

Rapid spatial risk modelling for invasion management under uncertainty

J.G Froese^a, G.S. Hamilton^a

^a *Queensland University of Technology, Brisbane, QLD 4001, Australia*

Email: jens.froese@qut.edu.au

Abstract: Across the globe, invasive alien plant species are a major threat to biodiversity and agricultural industries. Invasive plant spread is an ongoing process that plays out at different spatiotemporal scales. When an invasive ‘weed’ emerges in a new country, region or landscape, there is often a distinct lag phase during which population increase and invasive spread is slow. Rapid response by biosecurity managers to new incursions may bring the greatest opportunities with regard to facilitating eradication or effective containment as well as minimizing long-term control costs and negative impacts. An important consideration for allocating finite management resources is the spatial distribution of invasion risk. However, hard data is rarely available to assess risk, and decisions are consequently made under considerable uncertainty.

We developed a flexible, spatially-explicit methodology for rapidly modelling invasion risk using Bayesian networks and Geographic Information Systems (GIS). In this well-established modelling framework, the key processes of invasion and the major risk variables affecting these processes are represented as a Bayesian network. GIS analysis is used to link each risk variable contained in the model to one or several spatial data proxies and to map model predictions. Here, we present a simplified implementation that can be used rapidly and under limited knowledge, using the case study of Mexican bean tree (*Cecropia* spp.) in North Queensland. We modelled two aspects of invasion risk: (a) the potential, time-independent environmental suitability of a site for plant establishment and persistence, and (b) the actual susceptibility of suitable sites to propagule introduction from source infestations recorded during annual ‘detection periods’. We derived risk variables and spatial data proxies via desktop research and used simple mathematical functions to model the interplay between risk variables, invasion processes (establishment, persistence and propagule pressure), and invasion risk (suitability and susceptibility). We also evaluated the ability of our model to accurately predict the observed spatial progression of the *Cecropia* spp. invasion in North Queensland by validating mapped annual susceptibility against subsequently observed infestations.

The majority of North Queensland’s Wet Tropics was predicted to be potentially highly or moderately suitable for *Cecropia* spp. invasion. Actual susceptibility to invasion in a given fruiting season is more restricted due to dispersal constraints from source infestations. In each evaluated detection period, modelled susceptibility was a good predictor of actual *Cecropia* spp. detections in subsequent years. Annual susceptibility maps may be used for allocating surveillance and containment resources around existing infestations. Low-frequency long-distance dispersal or human-mediated translocations were not captured in the model. Further automation of risk factor / spatial proxy selection and data pre-processing may make our methodology for rapidly modelling invasion risk accessible to biosecurity agencies that typically have significant field knowledge and GIS capacities but may have limited expertise in applying complex modelling methods.

Keywords: *Emerging weeds, incursion response, Cecropia spp., Bayesian network, suitability, susceptibility*

1. INTRODUCTION

Across the globe, invasive alien plant species are a major threat to biodiversity and agricultural industries (Wilson et al. 2017). Invasive plant spread is an ongoing process that plays out at different spatiotemporal scales. When an invasive ‘weed’ emerges in a new country, region or landscape, there is often a distinct lag phase during which population increase and invasive spread is slow. This initial post-entry ‘incursion’ stage is followed by a phase of rapid ‘expansion’, and ultimately widespread ‘dominance’ when most negative impacts are fully realized (Wilson et al. 2017). Rapid response by biosecurity managers during the ‘incursion’ stage may bring the greatest opportunities with regard to facilitating eradication or effective containment as well as minimizing long-term control costs and negative impacts (Adams et al. 2015; Parry et al. 2013; Wilson et al. 2017). An important consideration for allocating finite management resources is the spatial distribution of invasion risk (Fletcher and Westcott 2013). Such information may be used to monitor and manage high-risk dispersal pathways, or to better delimit the total extent of the weed incursion (Panetta and Lawes 2005; Wilson et al. 2017). However, hard data is rarely available to assess risk, and decisions are consequently made under considerable uncertainty (Parry et al. 2013). Here, we developed a flexible, spatially-explicit methodology that can be used by biosecurity managers to rapidly model invasion risk for emerging weed species, and applied it to a case study of Mexican bean tree (*Cecropia* spp.) in North Queensland.

