climond.org) and biological (e.g. GBIF; https://www.gbif.org) databases (83%), but also from the literature (especially for process-based parameters; 42%).

4.2.3 Spread risk and impact

Estimates of spread risk were found in 22% of plant pest risk assessments (Fig. 4.4). Population and dispersal models were commonly used to estimate spread risk (Fig. 4.5). However, pathway models that encapsulated both entry and establishment risk attributes were also used. In most cases, two types of spread were encapsultated in these models: human-assisted spread (64%, Fig. 4.6) and natural spread (36%). Data used to parametrise these models came mainly from biological databases of pest occurence (e.g. National Gypsy Moth Slow the Spread (STS) program; http://www.gmsts.org, GBIF), and from literature review. Very few risk assessments for plant pests explicitly estimated impact (4% of RAs). When a risk assessment did incorporate impacts, they were mostly associated with economic impacts (i.e. crop yield and at industry level), followed by environmental (e.g. impacts on ecosystem services or biodiversity components) and societal impacts of plant pests (Fig. 4.6). Thus data used came mainly from literature review, and databases about land uses (e.g. ABARES catchment-scale land use dataset; http://www.agriculture.gov.au/abares/aclump/land-use/catchment-scale-land-use-reports).

Discussion

5.1 Accounting for conditional dependencies and impacts

A vast majority of RAs provides risk maps (81%). Ideally, pest risk maps used to inform post-border surveillance should 1) estimate likelihoods of establishment and subsequent spread conditioned on spatially explicit arrival probabilities, and 2) include measures of potential impact. In practice, such risk maps are exceedingly rare [6]. In our review we found that only 4% of maps met this standard. Specifically, we found that while most risk assessments incorporated some spatially explicit information about entry and establishment, most treated these two components independently and did not incorporate any measure of impact. For example, half the risk assessments reviewed here either inferred establishment risk directly from a niche model without any consideration of arrival rates or simply overlayed estimated arrival rates onto some form of habitat suitability model and inferred risk without any formal integration of these scores [50].

This approach to risk without accounting for the conditional dependencies between entry and establishment or accounting for impact is unsurprising [19]. The development of a comprehensive risk map requires integration of sophisticated modelling techniques from multiple disciplines. Specifically, a pathway analysis is required to estimate arrival rates at various points of entry and beyond. Following this, detailed biological knowledge and possibly habitat suitability models are required to infer likelihoods of establishment given estimated arrivals. And lastly, economic and spread modelling are needed to infer potential econonomic, societal and environmental impacts across space given estimated establishment likelihoods. Sourcing data and integrating these methods into a spatially explicit model is no easy task (but see [57]), which likely explains why few plant risk assessments appriopriately acknowledge conditionals along the invasion process or account for potential impacts.

5.2 Accounting for uncertainty

Risk uncertainties occur at all stages of the invasion process [68], and commonly arise due to knowledge gaps, model uncertainty and the robustness of assumptions made (see a list in [19]). Quantifying and understanding how these uncertainties propagate from

entry through to establishment, spread and impact is of vital importance in informing policy and management decisions in biosecurity.

We found that most plant pest risk assessments incorporated some measure of uncertainty when communicating the conclusion of the assessment. However, few accounted for uncertainties on derived risk maps (2%), and those that did focused mainly on model uncertainty. When uncertainty was accounted for on maps, three approaches were commonly used: 1) high uncertainty areas were masked (or highlighted) (e.g. [33]); 2) categorical scoring systems that included uncertainty levels were used (e.g. high risk; low uncertainty, high risk; high uncertainty) [57]; or 3) optimisation methods (e.g. mean-variance, hyper-volume based approach, second-degree stochastic dominance [52–55]). Each method has advantages: thresholding for mapping, and classification, are both relatively easy to implement, whereas results from optimisation techniques are simple to interpret.

5.3 Estimating entry risk

Understanding entry risk is of critical importance for managing and mitigating risk hitting the border, and for informing post-border surveillance [69]. However, 45% of risk assessments reported here did not explicitly estimate entry likelihoods. Entry risk was not assessed in RA when 1) the pest was already present in the territory (see guidelines [17]), 2) the pest was intentionally released for the control of another invasive pest [70], 3) the pest is not known to exist outside the examined territory [41], or 4) in academic studies interested in developing better analytical method to estimate estblishment or spread risks.

Unsurprisingly, entry via trade was the main factor used to estimate entry risk (51% of RAs). International trade is often considered a good predictor of the number of invasive species found within a country's borders [71–75]. Entry risk via trade was estimated at marine ports. This is likely because 90% of world trade is carried across sea [76], and consequently passes through marine ports. However, the increasing proportion of overseas shipments that are containerised and transferred "door to door" (i.e. they are sealed at origin, opened at the final destination, and they do not have to be opened or reloaded at a port) emphasises a need to consider additional entry points distributed away from ports. Thus, several RAs also distributed entry risk away from ports, by urban areas, distribution centres or markets within a territory.

Most of the quantitative information on trade flow or on pest interception rates comes from national or international database on agricultural (fruits, plants) or production (e.g. EUROSTAT, FAOSTAT, Army Corps of Engineer, Waterborne Commerce, Foreign Cargo Statistics, National transportation Database, Transport Canada, EURO-PHYT, Agricultural Quarantine Inspection Monitoring [AQIM]) databases (see Table 5.1). Studies distributing entry risk at a finer geographic scale required detailed knowledge on the movement of specific commodities within the territory. Yet, data on the volume of trade distributed by urban areas or distribution centres are not always accessible. Thus, in general, the distribution of entry risk within a territory required the use of proxies such as human population density [27, 33, 42, 77], housing density [62, 63, 66], or information on transport networks [77–79]. The critical assumption with this approach is that higher values of these metrics increase propagule pressure. Pests generally have a notable dispersal capacity. Yet, entry via natural spread (i.e. wind) was estimated in only 14% of the risk maps reported. We could hypothesise that in a majority of studies entry via natural spread was not considered because 1) natural spread was not expected to be a major pathway [17], or 2) the distance of the territory from a source of infestation was higher than the pest's dispersal range [33, 80]. When entry via natural spread was estimated, risk assessors had to 1) identify the factors that support migration, and 2) model the process, adding complexity and increasing the work load for the risk assessment. For instance, [33] studied the potential for Spodoptera frugiperda to arrive in Europe. In the course of their study, they performed a model entry via migration based on wind trajectories and corn crop availability from North African countries. This choice was based on an analysis of the literature review, suggesting that migration of S. frugiperda in the USA was supported by corn availability and wind patterns. Thus, they implemented a hysplit (Hybrid Single Particle Lagrangian Integrated Trajectory) trajectory model for S. frugiperda to describe where wind trajectories might carry species of concern from source to Europe and in what numbers. This model was spatially explicit because entry rate via migration was distributed in Europe according to the establishment index.

International travellers (i.e. airlines, rivers) and their baggage are known to be a potential pathway for pest species [81]. Yet, entry via international travellers was addressed in only 9% of the risk maps reported. We could hypothesise that in a majority of studies, entry via international travellers was not considered because 1) international travellers were not expected to be a major pathway for the pest, or 2) data on the rate of baggage interception was not always available (though most countries, including USA and Australia, have interception rates on airline baggages). For instance, [12], studied the risk for the establishment of nonindigenous species in Antarctica. They considered the entry risk via air passengers, but they estimated the number of propagules per visitor by collecting seeds from their outer clothing, footwear, walking poles, day packs, and camera bags. Entry risk via international travellers could be spatially distributed by airports, or along a distance from major transport axes (e.g. Danube river in [48]).

Three important factors were also often used to estimate entry risk: 1) pest traits (flight range, growth rate), 2) probability of pest transfer to a host during transport or at destination, and 3) probability of pest survival during harvesting, transportation, consignment and use (in 29%, 9%, and 5% of RAs reported, respectively). Yet, spatial variation in the value of these factors was never taken into account. At best, a centroid curvilinear distance decay value was assigned at points of entry (e.g. port or distribution centres) to delineate the potential flight range of the pest. But there was no geographical variation in pest traits (flight range) and consequently in the distance decay value across points of entry [e.g. 37]. In [82], the "probability of pest transfer to a host" was reported as spatially explicit (by regions). In fact, the probability of transfer depends on the vulnerability of the EU regions, and varied across the European territory according to: 1) the size of the area occupied by vulnerable hosts – the greater the area occupied, the more likely there is an opportunity for pest contact; and 2) the mobility of the pest – the greater the mobility of the pest, the greater is the opportunity for pest contact.

Pathways models were the main method used to estimates entry risk. This is coherent with a shift in the focus of plant biosecurity programs away from control at international borders and towards pathway-based pest risk assessment [83]. Pathway models varied greatly in mathematical form and level of detail (see [11]), but invariably they focused

on the means by which a species might arrive in the area of concern. For instance, EFSA often used pathway models in their risk assessment: the assessment of entry is made separately for each pathway for a pest, and the entry for each pathway is assessed in successive steps from the source abroad to one or more points within the European territory [e.g. 39].

Pathway models used in RAs could be either spatially explicit or implicit. In spatially explict pathway models, entry risk was distributed by regions [e.g. 46, 67, 84], or urban area [51]. In spatially implicit pathway models, regional differences could matter, but not the precise location of entry points. For instance, in [67] three classes of countries are distinguished according to the reported pest prevalence (higher prevalence, lower prevalence and absent), and calculations are made for each category. The spatial resolution is thus at the levels of the country class, but the precise location of the countries had no importance for the model.

5.4 Estimating establishment risk

Modelling establishment risk in a spatially explicit manner is very informative for risk managers because it clarifies in which areas establishment may occur. Establishment risk was almost always estimated in risk maps using habitat suitability, and often host distribution, abiotic factors and land cover (e.g. forest cover, urban forest).

Habitat suitability was modelled using either correlative distribution models, and/or process-based models that integrate distribution data with some expert estimates on physiological parameters (e.g. CLIMEX). The result was often a map that ranks each site in a geographical region by the relative suitability for the species. Effects of several environmental variables such as host distribution, or land cover were often combined in a summarised map. Factors were combined within a weighted overlay or simple multiplication (i.e. all the same weight), followed by a summation of the results [e.g. 37, 59, 85, 86].

It is surprising that a large range of countries (e.g. Australia, New Zealand, USA, South Africa, Europe) used species distribution modelling, despite a lack of data on species' distribution and environmental covariates [6]. In fact, while most of the data on pest species' distributions and environmental covariates came from databases (GBIF, CABI, EPPO Global Database, Early Detection and Distribution Mapping System [EDDMaps], Invasive Plant Atlas of the United States), some studies (42%) also integrated data from the literature. Risk assessors conducted literature searches and combined data from local surveys or museum occurrences datasets [e.g. 33, 64].

5.5 Estimating spread risk and impact

Spread was addressed in 22% of the risk maps. Natural and human-assisted spread were the most studied aspects of spread risk (in 36% and 64% RAs, respectively, Fig. 4.6). In some cases, spread risk was spatially explicit using dispersal models [46, 58] or fitting equations to historical data in order to estimate the rate of spread [41].

Economic (i.e. crop yield and at industry level) and environmental (e.g. impacts on ecosystem services or biodiversity components) impacts were addressed in less than

4% of the risk maps. A lot of information is required to estimate the impact of a pest. Indeed, impact depends on factors such as expected yield losses, change in crop quality, biodiversity losses, and also on pest abundance. However, in general, the geographical heterogeneity of impact risk was related to differences in distribution of crops/protected habitats within a territory.

5.6 Conclusion

Pest risk maps are a powerful decision support tool used in biosecurity to inform strategic and tactical decisions about pest invasion. However, complete, spatially explicit characterisation of risk is rare; most risk maps are still used to inform only a single step of the invasion process. In ideal cases, pest risk maps should 1) integrate all steps of the invasion process and their consequences, or at a minimum, entry and establishment; 2) take into account potential interdependencies between the main steps of the invasion process; and 3) simultaneously integrate and convey both risk and uncertainty into maps. Pest risk mapping remains challenging compared to spatially implicit or geographical neutral RAs, because it requires additional knowledge on the pest (e.g. distribution of entry within the territory), entails added complexity, and increases the work load. However, several techniques have been developed recently to help risk assessors to better integrate spatial components into RAs.

Table 5.1: Database

Risk of	Data for	Location	Database
Entry	trade volume and	Europe	EUROSTAT, EUROSTAT Nimexe database (1967-1987),
	movement		EUROSTAT Comext database (since 1988)
		Europe	FAOSTAT
		Europe	Comtrade
		Europe	ISEFOR trade data
		Europe	Cargo Handbook Transport of commodities
		Canada	Canadian roadside survey (CRS) database maintained by
			Transport Canada
		Canada	Canadian roadside survey (CRS) database maintained by
			Transport Canada
		Canada	StatsCan, Canadian Freight Analysis Framework
		USA	TrafficMetrix database
		USA	US Department of Transportation's freight analysis frame-
			WOIK (FAF)
		USA, Alaska	terborne Commerce
		USA, Alaska	U.S. National Transportation Atlas Database (NTAD)
	interception data	Europe	EUROPHYT
		USA	Animal and Plant Health Inspection Service (APHIS), U.S. Department of Agriculture
	location of businesses	USA, Alaska	ReferenceUSA database
	location of nursery stock, retailers	UK	Grid square agricultural census data for England, Scotland and Wales
	location of retailers	UK	Inspection records of Plant Health and Seeds Inspectorate (PHSI)
	location of marketing groups for growers	UK	2006 HortWeek Suppliers Guide
	location of cities	Worldwide	Environmental Systems Research Institute Inc. (ESRI)'s GIS mapping
	location of urban area	USA, Alaska	World at Night Environmental Research Institute (ESRI) in Redlands, California (see Imhoff et al. 1997)

Continuation of Table 5.1			
Risk of	Data for	Location	Database
		USA, Alaska	USA cities lity lights dataset National Oceanic and Atmo-
			spheric Administration's National Geophysical Data Cen-
			ter (NOAAINGDC) (see Imhoff et al. 1997)
	number of employees	USA, Alaska	North American Industry Classification System (NAICS),
	in business		Business and Industry, U.S. Census Bureau
	population density	Europe	EUROSIAI
		USA	US Census Bureau
		USA	California GAP dataset
	road density	USA	California GAP dataset
		USA	Tele Atlas Dynamap Transportation
	quantity of municipal solid waste	USA	USA New York Department of Sanitation
	mail data	USA	USA Forwarding data compiled by United States Postal Service (USPS)
	natural spread (wind trajectories)	Europe	International Research Institute for Climate and Society (Earth Institute) Columbia University, New York, NY, NOAA
Establishment	climatic data	Antarctica	NASA Global Modelling and Assimilation Office
		Australia,	CLIMEX
		Europe	
		Europe	European Climate data website
		Europe	Meteorological Data Base, from Join Research Center (JRC), Monitoring Agricultural Resources (MARS) unit, Euro- pean Commission
		Europe, worldwide	Climatic Research Unit (CRU) database
		South Africa	Computing Centre for Water Research (CCWR), University of Natal, Pietermaritzburg
		UK	UK meteorological office
		USA	USA California Geoportal
		USA	USADaymet model
		USA	NAPPFAST pest database (pest database threshold and
			growing degree days)

Continuation of Table 5.1			
Risk of	Data for	Location	Database
		USA	USA PRISM Climate Group from Spatial Climate Analysis
			Service at Oregon State University
		USA, Eu-	World Clim
		rope, UK	
		Antarctica	CSIRO Mk3.5 Climate Model
	host / pest distribution or information	UK	British Society of Britain and Ireland (BSBI)
		UK	maps from Little's Atlas of United States, trees' series
		USA	NYC Street Tree and Forest Surveys
		USA	Monterey Pine data, Remote Sensing Lab, Sacramento, CA, USDA Forest Service.
		USA	USDA National Arboretum
		USA	The Jepson Online Interchange for California Floristics
		USA	Biota of North America Program (BONAP), North American Plant Atlas
		USA	USA Early Detection and Distribution Mapping System (EDDMaps), from the University of Georgia - Center for Invasive Species and Ecosystem Health
		USA	Invasive Plant Atlas of the US
		USA, Alaska	Forest Inventory and Analysis (FIA) data, U.S. Forest Service
		USA, Alaska	USDA Plants Herbarium data
		USA, Alaska	USDA Plants database
		USA, Europe	Centre for Agriculture and Bioscience International (CABI)
		Europe	EPPO Global Database
		Europe	Fauna Europaea vector distribution
		UK, Europe, worldwide	Global Biodiversity Information facility (GBIF)
		worldwide	PRECIS database from South African National Biodiversity Institute (SANBI)
		worldwide	Seed Information Database
	forest damage	Alaska	Alaska Department of Natural Resources, USDA Forest Service Forest Health Protection, State and Private Forest

Continuation of Table 5.1		
Risk of Data for	Location	Database
	USA	USA National Park services
drought	USA	Forest Inventory and Analysis (FIA) data, U.S. Forest Ser-
		vice
fire	Alaska	Alaska Fire Service
	USA	NASA Goddard Space Flight Center
	USA	Program for Climate, Ecosystem and Fire Applications (CEFA) from Desert Research Institute (DRI)
lightening	USA	NASA Lightning Imaging Sensor (LIS) / Optical Transient Detector (OTD)
tornadoes and hurri canes and maximum winds	USA	National Climate Atlas of the United States
geochemical data	Europe	Geochemical Database of europe
hydric soil layer	USA	USDA, Natural Resources Conservation Service (NRCS)
incoming solar radia- tion	- USA	Surface Radar Topography Mission (SRTM), SRTM 90m DEM Digital Elevation Database
soil data	Europe	European soil database, from the Joint Research Center (JRC) and the european soil data centre (ESDAC)
	USA	Natural Resources Conservation Service (NRCS) soils data
soil wetness dryness	USA	State Soil Geographic (STATSGO) Database, from USDA
index		Natural Resources Conservation Service (NRCS)
surface hydrology data	Australia	Surface Hydrology Data, from the Australian National Surface Water Information, Geoscience Australia
topology	USA	California Geoportal
	USA	USA global digital elevation model (DEM) - Global 30 Arc- Second Elevation (GTOPO30)
	USA	PRISM Climate Group, from Spatial Climate Analysis Service at Oregon State University
	USA	PRISM Climate Group, from Spatial Climate Analysis Service at Oregon State University
	USA	Surface Radar Topography Mission (SRTM), SRTM 90m DEM Digital Elevation Database
	USA	US National Map, from U.S. Geological Survey's (USGS) National Geospatial Program

Continuation of Table 5.1			
Risk of	Data for	Location	Database
	land cover	Alaska	Alaska Geospatial Data Committee
		Antarctica	Antarctic Digital Database, from the Australian Antarctic
			Data Centre
		Europe	Corine Land Cover data
		Europe	European Space Agency's Global Land Cover 2000 project
		USA	CALVEG database
		USA	Global land cover facilities
		USA	Multiresolution Land Characteristics Consortium (MLRC)
		USA	National Agriculture Imagery Program (NAIP), from USDA's Farm Service Agency (FSA)
		USA	US National Land Cover Data (NLCD)
	vegetation index	USA	NASA MODIS Vegetation Index Products (NDVI and EVI)
	vegetation index	UK	Countryside Information System (CIS) Datasets, land clas- sification for the UK, England, Scotland and Wales
	phenology	USA	NASA MODIS land surface phenology
	protected area	Europe	Natura 2000 databases, European Commission
	-	UK	Site of Special Scientific interest, from Scottish Natural Her- itage
Spread		USA	National Gypsy Moth Slow the Spread (STS) program, U.S. Forest Service
Impact	land use (crop distri- bution, production)	Australia	ABARES catchment-scale land use dataset (ABARES), De- partment of Agriculture and Water resources, Australia
	-	Europe	EPPO Global Database
		Europe	EUROSTAT
		USA	Census of Agriculture, from USDA National Agricultural Statistics services

Acknowledgements

This report is a product of the Centre of Excellence for Biosecurity Risk Analysis (CE-BRA). In preparing this report, the authors acknowledge the financial and other support provided by the Australian Department of Agriculture, Water and the Environment and its predecessors, the New Zealand Ministry for Primary Industries, and the University of Melbourne.

Bibliography

- [1] Philip E Hulme. "Trade, transport and trouble: managing invasive species pathways in an era of globalization". In: *Journal of Applied Ecology* 46.1 (Feb. 2009), pp. 10–18.
- [2] PC Pheloung, PA Williams, and SR Halloy. "A weed risk assessment model for use as a biosecurity tool evaluating plant introductions". In: *Journal of Environmental Management* 57.4 (1999), pp. 239–251. ISSN: 0301-4797.
- [3] Robert C. Venette, Darren J. Kriticos, Roger D. Magarey, Frank H. Koch, Richard H. A. Baker, Susan P. Worner, Nadilia N. Gómez Raboteaux, Daniel W. McKenney, Erhard J. Dobesberger, Denys Yemshanov, Paul J. De Barro, William D. Hutchison, Glenn Fowler, Tom M. Kalaris, and John Pedlar. "Pest Risk Maps for Invasive Alien Species: A Roadmap for Improvement". en. In: *BioScience* 60.5 (May 2010), pp. 349–362. ISSN: 1525-3244, 0006-3568. DOI: 10.1525/bio.2010.60.5.5.
- [4] R. D. Magarey, D. M. Borchert, J. S. Engle, M. Colunga-Garcia, F. H. Koch, and D. Yemshanov. "Risk maps for targeting exotic plant pest detection programs in the United States: US risk maps for exotic plant pest detection". en. In: *EPPO Bulletin* 41.1 (Apr. 2011), pp. 46–56. ISSN: 02508052. DOI: 10.1111/j.1365-2338. 2011.02437.x.
- [5] A. Jiménez-Valverde, A. T. Peterson, J. Soberón, J. M. Overton, P. Aragón, and J. M. Lobo. "Use of niche models in invasive species risk assessments". en. In: *Biological Invasions* 13.12 (Dec. 2011), pp. 2785–2797. ISSN: 1387-3547, 1573-1464. DOI: 10.1007/s10530-011-9963-4.
- [6] R. C. Venette, ed. *Pest risk modelling and mapping for invasive alien species*. en. Wallingford: CABI, 2015. ISBN: 978-1-78064-394-6. DOI: 10.1079/9781780643946. 0000.
- [7] S Barry, Jane Elith, P Caley, M Kearney, P Tenant, and T Arthur. *Estimating probablies of pest absence across geographic space*. Tech. rep. 2018, pp. 1–39.
- [8] Jane Elith. "Predicting distributions of invasive species". In: *Invasive species risk assessment and management*. Ed. by A. P. Robinson, T Walshe, Mark A Burgman, and M Nunn. Cambridge, 2017, pp. 93–129.
- [9] Jane A Catford, Roland Jansson, and Christer Nilsson. "Reducing redundancy in invasion ecology by integrating hypotheses into a single theoretical framework". In: *Diversity and Distributions* 15.1 (Jan. 2009), pp. 22–40.
- [10] Brian Leung, Nuria Roura-Pascual, Sven Bacher, Jaakko Heikkilä, Lluis Brotons, Mark A. Burgman, Katharina Dehnen-Schmutz, Franz Essl, Philip E. Hulme, David M. Richardson, Daniel Sol, and Montserrat Vilà. "TEASIng apart alien species risk assessments: a framework for best practices". en. In: *Ecology Letters* 15.12 (Dec. 2012). Ed. by Marcel Rejmanek, pp. 1475–1493. ISSN: 1461023X. DOI: 10.1111/ele.12003.

- [11] J.C. Douma, M. Pautasso, R.C. Venette, C. Robinet, L. Hemerik, M.C.M. Mourits, J. Schans, and W. van der Werf. "Pathway models for analysing and managing the introduction of alien plant pests: an overview and categorization". en. In: *Ecological Modelling* 339 (Nov. 2016), pp. 58–67. ISSN: 03043800. DOI: 10.1016/j. ecolmodel.2016.08.009.
- [12] S. L. Chown, A. H. L. Huiskes, N. J. M. Gremmen, J. E. Lee, A. Terauds, K. Crosbie, Y. Frenot, K. A. Hughes, S. Imura, K. Kiefer, M. Lebouvier, B. Raymond, M. Tsujimoto, C. Ware, B. Van de Vijver, and D. M. Bergstrom. "Continent-wide risk assessment for the establishment of nonindigenous species in Antarctica". en. In: *Proceedings of the National Academy of Sciences* 109.13 (Mar. 2012), pp. 4938–4943. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1119787109.
- [13] Plant Health Australia. *Pest risk assessment for use in industry biosecurity plans*. Tech. rep. Canberra: Plant Health Australia, 2013, pp. 1–9.
- [14] DWAR- Department of Agriculture and Water Resources. *Biosecurity Import Risk Analysis Guidelines 2016: managing biosecurity risks for imports into Australia,* tech. rep. Canberra: Department of Agriculture and Water Resources, 2016.
- [15] EPPO. Guidelines on Pest Risk Analysis- Lignes directrices pour l'analyse du risque phytosanitaire. Tech. rep. European and Mediterranean Plant Protection Organization, 2011.
- [16] EPPO. "Decision-support scheme for an Express Pest Risk Analysis". en. In: *European and Mediterranean Plant Protection Organization (EPPO) Bulletin* 42.3 (Dec. 2012), pp. 457–462. ISSN: 02508052. DOI: 10.1111/epp.2591.
- [17] EFSA. "Guidance on a harmonised framework for pest risk assessment and the identification and evaluation of pest risk management options by EFSA". en. In: *EFSA Journal* 8.2 (Feb. 2010). ISSN: 18314732, 18314732. DOI: 10.2903/j.efsa. 2010.1495.
- [18] EFSA. "Guidance on Expert Knowledge Elicitation in Food and Feed Safety Risk Assessment: Guidance on expert knowledge elicitation". en. In: *European Food Safety Authority Journal*. European and Mediterranean Plant Protection Organization 12.6 (June 2014), p. 3734. ISSN: 18314732. DOI: 10.2903/j.efsa.2014.3734.
- [19] EFSA, Michael Jeger, Claude Bragard, David Caffier, Thierry Candresse, Elisavet Chatzivassiliou, Katharina Dehnen-Schmutz, Jean-Claude Grégoire, Josep Anton Jaques Miret, Alan MacLeod, Maria Navajas Navarro, Björn Niere, Stephen Parnell, Roel Potting, Trond Rafoss, Vittorio Rossi, Gregor Urek, Ariena Van Bruggen, Wopke Van Der Werf, Jonathan West, Stephan Winter, Andy Hart, Jan Schans, Gritta Schrader, Muriel Suffert, Virag Kertész, Svetla Kozelska, Maria Rosaria Mannino, Olaf Mosbach-Schulz, Marco Pautasso, Giuseppe Stancanelli, Sara Tramontini, Sybren Vos, and Gianni Gilioli. "Guidance on quantitative pest risk assessment". en. In: *EFSA Journal* 16.8 (Aug. 2018). ISSN: 18314732, 18314732. DOI: 10.2903/j.efsa.2018.5350.
- [20] USDA. Guidelines for Plant Pest Risk Assessment of Imported Fruit & Vegetable Commodities. Tech. rep. USDA- U.S. Department of Agriculture, 2012.
- [21] USDA. *Guidelines for the USDA-APHIS- PPQ Weed Risk Assessment Process*. Tech. rep. USDA- U.S. Department of Agriculture, 2016.
- [22] U.S. Environmental Protection Agency. *Guidelines for Ecological Risk Assessment*. Tech. rep. Washington, D.C: U.S. Environmental Protection Agency, 1998.
- [23] Helen E. Roy, Wolfgang Rabitsch, Riccardo Scalera, Alan Stewart, Belinda Gallardo, Piero Genovesi, Franz Essl, Tim Adriaens, Sven Bacher, Olaf Booy, Etienne

Branquart, Sarah Brunel, Gordon Howard Copp, Hannah Dean, Bram D'hondt, Melanie Josefsson, Marc Kenis, Marianne Kettunen, Merike Linnamagi, Frances Lucy, Angeliki Martinou, Niall Moore, Wolfgang Nentwig, Ana Nieto, Jan Pergl, Jodey Peyton, Alain Roques, Stefan Schindler, Karsten Schönrogge, Wojciech Solarz, Paul D. Stebbing, Teodora Trichkova, Sonia Vanderhoeven, Johan van Valkenburg, and Argyro Zenetos. "Developing a framework of minimum standards for the risk assessment of alien species". en. In: *Journal of Applied Ecology* 55.2 (Mar. 2018). Ed. by Ralph Mac Nally, pp. 526–538. ISSN: 00218901. DOI: 10.1111/1365-2664.13025.

