

ACCEPTED VERSION

Robert C. Cope, Joshua V. Ross, Talia A. Wittmann, Michael J. Watts, and Phillip Cassey
Predicting the risk of biological invasions using environmental similarity and transport network connectedness
Risk Analysis, 2019; 39(1):35-53

© 2017 Society for Risk Analysis

Which has been published in final form at <http://dx.doi.org/10.1111/risa.12870>

PERMISSIONS

<http://olabout.wiley.com/WileyCDA/Section/id-828037.html>

Funder Policies

Australian Research Council (ARC) and National Health and Medical Research Council (NHMRC)

Green open access

For ARC funded authors, the accepted version of the article will be made freely available on Wiley Online Library after a 12 month embargo period (starting with first publication online), in accordance with the Public Access Plan. Through CHORUS, ARC's public access interface will also link directly to the publicly available article on Wiley Online Library.

ARC and NHMRC funded authors may self-archive the accepted version of their article after a 12-month embargo period (starting with first publication online) in an open access institutional repository. If articles are made open access following payment of an article publication fee, it is not necessary to archive the accepted version of the article, but the metadata must be available in the institutional repository with a link to the final, published article on Wiley Online Library.

18 November 2019

<http://hdl.handle.net/2440/107341>

Predicting the risk of biological invasions using environmental similarity and transport network connectedness

Robert C. Cope^{1, *}, Joshua V. Ross¹, Talia A. Wittmann², Michael J. Watts³, Phillip Cassey⁴

Understanding the risk of biological invasions associated with particular transport pathways and source regions is critical for implementing effective biosecurity management. This may require both a model for physical connectedness between regions, and a measure of environmental similarity, so as to quantify the potential for a species to be transported from a given region and to survive at a destination region. We present an analysis of integrated biosecurity risk into Australia, based on flights and shipping data from each global geopolitical region, and an adaptation of the ‘range bagging’ method to determine environmental matching between regions. Here, we describe global patterns of environmental matching and highlight those regions with many physical connections. We classify patterns of global invasion risk (high to low) into Australian states and territories. We validate our analysis by comparison with global presence data for 844 phytophagous insect pest species, and produce a list of high-risk species not previously known to be present in Australia. We determined that, of the insect pest species used for validation, the species most likely to be present in Australia were those also present in geopolitical regions with high transport connectivity to Australia, and those regions that were geographically close, and had similar environments.

KEY WORDS: Biosecurity; biological invasions; climate similarity; flights; shipping; Australia.

1. INTRODUCTION

Growth in global trade and transport has greatly accelerated in recent decades, aiding the redistribution of many species across natural geographic barriers at increasing rates^(1,2,3). Establishment of species outside of their native ranges as a result of this transnational trade and traffic has led to serious

economic and environmental costs^(4,5). Damages caused by alien species in the United States of America (USA) are estimated to cost \$120 billion per annum⁽⁶⁾. For example, the Asian longhorned beetle (*Anoplophora glabripennis*) is one of the most destructive wood boring beetles, most likely introduced to North America and Europe via international trade containing wood packaging materials^(7,8). The red imported fire ant (*Solenopsis invicta*) alone, has cost the USA approximately \$1 billion per year, and economists estimate \$8.9 billion in damages over 30 years if uncontrolled in Australia^(9,6). Consequently, effective biosecurity directed towards reducing the risk of biological invasions is of primary concern to governments, industry stakeholders, and communities^(10,11). However, the resources available

¹School of Mathematical Sciences, The University of Adelaide, Adelaide, SA 5005, Australia

²School of Biological Sciences, The University of Adelaide, Adelaide, SA 5005, Australia

³Information Technology Programme, AIS St Helens, Auckland 1140, New Zealand

*Corresponding author email: robert.cope@adelaide.edu.au; phone: +61 8 8313 1082; fax: +61 8 8313 3696.

for biosecurity surveillance and control are limited, so biosecurity activity must incorporate cost-effective assignment of surveillance effort^(12,13,14). To this end, recent biosecurity policy efforts have included a focus on quantitative risk-based analyses, so that resources may be best targeted to those areas that generate the greatest risk⁽¹⁵⁾.

A key component of a risk-based biosecurity strategy is an understanding of the risk provided by particular transport pathways and source regions^(2,3,16,17), and the likelihood of establishment after release. The opportunity for a species to establish outside of its native range, firstly relies on the successful transition of a species through the initial stages of the invasion pathway (i.e., uptake, transport and release⁽¹⁸⁾). Once a species is released into a novel recipient location, many factors such as the suitability of the climate, habitat, dietary requirements and the presence or absence of predators, are likely to control whether a species can establish and spread⁽¹⁹⁾. Rather than considering factors that relate to the potential transport of specific individual species (e.g., the specific commodities alongside which species may be transported, or the presence of hosts for invasive parasites^(20,21)), we focus more broadly on geopolitical regions, and the transport pathways that connect them. We assume that alien species generally are more likely to survive and become established in destination regions that are climatically similar to their source regions. As such, a framework quantifying integrated invasion risk between geopolitical regions requires both a model for physical connectedness between source and destination (e.g., Cope *et al.*⁽²²⁾), along with a metric for environmental similarity between those regions. Transport connectivity and environmental similarity have been combined previously to assess invasion risk, with environmental matching limited to (scaled) Euclidean distance, and only considering the environment at, or around, each port or airport. For example, climate similarity between airports has been assessed alongside global air traffic volumes⁽²³⁾, and a number of studies that investigate global shipping networks have also included climate similarity^(16,24,25,26). However, few studies have considered multiple transport pathways simultaneously, nor considered environmental matches across whole, potentially environmentally diverse, geopolitical regions. This is critical given that export commodities (that may carry stowaway species unintentionally) may originate from any location within the source region from which they were

transported. The recently published ecological niche modelling algorithm ‘range bagging’⁽²⁷⁾ has been demonstrated to effectively and efficiently model species potential ranges on a global scale⁽²⁸⁾. In the present study, we adapt range bagging to determine environmental matches between geographic regions for the purpose of predicting invasion risk. In particular, this allows comparison of environmental conditions not only at a single location, but with environmental conditions present throughout entire source and destination regions. We describe global patterns and clusters of environmental match, and highlight those geopolitical regions with many high risk environmental matches. We then combine this environmental matching approach with an integrated analysis of physical transport, combining both flight and shipping networks⁽²²⁾, to assess overall invasion risk.

Using Australia as a biosecurity case study, we demonstrate the application of this approach for effective matching of source and destination regions to predict potential alien species establishment. Australia is an ideal case study due to its geographic isolation and border biosecurity policy being considered among the best in the world^(29,30). We highlight hotspot regions that are most likely to be the source of an alien species, which could survive and become established within Australia based on both environmental match and physical transport connectedness. We validate this approach by analysing global species aggregation data for 844 phytophagous insect pest species⁽³¹⁾, including 283 species that are present in Australia (and elsewhere). These two complementary approaches demonstrate the application of this approach to both border and post-border biosurveillance: environmental similarity and transport connectedness can inform the allocation of resources for border biosecurity, and the development of lists of high-risk species can inform further species-specific analyses, and post-border surveillance and planning.

2. METHODS

2.1 Environmental data

Environmental data were obtained from WORLDCLIM for 19 climate covariates, plus altitude data, at a resolution of 2.5 arc minutes⁽³²⁾. From these geospatial environmental layers, we extracted one grid point per pixel on terrestrial surfaces, resulting in $c.\sim 8.9$ million points. Of

these points, 500,000 were selected uniformly at random (i.e., selected with equal probability, without replacement) on which to perform our calculations, in order to balance resolution with computing time. Of these points, 27,395 were located within Australia.

2.2 Geopolitical regions

We divided the world hierarchically by country, and then state (or similar administrative region), using the geopolitical regions database in the World Administrative Divisions layer package⁽³³⁾, edited to suit the ISO 3166 country naming conventions. To obtain a more uniform (in size) collection of geopolitical regions, we constructed an aggregate database consisting of countries and large states, i.e. country-level aggregation was used, except for the following countries, in which state level aggregation was used: Australia, Brazil, Canada, China, United States of America, and the Russian Federation. Hereafter, *geopolitical regions* refers to this collection of regions, i.e., states from large countries, and countries otherwise. This resulted in a total of 392 geopolitical regions. Environmental points (above) were assigned to the geopolitical regions within which they were located. The largest region, India, contained 11999 climate points; and the median region size was 611 points.

Note that the scale at which geopolitical regions were chosen could be varied, e.g., Brazil could be retained as a single region, or India or Argentina could be split into subnational units, and this would have an impact on the number of connections between regions. Alternatively, smaller nations or states that are adjacent could be combined, e.g., in eastern Europe, or geographically close island nations. Ultimately, we chose a subdivision of the world that seemed appropriate for our goals, and other practitioners could similarly choose geopolitical assignments suited to the context of their research.

2.3 Environmental matching

Points from within each of the resulting geopolitical regions were used to calculate a range bagging score to each point within Australia and worldwide. Range bagging⁽²⁷⁾ has been demonstrated to effectively match environments for invasive species⁽²⁸⁾. In this context, the full set of environmental conditions within a geopolitical region are assumed to be

described by a many-dimensional convex hull (20 dimensions when based on the WORLDCLIM data), and our objective is to determine if a candidate point (i.e., a location in a different region with associated environmental measurements) is within this convex hull. Range bagging approximates this by constructing a series of lower-dimensional marginal convex hulls (in this case, 2-dimensional) based on random pairs of covariates. The result of range bagging is a score, calculating the proportion of these marginal convex hulls within which the candidate point is contained.

We then aggregated point-wise range bagging scores on a regional level to determine matches between source and destination geopolitical regions. Specifically, consider a source region and a destination region: the full set of environmental data for each point within the source region was used as a source environmental range, and each individual environmental point within the destination region was assessed for its range bagging score against that, i.e., calculating how well an individual point within the destination range matches to the environmental range of the entire source region. Note that the resulting distributions of scores of points between regions are not symmetrical, particularly when regions vary in size or environmental heterogeneity. For example, if region A contains a variety of environmental conditions while region B has a homogenous environment, it is possible that all of region B is a high scoring match to region A, but only a small portion of region A is a high scoring match to region B (e.g., Figure A4).