The persistence of any spatially distributed population essentially depends on two processes: local growth and connectivity between populations (Hastings 2014). These basic processes also underlie the spatial progression of biological invasions, and numerous theoretical discussions and applied models of invasive spread have focused on further refining, measuring and simulating them (Savage and Renton 2014; O'Reilly-Nugent et al. 2016). We follow a conceptual framework of plant invasions proposed by Smith et al. (2012) that considers two aspects of spatial risk: (a) the potential, time-independent environmental suitability of a site, which is driven by the local processes of plant establishment (i.e. germination and recruitment) and persistence (i.e. subsequent growth, reproduction and survival); and (b) the actual susceptibility of suitable sites within a certain time period, which is driven by the process of propagule introduction from previously infested sites.

2. METHODS

Mexican bean tree (and other closely related species of the genus *Cecropia*) is an archetypal neotropical pioneer species. Fast growing and highly fecund, *Cecropia* spp. have the capacity to invade and dominate early successional habitats in tropical and subtropical rainforests as well as other disturbed sites (Csurhes 2016). Mexican bean tree was initially introduced to Australia as an ornamental garden plant by private collectors. Naturalized specimens have been recorded at various locations in North Queensland since 2008. All *Cecropia* spp. detected in North Queensland are restricted invasive plants under the Queensland *Biosecurity Act 2014* (Department of Agriculture and Fisheries 2016).

To model suitability and susceptibility for *Cecropia* spp. invasion, we adopted a spatially-explicit methodology using Bayesian networks (BNs) and Geographic Information Systems (GIS). BN models are graphical influence diagrams, in which explanatory variables (i.e. risk factors) are linked by causal relationships to response variables. Causal relationships and interactions between several explanatory variables are quantified in conditional probability tables (CPT) behind each response variable. Because CPTs are formulated as probability distributions, BNs are often used to assess risk or evaluate decisions under uncertainty (Fenton and Neil 2013). Explanatory variables can also be linked to spatial data proxies, whose attributes are reclassified to match one of the mutually exclusive states each variable can take. Based on the state of each variable at a given study area pixel, the response at that pixel can then be modelled, and mapped. However, adequately defining model variables, representing them accurately by spatial proxies, and correctly calibrating CPTs can be complex. Common approaches to parameterizing Bayesian network models (Boets et al. 2015) are time-consuming and not always feasible (e.g. where detailed knowledge or data does not exist).

Here, we present a methodology that can be used rapidly and under limited knowledge. We derived explanatory variables and spatial data proxies via desktop research and used simple mathematical functions to model the interplay between risk variables, invasion processes (establishment, persistence and propagule pressure), and invasion risk (suitability and susceptibility). For this purpose, all model variables were converted into ‘ranked nodes’ and their states assigned with numerical values (for discrete explanatory variables) or intervals (for continuous response variables) on a scale from 0 to 100 (Fenton et al. 2007). CPTs were then computed as truncated normal distributions (TNormal) centered on the weighted mean of all influential explanatory variables (Fenton et al. 2007). The weighting of each explanatory variable or a specific state could be adjusted by altering the TNormal function or the state’s assigned numerical value. Uncertainty could be adjusted by altering the standard deviation assumed by the TNormal function (we used SD = 10, i.e. 1/10th of the total value range in explanatory variables). Finally, we computed suitability and susceptibility indices as model

expected values. The BN model of *Cecropia* spp. suitability and susceptibility is shown in Figure 1. Definitions of explanatory variables (risk factors), justifications for including them in the model, weightings applied, and methods for representing them by spatial data proxies are shown in Table 1.