- [24] Sunil K. Singh, Gavin J. Ash, and Mike Hodda. "Keeping 'one step ahead' of invasive species: using an integrated framework to screen and target species for detailed biosecurity risk assessment". en. In: *Biological Invasions* 17.4 (Apr. 2015), pp. 1069–1086. ISSN: 1387-3547, 1573-1464. DOI: 10.1007/s10530-014-0776-0.
- [25] Abraham Addo-Bediako, Baharnu Tameru, Louis E N Jackai, and Conrad K Bonsi. "Assessment of Risk of Introduction of Cylas formicarius elegantulus (Coleoptera: Brentidae) into Weevil-Free Areas in the Southern United States". en. In: JOURNAL OF ECONOMIC ENTOMOLOGY 100.2 (2007), p. 8. DOI: http: //dx.doi.org/10.1603/0022-0493(2007)100[315:A0R0I0]2.0.C0;2.
- [26] USDA, John Kliejunas, Harold H. Jr. Burdsall, Gregg A. DeNitto, Andris Eglitis, Dennis A. Haugen, Michael I. Haverty, and Jessie A. Micales. *Pest Risk Assessment* of the Importation into the United States of Unprocessed Pinus Logs and Chips from Australia. Tech. rep. USDA- U.S. Department of Agriculture, 2006.
- [27] USDA. *Risk assessment of Citrus spp. fruit as a pathway for the introduction of Guignardia citricarpa Kiely, the organism that causes Citrus Black Spot disease.* Tech. rep. USDA- U.S. Department of Agriculture, 2010.
- [28] USDA. Weed Risk Assessment for Delairea odorata Lem. (Asteraceae) Cape ivy. Tech. rep. USDA- U.S. Department of Agriculture, 2013.
- [29] USDA. Weed Risk Assessment for Hygrophila polysperma (Roxb.) T. Anderson (Acanthaceae) – Miramar weed. Tech. rep. USDA- U.S. Department of Agriculture, 2015.
- [30] USDA. Weed Risk Assessment for Tripleurospermum inodorum (L.) Sch. Bip. (Asteraceae) – Scentless mayweed. Tech. rep. USDA- U.S. Department of Agriculture, 2018, pp. 1–23.
- [31] Mathieu Rouget, Mark P. Robertson, John R. U. Wilson, Cang Hui, Franz Essl, Jorge L. Renteria, and David M. Richardson. "Invasion debt - quantifying future biological invasions". en. In: *Diversity and Distributions* 22.4 (Apr. 2016). Ed. by Ingolf Kühn, pp. 445–456. ISSN: 13669516. DOI: 10.1111/ddi.12408.
- [32] FAO. International standards for phytosanitary measures: Glossary of Phytosanitary Terms. Glossary of Phytosanitary Terms. ISPM No 5. Rome (Italy): International Plant Protection Convention (IPCC), 2016.
- [33] EFSA, Michael Jeger, Claude Bragard, David Caffier, Thierry Candresse, Elisavet Chatzivassiliou, Katharina Dehnen-Schmutz, Gianni Gilioli, Jean-Claude Grégoire, Josep Anton Jaques Miret, Maria Navajas Navarro, Björn Niere, Stephen Parnell, Roel Potting, Trond Rafoss, Vittorio Rossi, Gregor Urek, Ariena Van Bruggen, Wopke Van der Werf, Jonathan West, Stephan Winter, Roger Day, Regan Early, Allan Hruska, Rodney Nagoshi, Ciro Gardi, Olaf Mosbach-Schultz, and Alan MacLeod. "Pest risk assessment of Spodoptera frugiperda for the European Union". In: *EFSA Journal* 16.8 (Aug. 2018). ISSN: 18314732, 18314732. DOI: 10.2903/j.efsa.2018.5351.

- [34] Emily Haeuser, Wayne Dawson, Wilfried Thuiller, Stefan Dullinger, Svenja Block, Oliver Bossdorf, Marta Carboni, Luisa Conti, Iwona Dullinger, Franz Essl, Günther Klonner, Dietmar Moser, Tamara Münkemüller, Madalin Parepa, Matthew V. Talluto, Holger Kreft, Jan Pergl, Petr Pyšek, Patrick Weigelt, Marten Winter, Martin Hermy, Sebastiaan Van der Veken, Cristina Roquet, and Mark van Kleunen. "European ornamental garden flora as an invasion debt under climate change". en. In: *Journal of Applied Ecology* 55.5 (Sept. 2018). Ed. by Rafael Zenni, pp. 2386– 2395. ISSN: 00218901. DOI: 10.1111/1365-2664.13197.
- [35] Michael R. Kearney, Brendan A. Wintle, and Warren P. Porter. "Correlative and mechanistic models of species distribution provide congruent forecasts under climate change: Congruence of correlative and mechanistic distribution models". en. In: *Conservation Letters* 3.3 (June 2010), pp. 203–213. ISSN: 1755263X. DOI: 10. 1111/j.1755-263X.2010.00097.x.
- [36] Frank J Krist. A Multicriteria Framework for Producing Local, Regional, and National Insect and Disease Risk Maps. en. Tech. rep. U.S. Forest Services, 2012, p. 16.
- [37] FHTET, Marla C. Downing, Lead Daniel M Borchert, Donald A Duerr, Dennis A. Haugen, Frank H. Koch, Frank J. Krist Jr, Frank J. Sapio, Bill D. Smith, and Borys M. Tkacz. *Invasive Pest Risk Maps: Sirex Woodwasp Sirex noctilio*. USA, 2006.
- [38] Denys Yemshanov, Frank H. Koch, Mark Ducey, and Klaus Koehler. "Tradeassociated pathways of alien forest insect entries in Canada". en. In: *Biological Invasions* 14.4 (Apr. 2012), pp. 797–812. ISSN: 1387-3547, 1573-1464. DOI: 10.1007/ s10530-011-0117-5.
- [39] EFSA. "Scientific Opinion on the risk of *Phyllosticta citricarpa* (*Guignardia citricarpa*) for the EU territory with identification and evaluation of risk reduction options: Phyllosticta citricarpa pest risk assessment". en. In: *EFSA Journal* 12.2 (Feb. 2014), p. 3557. ISSN: 18314732. DOI: 10.2903/j.efsa.2014.3557.
- [40] Chris D. Stansbury, Simon J. McKirdy, Art J. Diggle, and Ian T. Riley. "Modeling the Risk of Entry, Establishment, Spread, Containment, and Economic Impact of *Tilletia indica*, the Cause of Karnal Bunt of Wheat, Using an Australian Context". en. In: *Phytopathology* 92.3 (Mar. 2002), pp. 321–331. ISSN: 0031-949X. DOI: 10.1094/PHYT0.2002.92.3.321.
- [41] EFSA, Michael Jeger, Claude Bragard, David Caffier, Thierry Candresse, Elisavet Chatzivassiliou, Katharina Dehnen-Schmutz, Gianni Gilioli, Josep Anton Jaques Miret, Alan MacLeod, Maria Navajas Navarro, Björn Niere, Stephen Parnell, Roel Potting, Trond Rafoss, Vittorio Rossi, Gregor Urek, Ariena Van Bruggen, Wopke Van Der Werf, Jonathan West, Stephan Winter, Domenico Bosco, Xavier Foissac, Gudrun Strauss, Gabor Hollo, Olaf Mosbach-Schulz, and Jean-Claude Grégoire. "Risk to plant health of Flavescence dorée for the EU territory". en. In: *EFSA Journal* 14.12 (Dec. 2016). ISSN: 18314732, 18314732. DOI: 10.2903/j.efsa.2016. 4603.
- [42] FHTET, Marla C. Downing, Ian I F Leinwand, John R Withrow, Gericke L Cook, Lisa F Kenneway, Catherine Jarnovich, and Frank J. Sapio. *Invasive Pest Risk Maps: European Gypsy Moth Lymantria dispar dispar*. English. USA, 2014.
- [43] Steven M. Bartell and Shyam K. Nair. "Establishment Risks for Invasive Species".
 en. In: *Risk Analysis* 24.4 (Aug. 2003), pp. 833–845. ISSN: 0272-4332, 1539-6924. DOI: 10.1111/j.0272-4332.2004.00482.x.
- [44] Thomas D. Harwood, Xiangming Xu, Marco Pautasso, Mike J. Jeger, and Michael W. Shaw. "Epidemiological risk assessment using linked network and grid based

modelling: Phytophthora ramorum and Phytophthora kernoviae in the UK". en. In: *Ecological Modelling* 220.23 (Dec. 2009), pp. 3353–3361. ISSN: 03043800. DOI: 10.1016/j.ecolmodel.2009.08.014.

- [45] Olav Skarpaas and Bjørn Økland. "Timber import and the risk of forest pest introductions". en. In: *Journal of Applied Ecology* 46.1 (Feb. 2009), pp. 55–63. ISSN: 00218901, 13652664. DOI: 10.1111/j.1365-2664.2008.01561.x.
- [46] EFSA, Michael Jeger, Claude Bragard, Elisavet Chatzivassiliou, Katharina Dehnen-Schmutz, Gianni Gilioli, Josep Anton Jaques Miret, Alan MacLeod, Maria Navajas Navarro, Björn Niere, Stephen Parnell, Roel Potting, Trond Rafoss, Gregor Urek, Ariena Van Bruggen, Wopke Van der Werf, Jonathan West, Stephan Winter, Alberto Santini, Panagiotis Tsopelas, Irene Vloutoglou, Marco Pautasso, and Vittorio Rossi. "Risk assessment and reduction options for Ceratocystis platani in the EU". en. In: *EFSA Journal* 14.12 (Dec. 2016). ISSN: 18314732, 18314732. DOI: 10.2903/j.efsa.2016.4640.
- [47] EFSA, Michael Jeger, Claude Bragard, David Caffier, Thierry Candresse, Elisavet Chatzivassiliou, Katharina Dehnen-Schmutz, Gianni Gilioli, Jean-Claude Grégoire, Josep Anton Jaques Miret, Alan MacLeod, Bjoern Niere, Stephen Parnell, Roel Potting, Trond Rafoss, Vittorio Rossi, Gregor Urek, Ariena Van Bruggen, Wopke Van Der Werf, Jonathan West, Stephan Winter, Filippo Bergeretti, Niklas Bjorklund, Olaf Mosbach-Schulz, Sybren Vos, and Maria Navajas Navarro. "Pest risk assessment of Eotetranychus lewisi for the EU territory". en. In: *EFSA Journal* 15.10 (Oct. 2017). ISSN: 18314732, 18314732. DOI: 10.2903/j.efsa.2017.4878.
- [48] L.R. Carrasco, J.D. Mumford, A. MacLeod, T. Harwood, G. Grabenweger, A.W. Leach, J.D. Knight, and R.H.A. Baker. "Unveiling human-assisted dispersal mechanisms in invasive alien insects: Integration of spatial stochastic simulation and phenology models". en. In: *Ecological Modelling* 221.17 (Aug. 2010), pp. 2068–2075. ISSN: 03043800. DOI: 10.1016/j.ecolmodel.2010.05.012.
- [49] Judith E Pasek, Harold H Burdsall, Joseph F Cavey, Andris Eglitis, and Robert A Haack. "Pest Risk Assessment for Importation of Solid Wood Packing Materials into the United States". en. In: (2000), p. 293.
- [50] Allan N. D. Auclair, G. Fowler, M. K. Hennessey, A. T. Hogue, M. Keena, D. R. Lance, R. M. McDowell, D. O. Oryang, and A. J. Sawyer. "Assessment of the Risk of Introduction of <I>Anoplophora glabripennis</I> (Coleoptera: Cerambycidae) in Municipal Solid Waste from the Quarantine Area of New York City to Landfills Outside of the Quarantine Area: A Pathway Analysis of the Risk of Spread and Establishment". en. In: *Journal of Economic Entomology* 98.1 (Feb. 2005), pp. 47–60. ISSN: 00220493, 00220493. DOI: 10.1603/0022-0493-98.1.47.
- [51] Manuel Colunga-Garcia, Robert A. Haack, and Adesoji O. Adelaja. "Freight Transportation and the Potential for Invasions of Exotic Insects in Urban and Periurban Forests of the United States". en. In: *Journal of Economic Entomology* 102.1 (Feb. 2009), pp. 237–246. ISSN: 00220493, 00220493. DOI: 10.1603/029.102.0133.
- [52] Denys Yemshanov, Frank H. Koch, D. Barry Lyons, Mark Ducey, and Klaus Koehler. "A dominance-based approach to map risks of ecological invasions in the presence of severe uncertainty: A dominance-based approach to map ecological invasions". en. In: *Diversity and Distributions* 18.1 (Jan. 2012), pp. 33–46. ISSN: 13669516. DOI: 10.1111/j.1472-4642.2011.00848.x.
- [53] Denys Yemshanov, Frank H. Koch, Mark Ducey, and Klaus Koehler. "Mapping ecological risks with a portfolio-based technique: incorporating uncertainty and
decision-making preferences". en. In: *Diversity and Distributions* 19.5-6 (May 2013). Ed. by Mark Burgman, pp. 567–579. ISSN: 13669516. DOI: 10.1111/ddi.12061.

- [54] Denys Yemshanov, Frank H. Koch, Bo Lu, Ronald Fournier, Gericke Cook, and Jean J. Turgeon. "A new hypervolume approach for assessing environmental risks". en. In: *Journal of Environmental Management* 193 (May 2017), pp. 188–200. ISSN: 03014797. DOI: 10.1016/j.jenvman.2017.02.021.
- [55] Juha Tuomola, Denys Yemshanov, Hanna Huitu, and Salla Hannunen. "Mapping risks of pest invasions based on the spatio-temporal distribution of hosts". en. In: *Management of Biological Invasions* 9.2 (2018), pp. 115–126. ISSN: 19898649. DOI: 10.3391/mbi.2018.9.2.05.
- [56] Denys Yemshanov, Frank H. Koch, Yakov Ben-Haim, and William D. Smith. "Detection capacity, information gaps and the design of surveillance programs for invasive forest pests". en. In: *Journal of Environmental Management* 91.12 (Dec. 2010), pp. 2535–2546. ISSN: 03014797. DOI: 10.1016/j.jenvman.2010.07.009.
- [57] Denys Yemshanov, Frank H. Koch, Daniel W. McKenney, Marla C. Downing, and Frank Sapio. "Mapping Invasive Species Risks with Stochastic Models: A Cross-Border United States-Canada Application for *Sirex noctilio* Fabricius". en. In: *Risk Analysis* 29.6 (June 2009), pp. 868–884. ISSN: 02724332, 15396924. DOI: 10.1111/j.1539-6924.2009.01203.x.
- [58] David J. Nowak, Judith E. Pasek, Ronaldo A. Sequeira, Daniel E. Crane, and Victor C. Mastro. "Potential Effect of <I>Anoplophora glabripennis</I> (Coleoptera: Cerambycidae) on Urban Trees in the United States". en. In: *Journal of Economic Entomology* 94.1 (Feb. 2001), pp. 116–122. ISSN: 00220493, 00220493. DOI: 10.1603/0022-0493-94.1.116.
- [59] FHTET, Marla C. Downing, Daniel M Borchert, Robert A Haack, Frank H. Koch, Roger D. Magarey, Bill D. Smith, and Frank J. Sapio. *Invasive Pest Risk Maps Oak Splendor Beetle - Agrilus biguttatus (Fabricus)*. English. USA, 2010.
- [60] Belinda Gallardo and David C. Aldridge. "Priority setting for invasive species management: risk assessment of Ponto-Caspian invasive species into Great Britain". en. In: *Ecological Applications* 23.2 (Mar. 2013), pp. 352–364. ISSN: 1051-0761. DOI: 10.1890/12-1018.1.
- [61] FHTET, Marla C. Downing, Frank H. Koch, Frank J. Krist Jr, Frank J. Sapio, and Bill D. Smith. *Invasive Pest Risk Maps- Asian Longhorned Beetle - Anoplophora glabripennis*. English. USA, 2009.
- [62] FHTET, Marla C. Downing, Lead Daniel M Borchert, Donald A Duerr, Dennis A. Haugen, Frank H. Koch, Frank J. Krist Jr, Frank J. Sapio, Bill D. Smith, and Borys M. Tkacz. *Invasive Pest Risk Maps: Phytophthora disease of alders - Phytophthora alni ssp alni (Conterminous United States)*. USA, 2007.
- [63] FHTET, Marla C. Downing, Lead Daniel M Borchert, Donald A Duerr, Dennis A. Haugen, Frank H. Koch, Frank J. Krist Jr, Frank J. Sapio, Bill D. Smith, and Borys M. Tkacz. *Invasive Pest Risk Maps: Phytophthora disease of alders - Phytophthora alni ssp alni (Alaska)*. USA, 2008.
- [64] Laura Arriaga, Alejandro E. Castellanos V., Elizabeth Moreno, and JesúS AlarcóN. "Potential Ecological Distribution of Alien Invasive Species and Risk Assessment: a Case Study of Buffel Grass in Arid Regions of Mexico: *Potential Invasion of Buffel Grass in Arid Regions*". en. In: *Conservation Biology* 18.6 (Dec. 2004), pp. 1504–1514. ISSN: 08888892. DOI: 10.1111/j.1523-1739.2004.00166.x.

- [65] Jenica M. Allen and Bethany A. Bradley. "Out of the weeds? Reduced plant invasion risk with climate change in the continental United States". en. In: *Biological Conservation* 203 (Nov. 2016), pp. 306–312. ISSN: 00063207. DOI: 10.1016/j.biocon. 2016.09.015.
- [66] Marc L. Imhoff, William T. Lawrence, David C. Stutzer, and Christopher D. Elvidge. "A technique for using composite DMSP/OLS "City Lights" satellite data to map urban area". en. In: *Remote Sensing of Environment* 61.3 (Sept. 1997), pp. 361–370. ISSN: 00344257. DOI: 10.1016/S0034-4257(97)00046-1.
- [67] EFSA, Michael Jeger, Claude Bragard, David Caffier, Thierry Candresse, Elisavet Chatzivassiliou, Katharina Dehnen-Schmutz, Gianni Gilioli, Jean-Claude Grégoire, Josep Anton Jaques Miret, Alan MacLeod, Maria Navajas Navarro, Björn Niere, Stephen Parnell, Roel Potting, Trond Rafoss, Vittorio Rossi, Ariena Van Bruggen, Wopke Van Der Werf, Jonathan West, Stephan Winter, Olaf Mosbach-Schulz, and Gregor Urek. "Risk to plant health of Ditylenchus destructor for the EU territory". en. In: *EFSA Journal* 14.12 (Dec. 2016). ISSN: 18314732, 18314732. DOI: 10.2903/j.efsa.2016.4602.
- [68] FAO. International standards for phytosanitary measures: Pest risk analysis for quarantine pests. Tech. rep. ISPM No 11. Rome (Italy): International Plant Protection Convention (IPCC), 2017.
- [69] Melodie A. McGeoch, Piero Genovesi, Peter J. Bellingham, Mark J. Costello, Chris McGrannachan, and Andy Sheppard. "Prioritizing species, pathways, and sites to achieve conservation targets for biological invasion". en. In: *Biological Invasions* 18.2 (Feb. 2016), pp. 299–314. ISSN: 1387-3547, 1573-1464. DOI: 10.1007/s10530-015-1013-1.
- [70] EFSA. "Risk to plant health in the EU territory of the intentional release of the bud-galling wasp Trichilogaster acaciaelongifoliae for the control of the invasive alien plant Acacia longifolia". en. In: EFSA Journal 13.4 (Apr. 2015). ISSN: 18314732, 18314732. DOI: 10.2903/j.efsa.2015.4079.
- [71] Jonathan M. Levine and Carla M. D'Antonio. "Forecasting Biological Invasions with Increasing International Trade". en. In: *Conservation Biology* 17.1 (Feb. 2003), pp. 322–326. ISSN: 0888-8892, 1523-1739. DOI: 10.1046/j.1523-1739.2003.02038.x.
- [72] Vladimir Hlasny and Michael J. Livingston. "Economic Determinants of Invasion and Discovery of Nonindigenous Insects". en. In: *Journal of Agricultural and Applied Economics* 40.01 (Apr. 2008), pp. 37–52. ISSN: 1074-0708, 1074-0708. DOI: 10.1017/ S1074070800027966.
- [73] Michael I. Westphal, Michael Browne, Kathy MacKinnon, and Ian Noble. "The link between international trade and the global distribution of invasive alien species". en. In: *Biological Invasions* 10.4 (Apr. 2008), pp. 391–398. ISSN: 1387-3547, 1573-1464. DOI: 10.1007/s10530-007-9138-5.
- [74] Hanno Seebens, Tim M. Blackburn, Ellie E. Dyer, Piero Genovesi, Philip E. Hulme, Jonathan M. Jeschke, Shyama Pagad, Petr Pyšek, Marten Winter, Margarita Arianoutsou, Sven Bacher, Bernd Blasius, Giuseppe Brundu, César Capinha, Laura Celesti-Grapow, Wayne Dawson, Stefan Dullinger, Nicol Fuentes, Heinke Jäger, John Kartesz, Marc Kenis, Holger Kreft, Ingolf Kühn, Bernd Lenzner, Andrew Liebhold, Alexander Mosena, Dietmar Moser, Misako Nishino, David Pearman, Jan Pergl, Wolfgang Rabitsch, Julissa Rojas-Sandoval, Alain Roques, Stephanie Rorke, Silvia Rossinelli, Helen E. Roy, Riccardo Scalera, Stefan Schindler, Kateřina

Štajerová, Barbara Tokarska-Guzik, Mark van Kleunen, Kevin Walker, Patrick Weigelt, Takehiko Yamanaka, and Franz Essl. "No saturation in the accumulation of alien species worldwide". en. In: *Nature Communications* 8 (Feb. 2017), p. 14435. ISSN: 2041-1723. DOI: 10.1038/ncomms14435.

- [75] Reid Tingley, Pablo García-Díaz, Carla Rani Rocha Arantes, and Phillip Cassey. "Integrating transport pressure data and species distribution models to estimate invasion risk for alien stowaways". en. In: *Ecography* 41.4 (Apr. 2018), pp. 635–646. ISSN: 09067590. DOI: 10.1111/ecog.02841.
- [76] International Maritime Organization. *International shipping facts and figures Information resources on trade, safety, security, environment.* Tech. rep. London, UK: IMO, 2012.
- [77] Shawna J. Dark. "The biogeography of invasive alien plants in California: an application of GIS and spatial regression analysis: Invasive alien plants in California". en. In: *Diversity and Distributions* 10.1 (Jan. 2004), pp. 1–9. ISSN: 13669516, 14724642. DOI: 10.1111/j.1472-4642.2004.00054.x.
- [78] Margaret L. Margosian, Karen A. Garrett, J. M. Shawn Hutchinson, and Kimberly A. With. "Connectivity of the American Agricultural Landscape: Assessing the National Risk of Crop Pest and Disease Spread". en. In: *BioScience* 59.2 (Feb. 2009), pp. 141–151. ISSN: 1525-3244, 0006-3568. DOI: 10.1525/bio.2009.59.2.7.
- [79] Natalie Clare Banks, Dean Ronald Paini, Kirsty Louise Bayliss, and Michael Hodda. "The role of global trade and transport network topology in the humanmediated dispersal of alien species". en. In: *Ecology Letters* 18.2 (Feb. 2015). Ed. by Montserrat Vila, pp. 188–199. ISSN: 1461023X. DOI: 10.1111/ele.12397.
- [80] R. W. Sutherst, G. F. Maywald, and W. Bottomley. "From CLIMEX to PESKY, a generic expert system for pest risk assessment". en. In: *EPPO Bulletin* 21.3 (Sept. 1991), pp. 595–608. ISSN: 0250-8052, 1365-2338. DOI: 10.1111/j.1365-2338.1991.tb01293.x.
- [81] Andrew M. Liebhold, Timothy T. Work, Deborah G. McCullough, and Joseph F. Cavey. "Airline Baggage as a Pathway for Alien Insect Species Invading the United States". en. In: *American Entomologist* 52.1 (2006), pp. 48–54. ISSN: 1046-2821, 2155-9902. DOI: 10.1093/ae/52.1.48.
- [82] J. Holt, A.W. Leach, J.D. Mumford, A. MacLeod, D. Tomlinson, R. Baker, M. Christodoulou, L. Russo, and A. Marechal. "Development of probabilistic models for quantitative pathway analysis of plant pest introduction for the EU territory". en. In: *EFSA Supporting Publications* 13.8 (Aug. 2016), 1062E. ISSN: 23978325. DOI: 10.2903/sp.efsa.2016.EN-1062.
- [83] P. E. Hulme, S. Bacher, M. Kenis, S. Klotz, I. Kühn, D. Minchin, W. Nentwig, S. Olenin, V. Panov, J. Pergl, P. Pyšek, A. Roques, D. Sol, W. Solarz, and M. Vilà. "Grasping at the routes of biological invasions: a framework for integrating pathways into policy". en. In: *Journal of Applied Ecology* 45.2 (Apr. 2008), pp. 403– 414. ISSN: 00218901, 13652664. DOI: 10.1111/j.1365-2664.2007.01442.x.
- [84] EFSA, Michael Jeger, Claude Bragard, David Caffier, Thierry Candresse, Elisavet Chatzivassiliou, Katharina Dehnen-Schmutz, Gianni Gilioli, Jean-Claude Grégoire, Josep Anton Jaques Miret, Alan MacLeod, Maria Navajas Navarro, Björn Niere, Stephen Parnell, Roel Potting, Trond Rafoss, Vittorio Rossi, Gregor Urek, Wopke Van Der Werf, Jonathan West, Stephan Winter, Ciro Gardi, Olaf Mosbach-Schulz, Ioannis Koufakis, and Ariena Van Bruggen. "Pest risk assessment of

Diaporthe vaccinii for the EU territory". en. In: *EFSA Journal* 15.9 (Sept. 2017). ISSN: 18314732, 18314732. DOI: 10.2903/j.efsa.2017.4924.