All of the resulting individual point-wise scores were aggregated to obtain overarching metrics of environmental match between source and destination regions by taking the 90th percentile score within the destination geopolitical region. Using the 90th percentile score means that at least 10% of the recipient region would have to match climatically to score highly. We also considered using the maximum match, which would indicate that there exists at least one location within the recipient geopolitical region with that level of match: if that maximum match is high, then a hypothetical species transported from the source region has somewhere that it could survive within the recipient region, and if that match is low, then there is nowhere at all in the recipient region environmentally similar to the source. However, the maximum had the potential to create matches at the rarest climatic extremes within destination

regions, possibly consisting of only a single location. For geographically large regions the probability of an alien species being transported to a single extreme point is likely low. This was of particular concern in this study as Australian states are large and contain diverse environments. Consequently, we wanted to use a more conservative threshold than the maximum, so that the score more reasonably represented areas of potential dispersal within the destination region, and a 10% cut-off satisfied this requirement. Alternatively, a fixed number of points (representing a fixed area of suitability), or a different quantile, would be reasonable.

We focus on aggregation at geopolitical regions because biosecurity and management decisions are typically made at this level. Specifically, we are most interested in the destination regions, on the assumption that biosecurity managers are most interested in preventing incursions from elsewhere. Climate matching methods like range bagging work by comparing candidate destination points against a source niche (a polygon in multidimensional climate space), so we need to partition the global land mass into a set of source regions against which points in the destination region of interest may be scored. Geopolitical regions are a natural choice for this partition, given that they may be directly associated with specific transport pathways. A reasonable alternative approach could be to divide the world into regions of equal size as source regions, but doing so has the trade-off of making the transport pathway side of the analysis much more difficult, given that international transport typically occurs through major hubs within geopolitical regions.

For destination geopolitical regions within Australia, we also calculated state-based matches specifically informed by the locations of the major (capital city) airport within these destination states. We chose the major airport so as to have a single main transport hub within each of these destination states. In each case (within Australia) these airports were also located near a major port location, and so provide a reasonable proxy for the location of incoming seaborne traffic (however, some states also contained distant ports associated with mining exports). We obtained weighted averages of scores for points within Australian destination regions, with weights decaying exponentially with distance from the major airport in each state (i.e., weight $w = \exp(-u \times \text{distance})$, with decay parameter u chosen in order to produce a half-life of 500 km).

In addition, we constructed a graph represent-

ing the links between geopolitical regions globally, using the open source GRAPHVIZ software (www.graphviz.org), by considering each geopolitical region as a node, and allowing edges between geopolitical regions if the 90th percentile environmental match, in either direction, between the two geopolitical regions was at least 80%. We use modularity-based community detection⁽³⁴⁾ via the `gmap` tool packaged with `graphviz`, to detect regions of the graph with high graph connectedness (i.e., many strong environmental matches between regions).

2.4 Risk model

We combined geopolitical region-based environmental matches with physical connectedness between geopolitical regions based on sea and air traffic (*sensu* Cope *et al.*⁽²²⁾) to determine an integrated model for transport and establishment risk. Specifically, we calculated the weighted cumulative transport pressure $I := w_p \times \text{the number of Passenger Flights} + w_c \times \text{the number of Cargo Flights} + w_s \times \text{the number of Ship voyages}$, for each geopolitical region into each state of Australia, where w_p, w_c, w_s denote the relative weightings of passenger flight, cargo flight, and shipping traffic, respectively. Weightings $w_p = w_c = 1, w_s = 10$ were used as a default, i.e., passenger and cargo flights were equally weighted, ships were 10 times more likely than flights to transport alien species. These values were chosen as the default given that the majority of commodities transported to Australia arrive by ship, but may easily be varied by users in the web app we developed (using the ‘Shiny’ package in R⁽³⁵⁾) to allow users to explore these data (https://robertcope.shinyapps.io/risk_calc/). We present environmental match against physical transport (along with total flights and ship voyages separately) so as to identify and highlight those source regions with low or high risk by either or both metrics.

2.5 Validation

We assessed the integrated risk presented by environmental match and physical transport against global species presence/absence data (at geopolitical scales) for 844 phytophagous insect crop pest species, previously used in studies by Gevery *et al.*⁽³⁶⁾, Worner *et al.*⁽³⁷⁾ and Watts *et al.*⁽³¹⁾. Those species that were not present in Australia in 2003, but were present in 2010, were assumed to be invaders, and

their environmental and transport characteristics were determined. Note that these data were compiled in 2003 and 2010, so we can not include more recent incursions, and may not provide a perfect account of species presence or absence in a region, as data are often sparse and may vary between sources. We used logistic regression to determine which predictors contribute to the presence of individual species within Australia in 2010, and also to model the proportion of species from each geopolitical region that were shared with Australia. While these species are not necessarily all invasive in Australia, presence in Australia is used as a proxy for invasion success.

Predictors considered in these logistic models included: transport volume, environmental match, geographic distance (both directly, and a term indicating exponential decay with distance, i.e., $\exp(-k \times \text{distance})$), size (i.e., geographic area of source regions), and number of insect species present (for regions), or number of regions in which the species is present (for species). For individual geopolitical regions, environmental match was assessed as the 90th percentile match to any Australian state; for species, we assessed both the maximum and median values of these scores across all states outside of Australia in which the species is present. Model selection was performed by choosing the model with minimum Akaike information criterion (AIC), and we report the coefficients and standardised odds ratios for each variable in the final model. We performed leave- k -out cross-validation to assess model fit by constructing random training sets consisting of 90% of points, with the remaining 10% of points forming test sets. The area under the receiver operator characteristic curve (AUC) was calculated for each test set based on models constructed on the training set. This was repeated 1000 times (i.e., 1000 random selections of test and training sets), and we reported the mean of these cross-validated AUC values. Note that AUC can be misleading in ecological applications (see e.g., Lobo et al.⁽³⁸⁾), however it is commonly used to assess predictive performance and thus allows for easy interpretation and comparison.

All unspecified calculations were performed in the R software environment (Version 3.2.0⁽³⁹⁾), with the ‘ggplot2’ package used to produce figures⁽⁴⁰⁾. R Code will be available online at publication.

3. RESULTS

3.1 Global environmental matching

At a global level, there were many strong environmental matches between geopolitical regions. Typically, strong environmental matches occurred between adjacent regions, and often between regions along similar latitudes (Figure A1). The source regions with the most frequent high-scoring (90th percentile scores of at least 80%) environmental matches to other regions were India and Mexico (Table 1), consistent with these being large countries straddling the ‘Tropic of Cancer’, such that they had environmental matches with many geopolitical regions throughout Asia, Africa, northern South America, southern North America, and northern Australia. Destination regions with strong environmental matches ($> 80\%$) were much less variable than source regions (Tables 1, 2).

Graphs were constructed, with geopolitical regions as nodes, and edges between nodes when 90th percentile scores were > 80 (Figure A1). Automatic community detection generally grouped geopolitical regions along latitudinal lines (Figure A1).

3.1.1 Australian case study

At the 90th percentile, ten geopolitical regions had environmental matches to each of Queensland and Western Australia for a threshold level of 80% (Table 3). The remaining states had fewer matching source regions. Note that when the maximum scores were considered instead, Queensland (38) and Western Australia (30) remained the Australian destination regions with the most frequent environmentally matching source regions. When scores were produced based on a weighted average of distance from major airports, so as to mimic potential dispersal for arriving species, the resulting scores were similar to those based on maximum or 90th percentile matches, but slightly lower for primarily desert or tropical regions (Figure 1). This is because in most Australian states the state capital (and thus the main airport) exists at one of the geographic extremes: the southern end of South Australia, Western Australia, and Queensland; the eastern extent of New South Wales, or the northern extent of the Northern Territory.

Table I .: Top 10 highest ranking source geopolitical regions. The “Number of regions as source” column lists the geopolitical regions with the most other regions as high environmental match destination regions (90th percentile match $> 80\%$). For example, there are 154 states or countries for which points within those states are a high range bagging environmental match with India as a source region, i.e., species found in India could likely survive in 154 other regions. The “Number of regions as destination” column lists the number of regions that are high environmental match source regions to this destination region.

Geopolitical region	Number of regions as source	Number of regions as destination
India	154	5
Mexico	101	11
Peru	80	11
Colombia	62	12
Venezuela	62	13
Ethiopia	57	8
Ecuador	55	10
Cameroon	54	20
Pakistan	53	8
Congo	52	18

Table II .: Top 10 highest ranking destination geopolitical regions. The “Number of regions as destination” column lists the the number of other geopolitical regions that are high environmental match source regions (90th percentile match $> 80\%$) to the given destination. For example, there are 26 states or countries for which points within the Gois state of Brazil are a high range bagging environmental match with the given state as a source region, i.e., species found in 26 regions could be considered likely to survive in Gois. The “Number of regions as source” column lists the number of regions that are high environmental match destination regions to this source region.

Geopolitical region	Number of regions as source	Number of regions as destination
Brazil – Gois	4	26
Brazil – Rondonia	1	25
Nicaragua	17	24
Cte d’Ivoire	9	23
Timor-Leste	1	23
Bolivia	32	22
Brazil – Acre	1	22
Brazil – Esprito Santo	1	21
Central African Republic	11	21
Gabon	4	21

Table III .: Number of source geopolitical regions for which each Australian state is a match, based on 90th percentile and maximum range bagging score $> 80\%$. For example, species found in 10-38 other geopolitical regions could be likely to survive within Queensland.