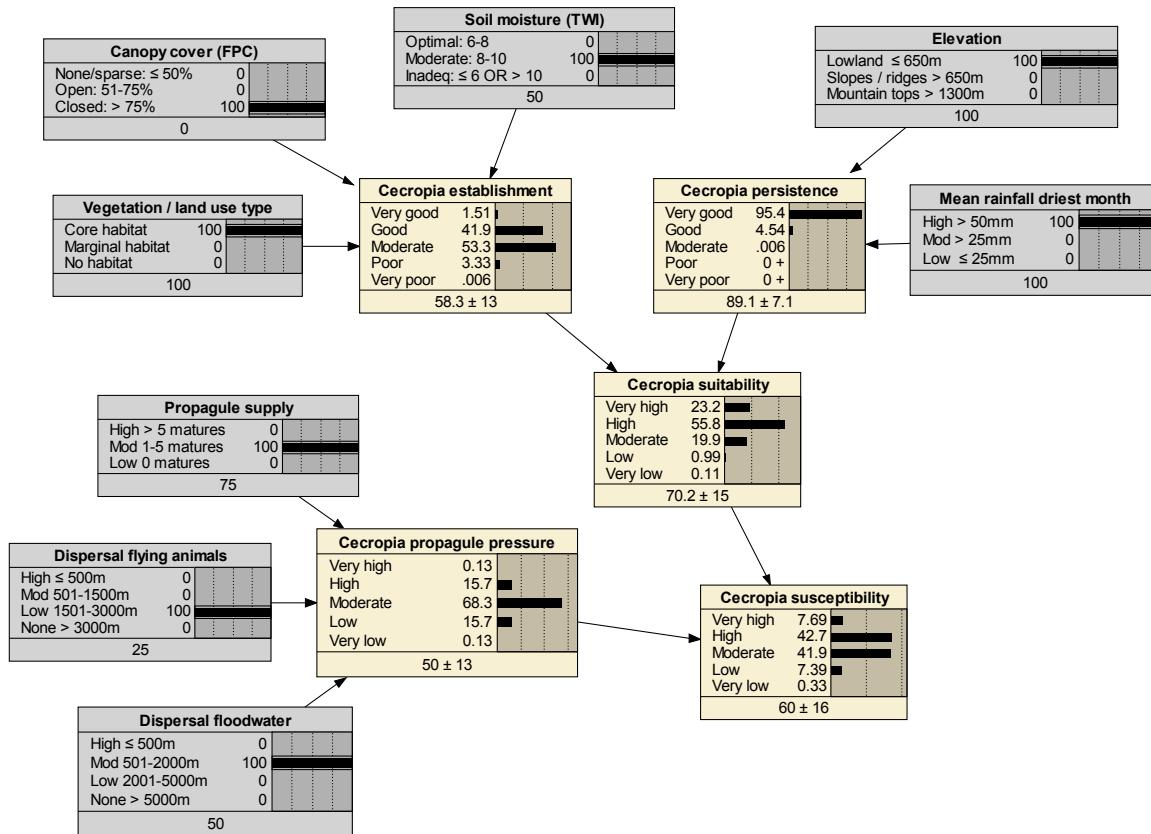


Figure 1. Bayesian network model of environmental suitability for, and susceptibility to invasion by, Mexican bean tree (*Cecropia* spp.). For illustration, we inserted evidence about the state of each explanatory variable (grey boxes) that matches a study area pixel in lowland rainforest that is located 2km away from a small existing infestation. For each response variable (yellow boxes), bar graphs show conditional probability distributions and values below graphs show model expected values ± standard deviation.

Table 1. Bayesian network model variables, with definitions, justification for inclusion in the risk model, weighting applied (colour coded), and spatial proxies used. Methods for reclassifying spatial data attributes to match one of the corresponding variable's three or four states (Figure 1) are also broadly outlined.