- [85] Ross Meentemeyer, David Rizzo, Walter Mark, and Elizabeth Lotz. "Mapping the risk of establishment and spread of sudden oak death in California". en. In: *Forest Ecology and Management* 200.1-3 (Oct. 2004), pp. 195–214. ISSN: 03781127. DOI: 10.1016/j.foreco.2004.06.021.
- [86] Frank J Jr Krist, James R Ellenwood, Meghan E Woods, Andrew J McMahan, John P Cowardin, Daniel E Ryerson, Frank J Sapio, Mark O Zweifler, and Sheryl A Romero. 2013-2027 National Insect and Disease Forest Risk Assessment. en. Tech. rep. U.S. Forest Services, 2014, p. 106.

B. Workshop

On the 11th of October 2018, CEBRA held a workshop in Canberra. In total 21 participants from various government agencies, academic institutions and private enterprises attended the workshop (Table B.1). These participants had a diverse range of expertise including: expert biological knowledge of various high priority plant pests and diseases, expertise in designing and implementing biosecurity surveillance programs, expertise in developing biosecurity policy and technical expertise in developing and using species distribution models. This diversity of expertise provided us with the ideal opportunity to elicit 1) how biosecurity practitioners currently estimate the potential distribution of pests and diseases that are not currently present in a country; and 2) what were the major areas of uncertainty they faced when developing and using such models to inform biosecurity policy and management.



Table B.1.: List of workshop attendees

Participant	Affiliation
Andrew Robinson	CEBRA
James Camac	CEBRA
Jane Elith	University of Melbourne
Elise Gould	University of Melbourne
Ranjith Subasinghe	Department of Agriculture
Kylie Calhoun	Department of Agriculture
Sophie Peterson	Department of Agriculture
Carrina Moeller	Department of Agriculture
Catherine Mathenge	Department of Agriculture
Christine Bennett	Department of Agriculture
Nathaniel Bloomfield	ABARES
Simon Barry	CSIRO
Peter Caley	CSIRO
Matt Hill	CSIRO
Louise Shuey	QDAFF
Rebecca Laws	QDAFF
Peter Whittle	AgKonect
John Weiss	DEDJTR
Martin Mebalds	DEDJTR
Craig Phillips	AgResearch



To elicit this information, we first asked all participants to identify the major uncertainties they faced at each of the five stages of model development and utility when estimating a pest's climate suitability. A summary of these uncertainties can be found within figure B.1.



Figure B.1.: The five stages of invasive species distribution model development and utility

Once major uncertainties were identified, participants were split into four roughly equal groups (three groups of 5 and one group of 6). To ensure each group contained a range of expertise and that discussions stayed on point, we ensured each group had a facilitator and at least one technical and policy person. Groups then discussed each point of uncertainty and presented these discussions to members of different groups. Group facilitators then presented a summary of these discussions to members of other groups, who provided feedback, and where warranted, requested clarification or greater justification for how various uncertainties are or could be tackled. In the following sections we provide a brief summary of these discussions associated with each point of uncertainty identified in Fig. B.1.

Data sourcing

Group discussions revealed that several data sources have either been used, or theoretically could be used, and each contained a variety of strengths and weaknesses (Table B.2). These included online databases, peer-reviewed and grey literature, expertderived data, surveillance data, museum and herbarium collections.

Online databases

General consensus among participants was that large online databases such as the Global Biodiversity Information Facility (GBIF - https://www.gbif.org) and the Centre for Agriculture and Bioscience International (CABI - https://www.cabi.org/isc) were commonly used for parametrising and validating potential distribution models informing pest climatic suitability. This was because they are easy to access, contained vast amounts of data for many species have large spatial coverage.

Participants highlighted that GBIF was most commonly used for sourcing data for parametrising correlative models or validating mechanistic models. The primary advantage of GBIF is that it is currently the most comprehensive global database with over 1 billion occurrence records from more than one million species (See Fig 3.2).



Moreover, GBIF contains highly accurate spatial coordinates, making it more useful for correlating occurrences with various climate rasters.

Information found in CABI was also commonly used in developing potential distribution models for pests and disease. CABI contains a large collection of pest data sheets that are compiled by experts. These data sheets provide up to date information on what is known about pest biology, distribution and spread pathways and many other risk factors. An example of the brown marmorated stink bug (BMSB) data sheet can be found at https://www.cabi.org/isc/datasheet/27377. Participants stated that CABI summaries were useful in finding relevant information for parametrising mechanistic models such as CLIMEX, or narrowing down the covariates used in correlative models. However, CABI distributional data was rarely used to parametrise correlative models because occurrence records were typically at country or county accuracy. Rather, participants tended to use CABI distributional data to examine the completeness of GBIF records (i.e. whether there were records in all countries with known established populations), as well as identifying records from countries with no known established populations that may require further scrutiny (see data cleaning).

Participants acknowledged that large biodiversity databases are commonly riddled with taxonomic and geographic errors and can sometimes source data from potentially dubious sources. However, most databases contain extensive record meta data that can be used to remove dubious or erroneous records. Moreover, because of the large userbase, there are many online tutorials and functions that have been developed to aid in flagging and removing erroneous or dubious records.

Table B.2.: Data sources identified by participants with associated strengths and weaknesses and whether they are commonly used								
Data source	Online Databases	Literature	Experts	Surveillance	Collections			
Example	GBIF & CABI	Scientific journals Government reports Grey literature	Expert elicitation	Interception & Eradication data (e.g. GERDA)	Museums & Herbariums			
Strengths	 Easily accessible Cover large extents Large volumes of data for many species Collation & standardisation of multiple datasets Some data cleaning & verification done High spatial resolution Historical records 	 Scientifically peer-reviewed Detailed pest biology & distribution data Peer-reviewed distribution models Published data 	 Readily obtainable Multiple elicitation methods exist Useful for filling data gaps Useful for informing parameter selection Useful for model validation 	 Records of pest establishment Identifies pathway risk Use for model validation 	 Historical records Contains specimens for verification Often catalogued in large online databases (e.g. GBIF) 			
Weaknesses	 Validity records not always known Can contain GPS & taxonomic errors Location data may at city or country level (e.g. CABI) Survey effort often unknown Geographically & temporally biased Records may not represent successful establishment Duplicate records 	 Can be difficult to reproduce (Raw data often not supplied with papers) Time consuming to conduct literature review Government agencies may not have access Literature may be restricted in space, time or scope Lack of standardisation Grey literature difficult to find & access Language barriers formida 	 Reliability unknown & difficult to verify May not contain uncertainty Difficulties in defining an expert Distrust in expert-derived data Answers are dependent on how questions are asked 	 Absences rarely recorded Restricted in space & time Reliability dependent on detection probabilities, pest prevalence & surveillance effort 	 Requires networking (i.e. Who to contact?) Locations may be inaccurate or coarse Survey effort unknown 			
Commonly used?	Yes extensively	Yes	Sometimes	Not often	Yes, but mostly from online databases			

~ L**I**• o d whether th المحديد معا J ----- 1. -Det : .1.

Camac et al. 2020

C e hear fer h*da* elysk



Literature

Participants all agreed that the first step in developing any type of climate suitability model for invasive species should be to first conduct a literature review of both the peer-reviewed and grey literature. This is because such reviews provides invaluable information on:

- Pest biology (e.g. physiological tolerances)
- Susceptible hosts
- Risk pathways
- Existing species distribution models; and
- Up to date occurrence records and physiological processes

Published species distribution models were considered to be particularly valuable to biosecurity practitioners. This was because they contain the additional scientific rigour of being peer-reviewed by experts in the field, and therefore, were generally held in higher regard when used to inform biosecurity policy and management decisions. Also, if two more more models are published, differences between outputs can be used to highlight uncertainties. Moreover, using published models ultimately reduces the financial and time burdens faced by government agencies when developing climate suitability models, especially if a department lacks the required expertise.

Grey literature, such as those found in academic theses or governments reports, were also considered an invaluable source of information and data. This is because these documents often contain important details associated with pest/disease biology or country status (i.e. presence/absence, localised or widespread) that may not be published in traditional academic journals.

However, participants raised several issues commonly faced when conducting literature reviews. First, extensive literature reviews often take several weeks to months to conduct, making them prohibitively time consuming when departments are required to build models and make decisions over very short time periods (e.g. following an outbreak).

Second, many government departments do not have access to pay walled peerreviewed literature. As such, government staff are often reliant on directly contacting authors¹ or relying on their networks within academic institutions to provide them with required materials. Access to grey literature, especially those authored by governments of other countries, was also considered difficult to find and access. Moreover, grey literature from other countries may require translation further adding to the barriers faced to information uptake. Again, government participants highlighted that it was of vital importance to maintain and build international networks with both academic and government institutions in order to remain up to date with latest grey literature.

Lastly, most reports and many scientific publications do not include the raw data required to validate physiological parameters or reproduce models². This is particularly problematic with species distribution papers because they generally summarise

¹Requiring author contact details provided on manuscripts to be up-to-date.

²Note that within the academic literature there has been a strong to improve the access to data with some journals requiring data to be uploaded to an online repository prior to publication.

results as maps and do not provide the mathematical functions used to derive these maps. Also, methods are often not adequately reported and so can't be scrutinised. As a consequence, biosecurity practitioners and academics are unable to reproduce these analyses at the desired resolution and extent required to inform management and policy.

Experts

Workshop participants indicated that both technical and biological experts are often consulted when developing species distribution models for use in informing biosecurity management. Specifically, experts were asked to provide advice on methodological approaches, covariate selection, whether model outputs were biologically plausible, and in some cases, make recommendations between competing models.

As a data source, experts were also commonly used to fill information gaps such as providing parameter estimates (or parameter uncertainty/ranges) in mechanistic models that that were otherwise unpublished or not known. In rare circumstances, when there is a paucity of data to parametrise correlative (e.g. few occurrence records) or mechanistic models (i.e. few parameter estimates), or when an appropriate congener species does not exist, potential distributions of pests and diseases can also be derived solely from expert judgements.

While expert-derived data is considered to be a useful and timely means of obtaining parameter estimates that are otherwise unknown, general consensus was that such data should only be used as a last resort – when there is insufficient data to fit a correlative or mechanistic model. Moreover, some participants highlighted that there are several issues related to using expert-derived data. These included:

- Difficulties in defining who is an expert
- Difficulties in examining expert reliability
- Expert answers may be highly dependent on how a question is asked; and
- A general distrust in the community of expert-derived data.

Expert elicitation is a science in of itself, and as such, several techniques have been developed increase scientific rigour and tackle some of these concerns (e.g. IDEA protocol; Hemming *et al.* 2017). Because of the issues raised above, several participants recommended that such techniques be used whenever eliciting information from experts.

Surveillance

Some participants raised that a wealth of information is collected as part of formal surveillance and eradication programs, and that these may provide an important data source for estimating the potential distribution of pests and diseases. For example, absence data derived from targeted surveillance programs can be used for determining the likelihood of pest absence in the region in which they were collected. The global eradication and response database (GERDA; http://b3.net.nz/gerda/faq.php) was also raised as an important data source on determining which countries currently have (or have had) an eradication program, and consequently, which countries are



cebra

- are typically highly restricted in space;
- are focused on delimiting incursion boundaries, eradication success or pest status, as opposed inferring potential distributions as a function of climate or habitat characteristics;
- are difficult to access from other countries (e.g. country targeted surveillance data);
- are highly dependent on surveillance effort and detection rates³.

Collections

Collections from museums, herbaria or experts were also considered an invaluable source of information that can be used to parametrise correlative models or validate mechanistic models. A major advantage of these data are is that they are either expert-verified or are often accompanied by voucher specimens that can be used to verify record taxonomy – especially geographic outliers. However, participants acknowl-edged sourcing these data directly is becoming less common due to the rise of large online databases such as GBIF which often source and collate such data. Moreover, data that is not currently present in online databases are often difficult to find and/or access, typically requiring substantial international networks with various museums, institutions and experts.

Data cleaning

The decisions made throughout the data cleaning process are likely to contribute to model uncertainty. This is because different decisions made throughout will ultimately emphasis different aspects of the data and lead to varying parametrisations of the models and consequently differing model outputs. To better understand how biosecurity practitioners process their data for use in correlative and mechanistic modelling, we asked them to describe the workflows they typically implement when cleaning data. We also asked participants whether they tended to parametrise correlative models using a subset of occurrence records (e.g. native only records), how they assessed data sufficiency and what they did in situations when there was inadequate data. A summary of these discussions are found below.

How are sourced data commonly cleaned by biosecurity practitioners?

Most discussions about data cleaning tended to focus on removing erroneous and irrelevant occurrence records, identifying and removing occurrence records that had inaccurate GIS coordinates, identifying geographic outliers and determining whether records were likely to be from established populations. Here, occurrence records are

³Technically this is true of other data sources too



directly used to parametrise correlative models (e.g. Maxent, GAMS, GLMS and climate matching algorithms) and can be used to validate the accuracy of mechanistic models.

A consolidated list of data cleaning steps commonly used by participants are listed below:

1. Remove erroneous occurrence records

- a) Records not on landmass (assuming terrestrial pest)
- b) Records not within or near documented country (if supplied)
- c) Records with high spatial error (if documented)⁴
- d) Records centred on country centroids (indicative of country-level spatial accuracy)
- e) Records in outlier/unusual locations with limited or no supportive information

2. Remove irrelevant records

- a) Documented fossil records
- b) Documented cultivated specimens
- c) Data from dubious/untrustworthy sources
- d) Duplicate records

3. Verify records are likely from established populations

- a) Cross reference records with CABI country status
- b) Cross reference occurrence records with biological experts

Participants highlighted that much of the occurrence data used to parametrise and validate species distribution models is often sourced directly from GBIF. In terms of data cleaning, a major advantage of these databases is that they commonly contain meta data associated with each record which can then be used to identify errors or possible issues. GBIF in particular has invested significant resources in developing algorithms for flagging potentially erroneous data and include this flag under the named field "issues" – A full list of GBIF issues can be found at:

https://gbif.github.io/gbif-api/apidocs/org/gbif/api/vocabulary/ OccurrenceIssue.html. Another advantage is that these databases have a large user-base, and as such comprehensive tutorials and functions for flagging and removing common errors are readily available. For example, a comprehensive tutorial on how to clean data from GBIF (and other databases) using R can be found at:

https://ropensci.github.io/CoordinateCleaner/articles/Cleaning_GBIF_ data_with_CoordinateCleaner.html.

Some participants also highlighted that occurrence records from databases such as GBIF should ideally be cross referenced against published CABI country pest statuses. The reasoning here is that CABI distributional data contains detailed information on how widespread a pest is within a given country, and thus, whether that country

⁴The definition of what is an acceptable error will depend on the spatial resolution one wishes to fit and project their model onto



has established populations. Consequently, some participants use CABI distributional data to remove GBIF records in countries with no known established population, thus, maximising the likelihood that occurrence records are representative of suitable locations for pest establishment.

Cleaning data for parametrising mechanistic models was not raised by directly participants. Presumably this could be due to these models being less frequently used by biosecurity practitioners, and because such data are generally sourced directly from published literature.

Should data be restricted to a particular region

Given that there are varied opinions as to whether correlative models should be parametrised using data from 1) the native regions only; 2) native + long-invaded regions; 3) invaded region; or 4) the entire known distribution, we asked participants whether they subsetted their distributional data prior to model fitting.

Many participants indicated that their preference was to use the entire distribution of cleaned known pest establishment records when parametrising correlative invasive species distribution models. The primary reason for this view was that such models would more closely approximate the fundamental niche of a species. However, technical experts highlighted that the use of occurrence records from the invaded range is often a double-edge sword. While they agreed that the inclusion of such records will likely increase the representation of suitable environments, it will also invariably create difficulties in determining appropriate background points for commonly used two-class models such as Maxent because of almost certain violations of the equilibrium assumption near invasion fronts and differential survey effort between native and invaded ranges. Technical experts then outlined a variety of sophisticated approaches that could be implemented to reduce sampling biases and assumption violations. These included:

- Using Target Group Sampling (TGS): Presence locations of taxonomically related species obtained using the same survey techniques and usually collated using the same database are used to estimate the geographic distribution of sampling bias;
- Sampling background points from a buffer around each and every occurrence data point;
- Sampling background points according to bioregions;
- Use ordination analysis to examine whether there were any significant groupings among occurrence records

How to assess data sufficiency for correlative models?

Participants raised several practical approaches for examining data sufficiency for fitting correlative models. Mostly, these approaches focused on examining the adequacy of geographic coverage, with participants highlighting that the extent of collated occurrences could be compared with published range maps or CABI distributional data. Some participants used rules of thumbs such as requiring a minimum of 30 occurrences distributed across the known species range before they would consider fitting correlative models. The presumption here was that this minimum number was required to

obtain a reasonable estimate of a species' response. However, it acknowledged that they were unaware where this rule of thumb originated and whether it had been scientifically verified. Lastly, some participants took a more statistical approach and either examined data sufficiency as a function of a model's capability of accurately predicting the suitability of locations that once contained established populations or how influential individual occurrence points were on model predictions.

What to do when there is too little data?

Participants provided various answers as to what they did when there was inadequate data to parametrise correlative or mechanistic models of pest/disease climate suitability. These included:

- Don't attempt to estimate climate suitability, but rather develop maps as a function of other attributes of risk (e.g. arrival rates or host presence);
- Estimate climate suitability using data from an exemplar species (i.e. a species from a similar taxonomic family or that has a similar suite of physiological or morphological traits);
- Use expert elicitation approaches to construct physiological models and define environmental thresholds;
- Develop sophisticated trait-based models whereby a pest's climate suitability is determined as a function of its morphological traits⁵.

Model development

Once data has been compiled and cleaned, modellers are faced with a range of decisions as to how to estimate the potential distribution of a pest or disease. Fundamental among these decisions is determining which modelling method to use⁶, and if using a correlative approach, which covariates to include.

As there is no single 'best' method or approach for modelling the potential distribution of invasive species (Barry *et al.*, 2015), we asked workshop participants to outline the methods (and associated strengths and weaknesses) they commonly used when estimating the potential distribution of a species.

Which modelling approaches are commonly used in biosecurity?

Participants identified three broad approaches that are (or have been) used by biosecurity agencies to estimate the potential distribution of pests and diseases (Table B.3).

First there were correlative methods which included regression approaches (e.g. GLMs, GAMs), machine learning algorithms (e.g. Maxent, Boosted Regression Trees (BRTs)), climatic envelopes (e.g. BIOCLIM) and climate matching algorithms (e.g. CLIMEX and CLIMATCH). Broadly, these methods work by relating occurrence records to spatially explicit covariates (commonly climatic variables). Both climate matching and

⁵The trait-based model was raised by some technical experts and but was rarely, if ever, used to inform biosecurity policy and management policy, presumably due to the additional expertise required to parametrise such a model

⁶And consequently, which assumptions they were willing to make



Maxent approaches were thought to be some of the most commonly used methods in Australian biosecurity because these methods are fast and easy to implement and because they can be parametrised using data readily available in online databases such as GBIF. However, while these models contain many strengths, they also contained several weaknesses including: issues with sampling biases, difficulties in selecting appropriate background points and covariates, and extrapolation issues when predicting into novel environments.

Participants also highlighted the increasingly use of the commercial software CLIMEX. Unlike most correlative methods, CLIMEX was specifically developed for modelling the potential distributions of invasive species (Kriticos et al., 2015). CLIMEX contains a sub-routine that conducts a climate matching analysis based on locations of persistent populations. However, it also contains a sophisticated sub-routine that allows one to specify physiological thresholds associated with temperature, moisture, radiation, substrate, light and diapause. Most participants generally considered CLIMEX's physiological threshold model to be an example of a process-based model. As such, we've included it as an example of process-based models. However, in practice CLIMEX models are not entirely process-based because they commonly require subjective tweaking of experimentally-derived stress parameters so that they are relevant to the long-term averaged climate data used in CLIMEX or parameters are inferred based on the known distribution of the pest of interest (Kriticos et al., 2015). Commonly, this is done by iteratively running the model and comparing outputs to occurrences of known persistent populations. The increasing use of CLIMEX in biosecurity is thought to be due its ease of use coupled with the perception that models with more process-based components will result in more biologically plausible models of a pest's fundamental niche. In fact, general consensus among participants was that process-based models such as CLIMEX should be preferentially selected ahead of other forms if there was sufficient data to estimate the relevant physiological parameters. However, some participants highlighted that these models also suffer from a variety of weaknesses such as the need for subjective tweaking of parameters, whether experimental-derived parameters are representative of a species response in the natural environment and that commonly there is insufficient physiological data to parametrise such models for most pests of concern.

Lastly, participants raised the utility of expert-based models. General perceptions were that these types of models were infrequently used to predict the potential distribution of pests. Mostly, this was because these models were considered to be a last resort, and only implemented when there was insufficient data to fit correlative or process-based models or an adequate congener species did not exist. However, some participants highlighted the utility of this approach in increasing stakeholder engagement in the modelling process and subsequently increasing stakeholder confidence/belief in the model outputs.

Method class	Correlative	Process-based	Experts-based
Commonly used	Maxent, GAMS, BIOCLIM, BRTs, CLIMATCH, CLIMEX	CLIMEX	Custom-built models
Strengths	 Easy & fast to implement Data often readily available from large online databases (e.g. GBIF) Wide variety of methods available 	 Models processes governing species distributions Theoretically more likely to approximate fundamental niche More likely to predict reliably in novel environmental conditions Perceived as more accurate & reliable relative to other methods Can easily be validated using existing occurrence records 	 Useful when data is unavailable to parametrise other methods Can readily incorporate unpublished information/knowledge Useful for approximating rare events that are otherwise difficult to estimate using empirical data Allows for a diversity of expert opinions to be accounted for Rigorous methods exist (e.g. IDEA protocol)
Weaknesses	 Predictions into novel environmental unreliable Proximal variables typically unavailable at relevant spatial scale or extent Susceptible to sampling biases Model complexity limited by amount of data Relies on unrealistic assumption of equilibrium Model outputs affected by choice of background points (two-class models) Some methods do not propagate uncertainty Typically do not account for adaptation 	 Difficult & slow to parametrise Requires high level of biological & technical knowledge Data only available for a limited number of species Parameter estimates often derived from experiments which may not be transferable to natural environment Parameter estimates may not be representative of species Often involves parameter tweaking Often do not contain parameter uncertainty Typically does not account for adaptation 	 Difficulties in defining an expert General distrust in expert-derived data and models Often categorical Typically does not account for adaptation

Table B.3.: Commonly used modelling methods for estimating the distribution of pest climatic suitability identified by workshop participants



How are covariates selected (correlative models)

Participants highlighted that the determination of which covariates to include in correlative models was a major source of model uncertainty. A primary reason for this is that in most cases the proximal variables governing a species distribution are not known, and when they are, they are often not available at the spatial resolution or extent required for estimating pest potential distributions (Barry *et al.*, 2015).

We therefore asked participants to outline how they commonly selected model covariates when fitting correlative species distribution models. Below is a brief summary of the standard approaches used by participants.

Participants all agreed that the first step would be to conduct a literature review to determine what factors are known (or considered) to govern species distributions or individual vital rates. In particular, CABI data sheets, published species distribution models and published experiments were considered a highly useful for informing the types of variables to include in a model.

In situations where there was either insufficient information, or there was no consensus on which factors govern a species distribution (e.g. multiple species distribution models using different covariates), participants used a number of approaches for identifying which parameters and interactions to include in a model. These methods ranged from directly seeking expert advice to conducting various statistical analyses. The primary methods used are summaries below:

- Selecting covariates based on expert biological advice;
- Using ordination techniques for selecting covariates and reducing dimensionality;
- Examining correlation between covariates and in situations where there are high correlations, select those that are considered to be more proximal based on either literature or expert advice;
- Fit combinations of simple (2 or 3 covariate models)⁷, then remove models that fail to predict the fitted or test data above some performance threshold and either ensemble the remaining models or use identified covariates in a range bagging (Drake, 2015) or convex hull approach (Blonder *et al.*, 2017);
- Use all available covariates and then conduct a backwards selection process;
- Fit all possible combinations in a cross-validated framework and then select the model(s) that best predict the test (i.e. heldout) data.

Model validation

Another area of uncertainty raised by participants was how to validate model outputs. This was of a particular concern among many biosecurity practitioners because they wished to use these methods to to predict into regions where the pest currently does not occur in order to inform where surveillance should occur for early detection.

Below is a summary of the types of validation approaches either recommended or used by participants for both correlative and process-based models:

Correlative Models:

⁷Simple models are less likely to over fit data and are often more conservative in their predictions



- Fit multiple plausible models and choose the model that best predicts the fitted data.⁸
- Fit competing models and use a cross-validation approach (e.g. k-fold) to compare model performance;
- Provide model outputs and meta data to experts to endorse or identify possible errors or areas of uncertainty;
- Identify areas of novel environment and mask (or identify) these areas on suitability maps.

CLIMEX (physiological model)

- Compare model output to contemporary occurrence data;
- Provide model output & meta data to experts for verification/endorsement

Model utility

Workshop participants also raised several important issues faced by policy staff when using models to inform biosecurity policy and management. Key among these were associated with uncertainties in regards to: 1) how to interpret and use the model; 2) Choosing between competing models; 3) Whether model outputs can be thresholded.

Communication of model to policy staff

A problem raised by some policy participants was that handover between model developers (internal or external) and policy staff was insufficient. Often handover was done as a once-off procedure involving a lengthy technical report. However, policy participants noted that often such models were done as external consultations and required additional work so that they could be fully integrated into departmental procedures and operations.