State	Number of regions 90th percentile score	Number of regions max score
Queensland	10	38
Western Australia	10	30
New South Wales	9	18
Northern Territory	8	24
South Australia	8	20
Victoria	7	12
Tasmania	4	7
Australian Capital Territory	3	3

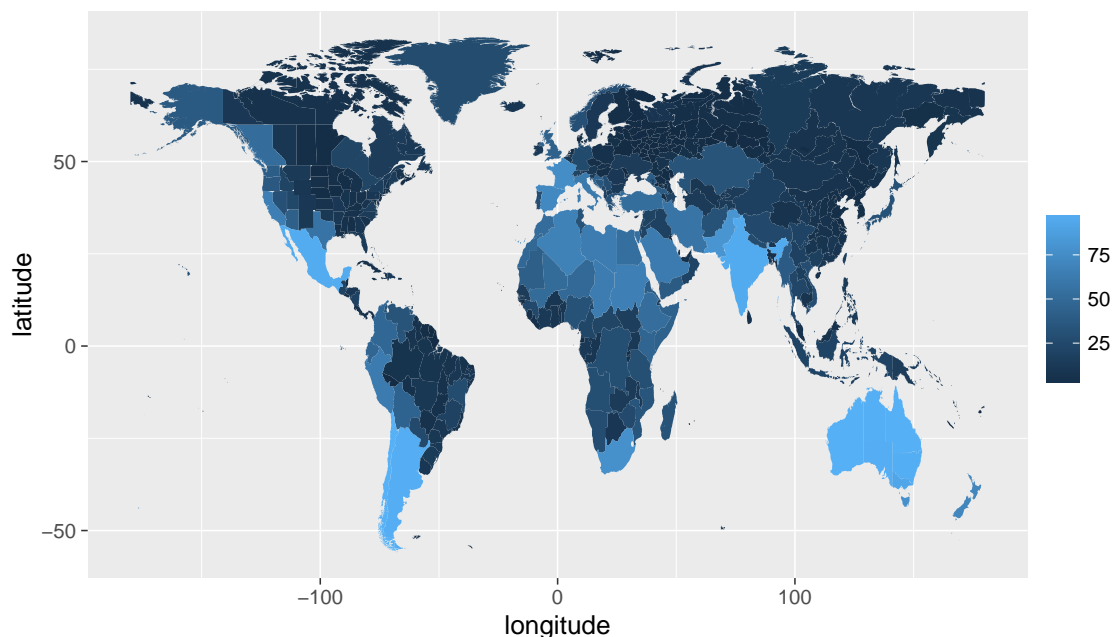


Fig. 1: Maximum environmental match to any Australian state when based on a weighted average of points within that state with weight decaying with distance from its major airport.

3.2 Risk model

When we compare physical connectedness between regions, i.e., the amount of air and sea traffic between them, with environmental match between those regions, we can categorize routes based on potential predicted risk. For example, when weighting integrated physical transport with weights $w_p = w_c = 1$, $w_s = 10$ (i.e., flights equally weighted, shipping traffic 10x greater risk), and using 90th percentile match between geopolitical regions, the highest risk pathway based on 2012 data was New Zealand to Victoria (Figure 2, also see web app). The majority of pathways had either low transport volumes, or low environmental match. These results may be investigated in detail using the web app available at https://robertcope.shinyapps.io/risk_calc/.

3.3 Validation

Four species were listed as not present in Australia in 2003, but present in 2010: *Nasonovia ribisnigri*, *Cerataphis lantaniae*, *Hyblaea puera*, and *Idioscopus niveosparus*. *H. puera* and *I. niveosparus* are found in south east Asia, and *N. ribisnigri* is found primarily in Europe and north America. *C. lantaniae* was found primarily in the Caribbean

during 2003, but its recorded range was substantially broader in 2010, including parts of the USA, and locations within India and south east Asia. In each of these cases, the species existed (by 2010) both somewhere with high environmental match to an Australian state ($\geq 80\%$ 90th percentile score), indicating likely survival potential, and somewhere with high volumes of transport to Australia, suggesting the potential for transport. These results hold more broadly. Of the 844 species in the database, 596 satisfied these two criteria: existing both in regions with environmental match $\geq 80\%$; and in regions with high volumes of physical transport to an Australian state. On a species level, the state-based environmental matches were not entirely informative - 822 of the 844 species exist in at least one state with an environmental match of 80% to an Australian state. Of the remaining 22 species, only the three that occur within Papua New Guinea are present in Australia. Note that there is a strong relationship between geographic proximity to Australia, and transport volume into Australia, and that the majority of species present in Australia were also present in a geographically close geopolitical region, with high transport volumes (Figure 3).

Geographic distance (i.e., distance from closest region to Australia) was consistently present in the

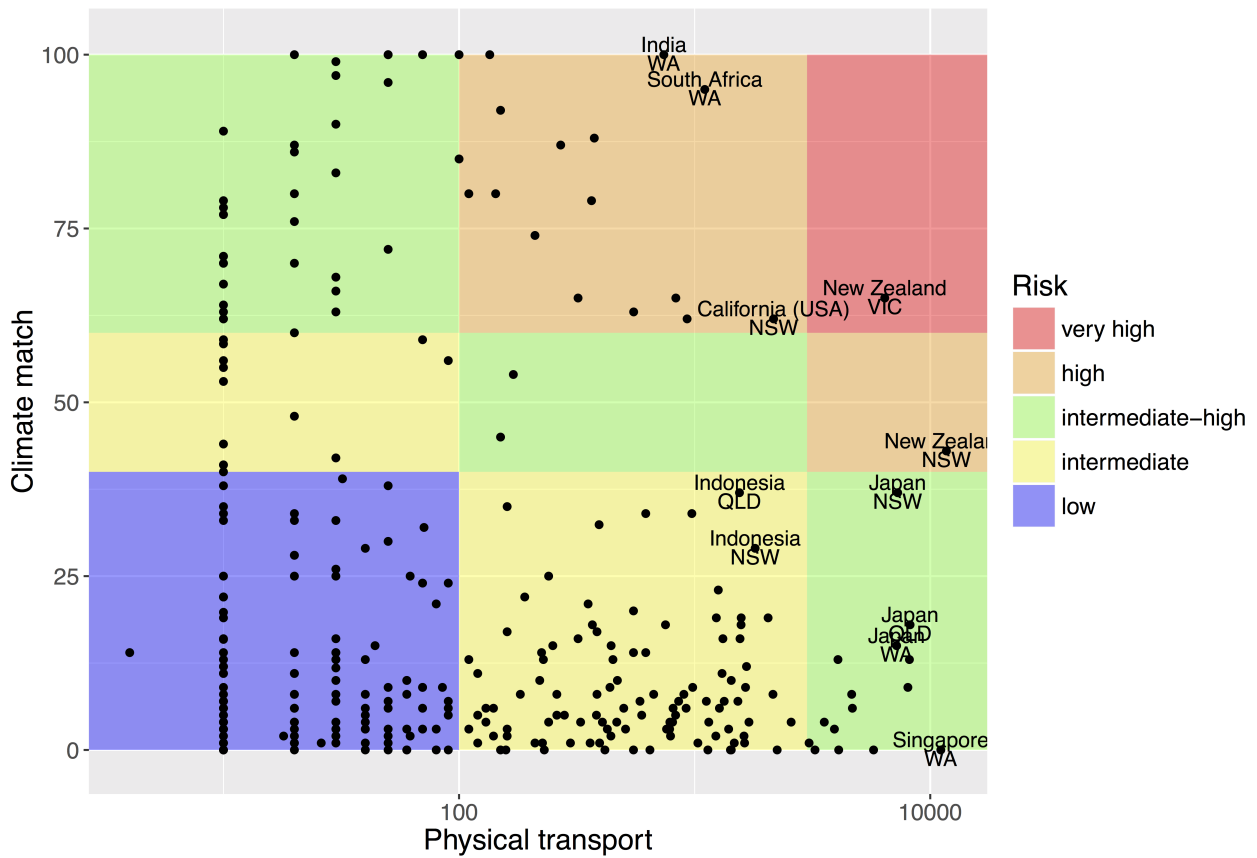


Fig. 2: Physical connectedness and environmental similarity between geopolitical regions overseas, and states within Australia, in 2012. Physical connectedness measured via the integrated physical network with $w_p = w_c = 1$, $w_s = 10$. Environmental match based on 90th percentile range bagging score from the source region to the Australian state. Note that risk region boundary choices are arbitrary. The web app (https://robertcope.shinyapps.io/risk_calc/) allows users to modify transport weights and boundaries, and identify individual points.

logistic regression models with minimum AIC (Table A1). When we instead included a term that decayed exponentially with geographic distance, this further improved model fit. Total transport volume over all regions in which the species exists; and the number of such regions, were also informative. An increase in each of these parameters resulted in an increased probability of the species being present in Australia (Table A2; Standardised odds ratios). Environmental match, when assessed as the maximum match across regions in which the species is present, was not informative - likely due to the fact that the vast majority of these species exist somewhere with a high environmental match to Australia (note that this is the maximum across regions of the 90th percentile score within each of those regions). However, when

environmental match was assessed as the median value over regions in which the species was present, it was informative, and was included in the final best model (Table A1). When tested via cross-validation (CV), this best model, including decayed geographic distance, total transport, number of regions, and median environmental match, produced an AUC value of 0.914 ± 0.032 . We used this best model to predict species not previously known to be in Australia that are likely to present as a high invasion risk (Table 4).

When logistic models were instead fitted with only transport volume, environmental match, and their interaction term, so as to correspond with the risk plot presented above (Figure 2), the resulting CV AUC values were 0.869 ± 0.038 when using maximum

per-species environmental match, or 0.882 ± 0.038 when using median per-species environmental match. A plot of the predicted probabilities from this model appears in the Appendix (Figure A2).

When we examined regions rather than species, distance to Australia again dominated, i.e., closer regions, geographically, were most likely to share high proportions of species with Australia (Figure 4). There was substantial variability in the proportion of species shared with regions, and only geographic distance (no other variables) were able to explain this variation using logistic regression models.

4. DISCUSSION

We have assessed a method of combining environmental similarity and physical connectedness for the prediction of biological invasion risk. Adapting the ‘range bagging’ method to geopolitical regions produces an effective and sensible measure of environmental similarity between these geopolitical regions. When combined with a physical transport network, we can produce an integrated picture of invasion risk globally, highlighting those pathways and hotspots that present the greatest risk and thus are key targets for biosecurity surveillance effort. In addition, using statistical models we can predict high-risk species of potential future concern, so that surveillance may be prioritised and detailed species-specific risk-analyses and action plans can be developed in preparation for potential incursions.