Explanatory variables		Justification, weighting applied in risk model, and spatial proxies used		
Name	Definition	Justification and weighting ^{a,b,c,d}	Spatial proxy	Classification into states
Vegetation / land use type (incl. level of disturbance)	Whether habitat supports dense seedling recruitment	<ul style="list-style-type: none"> Prefers early successional habitats in (sub-)tropical rainforests incl. disturbed gaps (cyclones, fire, clearing etc.) or margins Imported and planted as ornamental garden tree in residential areas Also establishes in other disturbed/ open sites incl. riparian areas, roadsides, and poorly managed or abandoned agricultural land Absent from closed canopy tall forests (see risk factor 'canopy') Highly important to establishment 	NVIS Major Vegetation Groups, v4.2 ^e Catchment scale Land Use of Australia, v2016 Queensland Geofabric Surf. Cartography v2.1	<ul style="list-style-type: none"> Good: rainforest Mod: other/regrowth forest Poor: dry forest, all else Good: agriculture (3.6/4.6), intensive (5.4/5.5/5.7) Mod: grazing (2.1/3.2/4.2) Poor: all else

Vegetation – canopy	Whether sufficient sunlight supports frequent seedling recruitment	<ul style="list-style-type: none"> Prefers full sunlight, esp. seedlings Tolerates some shade (e.g. riparian sites, open forests), due to rapid growth can outcompete other trees Highly important to establishment but lower weighting due to uncertainty in modelled spatial proxy 	Australian woody vegetation cover, imagery and v2.0 ⁱ	Foliage projective cover thresholds set from satellite imagery and ⁱ
				<ul style="list-style-type: none"> None/sparse: ≤ 50% Open: 51–75% Closed: > 75%
Soil – moisture	Whether soil drainage supports germination and vigorous seedling growth	<ul style="list-style-type: none"> Prefers well-draining soils Tolerates somewhat wet soils Saturated soils inhibit germination Absent from dry or limestone soils Moderately important to establishment but lower weighting due to uncertainty in modelled spatial proxy 	Topographic Wetness Index (TWI) derived from 1" SRTM DEM-H, v2 ^j	<ul style="list-style-type: none"> Quantile method with five classes, TWI thresholds rounded and adjusted from^j Optimal (Q2/3): 6.001–8 Mod (Q4): 8.001–10 Inadequate (Q1/5): ≤ 6 OR > 10
Climate – rainfall	Whether sufficient year-round rainfall supports tree growth and survival	<ul style="list-style-type: none"> Prefers wet (sub-)tropical rainforest climates with driest month > 60 mm Tolerates wet-dry (sub-)tropical savanna climates with driest month < 60 mm Moderately important to persistence 	Mean rainfall of driest month (August) ^k	<ul style="list-style-type: none"> Wet tropical threshold (> 60 mm) relaxed somewhat, then equal interval method High: > 50mm Mod: 26–50mm Low: ≤ 25mm
Topography – elevation	Whether trees are exposed to adverse climate and poor soil fertility at higher altitudes	<ul style="list-style-type: none"> Most <i>Cecropia</i> spp. prefer lowland environments May not prefer high slopes/ ridges May not tolerate exposed mountain tops Somewhat important to persistence 	SRTM derived Digital Elevation Model, v1 ^l	<ul style="list-style-type: none"> Mountain top threshold set from satellite imagery, then equal interval method Low: ≤ 650m Mod: 651–1300m High: > 1300m
Propagule supply	Amount of propagules potentially dispersed from source infestations within one fruiting season	<ul style="list-style-type: none"> Extremely high fecundity after > 3 years maturation If wind-pollinated by male plant, female produces millions of seeds (~20% viable) Fruits seasonally for 9 months, with a 4-month peak during the early wet Short seed longevity (< 6 months buried) Highly important to dispersal 	Pest Central - weed infestations (Qld series)	<ul style="list-style-type: none"> Computed number of mature individuals within maximum dispersal distance (set at 5km based on all dispersal modes) High: > 5 mature trees Mod: 1–5 mature trees Low: No mature trees (immature trees unlimited)
Dispersal – flying animals	Frequency and density of propagules dispersed from source infestations by bats and birds	<ul style="list-style-type: none"> Main dispersal mode: frugivorous bats & birds, gut passage enhances germination Generic distances for bird dispersal: 50% ~400m, 99% ~1500m Bat dispersal unknown but ≤ several km Highly important to dispersal 	Pest Central - weed infestations (Qld series)	<ul style="list-style-type: none"> Computed radial Euclidean distance from nearest source High: ≤ 500m Mod: 501–1500m Low: 1501–3000m None: > 3000m
Dispersal – (flood-) water	Frequency and density of propagules dispersed from source infestations along streams	<ul style="list-style-type: none"> Dropped seeds also carried by running water or floodwater Emergence has been observed along riverbanks after flooding Moderately important to dispersal 	<p>Pest Central - weed infestations (Qld series)</p> <p>Geofabric Surf. Cartography v2.1 – streams^j</p>	<ul style="list-style-type: none"> Weighted stream lines according to Euclidean distance from nearest source High: ≤ 500m Mod: 501–2000m Low: 2001–5000m None: > 5000m