Some participants recommended that both internal and external based projects should have a "decoder" (i.e. a staff member with both technical and policy experience) involved in quantitative projects to ensure it can be integrated into existing operations and procedures and so that fundamental technical details are appropriately communicated to policy makers. Specifically, on completion of the model, a short summary document is created that outlines 1) how the model was parametrised, 2) what are the key assumptions made; 3) how it can be incorporated into existing departmental systems; and 4) how to appropriately interpret the model output.

How to choose between competing models

Another difficulty raised by participants was associated with the uncertainty in choosing between competing published models. For many high profile plant pests, multiple

⁸Technical experts raised that while performance in the fitted region may be correlated with its predictive capacity in the predicted range, this isn't always the case because models may be over fitted (i.e. they may be modelling noise in the data, rather then species responses);



peer-reviewed distribution models may exist. These models are often parametrised using different tools (e.g. Maxent, CLIMEX, GAM, GARP), parameters/covariates, data sources and data cleaning methods.

Workshop participants suggested that this could be resolved by carefully summarising the methods associated with each published model. Specifically, a summary document that emphasises 1) where data was sourced and how it was cleaned; 2) whether the occurrence records and model predictions were representative of currently known distribution; 3) What tool was used to fit the model; 4) the covariates/parameters used; and 5) the key assumptions the model makes. This summary document could then be used by technical and biological experts in determining which model(s) are most appropriate for the context in which they are used⁹. In the cases, where experts believe multiple models are equally probable, models could be ensembles (i.e. model averaged) or used to propagate uncertainty in the potential distribution of the pest or disease of interest.

Should a model output be thresholded?

Multiple discussions were had by participants about the utility of thresholding continuous model outputs into a binary or categorical score system (e.g. Suitable/Not suitable or Low/Medium/High).

On one side of the argument, some participants indicated that categorised risk was more easily handled in various economic analyses and also simplified the decisionmaking process (e.g. where to put finite surveillance resources). More importantly, however, categorised risk was easier to communicate to the public, other government agencies and to stakeholders.

Technical experts, however, cautioned against the use of risk thresholding for various biosecurity decision-making. Their reasons included:

- Leads to a loss of information which hinders model utility in a variety of other decision-making scenarios (e.g. estimating proof of freedom, spread modelling, post-border prioritization)
- Degrades estimates of potential area of occupancy
- Not valid for informing optimal surveillance programs as it fails to acknowledge uncertainty in pest presence to evaluated trade-offs in expected costs.
- Tolerance to risk will vary over time, and as such the points which are used to define binary or categorical risk will likely change.

⁹In essence the same material required to appropriately communicate model to policy makers

C. Dealing with uncertainty in predictor selection



Dealing with uncertainty in predictor selection

Developing scientifically robust risk maps for priority plant pests

Matt Hill, Peter Caley, James Camac, Jane Elith, Simon Barry Matt.Hill@csiro.au November 2020



Summary

Predicting the novel ranges of non-native species is critical to understanding the biosecurity risk posed by pests and diseases. Species distribution models (SDMs) are typically employed to predict the potential ranges of non-native species in novel environments and geographic space. Researchers have focussed on issues around model complexity and specifics relevant to the data typically available for modelling, including the size of the geographic area to be considered and how the choice of model impacts results. These investigations are coupled with considerable examination of how model evaluation methods and test scores are influenced by these choices. An area that remains challenging and under-discussed is methods to choose predictive variables that will result in models that transfer between regions. This is challenging as non-causal variables may still show significant associations with observed species distributions.

Here we propose methods to finesse this problem by using multiple simple models to search for the most appropriate (transferable) variables from a given set, and then apply them as two variable envelopes in an ensemble approach. We apply these methods to both virtual species and real species data, and find that our method generally performs well against conventional approaches for statistically fitting numerous variables in a single model. While our methods only consider simple ecological relationships of species to environmental predictors, they allow for increased model transferability (ability of model to be extrapolated to a new region) as they largely avoid over-fitting, background selection and collinearity issues. Simple models are also likely to be more conservative (over-predict potential distributions) relative to complex models containing many covariates. The approach we have explored transforms a model selection problem, for which there is no true correct answer amongst the typically distal covariates on offer, to one of model uncertainty. We argue that increased model transferability at the expense of model interpretability is perhaps more important for effective rapid predictions and management of non-native species and biological invasions.

Contents

1	Intro	oduction	5
2	Met	hods	7
	2.1	Environmental predictors	7
	2.2	Model construction	7
		2.2.1 CHE: convex hull ensembles	7
		2.2.2 BBE: bounding box ensembles	9
		2.2.3 Maxent	9
		2.2.4 Range bagging	10
	2.3	Virtual species data	10
	2.4	Real species data	11
	2.5	Assessing model transferability	12
3	Resi	ults	13
	3.1	Virtual species	13
	3.2	Real species	14
		3.2.1 Case study 1: Aedes albopictus, Asian tiger mosquito	17
		3.2.2 Case study 2: Solenopsis invicta, red imported fire ant	19
		3.2.3 Case study 3: Bactrocera dorsalis, Oriental fruit fly	21
4	Disc	cussion	23
	4.1	Recommendations	25
5	Ackı	nowledgments	26

1 Introduction

Evaluating the risk of establishment and potential impact of exotic pest and pathogen species requires an accurate prediction of their potential distribution. A typical avenue for predicting the novel ranges of non-native or range-expanding species is using species distribution models (SDMs). Making inference on the distribution and abundance of a potentially invasive species in a new environment is a fundamental question in invasion ecology, as it is with ecology more generally (Krebs, 2009). Species distribution models aim to characterise the limits and response of species to a range of environmental covariates (or predictors), in order to predict whether a species can persist in a given environmental range. The diversity of SDM techniques available, and range of opinions on how to apply them, makes defining best practice across the variety of model applications challenging (Barry et al., 2015).

While SDM methods are particularly useful for many applications, model transferability is an area where they continue to perform unreliably. Transferability describes the ability of a model built for one situation (a given species, in a given area, for a given time span) to predict to another (e.g. under a climate change scenario, or in a new continent)(Sequeira et al., 2018)). For pest and pathogen species this may be thought of as extrapolating a model built on the native (or another established invasive range) to predict a novel area that may be at risk of that species establishing. Inherently, transferability is highly important to biosecurity, as the ability of models to predict under new scenarios forms the baseline of many risk assessments and for setting management priorities (e.g. surveillance allocation). There are many components that affect the ability of models to transfer well (as discussed in detail in Sequeira et al. (2018); Table 1), but one that is often not dealt with is choice of predictor variables.

For predicting the distribution of species to novel environmental space, such as for non-native species or under climate change scenarios, causal variables are ideally required (Varela et al., 2011; Austin and Van Niel, 2011; Fourcade et al., 2017). If the predictor variables chosen are only indirectly associated with defining the species' niche, then SDMs will be only be able to predict the populations they were characterised on (Austin, 2002), leading to poor transferability (Randin et al., 2006). Further, correlation structures between variables can change between geographic locations (Dormann et al., 2013), and therefore challenge model transferability. However, as we can never truly 'know' all the variables that define species distributions, determining which predictors should be used and how best to justify them can be a major challenge. Limited knowledge of the ecology and physiology of the species make its harder to identify which variables may be more proximal (how directly the organism responds to a given variable), and even then such variables may be unavailable for SDM purposes. A range of studies have attempted to determine which predictor variables are best to include, and in lieu of truly causal predictors, how proximal these variables may be (for a review, see Barry et al., 2015). Despite this, there has been little attempt to review or develop methods to simply identify "good predictive variables" from those typically available to modellers and how to best use them to reduce errors associated with model transferability.

Typically, species occurrence data used in SDMs are ad hoc observations of species presence,

with sampling biases due to a number of reasons that result in poor or patchy species records. For the typical amounts of species data available and for the typical modelling challenge of fitting a model to only part of a species' niche (that part observed in its native range), this is a strong argument for focussing on simple models. Complex models fitted to such data may be prone to fitting noise and describe erroneous patterns in data that may be due to limited or missing important relationships (often the case for species of biosecurity concern) (Merow et al., 2014). Envelope based methods provide the simplest forms of SDMs, with simple functions (e.g. bounds and step features) defining their response variables (Elith et al., 2005; Merow et al., 2014), and while they may not be as ecologically interpretable as fitted response curves in regression methods, they should not be as prone to overfitting.

To address these issues of selecting proximal variables and increasing transferability, it may be practical to select appropriate predictors based on how useful the variables are at defining the species distribution based on known presences, and then represent uncertainty across these putative proximal variable sets. In a recent paper, Breiner et al. (2015) used "ensembles of small models" to predict the distributions of rare species. Breiner et al. (2015) found that using an array of small models, each with only two variables, outperformed standard SDM methods. While these models were not tested for transferability into novel environments, it demonstrated that the use of many small models was able to outperform fitting a single, more complex, model. It follows then, that by "mining" for the most useful variables using simple models it should be possible to identify those predictors that perform better and select a subset of simple models to combine into an ensemble. This approach may avoid some of the issues associated with transferability, multi-dimensionality and correlation issues that "complex" SDMs often exhibit.

Here we examine the ability of an approach that incorporates an ensemble of simple models and envelopes to account for predictor uncertainty. Our approach ultimately recasts the process of constructing a useful SDM as a predictor selection uncertainty problem, as opposed to one of model selection (with the implicit paradigm that there exists a "best" model to be found). We first make use of virtual species so that we can define environmental limits and therefore assess model transferability against truth. We then extend this to look at data for a number of insect species (including plant pests) that have successfully invaded new geographic regions. These data allow us to examine transferability and quantify uncertainty, and to formulate recommendations for SDM methods when extrapolation is required.

2 Methods

2.1 Environmental predictors

There are a number of different environmental predictors convenient for use in SDMs, and many of these are based around climate. While different landscape and types of anthropogenic variables are often used too, climate serves as a particularly good indicator of species distributions at a broad geographic scale when little is known about the species biology and ecology. Employing climate in this way assumes that some of the climate variables are proximal to processes defining the species distribution, without knowing which variables, and how they affect those processes. Several databases provide long-term climatic records as gridded datasets of maximum and minimum temperatures, as well as precipitation, appropriate for species distribution modelling at large geographic scales. For this study, the environmental data obtained was the 19 bioclimatic variables from WorldClim 2.0 (Fick and Hijmans, 2017) at a 10' resolution. From here on we call these data the "bioclim" set, as they are derived as part of the BIOCLIM (Nix, 1986) package. BIOCLIM (from here on "BIOCLIM" refers to the modelling method) was instrumental in not only defining the role of SDMs, but also in the description of these variables that are used widely today (Booth et al., 2014), although now more typically outside of the BIOCLIM environment itself. These data provide an easily accessible, and therefore commonly employed modelling resource. The scale here is also relevant to the question of transferability between large, distinct, geographical regions.

2.2 Model construction

Here our goal was to build models to predict an invasive (or non-native) range from the native range data alone. While there are good arguments for using both native and invasive range data together to predict potential distributions, here we are concerned about the ability of different modelling approaches to predict distributions from native range data alone. To investigate transferability across a range of approaches, we constructed SDMs using a number of different methods, which fell into two broad categories: ensembles of simple models, and single multivariate models. We applied all methods to two sets of data, first virtual species data (simulation) and second, actual species data that has been used previously in the investigation of SDM performance for invasive species.

2.2.1 CHE: convex hull ensembles

To construct our ensembles, we chose simple "envelope" methods that classify environments (predictor variable ranges) as either suitable (inside the envelope) or unsuitable (outside the envelope). Generally, we constructed many two variable envelopes of species presence within their native range using pairwise combinations from the bioclim variable set. Instead of using all combinations of variables, we restricted the combinations to one temperature (bioclim 1-11) and one precipitation (bioclim 12-19) variable. The rationale here being that at least one aspect of temperature and one aspect of precipitation is likely to be limiting (Figure 2.1).

In order to find which variables appear most closely related to the distribution of the species in its native range, we fitted a regression model. More specifically, for the native range data we fitted many two-variable models, using a Generalized Additive Model (GAM) implementation of a Point Process Model (PPM) for presence-only data (Renner and Warton, 2013; Renner et al., 2015). GAMs were constructed as a down-weighted Poisson regression (DWPR) and smooth terms were restricted to the 2nd order polynomial at the most, to facilitate simple curves for fitted responses. Terms were either fitted additively or jointly. Joint fits allow interactions between predictors, and can be achieved in the mgcv package in R by fitting a smooth surface in two dimensions over the chosen variables rather than two one-dimensional fits, one for each variable. We built models using presence-background data, which are typically the most readily available data for non-native species. Background points provide a sample of the environment in the region from which presence points are available, here they were all the cells across the geographic extent of the calibration (or native) range.



Figure 2.1: Schematic of new ensemble modelling technique. A. Variable pairs of temperature and precipitation using the bioclim variable are used in simple GAMs and then assessed for relative model fit using AUC (and other metrics, see text). Green squares are the pairs kept, red squares with crosses are omitted. B. The variable pairs selected are used to create two-dimensional environmental envelopes capturing all the presence points (green points) inside all available environments (green + orange points), using a minimum convex polygon (convex hulls (CHE)), or by use of bounding boxes defined through the BIOCLIM algorithm (BBE). C. Each of the two-variable envelopes from B are projected to a new geographical surface (as a gridded raster) and then stacked (summed. The results are then averaged, and the resulting surface is a continuous scale from 0-1 reflecting the level of certainty (closer to 1) that a given raster cell falls in the environmental limits across the given predictor variables.

We evaluated the relative performance of GAMs constructed using the different variable pairs across all possible predictor combinations. That is to say, we are comparing the performance of variable pairs only against one another, and not on the absolute predictive performance of the model (which is akin to model selection). To determine performance of variable pairs, we calculated AUC for each variable pair, and then discounted the "worst" performing models by discarding the bottom quantile (25%) of variable pairs.

Using the retained variable pairs, we then used two different envelope methods to classify the environmental space within each predictor pair. The first of these approaches was to use convex hulls (CHE: convex hull ensembles). We created convex hulls around the presence points (of the training, native range data) in environmental space, and then projected these envelopes onto the geographic space of the native range, and the invasive (non-native) range. Each envelope gives binary values of 1 for inside the hull, and 0 for outside of the hull. The resulting geographic layers

for each model pair were combined together and then averaged to produce a single, continuous, prediction between 0 and 1 for the CHE. Here 0 means that a grid cell is outside all of the available pair-based envelopes (not suitable), and 1 means that the grid cell has environmental space inside all of the pair-based envelopes (highly suitable).

2.2.2 BBE: bounding box ensembles

To further examine the ensemble approach we used bounding boxes (BBE: bounding box ensembles), drawing from one of the early SDM methods, BIOCLIM (Nix, 1986). BIOCLIM works by creating a bounding box in multidimensional environmental space, and defines core and marginal environments for the species in relation to the predictors used by examining percentile distributions. Core environments are defined as the environments in which all points fall inside the 5 and 95 percentile range for all the predictors used (from an available set of 12 in the case of Nix (1986), later extended to 19 and then 35 (Booth et al., 2014)). Marginal environments are defined as where the values fall outside the 5 and 95 percentile range, but not the limits, for one or more of the predictors. In this way, the distribution is driven by the most limiting variable. BIOCLIM is typically used with several predictor variables, and here we started again with the 19 available from WorldClim (Fick and Hijmans, 2017). We used the implementation of BIOCLIM in the dismo package in R (Hijmans et al., 2015), to create multiple, two-variable models, using a pairwise approach across the bioclim variables available (one from 1-11, one from 12-19) (Figure 2.1). We used the AUC scores from the GAMs again to omit the bottom 25% of performing variable pairs.

To compare how the BBE approach may perform in regards to a 'default' BIOCLIM model, we also ran a BIOCLIM model using all 19 available predictors and projected this to the non-native range.

2.2.3 Maxent

To compare how our approaches stack up against a popular SDM approach, we ran three types of Maxent models Phillips et al. (2006) using version 3.4.0. Firstly, we ran a model with all 19 bioclim variables (for the virtual species, it was 17 as we left out the two causal variables). Maxent has internal predictor selection mechanisms, and although we expect a complex model like this to be overfitted, it should provide a good comparison based on making no choice in predictor inclusion. Secondly, we ran Maxent models using eight predictor variables that have been used elsewhere for invertebrates (see Hill et al. (2017)), sometimes referred to as "state of the art" (Petitpierre et al., 2017). While the fit and performance of Maxent models can be improved through exploration of different features and parameters, we chose to leave all at default for this exercise for two reasons: to provide a baseline, and because default settings are used throughout the literature. Thirdly, we used Maxent to build simple two-variable models in the same framework as our CHE and BBE approaches. In an attempt to avoid overfitting for these models, we did not allow hinge, product or threshold features (models could only use linear or quadratic terms). The same variable restrictions (one temperature, one precipitation) and performance assessment using AUC scores on the GAMs also applied, and the resulting single layer was an average across all retained Maxent models.

2.2.4 Range bagging

Range bagging was recently proposed by Drake (2015). This algorithm uses presence-only data to estimate the environmental limits of species habitat by subsetting the multidimensional environments (to user-defined levels of dimensionality), and then using convex hulls to estimate boundaries in each subset of environmental dimensions. Range bagging fits models to the individual samples and averages the outcome by using votes (how often a given environment occurs inside niche boundaries) on the ranges of convex hulls obtained from bootstrap samples across all the environmental dimensions. In this way, it is an approach that shares some similarities with our CHE and BBE methods, however an important difference here is that the range bagging method does not use the GAMs to rank variables.

The approach has seen recent applications to invasion biology, and appears promising for biosecurity applications. Part of the appeal for this approach is that only presence data are needed, and no absences or background data are required, and thus, reduces the number of subjective decisions required in the modelling process. We fit range bagging models using supplied code Drake (2015), using for the presence-only data of the native range, and set range bagging to subset using only two dimensions to be consistent with our other approaches. In addition to running range bagging at default settings (all variables pairwise, and in two dimensions) using code as supplied by Drake (2015), we made one modification to the range bagging method, and required it to use variable pairs in the same way as our convex hull and BIOCLIM ensemble methods (1 from the bioclim temperature variables 1-11, and 1 from the bioclim precipitation variables 12-19). We then projected our range bagging model to the non-native region for each species

2.3 Virtual species data

To test how different modelling approaches may perform in regards to predictor selection, we generated virtual species. The goal here was not to create 'realistic' species distributions, but rather geographical expressions of limits set within environmental ranges, to investigate how different modelling approaches perform for a predictor selection task. Defining virtual species allows us to know the "truth" behind a distribution in both "native" and "non-native" (i.e. invaded) ranges, and use SDM methods to model that niche and assess the transferability of the modelling methods. There are a number of ways for generating virtual species distribution across the landscape (e.g. Leroy et al. (2016), but as we were interested in creating bounding boxes in environmental space between pairs of variables, we simulated species occurrences by setting upper and lower limits for each variable. South America provides a range of climates that are both analogous and non-analogous to Australia. This allows us to test transferability across a range of environmental conditions and ensured that variables were not set outside of the range present across South America. We chose a temperature variable from the bioclim set (1-11) and a moisture variable (12-19). Virtual species distributions were created by taking the upper and lower limits of the entire surface area, and selecting random uniform distributions inside these to give upper and lower ranges for each variable. We enforced that the lower and upper bounds of the environmental range had to be taken from the respective sides of the mean, and at least one standard deviation existed between upper and lower bounds, to ensure there was at least some

expression of that niche in geographic space. Within these limits (a box in the environmental space of these two variables) the species is always capable of being present. By contrast, outside of the box the species is always absent. Thus these virtual species exhibit pure niches, with no biotic interactions, and they have no stochastic element, i.e. the probability of presence is zero or one. The limits were set based on a random selection across the global range of environments. These "niches" were then realised in South America and Australia, which makes it possible to examine geographic ranges of environmental similarity between ranges.

For assessing how well the ensemble approach could predict the distribution (or geographical representation of the true niche) in the non-native range (Australia), we followed the general framework but importantly, left out the two variables that were used to create the true niche. This means that we are acknowledging that we will never know the "true niche" but can only make approximations of it, based on variables that range from proximal to distal. As they are all different patterns they are likely to be highly correlated in some form.

2.4 Real species data

We also examined the performance of different modelling approaches using occurrence records for the 22 invasive insect species dataset from Hill et al. (2017) (see Table 3.1). Many of these species are priority plant pests, disease vectors or nuisance pests. These species were chosen as they have all successfully established in new geographic locations, and provide data for both "native" and "non-native" regions to train and test models with. While these species were chosen as they have occurrences for both native and non-native regions, the underlying data are inherently biased in the way they have been sampled across different regions. These biases include under-sampling across the total range of environments, and increased sampling in areas where the species are noted pests. The result is patchy data that poses challenges to fitting SDMs on a single range, but this is typical of many priority plant pests. We applied the same methodology for each of the species in order to see how different methods performed across a group of species of varying range size and data availability.

We defined backgrounds by selecting all unique biomes that occurence points intersected with (see Olson et al., 2001), as this has proved to be a useful extent for modelling these species (Hill et al., 2017). To help limit sampling bias, the presence data were rescaled to the level of the grid cell data (10'), so that there was only one occurrence per grid cell. The background was selected as all available cells in the model training region, or 50,000 random cells across this area, whichever was fewer.

To demonstrate how the different methods perform in practice, we selected three case studies from this dataset of species that have had numerous SDMs developed, and are considered highly invasive; the red imported fire ant, Solenpsis invicta; the Oriental fruit fly, Bactrocera dorsalis; and the Asian tiger mosquito, Aedes albopictus. These three species were chosen as they are widely distributed globally, and have all successfully arrived in Australia (and established in the case of Solenpsis invicta).

2.5 Assessing model transferability

To examine how models performed in predicting the non-native ranges, we employed three different measures. As we had independent datasets (non-native range) for each species to evaluate the models, we first employed AUC on this test dataset to assess performance. The AUC score was calculated using the non-native range distribution and the associated background (i.e. the geographical boundaries defined by the definitions of either continents, or biomes).

As different evaluation techniques may yield different results, we further looked at the Boyce index (Boyce et al., 2002), and sensitivity (see Petitpierre et al., 2017), as measures for the final surfaces projected into the novel ranges. The Boyce index is a presence-only evaluation technique and measures how much model predictions differ from a random distribution of observed presences across a prediction gradient (Hirzel et al., 2006). The Boyce index partitions the total habitat suitability range into a set of bins, and for each of these calculates the predicted to expected ratio, using the Spearman rank correlation coefficient (Hirzel et al., 2006). The Boyce index names near 0 indicating models no better than random and negative values indicating prediction of poor suitability where there are presences (Di Cola et al., 2017). Sensitivity is the percentage of true presences correctly predicted by the model.

3 Results

3.1 Virtual species

There was some variation in the ability of different modelling approaches trained in South America to transfer across to Australia. The different applications of Maxent (using all variables, using a pairwise approach) had high AUC and Boyce index scores, whereas there was more variation in these scores for the other approaches (Fig. 3.1). Sensitivity was highly variable for all of the approaches, although the two Maxent approaches gave the highest sensitivity values, indicating that they were correctly predicting true presences (sites inside the niche in Australia).



Figure 3.1: Evaluation scores for a comparison of the different SDM methods for virtual species: CHE = convex hulls ensembles; BBE = bounding box ensembles; BIOCLIM = BIOCLIM algorithm using all 19 bioclim variables (withtwo causal variables omitted); Range Bagging <math>[PW] = default Range Bagging algorithm with dimensions set at two and selection of only one temperature and one precipitation variable at a time, leaving out the two causal variables; Maxent [PW] = ensemble of simple Maxent models and selection of only one temperature and one precipitation variable at a time, leaving out the two casual variables; Maxent [All] = Maxent using default settings and all 19 bioclim variables

In addition to the two Maxent approaches, the BBE approach also worked well for these virtual species, with high Boyce and good AUC (Fig. 3.1). Similarly, the CHE approach also worked

reasonably well for the evaluation scores, but there was more variation for the Boyce scores than for BBE. The two range bagging approaches had the largest amounts of variation across the three evaluation techniques, and consequently the lowest mean scores (Fig. 3.1). BIOCLIM had good Boyce scores, however the AUC and Sensitivity were variable and lower than the best performing methods.

3.2 Real species

Tabl	e 3.1:	List	of species	used in	this s	study.	'Aus	native'	are	species	which	are	native	to	Australia.	'Aus	present'
and	'Aus a	absent	refer to	reported	l occu	rrence	s of t	he spec	ies i	n Austra	alia						

Common Name	Species	Aus native	Aus present	Aus absent
Asian tiger mosquito	Aedes albopictus			*
Asian longhorned beetle	Anoplophora glabripennis			*
Weevil species	Atrichonotus taeniatulus		*	
Weevil species	Aramigus tessellatus			*
Oriental fruit fly	Bactrocera dorsalis			*
Mediterranean fruit fly	Ceratitis capitata		*	
Light brown apple moth	Epiphyas postvittana	*		
Harlequin ladybeetle	Harmonia axyridis			*
Redlegged earth mite	Halotydeus destructor		*	
Colorado potato beetle	Leptinotarsa decemlineata			*
Argentine ant	Linepithema humile		*	
Asian needle ant	Pachycondyla chinensis			*
Cassava mealybug	Phenacoccus manihoti			*
African big-headed ant	Pheidole megacephala		*	
Red imported fire ant	Solenopsis invicta		*	
Marsh crane fly	Tipula oleracea			*
European crane fly	Tipula paludosa			*
Bronze bug	Thaumastocoris peregrinus	*		
Pavement ant	Tetramorium sp. E			*
Japanese pavement ant	Tetramorium tsushimae			*
Electric ant	Wasmannia auropunctata		*	
African fig fly	Zaprionus indianus			*

The real species data provides a test of projecting an incomplete understanding of the niche and the associated distributions to a novel geographic range. Seven of the 22 species have already established in Australia (Atrichonotus taeniatulus, Ceratitis capitata, Halotydeus destructor, Linepithema humile, Pheidole megacephala, Solenopsis invicta, Wasmannia auropunctata), and two (Epiphyas postvittana and Thaumastocoris peregrinus) originate from Australia (Table 3.1). The remaining 13 species have varying degrees of risk for establishment in Australia. Here the models were calibrated on only the native range and then projected to the invasive range. Normally it would be better practise to include the invasive range in the model training (Broennimann and Guisan, 2008) — however here we are providing a test of the ability of models to project from limited data rather than a description of the entire niche. Another important caveat is that Hill et al. (2017) found evidence of 'niche shift' (adapted or expanded to novel environments), in a number of these species, suggesting that the distributions in the novel ranges may extend outside of what is predictable from training on the native range alone. Bearing that in mind, the trends presented do encapsulate what may be expected from a dataset of actual species data, complete with inherent bias in sampling effort and geographical extent.