4.1 Global scale matching

At a global scale, the range bagging method provides valuable insights into the environmental similarity between geopolitical regions. We observed that matched environments primarily cluster along latitudinal lines (Figure A1), with clusters emerging broadly around the equator, the tropics, the arctic, etc. It is reassuring that these clusters emerge in a way that we would expect, based only on the range bagging environmental similarity between regions. However it is unsurprising that, for example, there are high matches between adjacent or geographically close geopolitical regions given that these regions will have geographically close environmental observations, which are often similar. We observed that the regions with the most environmental matches are typically those in tropical or equatorial areas, where there are a multitude of smaller geopolitical regions with which to connect. Large regions typically had

Table IV : Predicted high-risk species, i.e., species not currently present in Australia, but predicted to be based on the best logistic model. Lists all species with a predicted probability of presence in Australia ≥ 0.75 .

Species	maximum environmental match	median environmental match	total transport	distance (km)	number of states	predicted probability
<i>Xylosandrus compactus</i>	100.00	37.00	106214	2419	53	0.89
<i>Parabemisia myricae</i>	100.00	69.00	42343	2419	19	0.88
<i>Xyleborus ferrugineus</i>	100.00	43.00	34131	2419	79	0.88
<i>Agrotis segetum</i>	100.00	39.00	84627	3185	94	0.86
<i>Trichoplusia ni</i>	100.00	37.00	97844	3185	82	0.84
<i>Simoxylon conigerum</i>	100.00	67.50	84519	3185	22	0.84
<i>Aphis fabae</i>	100.00	43.00	76440	4220	118	0.81
<i>Pinnaaspis strachani</i>	100.00	41.00	39676	2419	48	0.80
<i>Aulacophora indica</i>	100.00	23.00	113879	2419	41	0.79
<i>Aleurocanthus woglumi</i>	100.00	27.00	75824	2419	53	0.78
<i>Aulacaspis tegalensis</i>	100.00	50.50	60898	2419	10	0.78
<i>Sesamia inferens</i>	100.00	23.00	124744	2419	31	0.78
<i>Pseudococcus jackbeardsleyi</i>	100.00	38.00	70542	2419	30	0.78

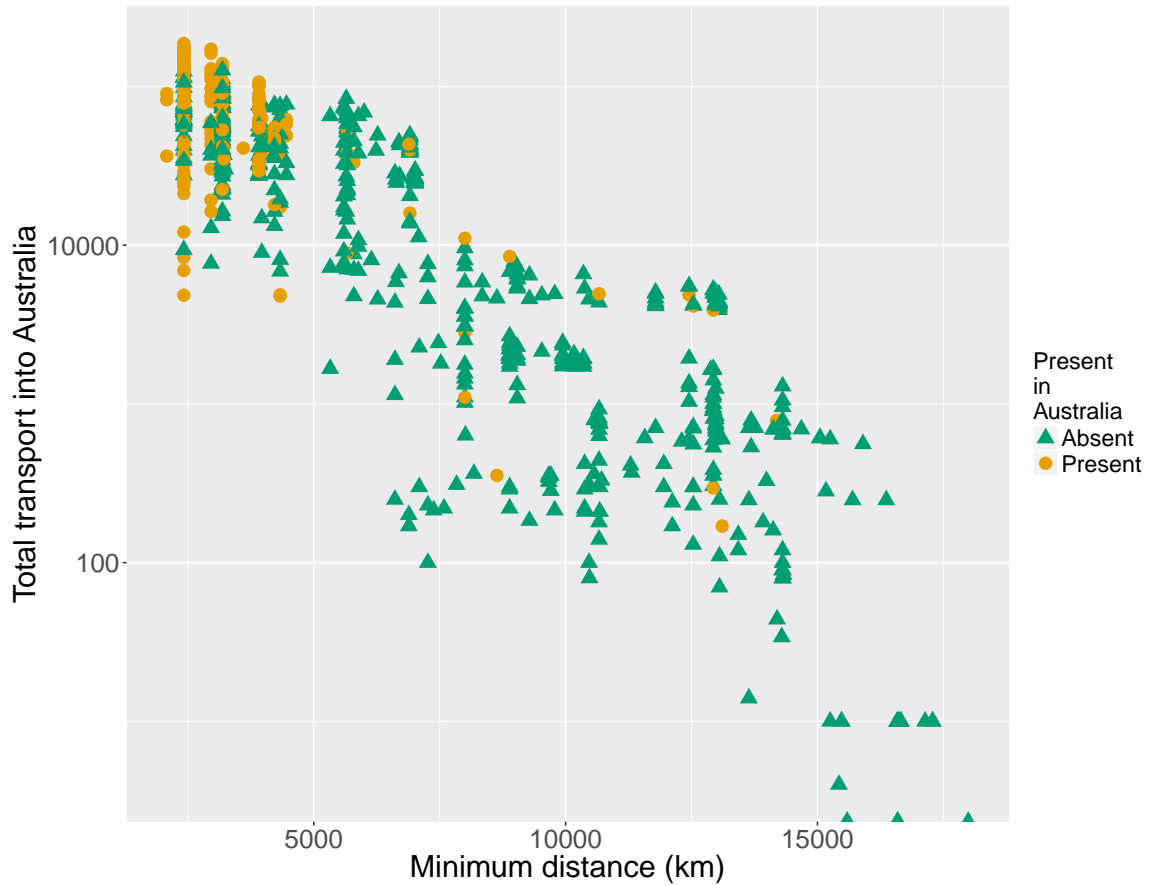


Fig. 3: For each insect pest species, the relationship between distance to Australia of its closest location, and total transport volume into Australia over all regions in which the species exists.

the most high-scoring environmental matches, as these regions contained more varied environmental conditions, e.g., India and Mexico. This does not necessarily mean that a species will be distributed throughout the whole country, but that species from many different locations may present risks to different parts of the country.

It should be noted that we chose a standard set of environmental variables with which to perform the analysis in this study, and this choice could reasonably be varied. In particular, the predictors used here are all abiotic factors, given that this analysis focused on the most general situation of transport and environment without considering specific species. Practitioners interested in specific species might also include biotic factors representing the presence of predation or competition, the availability of hosts for parasites or habitat and resources, etc., given that biotic factors are known to influence invasion success^(41,42,43). Any alternate predictors could be

built into the range bagging calculations with only minimal modifications. Further, if an analysis is for specific species for which range data are known, range bagging can be performed from the species range itself (rather than geopolitical regions), to candidate destinations anywhere in the world; an example of this appears for alien bird species in Dyer *et al.*⁽⁴⁴⁾.

4.2 Australian case study

Environmental matches into Australian states tend to follow the same patterns as are observed globally: the larger states, that are more tropical, i.e., Queensland and Western Australia, have the greatest numbers of potential source regions with high environmental matches. Smaller, more southern states (e.g., Tasmania) have fewer close environmental matches, with the closest match being New Zealand. However, the addition of transport data

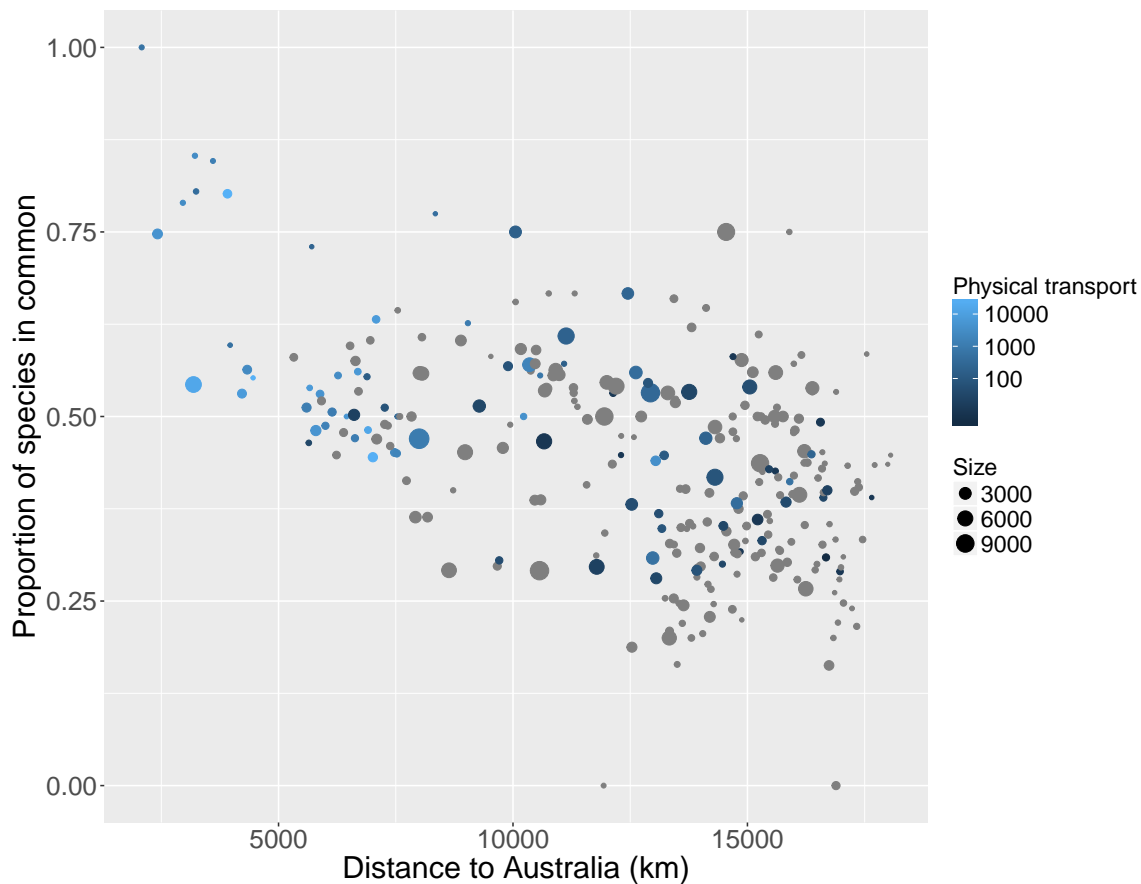


Fig. 4: For each geopolitical region, the relationship between distance to Australia and the proportion of pest insect species found in that region that are also found within Australia. Point size indicates the geographic area of the region (denoting the number of environmental grid points in the region), and colour indicates the volume of transport between the region and Australia. Grey points indicate the region had no direct transport to Australia.

suggests that the highest risks may be to those states with pathways having both high transport and high environmental matches, e.g., New Zealand to Victoria and New South Wales. The risk presented by particular pathways may be moderated by the source country in question, both in terms of the presence of species available to be transported, and biosecurity standards. In particular, given that New Zealand is known for high quality biosecurity standards, including a focus on pest-free export⁽²⁹⁾, it may present lower risks in practice than other nations despite being the source of high volumes of transport to Australia, though there have been recent interceptions of high-risk species transported from New Zealand to Australia, e.g., two interceptions of the Tomato-potato psyllid occurred between 2009–2012⁽⁴⁵⁾. It would be possible to modify this

framework to incorporate biosecurity within regions or along transport pathways, or to incorporate analyses of invasion debt (e.g., Rouget et al.⁽⁴⁶⁾) to discount regions when there are few new species that could be introduced from a given source.