Figure 3.2: Evaluation scores for a comparison of the different SDM methods for the 22 invasive insects species data: CHE = convex hulls ensembles; BBE = bounding box ensembles; BIOCLIM = BIOCLIM algorithm using all 19bioclim variables; Range Bagging = default Range Bagging algorithm with dimensions set at two; Range Bagging [PW] = default Range Bagging algorithm with dimensions set at two and selection of only one temperature and one precipitation variable at a time; Maxent [PW] = ensemble of simple Maxent models and selection of only onetemperature and one precipitation variable at a time; Maxent <math>[Expert] = Maxent using default settings and eightpredictor variables described as useful for predicting insect distributions; Maxent <math>[All] = Maxent using defaultsettings and all 19 bioclim variables

Figure 3.2 shows the evaluation scores for the different SDM techniques we employed for these real species data. While there is some variation, across the methods the distribution of AUC is fairly similar (assessed on the non-native distribution and range). Likewise, Sensitivity is above 0.8 for most species and methods (except BIOCLIM) which may be considered as "being transferable" (Petitpierre et al., 2017)). The greatest variation is in the Boyce score, which places the CHE and Maxent[PW] methods as better performing than the other approaches Fig. 3.2). On balance across the evaluation techniques, The Maxent[PW], CHE and BBE approaches appear to work well according to all the evaluation methods employed here (Fig. 3.2), although most of the approaches here could are performing except for BIOCLIM and perhaps Maxent [All] which also has low Boyce scores. The range bagging approaches were similar, with the default
implementation performing slightly better than the pairwise approach for AUC and sensitivity (Fig. 3.2).

Next, we look in detail at three case studies; A. albopictus, Solenopsis invicta, Bactrocera dorsalis. These represent three 'globally' invasive species, with S. invicta and A. albopictus both demonstrating niche shift in previous studies (see Hill et al. (2017)), whereas B. dorsalis has demonstrated broad niche conservatism. The species also differ in their results, highlighting the need for further scrutiny to be placed on model selection for these species.

3.2.1 Case study 1: Aedes albopictus, Asian tiger mosquito

Aedes albopictus is a globally invasive mosquito, and an important vector of human diseases, including dengue and chikungunya. Aedes albopictus has been able to establish in many regions of the world, by being transported as both eggs and larvae aboard ships (Tatem et al., 2006), and poses a current risk for Australia. The species has had numerous modelling attempts, and which amongst other things, has demonstrated both an expansion and a shift in the occupied environments

When assessed across the non-native range for A. albopictus, AUC and Sensitivity were fairly consistent across approaches, however there was substantial variation for the Boyce scores (Table 3.2). The CHE approach appears to have worked well for this species, with high AUC, Boyce and Sensitivity scores. Range bagging (PW) also performed reasonably well with high AUC and Sensitivity scores, but only moderate Boyce scores. BIOCLIM has the lowest performance scores for this species. Four of the different approaches (BBE, BIOCLIM, Maxent[Expert] and Maxent [All] had negative Boyce scores, indicating these models were not always capturing presences in the test dataset.

Method	AUC	Boyce	Sensitivity
CHE	0.82	0.87	0.82
BBE	0.78	-0.32	0.78
BIOCLIM	0.6	-0.14	0.6
Range Bagging	0.8	0.51	0.8
Range Bagging [PW]	0.81	0.62	0.81
Maxent [PW]	0.74	0.01	0.74
Maxent [Expert]	0.79	-0.67	0.79
Maxent [All]	0.82	-0.42	0.83

Table 3.2: Evaluation scores of the different modelling methods for Aedes albopictus developed in native ranges and projected to invaded range(s).

In terms of spatial predictions, the CHE and Range Bagging [PW] approaches gave similar patterns of prediction, and the BBE approach gave the narrowest prediction in terms of area (Figure 3.3). For the test data, particularly in North America, the CHE and range bagging [PW] gave broader predictions than the other approaches.





Figure 3.3: A. Distribution data for Aedes albopictus in its native (blue dots) and introduced (red dots) ranges. B. Model projections for different methods: CHE = convex hulls ensembles; BBE = bounding box ensembles; BIOCLIM = BIOCLIM algorithm using all 19 bioclim variables; Range Bagging [PW] = default Range Bagging algorithm with dimensions set at two and selection of only one temperature and one precipitation variable at a time; Maxent [PW] = ensemble of simple Maxent models and selection of only one temperature and one precipitation variable at a time; Maxent [PW] = Maxent [Expert] = Maxent using default settings and eight predictor variables described as useful for predicting insect distributions; Maxent [All] = Maxent using default settings and all 19 bioclim variables. Note: Suitability refers to slightly different processes as part of the respective modelling approaches - the scores are not directly comparable, but are useful for examining congruence in patterns of prediction.

3.2.2 Case study 2: Solenopsis invicta, red imported fire ant.

Solenopsis invicta is a highly successful invasive ant species that can have great ecological impacts through predation and outcompeting native species, as well as causing nuisance issues due to their bites. Solenopsis invicta is a useful test species due to the quality of data available, and distribution across four continents (North and South America (native range), Asia and Australia) as well as island populations. There have been a number of attempts to model the distribution of this species as it is a high profile invasive species with sometimes large impacts (Fitzpatrick et al., 2007; Broennimann et al., 2007; Hill et al., 2017).

Table 3.3 shows which of the methods performed best for S. invicta. Overall, all methods except BIOCLIM and Maxent [All] achieved a model with good AUC, Boyce and sensitivity scores. On balance between the three evaluation methods, the Maxent [PW] approach worked best for this species, the range bagging approaches also performed well, and the CHE and BBE approaches had good AUC and sensitivity, but lower Boyce scores (Table 3.3).

Table 3.3: Evaluation scores of the different modelling methods for Solenopsis invicta developed in native ranges and projected to invaded range(s).

Method	AUC	Boyce	Sensitivity
CHE	0.86	0.5	0.86
BBE	0.87	0.41	0.87
BIOCLIM	0.54	-0.38	0.55
Range Bagging	0.87	0.67	0.87
Range Bagging [PW]	0.87	0.66	0.87
Maxent [PW]	0.9	0.92	0.9
Maxent [Expert]	0.87	0.66	0.88
Maxent [All]	0.79	0.04	0.79

Spatially, the CHE, Range Bagging and Maxent [PW] approaches again gave broader predictions than the other two methods (Figure 3.4). These methods also captured the invasive distribution in North America better.





Figure 3.4: A. Distribution data for Solenopsis invicta in its native (blue dots) and introduced (red dots) ranges. B. Model projections for different methods: CHE = convex hulls ensembles; BBE = bounding box ensembles; BIOCLIM = BIOCLIM algorithm using all 19 bioclim variables; Range Bagging [PW] = default Range Bagging algorithm with dimensions set at two and selection of only one temperature and one precipitation variable at a time; Maxent [PW] = ensemble of simple Maxent models and selection of only one temperature and one precipitation variable at a time; Maxent [PW] = Maxent [Expert] = Maxent using default settings and eight predictor variables described as useful for predicting insect distributions; Maxent [All] = Maxent using default settings and all 19 bioclim variables. Note: Suitability refers to slightly different processes as part of the respective modelling approaches - the scores are not directly comparable, but are useful for examining congruence in patterns of prediction.

3.2.3 Case study 3: Bactrocera dorsalis, Oriental fruit fly

Bactrocera dorsalis is a highly polyphagous tropical fruit fly that has successfully established across Asia, Pacific islands and into Africa. It has been successfully eradicated following and outbreak in Australia. Bactrocera dorsalis is a good example of a species with a reasonably "stable" distribution, though it is widespread through tropical and subtropical areas. There have also been a number of attempts to model the distribution of this species (four or five to date). Table 3.4 shows that here that the Maxent [All] approach gives the best AUC, sensitivity, and a high Boyce score. In fact, most methods perform well for this species, even the poorest performing (BIOCLIM and BBE) yielded AUC and sensitivity evaluation scores that would be considered "acceptable", although the both had low Boyce scores, particularly BIOCLIM.

Table 3.4: Evaluation scores of the different modelling methods for Bactrocera dorsalis developed in native ranges and projected to invaded range(s).

Method	AUC	Boyce	Sensitivity
CHE	0.87	0.99	0.87
BBE	0.84	0.46	0.84
BIOCLIM	0.77	-0.01	0.77
Range Bagging	0.85	0.98	0.85
Range Bagging [PW]	0.86	0.97	0.86
Maxent [PW]	0.88	0.96	0.88
Maxent [Expert]	0.9	0.89	0.89
Maxent [All]	0.91	0.97	0.91

Overall all methods appear to do a decent job of predicting the non-native distribution of B. dorsalis and also the potential risk for Australia (where the species has been eradicated before) (Figure 3.5). Following on from the other two case studies, the CHE and range bagging approaches give broader predictions than the other two methods, and in this case Maxent[PW] also displays similar patterns.



в



Figure 3.5: A. Distribution data for Bactrocera dorsalis in its native (blue dots) and introduced (red dots) ranges. B. Model projections for different methods: CHE = convex hulls ensembles; BBE = bounding box ensembles; BIOCLIM = BIOCLIM algorithm using all 19 bioclim variables; Range Bagging [PW] = default Range Bagging algorithm with dimensions set at two and selection of only one temperature and one precipitation variable at a time; Maxent [PW] = ensemble of simple Maxent models and selection of only one temperature and one precipitation variable at a time; Maxent [PW] = Maxent [Expert] = Maxent using default settings and eight predictor variables described as useful for predicting insect distributions; Maxent [All] = Maxent using default settings and all 19 bioclim variables. Note: Suitability refers to slightly different processes as part of the respective modelling approaches - the scores are not directly comparable, but are useful for examining congruence in patterns of prediction.

4 Discussion

The discussion around "best practice" for using species distribution models in pest risk mapping is typically focused on model selection, with emphasis often placed on finding the best performing models (by some statistical measure), and using that information as the basis for projection to novel areas and risk assessment (e.g. Fan et al. (2018); Jiménez-Valverde et al. (2011)). By transforming a model selection problem to one of model uncertainty across predictors, we suggest alternative approaches that perform well against established methods, and provide ease of interpretation through an uncertainty framework. Specifically, when the choice of which available predictors to use is unknown or unguided, constructing an ensemble of small (two variable) models allows for model construction to be focussed on selection of "good" performing variables (ones that perform well in both native and non-native ranges), rather than relying on model fit as the indication of the "best" model. Our methods allow for putatively more proximal variables to be identified and to apply them in a simple manner to not only increase transferability (Petitpierre et al., 2017), but provide a more tractable approach to developing risk maps of potential non-native distributions.

Our CHE and the pairwise Maxent approaches were consistently good performers across virtual and real species models. Previously, using a ensemble of two-variable Maxent models has yielded useful results for predicting the distributions of rare species (Breiner et al., 2015), and here we provide evidence that this approach can be extended to a transferable modelling framework for non-native species. The true response to distal variables might also be complex (Merow et al., 2014), using an ensemble of simple models are somewhat able to avoid that, and the strength is model overfitting is avoided (Breiner et al., 2015). We also chose not to weight the different models or predictor pairs, as has been done in elsewhere (Breiner et al., 2015), as our goals were focussed on communicating uncertainty across available predictors in an attempt to identify "good" predictors.

There are some important differences between the CHE and pairwise Maxent approaches that are worth considering when making choices of appropriate modelling frameworks to apply to non-native species. Firstly, Maxent can be thought of as modelling closer to the realised distribution of the species, the CHE closer to the potential distribution (Jiménez-Valverde et al., 2011). This distiction is important when considering how the species-environment relationships are defined given the data available, and what this means for extrapolation. In our CHE approach the relationships are simplified to environmental limits and provide a final spatial prediction as a gridded surface that represents the frequency that a given cell fell within the environmental space as defined by the convex hulls. This is directly related to the uncertainty in predictor variables, with an increase in the number of times a grid cell falls within environmental space corresponding to an increase in certainty in the predictors retained. Alternatively, the pairwise Maxent approach models the species-environment relationships as functions in two-dimensions at a time, and then the averaged probabilities of those relationships used to predict distributions. Maxent fits the relationship to data, there is a chance that when projected to a new set of environments a combination of those variables form the training data does not exist. The CHE thus provides a more conservative approach to predicting distributions, with

simple environmental limits, whereas the pairwise Maxent approach allows for estimation of species' responses to environmental predictors using probabilistic functions which may be closer to the ecological interactions - but more prone to adverse effects of extrapolation if the training data does not allow for the characterisation of that relationship.

The other approach we investigated, range bagging, also performed well for the real species data, despite having highly variable performance for the virtual species. Range bagging shares some similarities to the CHE approach, in that the environmental relationships are estimated through the use of convex hulls across the different combinations of available predictors. This means that the resulting predictions are able to be interpreted in the same manner as for CHE. While CHE and rangebagging may lead to larger predictions (more conservative), this may be more preferable for invasive species (Kramer et al., 2017). Comparing modelling performance between methods is a lot more straightforward than explaining the reasons behind differences (Breiner et al., 2018), but the inconsistent performance for range bagging on the virtual species may be due to the difference in using GAMs (or Maxent) to rank variable pairs in the other approaches, versus the range bagging method voting for variable pairs based on iterations, especially when the causal variable pair is absent. When looking at performance for the virtual species it is important to note that the distributions that were created are not "realistic" species distributions, and instead are focussed on testing the ability to identify proximal variables in absence of the causal predictors for each given virtual species.

One of the biggest challenges of using real species datasets to evaluate modelling methods is "messy" datasets (as a result of sampling bias, under-representation of range, missing data, errors in data). For many non-native species the data available is not sampled with the same intensity and process across native and invasive ranges. Sampling processes under a real world situation may only yield few occurrence points across the actual distribution, with spatial bias in reporting. This translates into methods that may otherwise perform well, being penalised for characterising patterns in the data that are a result of some erroneous, unobservable, or incomplete process. This may help partly explain why there leads to a broad range of model performance, ranging from near perfect prediction, down to no better than random. Another issue that challenges SDM application is how different backgrounds (extent and scale) select different variables to be included in the process. The range-bagging technique avoids this, but future work would be to look at how sensitive the background selection is to changing modelled performance of the GAMs that rank model inclusion that require background environmental information. As we have focussed on non-native plant pests in this report, it is now important to examine more datasets to see if patterns hold across different taxa, such as other animal pests, weeds and perhaps pathogens.

4.1 Recommendations

While we have not performed an exhaustive investigation of all modelling parameters and decisions, our results suggest there are a few approaches that appear to be useful for predicting non-native distributions as part of evaluating risk posed by plant pests (and other species of concern). In particular, by being more "conservative" in the absence of knowledge of the most proximal variables, the ensemble methods of CHE and pairwise Maxent, and the range bagging approach, perform well. This supports that when no standardised set of predictors is required, using iterative approaches to predictor selection can increase predictive performance Petitpierre et al. (2017).

The decision to use CHE or pairwise Maxent may come down to the interpretation of the outputs, as both approaches seem viable options in the absence of complete knowledge of the most proximal variables. One advantage to using range bagging is that is does not require a background to be specified, which may be useful when there is limited or sparsely distribution occurrence data are available. Range bagging is also less computationally expensive than performing multiple GAMs to rank predictor pairs, but given an appropriate background, the convex hulls method outperforms on average. The goals of the risk mapping exercise are likely to define what trade-off of computational time and method exploration are appropriate (Breiner et al., 2018). An important caveat here is that for species where models projected badly, they tended to do so regardless of the modelling approach. In these cases it is hard to justify the use of any correlative or envelope methods to project them.

For all approaches, we also advocate for use of as much distribution data as possible, especially when the species may be present in multiple geographical regions, to define the environmental space more completely (Broennimann and Guisan, 2008; Hill et al., 2017). Of course data should always be examined for environmental outliers that would particularly alter results using convex hulls or other envelope approaches.

As the role of risk maps is in aiding industry preparedness, there is typically a preference of predictive accuracy. Therefore, increased model transferability at the expense of model interpretability is perhaps more important for effective rapid predictions (Merow et al., 2014), especially in the management of non-native species and biological invasions. This is especially important if the costs associated with overprediction (arising from more conservative models) translates to significantly less than the cost of the species invading (Jiménez-Valverde et al., 2011).

5 Acknowledgments

Daniel Heersink wrote earlier versions of some of the code used in this study. Funding was provided by Centre for Excellence in Biosecurity Risk Analysis.

Bibliography

- Austin, M. P. (2002). Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. Ecological modelling, 157(2-3):101–118. 5
- Austin, M. P. and Van Niel, K. P. (2011). Improving species distribution models for climate change studies: variable selection and scale: Species distribution models for climate change studies. Journal of Biogeography, 38(1):1–8. 5
- Barry, S., Elith, J., Heersink, D., Caley, P., Kearney, M., Tenant, P., and Arthur, A. (2015). Final report: CEBRA 1402b Tools and approaches for invasive species distribution modelling for surveillance. Technical report, CSIRO. 5
- Booth, T. H., Nix, H. A., Busby, J. R., and Hutchinson, M. F. (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. 7, 9
- Boyce, M. S., Vernier, P. R., Nielsen, S. E., and Schmiegelow, F. K. A. (2002). Evaluating resource selection functions. Ecological Modelling, 157(2):281–300. 12
- Breiner, F. T., Guisan, A., Bergamini, A., and Nobis, M. P. (2015). Overcoming limitations of modelling rare species by using ensembles of small models. Methods in Ecology and Evolution, 6(10):1210–1218. 6, 23
- Breiner, F. T., Nobis, M. P., Bergamini, A., and Guisan, A. (2018). Optimizing ensembles of small models for predicting the distribution of species with few occurrences. Methods in Ecology and Evolution. 24, 25
- Broennimann, O. and Guisan, A. (2008). Predicting current and future biological invasions: both native and invaded ranges matter. Biology Letters, 4(5):585–589. 14, 25
- Broennimann, O., Treier, U. A., Müller-Schärer, H., Thuiller, W., Peterson, A. T., and Guisan, A. (2007). Evidence of climatic niche shift during biological invasion. Ecology Letters, 10(8):701–709. 19
- Di Cola, V., Broennimann, O., Petitpierre, B., Breiner, F. T., D'Amen, M., Randin, C., Engler, R., Pottier, J., Pio, D., Dubuis, A., Pellissier, L., Mateo, R. G., Hordijk, W., Salamin, N., and Guisan, A. (2017). ecospat: an R package to support spatial analyses and modeling of species niches and distributions. Ecography, 40(6):774–787. 12
- Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J. R. G., Gruber, B., Lafourcade, B., Leitão, P. J., Münkemüller, T., McClean, C., Osborne, P. E., Reineking, B., Schröder, B., Skidmore, A. K., Zurell, D., and Lautenbach, S. (2013).
 Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. Ecography, 36(1):27–46. 5
- 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. 10
- Elith, J., Ferrier, S., Huettmann, F., and Leathwick, J. (2005). The evaluation strip: A new and robust method for plotting predicted responses from species distribution models. Ecological Modelling, 186(3):280–289. 6
- Fan, J. Y., Zhao, N. X., Li, M., Gao, W. F., Wang, M. L., and Zhu, G. P. (2018). What are the

best predictors for invasive potential of weeds? Transferability evaluations of model predictions based on diverse environmental data sets for Flaveria bidentis. Weed Research, 58(2):141–149. 23

- Fick, S. E. and Hijmans, R. J. (2017). WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas: NEW CLIMATE SURFACES FOR GLOBAL LAND AREAS. International Journal of Climatology. 7, 9
- Fitzpatrick, M. C., Weltzin, J. F., Sanders, N. J., and Dunn, R. R. (2007). The biogeography of prediction error: why does the introduced range of the fire ant over-predict its native range? Global Ecology and Biogeography, 16(1):24–33. 19
- Fourcade, Y., Besnard, A. G., and Secondi, J. (2017). Paintings predict the distribution of species, or the challenge of selecting environmental predictors and evaluation statistics. Global Ecology and Biogeography. 5
- Hijmans, R. J., Phillips, S., Leathwick, J., Elith, J., and Hijmans, M. R. J. (2015). Package 'dismo'. 9
- Hill, M. P., Gallardo, B., and Terblanche, J. S. (2017). A global assessment of climatic niche shifts and human influence in insect invasions: HILL et al. Global Ecology and Biogeography. 9, 11, 14, 16, 19, 25
- Hirzel, A. H., Le Lay, G., Helfer, V., Randin, C., and Guisan, A. (2006). Evaluating the ability of habitat suitability models to predict species presences. Ecological Modelling, 199(2):142–152. 12
- Jiménez-Valverde, A., Peterson, A. T., Soberón, J., Overton, J. M., Aragón, P., and Lobo, J. M. (2011). Use of niche models in invasive species risk assessments. Biological Invasions, 13(12):2785–2797. 23, 25
- Kramer, A. M., Annis, G., Wittmann, M. E., Chadderton, W. L., Rutherford, E. S., Lodge, D. M., Mason, L., Beletsky, D., Riseng, C., and Drake, J. M. (2017). Suitability of Laurentian Great Lakes for invasive species based on global species distribution models and local habitat. Ecosphere, 8(7):e01883. 24
- Krebs, C. J. (2009). The Experimental Analysis of Distribution and Abundance. Benjamin Cummings, 6th edition. 5
- Leroy, B., Meynard, C. N., Bellard, C., and Courchamp, F. (2016). virtualspecies, an R package to generate virtual species distributions. Ecography, 39(6):599–607. 10
- Merow, C., Smith, M. J., Edwards, T. C., Guisan, A., McMahon, S. M., Normand, S., Thuiller, W., Wüest, R. O., Zimmermann, N. E., and Elith, J. (2014). What do we gain from simplicity versus complexity in species distribution models? Ecography, 37(12):1267–1281. 6, 23, 25
- Nix, H. (1986). A biogeographic analysis of Australian elapid snakes. In Longmore, R., editor, Atlas of Elapid Snakes of Australia, pages 4–15. Australian Government Publishing Service, Canberra. 7, 9
- Olson, D. M., Dinerstein, E., Wikramanayake, E. D., Burgess, N. D., Powell, G. V. N., Underwood, E. C., D'amico, J. A., Itoua, I., Strand, H. E., Morrison, J. C., Loucks, C. J., Allnutt, T. F., Ricketts, T. H., Kura, Y., Lamoreux, J. F., Wettengel, W. W., Hedao, P., and

Kassem, K. R. (2001). Terrestrial Ecoregions of the World: A New Map of Life on Earth. BioScience, 51(11):933. 11

- Petitpierre, B., Broennimann, O., Kueffer, C., Daehler, C., and Guisan, A. (2017). Selecting predictors to maximize the transferability of species distribution models: lessons from cross-continental plant invasions: Which predictors increase the transferability of SDMs? Global Ecology and Biogeography, 26(3):275–287. 9, 12, 15, 23, 25
- Phillips, S. J., Anderson, R. P., and Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. Ecological Modelling, 190(3-4):231–259. 9
- Randin, C. F., Dirnböck, T., Dullinger, S., Zimmermann, N. E., Zappa, M., and Guisan, A. (2006). Are niche-based species distribution models transferable in space? Journal of Biogeography, 33(10):1689–1703. 5
- Renner, I. W., Elith, J., Baddeley, A., Fithian, W., Hastie, T., Phillips, S. J., Popovic, G., and Warton, D. I. (2015). Point process models for presence-only analysis. Methods in Ecology and Evolution, 6(4):366–379. 8
- Renner, I. W. and Warton, D. I. (2013). Equivalence of MAXENT and Poisson Point Process Models for Species Distribution Modeling in Ecology: Equivalence of MAXENT and Poisson Point Process Models. Biometrics, 69(1):274–281.
- Sequeira, A. M. M., Bouchet, P. J., Yates, K. L., Mengersen, K., and Caley, M. J. (2018). Transferring biodiversity models for conservation: Opportunities and challenges. Methods in Ecology and Evolution, 9(5):1250–1264. 5
- Tatem, A. J., Hay, S. I., and Rogers, D. J. (2006). Global traffic and disease vector dispersal. Proceedings of the National Academy of Sciences, 103(16):6242–6247. 17
- Varela, S., Lobo, J. M., and Hortal, J. (2011). Using species distribution models in paleobiogeography: A matter of data, predictors and concepts. Palaeogeography, Palaeoclimatology, Palaeoecology, 310(3-4):451–463. 5

CONTACT US

t 1300 363 400 +61 3 9345 2176 e enquiries@data61.csiro.ar

AT CSIRO WE SHAPE THE FUTURE

We do this by using science and technology to solve real issues. Our research makes a difference to industry, people and the planet.

FOR FURTHER INFORMATION

Matt Hill, Peter Caley, James Camac, Jane Elith, Simon Barry

e Matt.Hill@csiro.au

D. edmaps: an R package for creating maps for early detection

The framework described in Chapter 4 has been incorporated into an *R* package called edmaps (Early Detection Maps). edmaps greatly simplifies the computational steps required for creating maps of establishment likelihood. It achieves this by taking user-defined inputs in the form of a *Microsoft Excel* spreadsheet (see below) and creating a make-like workflow that automates the entire process from raw data processing through to the production of GeoTIFF rasters as well as presentation-quality static and interactive maps. Users greatly benefit from workflows using edmaps because they define the dependency structure of those objects, and ensure that after modifications (e.g., changes to files, *R* objects, function arguments), any affected dependants are updated. In order to create such workflows edmaps uses the drake package (Landau, 2018) for *R*.

In the following sections, we briefly outline the data and system requirements for running edmaps. We then provide details on how to specify the parameters for building edmaps workflows for multiple plant pests, and lastly, we provide details on three options for implementing edmaps on your system.

D.1. Software requirements

D.1.1. Minimum system requirements

- *R* version 4.0.0 (may also be compatible with more recent versions)
- *RStudio* (convenience, optional)
- Java (JDK)
- *pandoc* (allows standalone html maps; if missing, a directory of accessory files will be created alongside the html file.)
- 16 GB available RAM
- 10 GB available hard disk space
- Internet connection



D.1.1.1. Operating system specific requirements

Windows

• Rtools may be required to build packages from source.

macOS

- Command Line Tools to build packages from source when required.
- *macOS* is prone to errors when attempting to link *R* and *Java*. These errors can often be resolved by running sudo R CMD javareconf in a terminal (requires administrator privileges).

Linux

On Linux, system packages must be installed from source, meaning that additional software and system libraries are required. The list below gives *Ubuntu* package names. If using a different distribution, you will need to identify the corresponding package names (e.g. for some distributions, dev should be replaced with devel).