In addition, when considering integrated transport, results may vary based on weightings applied to different transport types, as often ships and flights originate from different geopolitical regions⁽²²⁾ and different modes of transport may present different risks⁽⁴⁷⁾. It should also be emphasised that many nations have land borders, providing additional transport pathways (both natural and through human transport) for the import of alien species, and this would need to be incorporated when applying this framework elsewhere. Understanding the links between transport pressure (as a reasonable

proxy for the number of species that might be unintentionally transported between regions) and environmental similarity (and hence the capacity of those species to potentially survive) is critical for effective border control when biosecurity resources need to be prioritised⁽⁴⁸⁾, and the approach we present here should provide valuable insight to this end.

4.3 Validation

We used a dataset of 844 phytophagous insect pest species to determine how the environmental match and physical transport networks might inform invasion risk. The overall trend from these species, when assessed both at a species level and between regions, was that geographic distance had the most dominant effect, i.e., that more species were shared with closer regions geographically. Transport volumes were also informative, as was environmental match when considered as the median value across regions in which a given species exists. It is intuitive that geographic distance is important, because the closest regions are those with the most transport, and shorter journeys may lead to greater survival. Small geographic distances also indicate the possibility of natural or historical dispersal (e.g., by wind), which is typically confined to close biogeographic regions, compared to anthropogenic dispersal which may occur at much greater distances⁽⁴⁹⁾. This is a particularly remarkable aspect of this study: it has allowed us to examine aspects of biosecurity that may normally be taken as assumptions. While all of these factors do appear to play a role, the importance of geographic distance is perhaps surprising, along with the strong relationship between geographic distance and overall transport, i.e., that the highest transport volumes come from the closest regions.

It is important to note that the dataset does not directly indicate invasion events, rather only shared species, and that it is restricted to only species that exist within at least 5% of geographic regions. This means that there are no species in the dataset that exist only in Australia, every species that is in Australia in these data is necessarily present somewhere else, including species that may have existed in both regions for long time periods (e.g., predating western colonisation of Australia). The presence of a species both in Australia and elsewhere may also be indicative of an Australian species that is invasive elsewhere. Further, the 5% minimum limit means that these species are generally widespread

geographically, which explains why the vast majority of these species exist in at least one region that is environmentally similar to an Australian state: the regions that are similar environmentally are broadly distributed throughout the world (present in each continent except Antarctica). Hence, it is unsurprising that widespread species would be present in one of these regions. On a species level, it is intuitive that the maximum environmental match would provide useful data as to the possible potential for the species to survive in the new location, as it indicates that there is somewhere within the species range that matches well, and it is unlikely that a species survivability is substantially different in different parts of its range. However, this effect may be moderated by the fact that match between regions is not equivalent to match between species ranges. For example, a species may be present at one extreme of a region, when it may be the other extreme of the region creating the high environmental match. Median environmental match provides one possible solution to this, a higher median suggests that more regions within the species range match Australia, and thus it is more likely that the effect is related to the species survivability.

In addition, the phytophagous insect pest species dataset was compiled based on the best information available to its authors at the time, but may not be a perfect representation of the actual presence or absence of these species. For example, recent additions to the Australian Faunal Directory (<https://biodiversity.org.au/afd/home>), indicates that some species we list as not present in Australia, may be present (including *Xyleborus ferrugineus*, *Pinnaspis strachani* and *Aulacaspis tegalensis*, which were identified as high risk), although these records and distributional information are very sparse. As this dataset was used primarily for validation (and was not the main focus of this study) we used the original full dataset as-is; we suggest that practitioners analysing individual species should seek the most up to date information based on all available resources.

4.3.1 High-risk insect pest species list

We were able to produce a list of species not previously present in Australia, but with a high predicted probability of invasion risk. This type of approach allows for the creation of priority species watch lists, which may inform both biosecurity managers and assist in the development of post-

border biosecurity policy in the near future. However, this framework considers only transport pathways and environmental conditions, not the characteristics of the individual species, or the specifics of how these species might interact with environments or transport pathways (e.g., the presence of hosts, or transport related to particular commodities). As such, the list provides a preliminary indication of species that warrant further risk analyses incorporating these species and pathway specific details⁽⁵⁰⁾, along with an analysis of potential impacts (e.g., Table 5).

Of the top 7 (predicted probability > 0.80) ranked high invasion risk species (Table 4), only *Sinoxylon conigerum* and *Aphis fabae* have been intercepted on numerous occasions by Australian government agencies since 2004 (with 500+ and 113 interceptions, respectively). Of the remaining species, *Xyleborus ferrugineus* has been intercepted twice, *Xylosandrus compactus* and *Trichoplusia ni* have been intercepted once, and *Parabemisia myricae* and *Agrotis segetum* have not been intercepted (A. Broadley, personal communication, May 2017).

It should be emphasised that many of the 844 insect pest species considered showed some potential to be transported to and survive in Australia, and most would present substantial risk to Australian agriculture were they to become established, so border biosecurity efforts should continue to aim to prevent all incursions of any species. Lists of high-risk species may instead be used to prioritise post-border biosecurity efforts and assist with decision making related to additional surveillance, preventative strategies, or contingency planning⁽⁵¹⁾.

4.4 Conclusions

In this study, we have demonstrated the combination of environmental similarity with physical transport on a geopolitical level to determine biological invasion risk. This approach is valuable for four key reasons. First, trade occurs at geopolitical scales, between nations or states, and thus assessing environmental match between these geopolitical regions allows direct assessment of invasion risk associated with pathways and hotspots at the same scale. Second, it is scalable, so the same technique could be applied to compare countries, states or other sub-national units, or smaller regions again. Alternatively, more diverse geographic regions, e.g., national parks or environmental sanctuaries, or agricultural regions, could be compared as source

or destination regions. Third, it is species agnostic, and thus avoids assumptions around species abilities to survive in new ranges; rather it deals directly with environmental similarity between geopolitical regions. Finally, the approach can be used both to specify the risk presented by specific regions, and to construct lists of high-risk species (i.e., species not currently present, but with a high probability of being so), that can then be further analysed considering species-specific risk factors. This means that the approach can inform both border and post-border biosecurity efforts. Using Australia as a case study, we found that a variety of regions in diverse locations throughout the world were potential sources of invasive species, due to high environmental matches. We provide an interactive web app (https://robertcope.shinyapps.io/risk_calc/) so that users may observe how risk changes with weightings and identify individual pathways. We determined that, for 844 phytophagous insect species, geographic distance was an important predictor, in addition to transport and environmental similarity. Given the growth in global trade and transport over recent decades, and the associated risk of biological invasions, it is critical that tools and technologies to quantitatively assess risk and inform biosecurity surveillance and management be developed. We believe the combination of environmental matching via range bagging with integrated physical transport connectedness achieves this goal.

ACKNOWLEDGEMENTS

We would like to thank Adam Broadley from the Australian Government Department of Agriculture and Water Resources for providing data on recent species interceptions, and for helpful comments on the manuscript. This study was supported by an ARC Discovery grant (DP140102319) to P.C. and J.V.R., and Future Fellowships to P.C. (FT0991420) and J.V.R. (FT130100254). We also thank Pablo Garcia Diaz and Thomas Prowse for helpful comments on the manuscript.

REFERENCES

1. Meyerson LA, Mooney HA. Invasive alien species in an era of globalization. *Front Ecol Environ*. 2007;4:199–208.
2. Hulme PE, Bacher S, Kenis M, Klotz S, Kühn I, Minchin D, et al. Grasping at the routes of biological invasions: a framework for integrating pathways into policy. *Journal of Applied Ecology*. 2008;45(2):403–

Table V : Six phytophagous insect pest species not established in Australia, predicted to have high invasion risk.