- libxml2-dev
- libudunits2-dev
- libgeos-dev
- libgdal-dev
- default-jdk
- liblzma-dev
- libbz2-dev
- libicu-dev
- libpcre3-dev
- libz-dev
- libcairo2-dev
- gdal-bin



D.1.2. Required *R* packages

The edmaps package depends upon a set of other *R* packages (Table D.1) in order to calculate likelihoods of establishment. Package functionality often changes as packages evolve, so specific versions of these dependencies (i.e. current versions as at 30 September, 2019; see Table D.1) were used in the development of edmaps and its application to the case study species explored in this report. Binaries or source tarballs of these package versions are archived by the MRAN (Microsoft R Application Network) Time Machine, and are available at http://cran.microsoft.com/snapshot/2020-07-14/. To ensure expected behaviour and accurate reproduction of outputs, these package versions are recommended. Similarly, recursive dependencies (i.e., dependencies of the direct dependencies, and so on) should also be installed from the above snapshot. Installation of the appropriate versions of dependencies can be automated, as described below.

D.1.3. Data directory

edmaps requires a data directory (approximately 0.5 GB in size, supplied with this report) to be present on your computer. This data directory contains all the raw spatial layers and data needed to create maps of establishment likelihood for Australia. The general structure and content of this directory are as follows:

- risk_layers/
 - abiotic/occurrences/ (user-collated occurrence records and CABI data)¹
 - biotic/raw_data/
 - ACLUM/ (Australian landuse raster);
 - NDVI/ (NDVI raster);
 - NVIS_5.1/ (NVIS raster);
- pathway/raw_data/
 - Containers/ (contains POA shapefile, containers_bypostcode.xls)
 - Fertiliser/ (contains fertiliser use csv file, NRM shapefile)
 - Major_Airports/ (Australian airport locations)
 - Population/ (human population raster)
 - Ports/ (port locations/use csv file)
 - Tourist_Beds/ (tourist bed shapefile)
- user_input/parameters.xlsx (file for specifying global and species parameters)
- edmaps_1.6.3.tar.gz (tarball for installing edmaps)
- make.R (script for running edmaps without renv)
- make_renv.R (script for running edmaps with renv)
- renv.lock (renv lock file containing package dependency details)

¹Generally set up as a sub folder for each species. CABI distribution data is downloaded from CABI datasheets in csv format. For example see here.



Mandling new data

This data repository currently contains the latest available spatial layers (see Chapter 4 for details). However, we envisage that some of these spatial layers will be updated as new versions become available. For example, both CLUMC and populations layers can be replaced when newer versions become available. Furthermore, different NDVI layers, using different temporal periods, can be added to this data directory for use with different pests. If layers are to be replaced it is critical that they are in the same format as the ones currently supplied in the directory.



Package	Version
countrycode	1.2.0
CoordinateCleaner	2.0-15
dplyr	1.0.0
drake	7.12.4
fasterize	1.0.2
furrr	0.1.0
gdalUtilities	1.1.0
geometry	0.4.5
ggplot2	3.3.2
glue	1.4.1
htmlwidgets	1.5.1
leaflet	2.0.3
leaflet.opacity	0.1.0
leafem	0.1.1
magrittr	1.5
mapedit	0.6.0
purrr	0.3.4
raster	3.3-7
rasterVis	0.48
readr	1.3.1
readxl	1.3.1
rgbif	3.1.0
rlang	0.4.7
rnaturalearth	0.1.0
sf	0.9-4
sp	1.4-2
stars	0.4-3
tidyr	1.1.0
tmap	3.0
tmaptools	3.0
viridis	0.5.1
knitr	1.29
rmarkdown	2.3
testthat	2.3.2
future.callr	0.5.0
rnaturalearthdata	0.1.0
lubridate	1.7.9
styler	1.3.2
OpenStreetMap	0.3.4
rnaturalearthhires	0.2.0

 Table D.1.: Direct R package dependencies of the edmaps R package.



D.2. Parametrising edmaps

edmaps automates the many steps that need to be taken in order to create pest-specific maps of establishment likelihoods (Chapter 4). Broadly, these steps are grouped as follows:

- 1. Download global climate data from WorldClim;
- 2. Restructure large categorical raster datasets (landuse and NVIS) for efficient access;
- 3. Prepare supporting datasets (e.g. a raster defining the Australian land mass);
- 4. Create rasters describing the geographic distribution of arrivals for each pathway relevant to a pest;
- 5. Create a raster defining the expected number of arrivals across all relevant pathways;
- 6. Create a raster defining biotic suitability for the pest (i.e. geographic distribution of host material);
- 7. Create a raster defining abiotic suitability for a pest (if relevant). This may involve using a climate suitability layer provided by the user, or if not supplied, occurrence data will be sourced (from user-supplied files, from GBIF, or both), automatically cleaned, and a range bagging model will be fit and projected across the landscape to estimate climatic suitability;
- 8. Create a raster defining the establishment likelihood;
- 9. Create a raster at coarser resolutions (for management or display purposes);
- 10. Create publication quality national and city-scale maps of pest establishment likelihoods and other intermediate maps (e.g. climate suitability, host distributions etc.);
- 11. Create national interactive html maps of estimated total pest arrivals, environmental suitability (abiotic × biotic) and establishment likelihoods; and
- 12. Create GeoTiffs of all output rasters.



Which steps are implemented and how edmaps implements them is all dictated by a single *Microsoft Excel* workbook found within the data subdirectory,

user_input/parameters.xlsx. This workbook contains two tabs. The first (*Global variables*) defines the set of parameters that are relevant to all pests considered (Figure D.1). These include various input data paths², the number of carriers entering Australia per annum for each pathway, air passenger distance decay penalties, and whether interactive html maps should be produced.

The second tab (*Species-specific parameters*) contains pest-specific parameters, where each pest is a separate row in the spreadsheet (Figure D.2). These parameters encompass:

- which pathways are relevant to the pest;
- whether to incorporate abiotic suitability (may not be relevant for some pests);
- whether NDVI and/or NVIS should be incorporated when estimating biotic suitability;
- file paths to occurrence datasets and CABI distributional data or a user's own climate suitability layer;
- whether to exclude BIOCLIM variables from range bagging model (only relevant if climate suitability file not specified);
- the probability of a pathway carrier containing the pest for each considered pathway;
- a likelihood penalty defined by distance from ports; and
- the scale at which outputs should be aggregated.

Note:

In order to facilitate users in filling in the species-specific parameters, the spreadsheet contains conditional formatting that will highlight columns in red if they are undefined but required in order for edmaps to design a workflow for that pest. Columns that are white are optional and columns that are grey are not applicable.

Supplying your own climate suitability layer

The Excel spreadsheet also allows users to feed in their own climate suitability layer, as opposed to edmaps estimating this for you. This can be done by passing the path to the layer into parameters.xlsx via the column Climate suitability path. However, if this is done it is critical it meets edmaps requirements. Specifically, that its score ranges between 0 and 1, it is projected to Australian Albers (EPSG:3577), has 1000 m resolution and has the following extent: xmin = -1888000, xmax = 2122000, ymin = -4847000, ymax = -1010000.

²Paths may change as newer versions of data become available.



2	А	В	C
1	GLOBAL VARIABLES		
2	Variable	Used for	Value
3	Make interactive maps?	Specifies whether interactive maps should be produced	TRUE
4	CLUM raster path	Defining distribution of host land use	risk_layers/biotic/raw_data/ACLUM/clum_50m1 218m.tif
	NVIS raster path	Defining distribution of vegetation classes in which the pest might	risk_layers/biotic/raw_data/NVIS_5.1/aus5_1e_ mvs
5	NDVI raster path	Optional component of habitat	risk_layers/biotic/raw_data/NDVI/NDVI_Oct18_ Mar19.grid
7	Postcode shapefile path	Defines container weights (containers pathway)	risk_layers/pathway/raw_data/Containers/postal areas/POA 2011 AUST.shp
8	Containers data path	Defines container weights (containers pathway)	risk_layers/pathway/raw_data/Containers/contai ners_bypostcode.xls
9	Marine ports data path	Defines port weights (vessels pathway)	risk_layers/pathway/raw_data/Ports/ports.csv
10	Fertiliser data path	Defines distribution of fertiliser (fertiliser pathway)	risk_layers/pathway/raw_data/Fertiliser/fertilise r2016_17.csv
11	NRM shapefile path	Defines distribution of fertiliser (<i>fertiliser</i> pathway)	risk_layers/pathway/raw_data/Fertiliser/nrm_re gions/NRMR_2016_AUST.shp
12	Airport distance penalty (tourists)	Defines decrease in density of international tourists with distance from international airports (<i>tourists</i> pathway)	-0.003465736
13	Airport distance penalty (Torres)	Defines decrease in density of passengers arriving from Torres Strait Islands with distance from Cairns airport (<i>torres</i> pathway)	-0.069314718
14	Total tourists	Total number of international tourists arriving via air passage per annum (<i>tourists</i> pathway)	10633690
15	Total returning residents	Total number of residents returning from overseas via air passage per annum (<i>residents</i> pathway)	12722820
16	Total passengers from TSI	Total number of air passengers arriving from Torres Strait Islands per annum (torres pathway)	51000
17	Total mail	Total number of mail items arriving in Australia per annum (<i>mail</i> pathway)	15200000
18	Total vessels	Total number of vessels arriving at Australian ports per annum (<i>vessels</i> pathway)	18612
19	Total fertiliser	Total volume of fertiliser arriving in Australia per annum (<i>fertiliser</i> pathway)	6907
20	Total machinery	Total number of items of machinery arriving in Australia per annum (<i>machinery</i> pathway)	1216726
21	Total nurserystock	Total volume of nurserystock arriving in Australia per annum (<i>nurserystock</i> pathway)	4475
22	Total food	Total volume of food arriving in Australia per annum (<i>food</i> pathway)	15832
23	Global variables	Species-specific variables	·

Figure D.1.: *Global parameters* tab in parameters.xlsx. Used for defining parameters that are relevant to all pests considered.

1	A	В	C	D	E	F	G	Н	1	1	K	L	М	N	0	P	Q	R	5	Ţ	U	٧	W	Х	Y
1	Species	Pathways	Include abiotic weight?	Include NDVI?	Include NVIS?	CLUM classes	NVIS classes	GBIF species name(s)	GBIF min year	Occurrence path	CA8I path	Climate suitability path	Exclude BIOCUM vars	Prob tourists	Prob returning resident	Prob Torres passenger	Prob mail	Prob vessels	Distance penalty (ports)	Prob fertiliser	Prob machinery	Prob containers	Prob nurserystock	Prob food	Aggregated res
2	oriental fruitfly	tourists, residents, torres	TRUE	TRUE	FALSE	340-342, 344, 345, 347-351, 353, 365, 440-442, 444, 445, 447-451, 453, 540- 544		Bactrocera dorsalis, Bactrocera invadens, Bactrocera papayae, Bactrocera philippinensis	1970	risk_layers/abiotic /occurrences/orien tal_fruitfly/dorsali s_occurrences.csv	risk_layers/abiotic /occurrences/orien tal_fruitfly/cabi_d orsalis_20April19.c sv			1.5877E-06	7.9386E-07	4.1176E-04									5000
3	khapra beetle	tourists, residents, mail, vessels, fertiliser	FALSE	FALSE	FALSE	520-524, 526-528, 530-534, 538, 540- 545, 574								1.0938E-06	5.4688E-07		7.9710E-08	1.6119E-04	-0.6931	4.3434E-04		5	5		5000
4	gypsy moth	vessels, machinery, container	TRUE	TRUE	TRUE	110-117, 120-125, 130-134, 210, 220- 222, 310-314, 332, 341, 343, 345, 346, 410-414, 432, 441, 443, 445, 446, 541 543	1-11, 13, 14, 16, 18-25, 27-29, 45, 47, 48, 51-56, 59- 62, 65-69, 79, 80, 90, 91, 96-99	Lymantria dispar	1970		risk_layers/abiotic /occurrences/gyps y_moth/cabi_lyma ntria_dispar_6lune 19.csv							3.4548E-05	-0.0693		1.1506E-05	7.3590E-07			5000
5	BMSB	tourists, residents, machinery, nursery, food	TRUE	TRUE	FALSE	331, 332, 338, 340 346, 348-353, 431, 432, 438, 440-453, 465, 510-515, 540- 544		Halyomorpha halys	1970	risk_layers/abiotic /occurrences/BMS B/bmsb_presences _iprrg.csv	risk_layers/abiotic /occurrences/BMS B/cabi_halyomorp ha halys_9Jun19.csv			1.5877E-06	7.9386E-07						1.6438E-09		4.4693E-04	1.2633E-04	5000

Figure D.2.: Species-specific parameters tab in parameters.xlsx. Used for defining species-specific parameters. Each species is a row.



D.3. Running edmaps

As highlighted in the above sections, edmaps requires multiple software and *R* packages to be installed. While these can be manually installed, we advise that this is done as a last resort. This is because each computer is different. Operating systems and software versions are likely to vary substantially between computers. As such it is extremely difficult ensure edmaps will run in all settings.

For this reason, and to ensure edmaps functionality into perpetuity, we strongly recommend that edmaps is implemented in a *Docker* virtual machine. *Docker* is the world's leading software container platform that is used to create lightweight, self-contained virtual Linux systems that contain all relevant open source software required to run developed software. Unlike other virtual machines, *Docker* does not bundle a full operating system. Rather it only installs libraries and settings required to make the developed software work. This means that *Docker* ensures software works as expected on any machine. The other major advantage is that it removes the need for users to install software dependencies, as the hard work is already done.

In the following sections we outline three methods for running edmaps, given that parameters.xlsx has been filled out. While we recommend the *Docker* approach, we acknowledge that not all users will have the appropriate permissions to do so. As such we also outline two additional approaches for implementation: one that assumes software dependencies are correctly installed, but handles *R* package dependencies (renv approach), and one that assumes all dependencies are installed, and focuses on installation and use of edmaps.



D.3.1. Using edmaps with Docker (recommended)

We have created a *Docker* image that contains all the system libraries, software (e.g. *R*, *Java*, *pandoc*) and *R* packages required to install and run workflows produced by edmaps.

To use the image, first install *Docker* onto your machine. When installing *Docker* on *Windows*, you will be prompted to select whether to use Linux or *Windows* containers. Leave this at it's default (i.e. to use Linux containers). Once installed, we recommend users set *Docker* settings such that containers have access to *at least* 10 GB of RAM, and a specified number of CPUs³.

Next we can download an edmaps virtual machine by entering the following into the command line using Command Prompt or the terminal (requires internet connection):

docker pull jscamac/edmaps

Once the *Docker* image has been successfully downloaded, use the Command Prompt or terminal to navigate to the local copy of the data directory outlined in section D.1.3. Next, run the system-specific command:

macOS & Linux

docker run -d -v \$(pwd):/home/ jscamac/edmaps R CMD BATCH --vanilla make.R make.log

Windows

docker run -d -v %cd%:/home/ jscamac/edmaps R CMD BATCH --vanilla make.R make.log

The above command will launch the virtual machine and run make.R in non-interactive mode. This means that you may close the terminal or shell window and the virtual machine will continue running until all tasks have been completed upon which it will automatically shut down. The status and any error messages (if encountered) associated with the workflow progression can be examined by opening the log file make.log. This file will be saved in the data directory, alongside make.R. Once completed, all static and interactive maps as well as GIS-compatible rasters (GeoTIFF) can be found within the outputs directory.

Note:

The flag -v mounts the current directory (i.e. the data directory) into the container's home directory. This allows for any results produced within the container to be reflected within the local directory. This means that you can play with the results, data and figures outside the Docker container later.

³We recommend that users allow Docker to access up to all but one core for processing workflows derived from edmaps, especially if being used for many pests.



D.3.2. Using edmaps with renv

If insufficient privileges exist to use *Docker*, an alternative is to use a renv (reproducible **env**ironments) workflow (Ushey, 2019). The renv *R* package reproduces a pre-defined package environment, ensuring that specified package versions are used. To do this it creates a private, isolated package library for an *R* project, and obtains defined versions of packages from defined repositories. Versions and repositories are specified in a "lock file" (renv.lock), and the lock file describing the versions used for the case studies has been provided with this report. Since the renv package library is project-specific, it will not overwrite or interfere with package versions used in other projects.

The most important difference between the *Docker* approach described above, and renv, is that the latter manages *R* packages only; required system libraries and tools must be installed manually.

Assuming software dependencies (e.g. *Java*, *R*) are available, recreating the package environment with renv is straightforward.

First, navigate to the data directory (Section D.1.3) using Command Prompt or the terminal. Then run:

R CMD BATCH --vanilla make_renv.R make.log

This will create the local package environment, and commence the process of rebuilding the output datasets. When this process is underway, the file make.log will be created and will contain a live log of progress. If any errors are encountered, the error messages will be logged to this file. As the build progresses, files will be created within various subdirectories (e.g. the outputs directory).

If you see an error message stating that R is not a recognized command, you may need to provide the full path to the *R* executable, for example:

c:/Program Files/R/R-4.0.2/bin/R CMD BATCH --vanilla make_renv.R make.log

D.3.3. Using edmaps by itself

We recommend that this mode of running edmaps is only used as a last resort, or if one can be sure that the appropriate versions of all software and *R* package dependencies have been correctly installed.

Assuming all dependencies have been correctly installed, one need only install edmaps. This can be done by using Command Prompt or the terminal to navigate to the data directory and running:

```
R CMD INSTALL edmaps_1.3.0.tar.gz
```

Alternatively, the latest version of edmaps can be directly downloaded from GitHub by running the following in R (assuming devtools is installed):

```
devtools::install_github("jscamac/edmaps")
```

Once edmaps has been installed, a workflow can be initiated by running:

R CMD BATCH --vanilla make.R make.log

As with other approaches, a log file (make.log) will be saved to the data directory and can be interrogated to assess analysis progression or any errors that may have been encountered.

E. Reference manual for edmaps R package

Package 'edmaps'

November 3, 2020

Type Package

Title Estimate Likelihood of Pest Establishment

Version 1.6.3

Maintainer James Camac <james.camac@gmail.com>

Description This package combines spatial environmental data (e.g. distance from airports/marine ports, distribution of pest habitat/hosts) with biosecurity leakage rates (e.g. expert-elicited) to generate maps of the likelihood of pest establishment across the landscape. Final outputs include raster datasets (GTiff) indicating establishment likelihood across the area of interest, as well as interactive (html) and static (pdf) versions of these maps.

Depends R (>= 4.0)

SystemRequirements Java (>= 1.5), JRI, GNU make

Imports countrycode, CoordinateCleaner, dplyr, drake, fasterize, furrr, gdalUtilities, geometry, ggplot2, glue, htmlwidgets, httr, leaflet, leaflet.opacity, leafem, magrittr, mapedit, methods, purrr, raster, rasterVis, readr, readxl, rgbif, rlang, rnaturalearth, sf, sp, stars, tidyr, tmap, tmaptools, viridis

Suggests knitr, rmarkdown, testthat, future.callr, rnaturalearthdata, lubridate, styler, OpenStreetMap, rnaturalearthhires

VignetteBuilder knitr

License MIT + file LICENSE

Encoding UTF-8

LazyData true

RoxygenNote 7.1.1

ByteCompile true

Author James Camac [aut, cre], John Baumgartner [aut]

RemoteType github

RemoteHost api.github.com

RemoteRepo edmaps

RemoteUsername jscamac

RemoteRef master

RemoteSha 80d25701fc649174d397288d0ff72eeebac1df02

GithubHost api.github.com GithubRepo edmaps GithubUsername jscamac GithubRef master GithubSHA1 80d25701fc649174d397288d0ff72eeebac1df02

R topics documented:

aggregate_raster	3
arrivals_by_containers	4
arrivals_by_fertiliser	5
arrivals_by_food	6
arrivals_by_machinery	6
arrivals_by_mail	7
arrivals_by_nurserystock	8
arrivals_by_residents	9
arrivals_by_torres	10
arrivals_by_tourists	11
arrivals_by_vessels	12
pinarize_and_aggregate	12
calc_proportion	13
captured_by_ncells	14
combine_arrivals	14
container_weights	15
lisperse_arrivals	16
download_worldclim2	16
establishment_likelihood	17
excel_to_plan	18
exp_function	18
extract_highest_ncells	19
extract_worldclim2	19
fertiliser_by_nrm	20
fertiliser_weight	21
fill_na	21
gdal_reproject	22
get_airport_dist	24
get_gbif_records	25
nitialise_raster	26
nteractive_map	27
max_normalize	28
min_max_normalize	29
na_mask	29
ncells_4_threshold	30
plan_globals	30
plot_establishment_captured	31
plot_raster	32
port_weights	33
range_bag	34
asterize_vector	35
record_flagger	36

2

aggregate_raster

rle_compress	
species_plan	
static_map	
step_function	
suitability	
surveillance_coverage	
threshold_raster	
weight_airport_dist	

aggregate_raster Downscale raster resolution

Description

Aggregate raster cells (and optionally layers) to coarser resolution.

Usage

```
aggregate_raster(
  rast,
  outfile,
  aggregate_factor,
  fun = sum,
  return_rast = FALSE
)
```

Arguments

rast	Raster* object or file path to a raster file.
outfile	Character. Output raster file path. If not provided, object will be returned to R. Directory will be created recursively if it does not exist.
aggregate_facto	r
	Integer. Aggregation factor expressed as number of cells in each direction (hor- izontally and vertically). Or a vector of two integers (horizontal and vertical aggregation factors) or three integers (when also aggregating over layers).
fun	Function. Function used to aggregate values. Default is sum.
return_rast	Logical. Should the resulting raster be returned to R? Ignored if outfile is not provided.

Value

If outfile is specified, the resulting raster is saved as a geotiff to that path. If return_rast is TRUE or outfile is not specified the resulting raster is returned, otherwise NULL is returned invisibly.

arrivals_by_containers

```
Estimates pest arrivals by containers
```

Description

Estimates pest arrivals by containers.

Usage

```
arrivals_by_containers(
  container_weights,
  port_data,
  template_raster,
  probability,
  outfile,
  return_rast = FALSE
)
```

Arguments

container_weigh	its
	An sf object or file path to a shapefile supported by OGR, as produced by container_weights.
port_data	Character. Path to csv file containing Port Names, Port Codes, Longitude, Lati- tude and Container volumes.
template_raster	
	RasterLayer or file path to a raster file. This is used to define the extent and resolution of output. Must be in CRS EPSG:3577.
probability	Numeric. The probability that a unit carries the pest.
outfile	Character. Output raster file path. If not provided, raster object will be returned to R.
return_rast	Logical. Should the raster object be returned to R? Ignored if outfile is not provided.

Value

If outfile is specified, the resulting raster is saved as a to that path. If return_rast is TRUE or outfile is not specified the resulting raster is returned, otherwise NULL is returned invisibly.

See Also

```
Other functions estimating arrivals: arrivals_by_fertiliser(), arrivals_by_food(), arrivals_by_machinery(), arrivals_by_mail(), arrivals_by_nurserystock(), arrivals_by_residents(), arrivals_by_torres(), arrivals_by_tourists(), arrivals_by_vessels()
```

arrivals_by_fertiliser

Estimates pest arrivals by fertiliser imports

Description

Estimates pest arrivals by fertiliser imports.

Usage

```
arrivals_by_fertiliser(
   fertiliser_weight,
   fertiliser_units,
   probability,
   outfile,
   return_rast = FALSE
)
```

Arguments

fertiliser_weig	ht
	A RasterLayer or file path to a raster file containing weights to distribute arrivals by.
fertiliser_unit	S
	Integer. The total number of fertiliser units entering country.
probability	Numeric. The probability a unit carries pest.
outfile	Character. Output raster file path. If not provided, raster object will be returned to R.
return_rast	Logical. Should the raster object be returned to R? Ignored if outfile is not provided.

Value

If outfile is specified, the resulting raster is saved as a to that path. If return_rast is TRUE or outfile is not specified the resulting raster is returned, otherwise NULL is returned invisibly.

See Also

Other functions estimating arrivals: arrivals_by_containers(), arrivals_by_food(), arrivals_by_machinery(), arrivals_by_mail(), arrivals_by_nurserystock(), arrivals_by_residents(), arrivals_by_torres(), arrivals_by_tourists(), arrivals_by_vessels()

arrivals_by_food

Description

Estimates pest arrivals through imported food as a function of population density.

Usage

```
arrivals_by_food(
   pop_density,
   total_imports,
   probability,
   outfile,
   return_rast = FALSE
)
```

Arguments

pop_density	A RasterLayer or file path to a raster file containing population density.
total_imports	Integer. Amount of imported food entering country.
probability	Numeric. The probability a imported food item contains pest
outfile	Character. Output raster file path. If not provided, raster object will be returned to R.
return_rast	Logical. Should the raster object be returned to R? Ignored if outfile is not provided.

Value

If outfile is specified, the resulting raster is saved as a to that path. If return_rast is TRUE or outfile is not specified the resulting raster is returned, otherwise NULL is returned invisibly.

See Also

Other functions estimating arrivals: arrivals_by_containers(), arrivals_by_fertiliser(), arrivals_by_machinery(), arrivals_by_mail(), arrivals_by_nurserystock(), arrivals_by_residents(), arrivals_by_torres(), arrivals_by_tourists(), arrivals_by_vessels()

arrivals_by_machinery Estimates pest arrivals through imported machinery

Description

Estimates pest arrivals through imported machinery as a function of population density.

arrivals_by_mail

Usage

```
arrivals_by_machinery(
   pop_density,
   total_machinery,
   probability,
   outfile,
   return_rast = FALSE
)
```

Arguments

<pre>pop_density total_machiner;</pre>	A RasterLayer or file path to a raster file containing population density.
	Integer. Amount of machinery entering country.
probability	Numeric. The probability a machinery item contains pest.
outfile	Character. Output raster file path. If not provided, raster object will be returned to R.
return_rast	Logical. Should the raster object be returned to R? Ignored if outfile is not provided.

Details

This pathway is used under the assumption that the vast majority of imported machinery comprises new and used motor vehicles.

Value

If outfile is specified, the resulting raster is saved as a to that path. If return_rast is TRUE or outfile is not specified the resulting raster is returned, otherwise NULL is returned invisibly.

See Also

```
Other functions estimating arrivals: arrivals_by_containers(), arrivals_by_fertiliser(), arrivals_by_food(), arrivals_by_mail(), arrivals_by_nurserystock(), arrivals_by_residents(), arrivals_by_torres(), arrivals_by_tourists(), arrivals_by_vessels()
```

arrivals_by_mail Estimates pest arrivals through mail

Description

Estimates arrival rate attributable to mail as a function of population density.