Species	Range	Closest presence to Australia	Transport potential	Impacts	Interceptions 2004–2017
<i>Xylosandrus compactus</i>	Native to Asia, introduced to parts of America, Africa, and some oceanic countries ⁽⁵²⁾ .	Fiji, Papua New Guinea, Samoa and the Solomon islands ⁽⁵²⁾ .	Imported plant parts like bark, stems, seedlings and wood packaging ⁽⁵²⁾ . Strong wind dispersal ability ^(53,52) .	Economic damage to tea, coffee, cocoa and avocado crops ^(54,55,52) .	1
<i>Parabemisia myricae</i>	Asian origin, also present in parts of Europe, Africa, America and parts of Oceania (Papua New Guinea) ⁽⁵²⁾ .	Papua New Guinea ⁽⁵²⁾ .	Plant materials of host plants (particularly citrus) should be considered a quarantine risk to Australia ^(56,52) . Infested <i>Gar-denita</i> was the likely vector of distribution in California ⁽⁵⁷⁾ .	In California it has been known to cause serious defoliation of lemons and oranges ⁽⁵⁷⁾ . Also a known citrus pest in Turkey, Israel and Algeria ^(58,59,60) .	0
<i>Xyleborus ferrugineus</i>	Tropical regions worldwide (including Asia, tropical Africa, North America) ⁽⁵⁶⁾ .	Indonesia, Fiji ⁽⁵⁶⁾ .	Imported plant parts (bark, stems, seedlings) ^(56,52) .	Attacks a range of plants including Mango, Avocado, and Cashew ⁽⁵⁶⁾ .	2
<i>Agrotis segetum</i>	Widely spread globally (Asia, Africa, and Europe), but currently absent from America and Australia ⁽⁵²⁾ .	Indonesia ⁽⁵²⁾ .	Plant and soil materials, potatoes and flowers ⁽⁶¹⁾ . Wind assisted flight migration ^(62,61) .	Attacks cultivated plants of approximately 15 families, and has been known to reduce crop production in potatoes by 33% in Nepal, and 34% of beetroot crops were damaged when larvae were at a density of 14 larvae/m ² in the UK ^(63,64,52) .	0
<i>Trichoplusia ni</i>	Native to North America and introduced to Europe ⁽⁶⁵⁾ . CABI distribution maps also indicate presence in Central and southern America, Asia and Africa ⁽⁵²⁾ .	Indonesia ⁽⁵²⁾ .	Vegetables (> 160 host species, but favours cultivated crucifers), and ornamental cut flowers ^(66,67) .	Reductions in crop yield and quality of produce ⁽⁵²⁾ .	1
<i>Simoxylon conigerum</i>	First described from India and is widespread in Oriental and Afrotropical regions and also in Yemen and China ⁽⁶⁸⁾ . Introduced to USA, South America and Europe ⁽⁵²⁾ .	Indonesia, American Samoa, Niue ⁽⁵²⁾ .	Wood packaging and dunnage via shipping networks ⁽⁵²⁾ .	Potential to kill tress and reduce timber market value due to loss in timber weight as a result of boring ^(52,69) .	500+
<i>Aphis fabae</i>	Native to Europe, and introduced to parts of Asia, Africa, and America ⁽⁵²⁾ .	Malaysia and the Philippines ⁽⁵²⁾ .	Plant material of crops (i.e. infested bean plants imported to Iraq have been reported) ⁽⁵²⁾ .	Crop yield loss due to direct feeding damage, and indirect impacts on plants via disease transmission ^(70,52) .	113

414. Available from: <http://dx.doi.org/10.1111/j.1365-2664.2007.01442.x>.
3. Hulme PE. Trade, transport and trouble: managing invasive species pathways in an era of globalization. *Journal of Applied Ecology*. 2009;46:10–18.
 4. Richardson DM, Ricciardi A. Misleading criticisms of invasion science: a field guide. *Diversity and Distributions*. 2013;19:1461–1467.
 5. Simberloff D, Martin JL, Genovesi P, Maris V, Wardle DA, Aronson J, et al. Impacts of biological invasions: what's what and the way forward. *Trends in Ecology & Evolution*. 2013;28(1):58–66. Available from: <http://www.sciencedirect.com/science/article/pii/S0169534712001747>.
 6. Pimentel D, Zuniga R, Morrison D. Update on the environmental and economic costs associated with alien-invasive species in the United States. *Ecological Economics*. 2005;52:273–288.
 7. Haack RA, Hérard F, Sun J, Turgeon JJ. Managing invasive populations of Asian longhorned beetle and citrus longhorned beetle: a worldwide perspective. *Annual review of entomology*. 2009;55(1):521.
 8. Nowak DJ, Pasek JE, Sequeira RA, Crane DE, Mastro VC. Potential effect of Anoplophora glabripennis (Coleoptera: Cerambycidae) on urban trees in the United States. *Journal of Economic Entomology*. 2001;94(1):116–122.
 9. Department of Environment and Heritage. Threat abatement plan to reduce the impacts of tramp ants on biodiversity in Australia and its territories. Australian Government, Canberra; 2006.
 10. Lodge DM, Williams S, MacIsaac HJ, Hayes KR, Leung B, Reichard S, et al. Biological invasions: recommendations for US policy and management. *Ecological Applications*. 2006;16(6):2035–2054.
 11. Pimentel D. Biological invasions: economic and environmental costs of alien plant, animal, and microbe species. CRC Press; 2011.
 12. Waage J, Mumford J. Agricultural biosecurity. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*. 2008;363(1492):863–876.
 13. Barrett S, Whittle P, Mengersen K, Stoklosa R. Biosecurity threats: the design of surveillance systems, based on power and risk. *Environmental and ecological statistics*. 2010;17(4):503–519.
 14. Rout TM, Moore JL, Possingham HP, McCarthy MA. Allocating biosecurity resources between preventing, detecting, and eradicating island invasions. *Ecological Economics*. 2011;71:54–62.
 15. Lodge DM, Simonin PW, Burgiel SW, Keller RP, Bossenbroek JM, Jerde CL, et al. Risk Analysis and Bioeconomics of Invasive Species to Inform Policy and Management. *Annual Review of Environment and Resources*. 2016;41(1).
 16. Faulkner KT, Robertson MP, Rouget M, Wilson JR. Understanding and managing the introduction pathways of alien taxa: South Africa as a case study. *Biological Invasions*. 2016;18:73–87.
 17. Colunga-Garcia M, Haack RA, Magarey RD, Borchert DM. Understanding trade pathways to target biosecurity surveillance. *NeoBiota*. 2013;18:103–118.
 18. Blackburn TM, Pysek P, Bacher S, Carlton JT, Duncan RP, Jarosik V, Wilson JR, Richardson DM. A proposed unified framework for biological invasions. *Trends in Ecology & Evolution*. 2011;26:333–339.
 19. Williamson M. Biological invasions. Springer Science and Business Media; 1996.
 20. Bacon SJ, Bacher S, Aebi A. Gaps in border controls are related to quarantine alien insect invasions in Europe. *PLoS ONE*. 2012;7:e47689.
 21. Bacon SJ, Aebi A, Calanca P, Bacher S. Quarantine arthropod invasions in Europe: the role of climate, hosts and propagule pressure. *Diversity and Distributions*. 2014;20:84–94.
 22. Cope RC, Ross JV, Wittmann TA, Prowse TAA, Cassey P. Integrative Analysis of the Physical Transport Network into Australia. *PLoS ONE*. 2016 02;11(2):1–16. Available from: <http://dx.doi.org/10.1371/journal.pone.0148831>.
 23. Tatem AJ, Hay SI. Climatic similarity and biological exchange in the worldwide airline transportation network. *Proceedings of the Royal Society of London B: Biological Sciences*. 2007;274(1617):1489–1496. Available from: <http://rsos.royalsocietypublishing.org/content/274/1617/1489>.
 24. Keller RP, Drake JM, Drew MB, Lodge DM. Linking environmental conditions and ship movements to estimate invasive species transport across the global shipping network. *Diversity and Distributions*. 2011;17:93–102.
 25. Seebens H, Gastner MT, Blasius B. The risk of marine bioinvasion caused by global shipping. *Ecology Letters*. 2013;16:782–790.
 26. Xu J, Wickramaratne TL, Chawla NV, Grey EK, Steinhäuser K, Keller R, et al. Improving management of aquatic invasions by integrating shipping network, ecological, and environmental data: data mining for social good. *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*. 2014;p. 1699–1708.
 27. Drake JM. Range bagging: a new method for ecological niche modelling from presence-only data. *J R Soc Interface*. 2015;12:<http://dx.doi.org/10.1098/rsif.2015.0086>.
 28. Cope RC, Ross JV, Drake JM, Wittmann TA, Dyer EE, Blackburn TM, West P, Cassey P. Assessment of range bagging for the prediction of potential invasive ranges. *Biological Invasions*. Submitted 2017;.
 29. Meyerson LA, Reaser JK. Biosecurity: Moving toward a Comprehensive Approach. *BioScience*. 2002;52(7):593–600. Available from: <http://bioscience.oxfordjournals.org/content/52/7/593.short>.
 30. Wilson D, Weber L. Surveillance, risk and preemption on the Australian border. *Surveillance & Society*. 2002;5:124–141.
 31. Watts MJ, Worner S. Estimating the risk of insect species invasion: Kohonen self-organising maps versus k-means clustering. *Ecological Modelling*. 2009;220(6):821–829.
 32. Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*. 2005;25:1965–1978.
 33. ESRI. World administrative divisions layer package; 2012.
 34. Newman MEJ. Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*. 2006;103(23):8577–8582. Available from: <http://www.pnas.org/content/103/23/8577.abstract>.
 35. RStudio, Inc. shiny: Web Application Framework for R; 2014. R package version 0.10.2.1. Available from: <http://CRAN.R-project.org/package=shiny>.
 36. Gevrey M, Worner S. Prediction of global distribution of insect pest species in relation to climate by using an ecological informatics method. *Journal of economic entomology*. 2006;99(3):979–986.
 37. Worner S, Gevrey M. Modelling global insect pest species assemblages to determine risk of invasion. *Journal of Applied Ecology*. 2006;43(5):858–867.
 38. Lobo JM, Jiménez-Valverde A, Real R. AUC: a misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography*.