Usage

```
arrivals_by_mail(
   pop_density,
   total_mail,
   probability,
   outfile,
   return_rast = FALSE
)
```

Arguments

pop_density	A RasterLayer or file path to a raster file containing population density.
total_mail	Integer. Amount of mail.
probability	Numeric. The probability a mail item contains pest.
outfile	Character. Output raster file path. If not provided, raster object will be returned to R.
return_rast	Logical. Should the raster object be returned to R? Ignored if outfile is not provided.

Value

If outfile is specified, the resulting raster is saved as a to that path. If return_rast is TRUE or outfile is not specified the resulting raster is returned, otherwise NULL is returned invisibly.

See Also

Other functions estimating arrivals: arrivals_by_containers(), arrivals_by_fertiliser(), arrivals_by_food(), arrivals_by_machinery(), arrivals_by_nurserystock(), arrivals_by_residents(), arrivals_by_torres(), arrivals_by_tourists(), arrivals_by_vessels()

arrivals_by_nurserystock

Estimates pest arrivals through imported nursery stock

Description

Estimates pest arrivals through imported nursery stock as a function of population density.

Usage

```
arrivals_by_nurserystock(
  pop_density,
  total_imports,
  probability,
  outfile,
  return_rast = FALSE
)
```

Arguments

pop_density	A RasterLayer or file path to a raster file containing population density.
total_imports	Integer. Amount of imported nursery stock entering country.
probability	Numeric. The probability a imported food item contains pest.
outfile	Character. Output raster file path. If not provided, raster object will be returned to R.
return_rast	Logical. Should the raster object be returned to R? Ignored if outfile is not provided.
arrivals_by_residents

Value

If outfile is specified, the resulting raster is saved as a to that path. If return_rast is TRUE or outfile is not specified the resulting raster is returned, otherwise NULL is returned invisibly.

See Also

```
Other functions estimating arrivals: arrivals_by_containers(), arrivals_by_fertiliser(), arrivals_by_food(), arrivals_by_machinery(), arrivals_by_mail(), arrivals_by_residents(), arrivals_by_torres(), arrivals_by_tourists(), arrivals_by_vessels()
```

arrivals_by_residents Estimates pest arrivals by returning residents

Description

Estimates arrival rate attributable to returning residents as a function of population density.

Usage

```
arrivals_by_residents(
   pop_density,
   total_returning,
   probability,
   outfile,
   return_rast = FALSE
)
```

Arguments

pop_density	A RasterLayer or file path to a raster file containing population density.
total_returning	
	Integer. The number of returning residents.
probability	Numeric. The probability a returning resident carries pest
outfile	Character. Output raster file path. If not provided, raster object will be returned to R.
return_rast	Logical. Should the raster object be returned to R? Ignored if outfile is not provided.

Value

If outfile is specified, the resulting raster is saved as a to that path. If return_rast is TRUE or outfile is not specified the resulting raster is returned, otherwise NULL is returned invisibly.

See Also

```
Other functions estimating arrivals: arrivals_by_containers(), arrivals_by_fertiliser(), arrivals_by_food(), arrivals_by_machinery(), arrivals_by_mail(), arrivals_by_nurserystock(), arrivals_by_torres(), arrivals_by_tourists(), arrivals_by_vessels()
```

arrivals_by_torres Estimates pest arrivals due to Torres Strait air passengers

Description

Estimates pest arrivals due to Torres Strait air passengers coming into Cairns.

Usage

```
arrivals_by_torres(
   pop_density,
   airport_weight,
   total_passengers,
   probability,
   outfile,
   return_rast = FALSE
)
```

Arguments

pop_density	A RasterLayer or file path to a raster file containing population density.	
airport_weight	A RasterLayer or file path to a raster file containing distance from Cairns airport weights.	
total_passengers		
	Integer. The number of passengers arriving to Cairns from Torres Strait.	
probability	Numeric. The probability a Torres Strait passenger carries a pest.	
outfile	Character. Output raster file path. If not provided, raster object will be returned to R.	
return_rast	Logical. Should the raster object be returned to $R?\ Ignored \ if \ outfile \ is not provided.$	

Value

If outfile is specified, the resulting raster is saved as a to that path. If return_rast is TRUE or outfile is not specified the resulting raster is returned, otherwise NULL is returned invisibly.

See Also

```
Other functions estimating arrivals: arrivals_by_containers(), arrivals_by_fertiliser(), arrivals_by_food(), arrivals_by_machinery(), arrivals_by_mail(), arrivals_by_nurserystock(), arrivals_by_residents(), arrivals_by_tourists(), arrivals_by_vessels()
```

arrivals_by_tourists Estimates pest arrival due to tourists

Description

Estimates pest arrival due to tourists as a function of distance from airport and tourist accommodation.

Usage

```
arrivals_by_tourists(
   tourist_beds,
   airport_weights,
   total_tourists,
   probability,
   outfile,
   return_rast = FALSE
)
```

Arguments

tourist_beds	A RasterLayer or file path to a raster file containing tourist bed numbers.	
airport_weights		
	A RasterLayer or file path to a raster file containing airport distance weights.	
total_tourists	Integer. The number of tourists.	
probability	Numeric. The probability a tourist carries a pest.	
outfile	Character. Output raster file path. If not provided, raster object will be returned to R.	
return_rast	Logical. Should the raster object be returned to R? Ignored if outfile is not provided.	

Value

If outfile is specified, the resulting raster is saved as a to that path. If return_rast is TRUE or outfile is not specified the resulting raster is returned, otherwise NULL is returned invisibly.

See Also

```
Other functions estimating arrivals: arrivals_by_containers(), arrivals_by_fertiliser(), arrivals_by_food(), arrivals_by_machinery(), arrivals_by_mail(), arrivals_by_nurserystock(), arrivals_by_residents(), arrivals_by_torres(), arrivals_by_vessels()
```

arrivals_by_vessels Estimates pest arrivals by vessels

Description

Estimates pest arrivals by vessels.

Usage

```
arrivals_by_vessels(
   port_weight,
   n_vessels,
   probability,
   outfile,
   return_rast = FALSE
)
```

Arguments

port_weight	A RasterLayer or file path to a raster file containing weights to be used to distribute arrivals.
n_vessels	Integer. The total number of vessels entering country.
probability	Numeric. The probability a unit carries pest.
outfile	Character. Output raster file path. If not provided, raster object will be returned to R.
return_rast	Logical. Should the raster object be returned to $R?$ Ignored if outfile is not provided.

Value

If outfile is specified, the resulting raster is saved as a to that path. If return_rast is TRUE or outfile is not specified the resulting raster is returned, otherwise NULL is returned invisibly.

See Also

```
Other functions estimating arrivals: arrivals_by_containers(), arrivals_by_fertiliser(), arrivals_by_food(), arrivals_by_machinery(), arrivals_by_mail(), arrivals_by_nurserystock(), arrivals_by_residents(), arrivals_by_torres(), arrivals_by_tourists()
```

binarize_and_aggregate

Binarize a categorical raster and optionally aggregate

Description

Binarize a categorical raster and optionally aggregate and/or resample.

calc_proportion

Usage

```
binarize_and_aggregate(
    infile,
    rle,
    outfile,
    extent,
    res,
    categories,
    overwrite = FALSE,
    return_rast = FALSE,
    quiet = FALSE
)
```

Arguments

infile	Character. File path to a categorical raster. If rle is provided, infile must not be provided.
rle	A raster_rle object generated by rle_compress or a file path to such an object saved as .rds.
outfile	Character. The target file path for the binarized raster.
extent	Output extent as an Extent object or an object from which an Extent object can be extracted. If not provided, extent will be taken from input. If extent differs from that of infile, new cells will be assigned value 1 if any original cells belonging to categories have their centroids within the new cell.
res	output resolution as a numeric vector of 1 or 2 elements, or a Raster* object from which resolution can be extracted. If not provided, resolution will be taken from input. If res differs from that of infile, new cells will be assigned value 1 if any original cells belonging to categories have their centroids within the new cell.
categories	Integer or numeric vector of class values to be laballed as 1 in the target raster.
overwrite	Logical. Should outfile be overwritten if it exists?
return_rast	Logical. Should the target raster be returned to R as a Raster layer (TRUE) or not returned (FALSE)?
quiet	Logical. Should progress messages be suppressed?

Value

A binarized raster layer is written to outfile, and if $return_rast$ is TRUE, the raster is additionally returned to R as a Raster layer.

calc_proportion

Calculate raster cell values as proportions

Description

Calculates raster cell values as a proportion of the sum of all cells' values.

Usage

calc_proportion(rast)

Arguments

rast Raster object.

Value

 $A \; {\tt RasterLayer}.$

captured_by_ncells Calculate establishment likelihood captured in top n cells

Description

Calculate the proportion of establishment likelihood captured in top n cells.

Usage

```
captured_by_ncells(infiles, names, n_cells, all = TRUE)
```

Arguments

infiles	Character vector. File path(s) to one or more raster files.
names	Character vector. Names corresponding to infiles.
n_cells	Integer. The number of cells to consider.
all	Logical. If TRUE, return the proportion captured from 1 to n cells. If FALSE, return the cumulative proportion.

Value

Proportion of establishment likelihood captured, or a vector of cumulative proportions.

combine_arrivals Sum arrivals across entry pathways

Description

Sums estimated arrivals rates across all entry pathways.

Usage

```
combine_arrivals(x, outfile, return_rast = FALSE)
```

14

Arguments

x	Character vector giving file path(s) to rasters to be included in the summation, or a RasterStack containing those rasters. If a character vector of length > 1 is provided, all rasters must have the same extent and resolution.
outfile	Character. Output raster file path. If not provided, the RasterLayer will be returned to R.
return_rast	Logical. Should the RasterLayer be returned to R? Ignored if outfile is not provided.

Value

If outfile is specified, the resulting RasterLayer is saved to outfile. If return_rast is TRUE or outfile is not specified, the resulting RasterLayer is returned, otherwise NULL is returned invisibly.

container_weights Creates a weight by postcode for each major port

Description

Creates a weight by postcode for each major port.

Usage

```
container_weights(
   path,
   sheet_nums,
   range = "A7:M2217",
   postcode_shp,
   na = c("", "-", "np"),
   outfile,
   return_sf = FALSE
)
```

path	Character. File path to xls file containing containers by postcode for each port.
sheet_nums	Integer. Vector of integers signifying the sheet numbers to read in.
range	A cell range to read from, as described in cell-specification. Includes typ- ical <i>Excel</i> ranges such as "B3:D87", possibly including the sheet name like "Budget!B2:G14", and more. Interpreted strictly, even if the range forces the inclusion of leading or trailing empty rows or columns.
postcode_shp	Character. File path to postcode shape file.
na	Character vector of strings to interpret as missing values. By default, readx1 treats blank cells as missing data.
outfile	Character. Name of shapefile where output will be saved. If not provided, sf object will be returned to R.
return_sf	Logical. Should the sf object be returned to R? Ignored if outfile is not provided.

Details

For the purposes of this analysis missing data (i.e. NAs) will be treated as zeroes.

Value

An sf object or shapefile export.

disperse_arrivals Disperse pest arrivals

Description

Disperses pest arrivals based on input raster.

Usage

disperse_arrivals(rast, n_vectors, probability)

Arguments

rast	Raster object containing dispersal weights.
n_vectors	Numeric. The number of transmission vectors (e.g. passengers).
probability	Numeric. The probability a vector contains pest.

Value

A raster object.

download_worldclim2 Download climate layers from WorldClim 2.0

Description

Download climate layers from WorldClim 2.0 (current climate only).

Usage

```
download_worldclim2(outfile, variable, resolution)
```

Arguments

outfile	Character. Target file for downloaded .zip archive.
variable	Character. Can be one of "bio" (19 standard bioclim variables), "tmin", "tmax", "tavg", "srad", "wind" or "vapr".
resolution	Character. Can be one of "10m", "5m", "2.5m", or "30s".

Value

A zipfile is downloaded and NULL is returned invisibly.

16

establishment_likelihood

See Also

extract_worldclim2

Examples

```
## Not run:
download_worldclim2('bioclim_10m', 'bio', '10m')
```

End(Not run)

establishment_likelihood

Estimate establishment likelihood

Description

Estimate the likelihood of pest establishment based on total arrivals and environmental suitability.

Usage

```
establishment_likelihood(
  total_arrivals,
  suitability,
  outfile,
  return_rast = FALSE
)
```

Arguments

total_arrivals	A RasterLayer (or path to raster file) containing total arrival estimates.
suitability	A RasterLayer (or path to raster file) containing suitability scores.
outfile	Character. Output raster file path. If not provided, the RasterLayer will be returned to \ensuremath{R} .
return_rast	Logical. Should the RasterLayer be returned to $R?\ Ignored\ if\ outfile\ is\ not\ provided.$

Value

If outfile is specified, the resulting RasterLayer is saved as to that path. If return_rast is TRUE or outfile is not specified the resulting RasterLayer is returned, otherwise NULL is returned invisibly.

excel_to_plan

Description

Import parameters from an Excel xlsx file and create a drake plan to estimate establishment likelihoods.

Usage

excel_to_plan(file)

Arguments

file

Character. File path to an xlsx file containing required parameters for creating a species plan. See Details and species_plan for further information.

Details

To simplify reproducibility, edmaps provides an *Excel* interface for specifying species parameters relevant to estimating establishment likelihood. An example spreadsheet is bundled with the package, available at the path given by system.file('extdata/parameters.xlsx', package='edmaps'). The spreadsheet has two sheets, the first specifying "global" parameters that will apply to all species (e.g. file paths to rasters that will), be used regardless of species identity and the second specifying parameters that can vary by species. In the second sheet, each row corresponds to a separate species. Tooltips and data validation guide the user with respect to expected/allowable data.

Value

A drake plan with a combined workflow for all species.

exp_function Exponentiate raster values

Description

Apply exponential function to raster values.

Usage

```
exp_function(rast, beta)
```

Arguments

rast	Raster object.
beta	Numeric. The beta coefficient of a standard exponential function.

Value

A Raster object.

extract_highest_ncells

Returns a raster with values for the n cells with highest establishment likelihood

Description

Returns a raster with values only for the n cells with highest establishment likelihood.

Usage

```
extract_highest_ncells(infile, n_cells, outfile, return_rast = FALSE)
```

Arguments

infile	File path to a raster containing estimated likelihoods of establishment or arrival.
n_cells	Integer. The number of cells to return.
outfile	Character. Output raster file path. If not provided, the RasterLayer will be returned to R.
return_rast	Logical. Should the RasterLayer be returned to R? Ignored if outfile is not provided.

Details

This function returns a raster containing values for only the top n cells.

Value

If outfile is specified, the resulting RasterLayer is saved as to that path. If return_rast is TRUE or outfile is not specified, the resulting RasterLayer is returned, otherwise NULL is returned invisibly.

Warning

This function uses the quantile approach for determining the top n cells. As such, if risk is highly aggregated this function may run into issues when n_cells is high.

extract_worldclim2 Extract WorldClim 2.0 data

Description

Extract WorldClim 2.0 data

Usage

extract_worldclim2(path_2_zip, outdir)

Arguments

path_2_zip	Character. Path to .zip archive downloaded by download_worldclim2.	
outdir	Character. File path to which contained files should be extracted. Will be created (recursively) if necessary.	

Value

Raster data are extracted to outdir and NULL is returned invisibly.

See Also

download_worldclim2

Examples

```
## Not run:
download_worldclim2('bioclim_10m.zip', 'bio', '10m')
extract_worldclim2('bioclim_10m.zip', outdir='bioclim')
```

End(Not run)

fertiliser_by_nrm Create fertiliser by nrm sf object

Description

Create fertiliser by nrm sf object.

Usage

```
fertiliser_by_nrm(abs_data, nrm_shapefile, outfile, return_sf = FALSE)
```

Arguments

abs_data	Character. File path to ABS .csv file.
nrm_shapefile	Character. File path to NRM shapefile.
outfile	Character. Name of shapefile (or other vector data format supported by OGR) where output will be saved. If not provided, sf object will be returned to R.
return_sf	Logical. Should the sf object be returned to R? Ignored if outfile is not provided.

Value

An sf object or vector data export.

fertiliser_weight Create a fertiliser weight raster

Description

Create a fertiliser weight raster as a function of estimated nrm fertiliser tonnes and landuses.

Usage

```
fertiliser_weight(fert_nrm, fert_landuses, outfile, return_rast = FALSE)
```

Arguments

fert_nrm	A file path to vector data or an sf object.
fert_landuses	A file path to raster file or a RasterLayer object.
outfile	Character. Output raster file path. If not provided, RasterLayer will be returned to R.
return_rast	Logical. Should the RasterLayer be returned to R? Ignored if outfile is not provided.

Value

If outfile is specified, the resulting RasterLayer is saved as to that path. If return_rast is TRUE or outfile is not specified, the resulting RasterLayer is returned, otherwise NULL is returned invisibly.

fill_na	Modify NAs within neighbourhood of data cells
	<i>J J J J J J J J J J</i>

Description

Apply a function to NA cells within a moving window.

Usage

fill_na(x, fun, w, outfile, return_rast = FALSE, overwrite = FALSE)

х	A Raster* object.
fun	The function (name or symbol) to apply to the moving windows. First argument should represent the vector of cells contained in the focal window. Only non-NA cell values will be passed to the function.
w	A weights matrix (see focalWeight and focal defining the focal window to which fun will be applied. Note that all contributing cells will be given equal weight (i.e. varying weights are not respected - non-zero & non-NA weights will be replaced by 1).
outfile	Character. File path to an output raster file. If missing, a temporary file will be used.
return_rast overwrite	Logical. Should the RasterLayer be returned to R? Logical. Should outfile be replaced if it already exists?

Details

NA values within focal blocks will be ignored. Raster edges will be NA-padded to allow focal computations at the edge (see focal.

Value

Returns the resulting RasterLayer if return_rast is TRUE. Returns the output file path otherwise.

gdal_reproject Reproject and resample a raster

Description

This function can be used to change resolution, projection and extent of a raster.

Usage

```
gdal_reproject(
    infile,
    outfile,
    src_proj,
    tgt_proj,
    res,
    resampling_method = "near",
    tgt_extent,
    buffer,
    src_nodata,
    datatype = "Float32",
    return_rast = FALSE,
    overwrite = TRUE
)
```

infile	Character. File path to a raster file.
outfile	Character. Output raster file path.
src_proj	Character. The source file coordinate system. Only needs to be set if you wish to reproject object and it is not specified in the infile. The coordinate systems that can be passed includes EPSG PCS and GCSes (i.e. EPSG:4296), PROJ.4 declarations (as above), or the name of a .prj file containing well known text. Starting with GDAL 2.2, if the SRS has an explicit vertical datum that points to a PROJ.4 geoidgrids, and the input dataset is a single band dataset, a vertical correction will be applied to the values of the dataset.
tgt_proj	Character. The target coordinate system. See src_proj for details. If not speci- fied, outfile will inherit the coordinate system of infile.
res	Numeric vector containing x and y resolution. e.g. c(xres, yres). Must be specified in the units of tgt_proj.

resampling_method		
	Character. One of "near", "bilinear", "cubic", "cubicspline", "lanczos", "average", "mode", "max", "min", "med", "q1", "q3". See Details. Default is "near".	
tgt_extent	Numeric vector containing corner coordinates specified as c(xmin,ymin,xmax,ymax). Must be specified in the units of tgt_proj.	
buffer	Numeric. Add buffer (specified in units of tgt_proj). Useful for adding buffers around coasts, and for filling in small gaps within raster specified by infile.	
<pre>src_nodata</pre>	Integer. The nodata value for an input file.	
datatype	A character string indicating the output data type. See the gdalwarp docs for more information.	
return_rast	Logical. Return RaserLayer to R?	
overwrite	Logical. Should outfile be overwritten if it already exists?	

Details

The resampling methods available are as follows:

- near: nearest neighbour resampling (default, fastest algorithm, worst interpolation quality).
- bilinear: bilinear resampling.
- cubic: cubic resampling.
- cubicspline: cubic spline resampling.
- lanczos: Lanczos windowed sinc resampling.
- average: average resampling, computes the average of all non-NODATA contributing pixels. (GDAL >= 1.10.0).
- mode: mode resampling, selects the value which appears most often of all the sampled points. (GDAL >= 1.10.0).
- max: maximum resampling, selects the maximum value from all non-NODATA contributing pixels. (GDAL >= 2.0.0).
- min: minimum resampling, selects the minimum value from all non-NODATA contributing pixels. (GDAL >= 2.0.0).
- med: median resampling, selects the median value of all non-NODATA contributing pixels. (GDAL >= 2.0.0).
- q1: first quartile resampling, selects the first quartile value of all non-NODATA contributing pixels. (GDAL >= 2.0.0).
- q3: third quartile resampling, selects the third quartile value of all non-NODATA contributing pixels. (GDAL >= 2.0.0).

Value

_rast A raster file is produced on disk. Additionally, if return_rast is TRUE a RasterLayer object is returned to R.

get_airport_dist

Description

Generates a raster proximity map indicating the distance from the center of each pixel to the centre of the nearest pixel containing a major airport. Airport coordinates are in the source raster.

Usage

```
get_airport_dist(
   vector_data,
   outfile,
   template_raster,
   extent,
   res,
   crs,
   airport_codes,
   return_rast = FALSE,
   overwrite = FALSE
)
```

vector_data	Character. Path to airport data (available here).
outfile	Character. Output raster file path. Directory will be created recursively if it does not exist.
template_raster	
	Optional. Raster* object or a file path to template raster. If this is provided, extent, res, and crs will be taken from this raster unless they are also passed to this function. If template_raster is not provided, then extent and res must be provided. The template raster will also be used to mask the output raster so that NA cell values will be propagated.
extent	Either a character path to a raster file, an Extent object (or an object from which such an extent can be extracted), or a numeric vector with four elements giving xmin, xmax, ymin, ymax.
res	Numeric or integer vector giving the horizontal and vertical spatial resolution, in units of crs. If a single value is given, it will be used for both horizontal and vertical resolution.
crs	Target coordinate reference system as a PROJ string (character) or an object of class CRS.
airport_codes	Numeric. Airport codes to be used as targets in order to estimate cell proximity.
return_rast	Logical. Return RasterLayer to R?
overwrite	Logical. Should outfile be removed if it already exists?

get_gbif_records

Value

A proximity raster is written to outfile. If return_rast is TRUE, the raster object is also returned to R, otherwise NULL is returned invisibly. This function assumes that crs is either an unprojected coordinate system, or that the units of crs are metres. Resulting distances are expressed in kilometres.

get_gbif_records Downloads GBIF records of species

Description

Downloads GBIF records of species, removing auxiliary columns and records recorded prior to a specified year, or that have coordinate uncertainty above a specified amount.

Usage

```
get_gbif_records(
  taxon,
  min_year,
  coord_uncertainty,
  method = c("search", "download"),
  username,
  pwd,
  email,
  retries = 10,
  cleanup = TRUE
)
```

taxon	Character. Species taxonomic name.
min_year	Integer. The minimum year for which records should be collated. Default is 'NULL', i.e. no minimum.
coord_uncertair	nty
	Integer. The maximum allowable documented coordinate uncertainty (in me- tres). If specified, this argument will result in removal of any records that have a documented uncertainty above this limit. Records with no uncertainty docu- mented will be returned regardless of the value of 'coord_uncertainty'.
method	Either ''search'' (uses the GBIF '/occurrence/search' API endpoint) or ''down- load'' (uses the GBIF '/occurrence/download'API endpoint). The former makes paginated queries to the API, while the latter performs an asynchronous query (but waits for the resulting dataset to be ready for download). The ''search'' method is limited to 100,000 records; for large datasets, consider using ''down- load''. When using 'method='download'', the arguments 'username', 'pwd', and 'email' must be provided.
username	GBIF username, required when method is ''download''.
pwd	GBIF password, required when method is ''download''.
email	Email address, required when 'method = 'download''. This _may_ be used to notify user when download is ready.

retries	If 'method='download' ' and file download fails, how many additional attempts should be made to download the file?
cleanup	Logical. Should temporary files associated with 'method='download' be deleted? Default is 'TRUE'.

Details

This function is a wrapper of 'rgbif' such that it can be readily used with the 'CoordinateCleaner' package.

Value

A 'data.frame' of species occurrence records.

initialise_raster Create an empty raster

Description

Create an empty or constant raster with specified attributes.

Usage

```
initialise_raster(
    x,
    outfile,
    extent,
    res,
    crs,
    init = NA,
    datatype = "FLT4S",
    overwrite = TRUE,
    return_rast = FALSE
)
```

x	Optional. Raster* object or a file path to template raster. If this is provided, extent, res, and crs will be taken from this raster unless they are also passed to this function. If x is not provided, then extent and res must be provided.
outfile	Target raster file path. Directory will be created (recursively) if it doesn't exist.
extent	Either a character path to a raster file, an Extent object (or an object from which such an extent can be extracted), or a numeric vector with four elements giving xmin, xmax, ymin, ymax.
res	Numeric or integer vector giving the horizontal and vertical spatial resolution of the target raster, in units of crs. If a single value is given, it will be used for both horizontal and vertical resolution.
crs	Target coordinate reference system as a PROJ string (character) or an object of class CRS. If missing and x is supplied, the crs of x will be used.
init	Numeric. A value assigned to all cells of the created raster.

interactive_map

datatype	Character. Data type for the created raster. See dataType.
overwrite	Logical. Should outfile be overwritten if it exists?
return_rast	Logical. Should the resulting raster be returned?

Value

An empty raster is created at outfile, and the corresponding <code>RasterLayer</code> is returned if <code>return_rast</code> is <code>TRUE</code>.

interactive_map Produce an interactive html map

Description

Produce an interactive html map.

Usage

```
interactive_map(
  ras,
  layer_name = NULL,
  palette = "inferno",
  transparency = 0.8,
  legend = TRUE,
  set_value_range = NULL,
  discrete = FALSE,
  scale_type = "none",
  outfile = NULL,
  surveillance_locs = NULL,
  pt_col = "red",
  cleanup = FALSE
)
```

ras	A RasterLayer or file path to a raster file.	
layer_name	Character. An optional name to assign to ras.	
palette	Either a vector of 2 or more colours (e.g. as hex codes or colour names) or the name of a palette function supported by tmap (see palette_explorer and tm_raster).	
transparency	Numeric. Value between 0 and 1 defining the opacity of the plotted raster data $(1 = $ fully opaque; $0 = $ fully transparent).	
legend	Logical. Should a legend be plotted?	
set_value_range		
	A numeric vector giving upper and lower limits for raster values. Values outside this range (including the limits) will be set to NA.	
discrete	Logical. Are the values of ras discrete (categorical)?	

scale_type	Character. Can be one of: "none" (raw data, no rescaling), "log10", "max normalize" (proportional to maximum value), "minmax normalize" (rescale values to be between 0 and 1 based on min and max) or "logit". Note that if "log10" or "logit" is used, 0 or 1 values must be masked (using set_value_range) or rescaled before passing to this function. scale_type is ignored if discrete is TRUE.	
outfile	Character. If NULL, map will be returned to R and not saved. Otherwise, map will be exported as a html file. Full path address must be used. If pandoc is available, a standalone html file is created (see details).	
surveillance_locs		
	A spatial object or a path to a .csv file containing columns named "Latitude" and "Longitude".	
pt_col	Character. The plotting colour for surveillance points.	
cleanup	Logical. If a standalone html file is created, should accessory files be removed after the standalone file is generated? This will be a folder created within tempdir().	