- 2008;17(2):145–151. Available from: <http://dx.doi.org/10.1111/j.1466-8238.2007.00358.x>.
39. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria; 2015. Available from: <https://www.R-project.org/>.
 40. Wickham H. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York; 2009. Available from: <http://had.co.nz/ggplot2/book>.
 41. Kennedy TA, Naeem S, Howe KM, Knops JM, Tilman D, Reich P. Biodiversity as a barrier to ecological invasion. *Nature*. 2002;417(6889):636–638.
 42. Levine JM, Adler PB, Yelenik SG. A meta-analysis of biotic resistance to exotic plant invasions. *Ecology letters*. 2004;7(10):975–989.
 43. Theoharides KA, Dukes JS. Plant invasion across space and time: factors affecting nonindigenous species success during four stages of invasion. *New Phytologist*. 2007;176(2):256–273.
 44. Dyer EE, Franks V, Cassey P, Collen B, Cope RC, Jones KE, et al. A global analysis of the determinants of alien geographical range size in birds. *Global Ecology and Biogeography*. 2016;25(11):1346–1355.
 45. Department of Agriculture, Fisheries and Forestry. Draft report for the review of import conditions for fresh potatoes for processing from New Zealand. Canberra: Australian Government; 2012.
 46. Rouget M, Robertson MP, Wilson JR, Hui C, Essl F, Renteria JL, et al. Invasion debt: quantifying future biological invasions. *Diversity and Distributions*. 2016;22:445–456.
 47. Pyšek P, Jarošík V, Pergl J. Alien Plants Introduced by Different Pathways Differ in Invasion Success: Unintentional Introductions as a Threat to Natural Areas. *PLoS ONE*. 2011 09;6(9):e24890. Available from: <http://dx.doi.org/10.1371/journal.pone.0024890>.
 48. Faulkner KT, Robertson MP, Rouget M, Wilson JR. Border control for stowaway alien species should be prioritised based on variations in establishment debt. *Journal of environmental management*. 2016;180:301–309.
 49. Capinha C, Essl F, Seebens H, Moser D, Pereira HM. The dispersal of alien species redefines biogeography in the Anthropocene. *Science*. 2015;348(6240):1248–1251. Available from: <http://science.sciencemag.org/content/348/6240/1248>.
 50. Wilson JR, Panetta F, Lindgren C. Detecting and responding to alien plant incursions. Cambridge University Press; 2017.
 51. Faulkner KT, Robertson MP, Rouget M, Wilson JR. A simple, rapid methodology for developing invasive species watch lists. *Biological Conservation*. 2014;179:25–32.
 52. CABI. Crop Protection Compendium; 2016.
 53. Entwistle PF. Pests of Cocoa. London: Longman; 1972.
 54. Greco E, Wright M. Ecology, biology, and management of *Xylosandrus compactus* (Coleoptera: Curculionidae: Scolytinae) with emphasis on coffee in Hawaii. *Journal of Integrated Pest Management*. 2015;6(1):7.
 55. Kaneko T. Biology of some scolytid ambrosia beetles attacking tea plants. I. Growth and development of two species of scolytid beetles reared on sterilized tea plants. *J Appl Entomol Zool*. 1965;9:211–216.
 56. Plantwise. Plantwise Technical Factsheet, bayberry whitefly (*Parabemisia myricae*); 2016.
 57. Rose M, Debach P. Biological control of *Parabemisia Myricae* (Kuwana) (Homoptera: Aleyrodidae) in California. *Israel Journal of Entomology*. 1992;p. 73–95.
 58. Swirski E, Blumberg D, Wysoki M, Izhar Y. Biological control of the Japanese bayberry whitefly *Parabemisia myricae* (Kuwana) (Homoptera: Aleyrodidae) in Israel. *Israel Journal of Entomology*. 1987;21:11–18.
 59. Berkani A, Dridi B. *Parabemisia myricae* Kuwana (Homoptera: Aleyrodidae), a new white fly attacking citrus in Algeria. *Fruits (France)*; 1992.
 60. Sengonca C, Uygun N, Kersing U, Ulusoy MR. Successful colonization of *Eretmocerus debachi* (Hym.: Aphelinidae) in the eastern mediterranean citrus region of Turkey. *Entomophaga*. 1993;38:383–390.
 61. Plant Health Australia. Threat specific contingency plan: Turnip moth 2011; 2011.
 62. Fox KJ. Trans-oceanic dispersal of insects to New Zealand. *New Zealand Entomologist*. 1973;5(3):240–243.
 63. Neupane FP. Assessment of damage due to cutworms in potato. *Nepalese Journal of Agriculture*. 1971;6:153–155.
 64. Kay RH, Wheatley GA. Some observations on the biology of the turnip moth (*Agrotis segetum*) relevant to its control with insecticides. In: British Crop Protection Council: Proceedings of the 1979 British Crop Protection Conference-Pests and Diseases (10th British Insecticide and Fungicide Conference) 19th to 22nd November 1979. Brighton, England.: British Crop Protection Council; 1980.
 65. Cox G. Alien Species and Evolution: The Evolutionary Ecology of Exotic Plants, Animals, Microbes, and Interacting Native Species. Washington: Island Press; 2004.
 66. Sutherland DWS, Greene GL. Cultivated and wild host plants. Suppression and management of cabbage looper populations. USDA Technical Bulletin. 1984;p. 1–13.
 67. Poe SL, Workman RB. Potential for dispersal by winter shipments of vegetables, ornamental plants and cut flowers; 1984.
 68. Savoldelli S, Regalin R. Infestation of wood pallets by *Sinoxylon unidentatum* (Fabricius) (Coleoptera Bostrichidae) in Italy. *Bollettino di Zoologia Agraria e di Bachicoltura*. 2009;41:235–238.
 69. Plantwise. Plantwise Technical Factsheet, conifer auger beetle (*Sinoxylon conigerum*); 2016.
 70. Shannag H, Ababneh J. Biometry and responses of faba bean varieties to black bean aphid, *Aphis fabae* Scopoli. *Am-Eurasia J Agric Environ Sci*. 2007;2:328–334.
 71. Burnham KP, Anderson DR. Model selection and multimodel inference: a practical information-theoretic approach. Springer Science and Business Media; 2002.

Table A1.: Summary of logistic model AIC, when predicting the presence or absence of a species within Australia. distDecay is $\exp(-k \times \text{minDist})$ with k parameter chosen to minimise AIC, resulting in $k = 2.38 \times 10^{-7}$. All other parameter combinations result in higher AIC values. Δ AIC and Akaike weights for each model⁽⁷¹⁾ are also presented.

model	AIC	Δ AIC	Akaike weight
medEnv + sumTransport + distDecay + noStates	590	0	1.0
maxEnv + sumTransport + minDist + noStates	644	54	1.8×10^{-12}
maxEnv + sumTransport + minDist	663	73	1.4×10^{-16}
sumTransport + minDist + noStates	643	53	3.1×10^{-12}
maxEnv + minDist + noStates	655	65	7.7×10^{-15}
maxEnv + sumTransport + noStates	711	121	5.3×10^{-27}

Table A2.: Coefficients and standardised odds ratios for each parameter in the best performing logistic model predicting presence or absence of a species within Australia. Odds ratios were standardised by first transforming the data to have mean zero and standard deviation 1, with the odds ratios then calculated by performing logistic regression on these transformed data. Note that an increase in distDecay corresponds to a closer geographic region.

parameter	coefficient (not standardised)	standardised odds ratio	confidence interval for odds ratio
(Intercept)	-6.87	0.2532	(0.1938, 0.3308)
sumTransport	1.14×10^{-05}	1.6383	(1.1349, 2.3648)
distDecay	9.14	6.3642	(4.3017, 9.4156)
noStates	1.92×10^{-02}	1.8219	(1.3922, 2.3842)
medEnv	4.21×10^{-02}	1.8190	(1.4672, 2.2552)

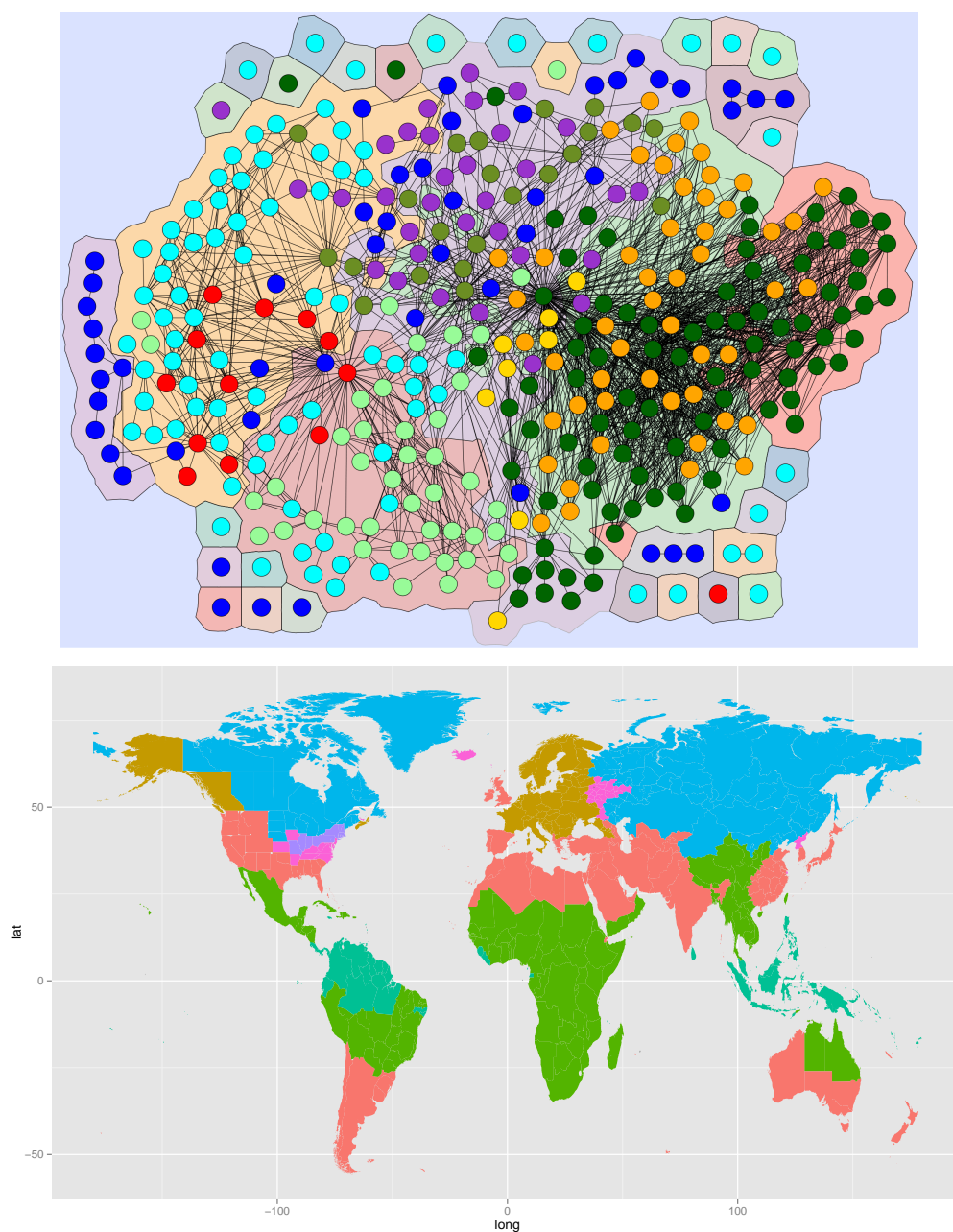


Fig. A1: (a) Environmental match between geopolitical regions, based on 90th percentile match scores. Nodes represent geopolitical regions, colour coded broadly by country/continent. Edges indicate that, in either direction, the 90th percentile range bagging environmental match between the two regions is at least 80%. Background colour denotes modularity-based community detection on this graph: the resulting communities are shown on the map (b), each colour represents an automatically detected community.

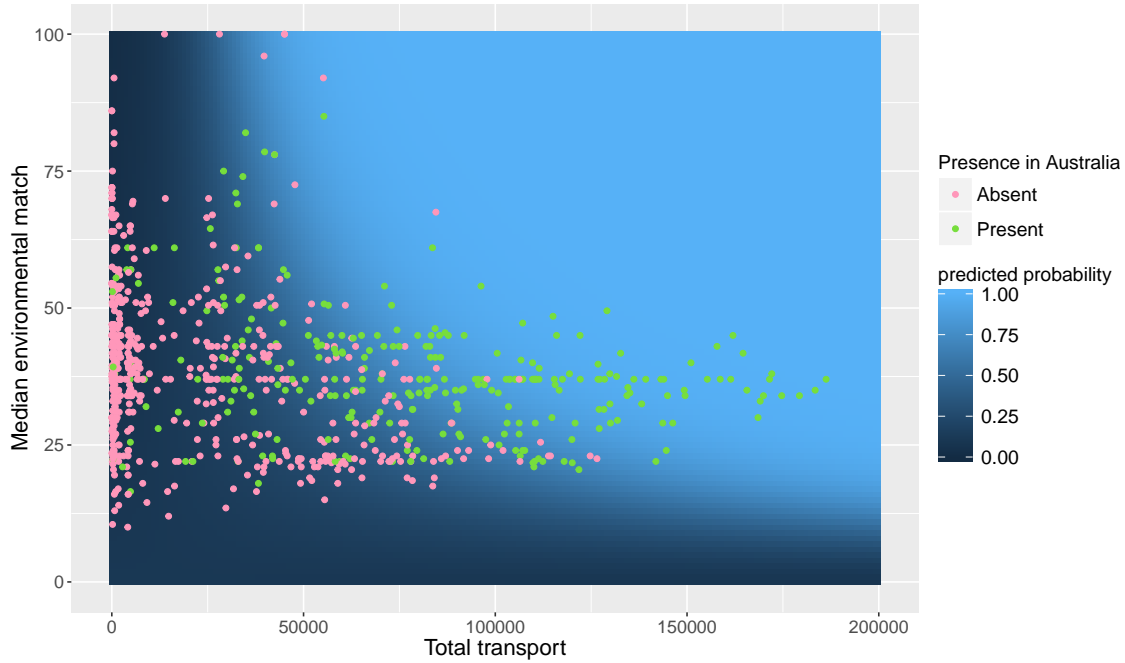


Fig. A2: Predicted probability of presence in Australia for phytophagous insect species based solely on total transport from countries in which the species is present, and median region-based environmental match. Points show actual presence or absences in Australia.

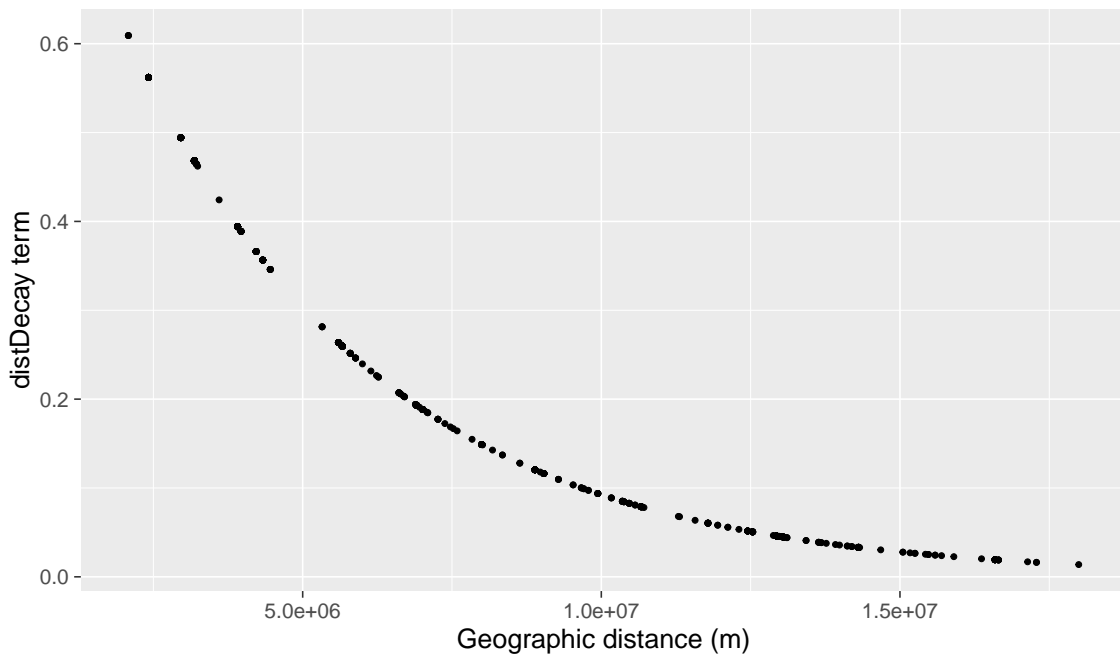


Fig. A3: Relationship between geographic distance and distDecay regression parameter, i.e., $\text{distDecay} = \exp(-k \times \text{distance})$, with $k = 2.38 \times 10^{-7}$. Each point represents a country.

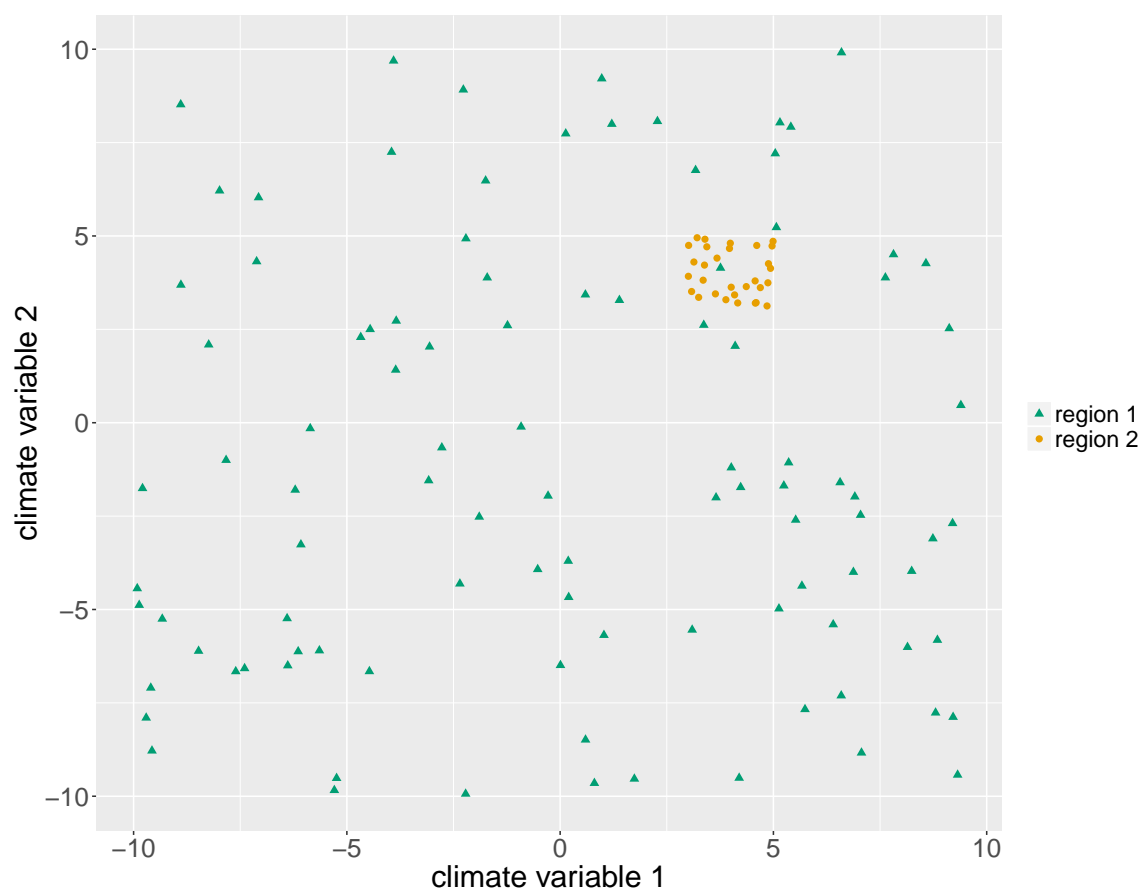


Fig. A4: Example (synthetic) climate observations from two regions, in two climate variables. This demonstrates how range bagging between two regions is asymmetric: every point in region two would be contained within a marginal niche generated from the points in region one, whereas only one of the points from region one would be within a marginal niche generated from the points in region two. Note that this is only one pair of climate variables; marginal niches in other variables may differ, and the range bagging score for a point is determined from many pairs of climate variables.