Details

To create a standalone html file, the pandoc software must be installed and available to R. If pandoc is unavailable, the html file will be accompanied by a folder of accessory files.

Value

A html map.

max_normalize Raster max normalization

Description

Rescales raster values such that maximum is 1.

Usage

```
max_normalize(rast, outfile)
```

Arguments

rast	Raster object.
outfile	Character. Output raster file path. If not specified, the resulting RasterLayer will be returned to R. Directory will be created recursively if required.

Value

A normalised RasterLayer will be written to outfile if specified, and returned to R otherwise.

min_max_normalize Raster min-max normalization

Description

Performs a min-max normalization on a raster object.

Usage

```
min_max_normalize(rast, outfile)
```

Arguments

rast	Raster object.
outfile	Character. Output raster file path. If not specified, the resulting RasterLayer
	will be returned to R. Directory will be created recursively if required.

Value

A normalised RasterLayer will be written to outfile if specified, and returned to R otherwise.

na_mask

Create a mask raster

Description

Create a mask raster, retaining NA and setting all non-NA cells to 1. Input raster will be resampled if necessary, using "max" resampling (i.e. if any contributing cells are not NA, the new cell will be set to 1).

Usage

na_mask(infile, outfile, res, extent)

Arguments

infile	Character. File path to input raster file.
outfile	Character. Output raster file path. Directory will be created (recursively) if it doesn't exist.
res	Numeric. A vector of one or two numbers giving the desired output resolution. If missing, the input resolution will be used. If a single number is provided, it will be used for both horizontal and vertical resolution.
extent	One of of: a numeric vector giving the desired extent of the output raster as c(xmin, xmax, ymin, ymax); an Extent object; or an object from which such an Extent object can be extracted.

Value

A raster is written to outfile, with NA cells transferred from infile and all other cells set to 1.

ncells_4_threshold Extract number of cells required to meet threshold proportion of risk

Description

Extract number of cells required to meet threshold proportion of risk.

Usage

```
ncells_4_threshold(
   risk_rasters,
   names,
   proportion_captured = c(0.6, 0.8, 0.9, 0.95)
)
```

Arguments

risk_rasters	Character. File path(s) to rasters to be loaded.	
names	Character. Names corresponding to infiles.	
proportion_captured		
	Numeric vector. Proportion(s) of risk to be captured.	

Value

A data.frame containing the number of cells to be trapped to capture given proportion(s) of total risk across multiple input files.

plan_globals Create a drake plan describing targets relating to global parameters

Description

Create a drake plan that describes targets relating to global parameters to be used by individual species plans.

Usage

```
plan_globals(
    clum_path,
    nvis_path,
    ndvi_path,
    fertiliser_data_path,
    nrm_path,
    containers_data_path,
    postcode_path,
    airport_beta,
    airport_tsi_beta
)
```

Arguments

clum_path	Character. File path to Catchment Scale Land Use of Australia (ACLUM) raster.	
nvis_path	Character. File path to the National Vegetation Information System (NVIS) raster dataset.	
ndvi_path	Character. File path to the Normalised Difference Vegetation Index (NDVI) raster dataset.	
fertiliser_data	_path	
	Character. File path to a csv containing data describing fertiliser use, available here.	
nrm_path	Character. File path to shapefile of Natural Resource Management Regions.	
containers_data	_path	
	Character. File path to xlsx file containing data about shipping container move- ments, available here.	
postcode_path	Character. File path to postal areas (i.e. post codes) shapefile, available here.	
airport_beta	Numeric. Parameter controlling the distribution of tourists (international air passengers) around Australian international airports. Distance to nearest airport is multiplied by this value and exponentiated to give the relative density of tourists at a location. To generate a distribution that ensures proportion p of tourists within distance d of nearest airport, specify airport_beta=log(p)/d (e.g. to have 50 of an airport, use log(0.5)/200).	
airport_tsi_beta		
	Numeric. Interpretation is as for airport_beta, but applies to air passengers arriving at Cairns International Airport (CNS) from the Torres Strait Islands.	

Value

A drake plan containing targets that generate objects used across species.

plot_establishment_captured

Plots the cumulative proportion of establishment likelihood captured

Description

Plots the cumulative proportion of establishment likelihood captured as a greater number of top ranking risk cells is considered.

Usage

```
plot_establishment_captured(
   data,
   xlab = "Number of cells",
   ylab = "Proportion",
   legend_title = NULL,
   legend = "right",
   prop_line,
   y_limit = NULL,
   width = NA,
```

```
height = NA,
units = c("in", "cm", "mm"),
outfile
)
```

Arguments

data	A data.frame as derived from captured_by_ncells.
xlab	Character. x axis label.
ylab	Character. y axis label.
legend_title	Character. Optional legend title.
legend	Legend position. Either a position name (one of "none", "top", "bottom", "left", or "right", or a vector of two normalised coordinates ranging from 0 to 1, e.g. $c(1,0)$ (for bottom-right).
prop_line	Numeric. Proportion in which to add a vertical line. Default is NULL (no line).
y_limit	Numeric vector giving the minimum and maximum y-axis limits. If omitted this will be determined based on the data.
width	Width of plot. If not defined will use size of current graphic device.
height	Height of plot. If not defined will use size of current graphic device.
units	Character. Units corresponding to height and width. Can be "in", "cm", or "mm". Default is inches ("in").
outfile	Character. Output image file path. Containing directory will be created recursively if it does not already exist.

Value

An image is written to outfile if provided, and otherwise a ggplot object is returned.

plot_raster

Plot a raster

Description

Plot a raster.

Usage

```
plot_raster(
   object,
   legend_title,
   occurrence_data = NULL,
   pt_col = "red",
   height,
   units = c("in", "cm", "mm"),
   compass = FALSE,
   outfile
)
```

32

port_weights

Arguments

object	A RasterLayer or file path to raster file.
legend_title	Character. If missing, the name of the raster layer will be used.
occurrence_data	I Contraction of the second
	A data.frame, sf object, SpatialPointsDataFrame object, or path to a .csv file containing columns named "Latitude" and "Longitude". If NULL, no points will be plotted.
pt_col	Character. Colour of points (if plotted).
height	Height of plot, specified in units. Required if outfile is provided.
units	Character. Units corresponding to height and width. Can be "in", "cm", or "mm". Default is inches ("in").
compass	Logical. Should a North arrow be shown?
outfile	Character. Path to save output.

Value

A tmap object. If outfile is provided, a map will also be written to that file.

port_weights	Create raster giving the establishment likelihoods relating to ports.
--------------	---

Description

Creates spatial port establishment likelihood raster as a function of port container volumes and distance from port.

Usage

```
port_weights(template_raster, port_data, beta, outfile, return_rast = FALSE)
```

Arguments

template_raster	
	A RasterLayer or path to supported raster file.
port_data	Character. Path to csv file containing port container volumes and named Latitude and Longitude (in WGS84) for each port of interest.
beta	Numeric. The beta coefficient exponential function of how risk changes with distance from port. beta should be in units per 1km.
outfile	Character. Output raster file path. Directory will be created if it does not exist. If not provided, object will be returned to R.
return_rast	Logical. Should the resulting raster be returned to R? Ignored if outfile is not provided.

Value

If outfile is specified, the resulting raster is saved as a geotiff to that path. If return_rast is TRUE or outfile is not specified the resulting raster is returned, otherwise NULL is returned invisibly.

range_bag

Description

Fit and project range bag model.

Usage

```
range_bag(
    occurrence_data,
    bioclim_dir,
    n_dims = 2,
    n_models = 100,
    p = 0.5,
    exclude_vars = NULL,
    outfile
)
```

Arguments

occurrence_data

	sf object, data.frame or character path to a csv file containing occurrence coordinates (must contain columns named "Latitude" and "Longitude").
bioclim_dir	Path. Path to directory containing WorldClim raster data.
n_dims	Integer. The number of dimensions ranges to bag.
n_models	Integer. The number of bootstrapped model ensembles to run.
р	Numeric between 0 and 1. The proportion of occurrence records to include in bootstrapping .
exclude_vars	Character vector. A vector of bioclim variables to exclude from analysis. Default is NULL.
outfile	Character. Output raster file path. Parent directory will be created recursively if required. If NULL, the RasterLayer will be returned in R.

Value

A RasterLayer of model predictions is written to outfile if provided, and returned to R otherwise. The raster's extent, resolution and CRS are taken from the raster data in bioclim_dir. Cell values give the fraction of bootstrapped models for which the cell's environment fell within the species' modelled climate envelope.

References

This function is a modified version of the rb function provided in Drake, J.M. & Richards, R.L. (2019) Data from: Estimating environmental suitability. Dryad, Dataset, doi:10.5061/dryad.g5p7d1c.

See also: 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. doi:https://doi.org/10.1098/rsif.2015.008

rasterize_vector Rasterize spatial vector dataset

Description

Converts vector object to a raster with specified extent and resolution.

Usage

```
rasterize_vector(
  vector_data,
  outfile,
  template_raster,
  extent,
  res,
  crs,
  field,
  burn_value,
  datatype = "Float32",
  overwrite = FALSE,
  return_rast = FALSE
)
```

vector_data	Character. File path to vector spatial data.
outfile	Raster output file path. Parent directory will be created recursively if required.
template_raster	
	Optional. Raster* object or a file path to template raster. If this is provided, extent, res, and crs will be taken from this raster unless they are also passed to this function. If x is not provided, then extent and res must be provided.
extent	Either a character path to a raster file, an Extent object (or an object from which such an extent can be extracted), or a numeric vector with four elements giving xmin, xmax, ymin, ymax.
res	Numeric or integer vector giving the horizontal and vertical spatial resolution of the target raster, in units of crs. If a single value is given, it will be used for both horizontal and vertical resolution.
crs	Target coordinate reference system as a PROJ string (character) an object of class CRS. If missing and template_raster is supplied, the crs of template_raster will be used. If template_raster is also not supplied, the CRS of vector_data will be used.
field	Character. Name of attribute to be burned into raster.
burn_value	Numeric. A constant value to burn into raster.
datatype	Character. Output data type (see gdal_rasterize documentation).
overwrite	Logical. Should outfile be overwritten if it exists?
return_rast	Logical. Return object to R?

Value

A binarized raster is written to outfile, and returned to R as a <code>RasterLayer</code> if <code>return_rast</code> is <code>TRUE</code>.

record_flagger

Flags records that occur in countries that are not reported in either infected_countries or CABI datasheet

Description

'record_flagger' cross references occurrence records against either a vector of country names with established populations specified in 'infected_countries' or those presented in a CABI distribution file. It will then either flag or automatically remove occurrence records that occur in countries outside one of these lists.

Usage

```
record_flagger(
    occurrence_records,
    infected_countries,
    cabi_ref,
    manual_check = FALSE,
    return_df = FALSE
)
```

Arguments

occurrence_records		
	A 'data.frame', 'sf' object or 'SpatialPoints*' object, or a path to a .csv file containing the locations of species of interest. (If a 'data.frame' is provided, it must contain the named columns "Latitude" and "Longitude").	
infected_countr	ries	
	Character vector of countries with known established populations.	
cabi_ref	Character. The path to a .csv file downloaded from a CABI datasheet containing the country-scale distributional data for the pest. Note that if infected_countries the CABI data will not be used.	
manual_check	Logical. Allows interactive selection of which unverified points to retain. In- teractive map will only be produced if unverified records are present and if this argument is set to 'TRUE'.	
return_df	Logical. Return the object as a 'data.frame' or as an 'sf' object (i.e. 'TRUE', the default).	

Value

A 'data.frame' or an 'sf' points object flagging records found in countries in which do not occur in infected_countries OR is not reported as being present in CABI.

36

rle_compress

Description

Compress categorical raster data using run length encoding.

Usage

rle_compress(x, outfile, quiet = FALSE)

Arguments

х	File path to the categorical raster to be compressed, or a Raster* object.
outfile	Character (optional). Path to target .rds file that will store RLE results. Directory will be created recursively if it doesn't exist.
quiet	Logical. Should progress messages be suppressed?

Value

A list with five elements:

starts Cell numbers corresponding to run starts

lengths Run lengths

values Run values

extent Raster extent

res Raster resolution

This object is additionally saved in rds format to outfile, if provided.

species_plan

Generate a drake plan for a species

Description

Generate a drake plan that facilitates reproducible generation of species outputs.

Usage

```
species_plan(
   species,
   clum_classes,
   nvis_classes,
   pathways,
   include_abiotic_weight = TRUE,
   climate_suitability_path,
   exclude_bioclim_vars = NULL,
   include_ndvi = TRUE,
```

```
aggregated_res = c(5000, 5000),
  make_interactive_maps = TRUE,
  clum_path,
  nvis_path,
  ndvi_path,
  airport_beta = log(0.5)/200,
  airport_tsi_beta = log(0.5)/10,
  port_data_path,
  port_weight_beta,
  fertiliser_data_path,
  nrm_path,
  containers_data_path,
  postcode_path,
  occurrence_path,
  infected_countries,
  cabi_path,
  use_gbif = FALSE,
  gbif_species,
  gbif_min_year = 1970,
  gbif_max_uncertainty = 20000,
  manual_check_flagged_records = FALSE,
  total_tourists,
  prob_tourists,
  total_returning,
  prob_returning,
  total_torres,
  prob_torres,
  total_mail,
  prob_mail,
  total_vessels,
  prob_vessels,
  total_fertiliser,
  prob_fertiliser,
  total_machinery,
  prob_machinery,
  prob_containers,
  total_nurserystock,
  prob_nurserystock,
  total_food,
  prob_food
)
```

Arguments

species	The name of the species. This will be used for naming output files and folders.
clum_classes	An integer vector indicating which ACLUM classes are considered host plants for species. Either clum_classes or nvis_classes (or both) must be provided.
nvis_classes	An integer vector indicating which NVIS classes are considered host plants for species. Either clum_classes or nvis_classes (or both) must be provided.
pathways	A character vector of invasion pathways that should be included. Can be one or more of: 'containers', 'fertiliser', 'food', 'machinery', 'mail',

38

ingluda abiatia	<pre>'nurserystock', 'residents', 'torres', 'tourists', 'vessels'.</pre>
include_ablotic	Logical. Should suitability be dependent on climate? Considered TRUE if
climate suitabi	lity path
ciillate_Suitabi	Optional file path to a raster describing climatic suitability across the landscape
	If provided, the raster must be have the Australian Albers coordinate system (EPSG:3577), spatial resolution of 1000 m, and must have $xmin = -1888000$, $xmax = 2122000$, $ymin = -4847000$, $ymax = -1010000$. If not provided and include_abiotic_weight is TRUE, a range bag model will be fit to estimate
	climatic suitability.
<pre>exclude_bioclim</pre>	Lvars
	Character vector of bioclim variables that should not be used when fitting a range bag model (see range_bag) of climatic suitability. Variables should be specified as, e.g., c("bio01", "bio12"). Ignored if climate_suitability_path is provided.
include_ndvi	Logical. Should biotic suitability be dependent on NDVI?
aggregated_res	A numeric vector of 2 elements, indicating the desired resolution of aggregated establishment likelihood rasters, in metres.
<pre>make_interactiv</pre>	e_maps
	Logical. Should interactive html maps be generated?
clum_path	Path to the ACLUM raster.
nvis_path	Path to the NVIS raster.
ndvi_path	Path to the NDVI raster.
airport_beta	Numeric. Parameter controlling the distribution of international tourists passengers around international airport. Default is $log(0.5)/200$ (i.e. 50% of passengers within 200km of airport).
airport_tsi_bet	a
	Numeric. Parameter controlling the distribution of Torres Strait passengers around Cairns airport. Default is $log(0.5)/10$ (i.e., 50% of passengers within 10km of Cairns airport).
<pre>port_data_path port_weight_bet</pre>	File path to the marine ports .csv file.
por t_weight_bet	Numeric. Defines the decay rate of an exponential model. In the context of
	pests entering via the vessel pathway, this reflects the decrease in the relative likelihood of pest arrival at locations distant from marine ports. For example, prob_weight_beta=log(0.5)/10 would lead to distance-decay that leads to 50% (i.e. 0.5) of establishment likelihood (prior to considering other relevant pathways) within a distance of 10 map units (i.e., 10 kilometres when res is 1000).
fertiliser_data	_path
	File path to a csv file containing information about fertiliser usage by NRM.
<pre>nrm_path containers_data</pre>	File path to a polygon shapefile of NRMs (natural resource management areas)path
	File path to the dataset giving the distribution of containers by postcode.
<pre>postcode_path occurrence_path</pre>	File path to postal areas shapefile.
	Path to a .csv file containing occurrence data. Must include columns Longitude, Latitude, and Species.

infected_countries		
	A character vector of countries within which the species occurs. Ignored if climate_suitability_path is provided.Only one of infected_countries or cabi_path should be provided.	
cabi_path	Path to a .csv file downloaded from CABI, indicating the countries within which the species occurs. Download links to these files can be found at the bottom of CABI species datasheet webpages, e.g. https://www.cabi.org/isc/datasheet/17685. Ignored if climate_suitability_path is provided. Only one of infected_countries or cabi_path should be provided.	
use_gbif	Logical. Should species occurrence records be sourced from GBIF? Ignored if climate_suitability_path is provided.	
gbif_species	Character vector. Taxon names to use when querying GBIF. Ignored if climate_suitability_path is provided.	
gbif_min_year	Integer. The minimum year (yyyy) to be included when downloading GBIF data. Ignored if climate_suitability_path is provided.	
gbif_max_uncer	tainty	
	Numeric. The maximum permissable coordinate uncertainty for GBIF records. Ignored if climate_suitability_path is provided.	
<pre>manual_check_f</pre>	lagged_records	
	Logical. Should an interactive map be used for manually checking flagged oc- currence records? If TRUE, the user will have the opportunity to select dubi- ous points (i.e. occurrences in countries for which CABI has no record of the species' establishment), to be retained. If FALSE (the default), all such dubious points will be excluded. Ignored if climate_suitability_path is provided.	
	required interactivity will interrupt plan processing.	
total_tourists	Numeric. The total number of tourists entering Australia.	
prob_tourists	Numeric. The per capita rate of pest entry that applies to tourists.	
total_returning	g	
	Numeric. The total number of returnig residents entering Australia.	
prob_returning	Numeric. The per capita rate of pest entry that applies to returning residents.	
total_torres	Numeric. The total number of passenges entering Australia via the Torres Strait Islands pathway.	
prob_torres	Numeric. The per capita rate of pest entry that applies to passengers entering via the Torres Strait Islands pathway.	
total_mail	Numeric. The total volume of mail entering Australia.	
prob_mail	The rate of pest entry per unit volume of mail.	
total_vessels	Numeric. The total volume of marine vessels entering Australia.	
prob_vessels	The rate of pest entry per vessel.	
total_fertilis	er	
	Numeric. The total volume of fertiliser entering Australia.	
prob_fertilise	r The rate of pest entry per unit volume of fertiliser.	
total_machiner	у	
	Numeric. The total volume of machinery entering Australia.	
prob_machinery	The rate of pest entry per unit volume of machinery.	

static_map

prob_containers	5
	The rate of pest entry per container.
total_nurserys	tock
	Numeric. The total volume of nursery stock entering Australia.
prob_nurserysto	ock
	The rate of pest entry per unit volume of nursery stock.
total_food	Numeric. The total volume of food entering Australia.
prob_food	The rate of pest entry per unit volume of food.

Details

To simplify reproducibility, edmaps provides an *Excel* interface for specifying species parameters relevant to estimating establishment likelihood. An example spreadsheet is bundled with the package, available at the path given by system.file('extdata/parameters.xlsx',package='edmaps'). The spreadsheet has two sheets, the first specifying "global" parameters that will apply to all species (e.g. file paths to rasters that will), be used regardless of species identity and the second specifying parameters that can vary by species. In the second sheet, each row corresponds to a separate species. Tooltips and data validation guide the user with respect to expected/allowable data.

See Also

excel_to_plan

static_map

Produce a static map of establishment likelihood

Description

Produce a static map of establishment likelihood, with OpenStreetMap base layer.

Usage

```
static_map(
 ras,
  xlim,
 ylim,
  layer,
 legend_title,
  set_value_range,
  scale_type = "none",
  transparency = 0.7,
  colramp_entire_range = TRUE,
  surveillance_locs,
 pt_col = "red",
  aggregate_raster,
 height,
 units = c("in", "cm", "mm"),
 outfile
)
```

Arguments

ras	A RasterLayer or file path to a raster file.	
xlim	Numeric vector. The longitudinal extent of the area to plot.	
ylim	Numeric vector. The latitudinal extent of the area to plot.	
layer	Character. A layer name to be plotted. Only relevant if loading a RasterStack or RasterBrick.	
legend_title	Character. Legend title.	
<pre>set_value_range</pre>		
	A numeric vector containing an upper and lower bound value (in units of raster). Values outside this range (including values falling on the boundaries) will be masked.	
scale_type	Character. Can be: "none" (raw data, no rescaling), "log10", "max normalize" (proportional to maximum value), "minmax normalize" (rescale values to be between zero and 1 based on min and max), or "logit". Note that if "log10" or "logit" is used 0 or 1 values. must be masked (using set_value_range) or rescaled outside of this function.	
transparency	Numeric. Transparency of raster, between 0-1.	
colramp_entire_range		
	Logical. Whether to set colour ramp limits based on national risk range (TRUE) or by risk range present in region specified by xlim and ylim.	
surveillance_locs		
	A spatial object or a path to a .csv file containing columns named "Latitude" and "Longitude".	
pt_col	Character. The plotting colour for surveillance points.	
aggregate_raster		
	NULL or a list containing the aggregation factor (i.e. number of raster cells to aggregate) and the aggregation function e.g. $list(10, sum)$.	
height	height of plot. If not defined will use size of current graphic device. Width will be determined automatically, based on the aspect ratio given by the plotting extent.	
units	Character. Units in which height is specified. Can be "in" (inches), "cm", or "mm". Default is "in".	
outfile	Character. File path to save map.	

Details

This function relies on the OpenStreetMap package to obtain base layer tiles. This in turn requires Java to be installed, and linking R with Java can sometimes prove challenging. On macOS, Java errors can often be resolved by entering sudo R CMD javareconf in a terminal (which updates R's Java configuration variables). On Windows, ensure that the Java architecture (32-bit/64-bit) matches that of R. Additionally, some Java errors arise when using RStudio but not when using R.

Value

If outfile is provided, a map is saved to that file. Otherwise, a tmap object is returned to R.

step_function Apply a step function to raster values

Description

Reclassify raster values either side of a defined threshold value.

Usage

```
step_function(rast, threshold, lower_value, upper_value)
```

Arguments

rast	Raster object
threshold	Numeric. The threshold value.
lower_value	Numeric. The value to set below the threshold.
upper_value	Numeric. The value to set above the threshold.

Value

A raster object

suitability Combine abiotic and biotic layers to create suitability ra	ıster
--	-------

Description

Combine biotic (and abiotic) layers into a single suitability raster to be used to scale arrival rates.

Usage

suitability(x, outfile, return_rast = FALSE)

Arguments

x	One of: a RasterStack; a RasterBrick; a list of RasterLayer objects, or a vector of file paths to raster files.
outfile	Character. Name of geotiff where output will be saved. If not provided, raster object will be returned to R.
return_rast	Logical. Should the raster object be returned to R? Ignored if outfile is not provided.

Value

If outfile is specified, the resulting raster is saved as a geotiff to that path. If return_rast is TRUE or outfile is not specified the resulting raster is returned, otherwise NULL is returned invisibly.

surveillance_coverage Calculate the proportion of total establishment likelihood captured by current surveillance

Description

Calculate the proportion of total establishment likelihood captured by current surveillance.

Usage

```
surveillance_coverage(
   establishment_rasters,
   layer_names,
   surveillance_locations
)
```

Arguments

establishment_r	asters
	Character vector. A vector of one or more file path to rasters describing estab- lishment likelihood.
layer_names	Character. Name to be assigned to each establishment raster. If not specified, names of the raster layers will be used.
surveillance_lo	cations Character. Path to a csv file containing surveillance locations, including columns titled "Longitude" and "Latitude" with coordinates given in decimal degrees (GDA94).
layer_names surveillance_lo	lishment likelihood. Character. Name to be assigned to each establishment raster. If not specifinames of the raster layers will be used. Incations Character. Path to a csv file containing surveillance locations, including colur titled "Longitude" and "Latitude" with coordinates given in decimal degr (GDA94).

Value

The proportion of establishment likelihood captured by current surveillance.

threshold_raster Threshold raster

Description

Threshold raster.

Usage

```
threshold_raster(rast, threshold, value = 0, outfile)
```
Arguments

rast	A file path to a raster file.
threshold	A named list giving the minimum and/or maximum values defining the range of values to retain. Values outside this range will be replaced with value. Can be list(min = Y, max = Z) or list(min = Y) or list(max = Z).
value	Numeric. The value supplied to cells beyond the threshold(s).
outfile	Character. Output raster file path. Parent directory will be created recursively if required. If outfile is not provided, the resulting RasterLayer will be returned to R.

Value

A RasterLayer will be written to outfile if provided, and returned to R otherwise.

weight_airport_dist Transform airport distances

Description

Transform airport distances according to a negative exponential function.

Usage

```
weight_airport_dist(
    airport_dist,
    beta = log(0.5)/200,
    outfile,
    overwrite = FALSE,
    return_rast = FALSE
)
```

Arguments

airport_dist	Character. File path to a raster file containing proximity to airports. Map units are expected to be kilometres.
beta	Numeric. Parameter passed to the exponential function. Distance to nearest airport is multiplied by this value and exponentiated to give the relative density of tourists at a location. To generate a distribution that ensures proportion p of tourists within distance d of nearest airport, specify airport_beta=log(p)/d (e.g. to have 50 tourists within 200 km of an airport, use log(0.5)/200).
outfile	Character. Output raster file path. If missing, object will be returned to R.
overwrite	Logical. Overwite the target raster if it already exists?
return_rast	Logical. Should the raster object be returned to R?

Value

If return_rast is TRUE, or if outfile is missing, the resulting RasterLayer object will be returned. Otherwise NULL is returned invisibly.