3x weighting = highly important

2x weighting = moderately important

1x weighting = somewhat important

References and spatial data sources:

- ^a CABI (2016). *Datasheet: Cecropia peltata (trumpet tree)*. In Invasive Species Compendium. CAB International, Wallingford, UK. URL <http://www.cabi.org/isc/datasheet/11955>
- ^b Csurhes S (2016). *Invasive plant risk assessment: Mexican bean tree (Cecropia spp.)*. State of Queensland, Brisbane.
- ^c Department of Agriculture and Fisheries (2016). *Fact sheet: Mexican bean tree (Cecropia spp.)*. State of Queensland.
- ^d Vittoz P and Engler R. (2007). Seed dispersal distances: a typology based on dispersal modes and plant traits. *Botanica Helvetica* 117(2): 109–124. doi: <https://doi.org/10.1007/s00035-007-0797-8>
- ^e Department of the Environment and Water Resources (2007). *Australia's native vegetation: a summary of Australia's major vegetation groups, 2007*. Commonwealth of Australia, Canberra.
- ^f Australian Bureau of Agricultural and Resource Economics and Sciences (2011). *Guidelines for land use mapping in Australia: principles, procedures and definitions, fourth edition*. Commonwealth of Australia, Canberra.
- ^g Environmental Protection Agency (2005). *Wetland mapping and classification methodology. Overall framework: a method to provide baseline mapping and classification for wetlands in Queensland*, v1.2. State of Queensland, Brisbane.

^h Bureau of Meteorology (2012). *Australian hydrological geospatial fabric (geofabric) data product specification: surface cartography Version 2.1*. Bureau of Meteorology, Canberra.

ⁱ Gill T, Johansen K, Phinn S, Trevithick R, Scarth P, et al. (2017). A method for mapping Australian woody vegetation cover by linking continental-scale field data and long-term Landsat time series. *International Journal of Remote Sensing* 38(3): 679-705. doi: <http://dx.doi.org/10.1080/01431161.2016.1266112>

^j Gallant JC and Austin JM (2015). Derivation of terrain covariates for digital soil mapping in Australia. *Soil Research* 53(8): 895-906. doi: <https://doi.org/10.1071/SR14271>

^k Bureau of Meteorology (2015). *Mean monthly and mean annual rainfall data*. Bureau of Meteorology, Canberra.

^l Gallant JC, Dowling TI, Read AM, Wilson N, Tickle P, et al. (2011). *1 second SRTM derived products user guide, Version 1.0.4*. Geoscience Australia, Canberra.

While habitat suitability is a static attribute of the environment (albeit influenced by changing conditions), susceptibility measures invasion risk with reference to a specific time period. Here, we divided a dataset of 320 *Cecropia* spp. detection records collected between 2008 and 2016 into 9 annual ‘detection periods’, which broadly captured the species’ fruiting season. We modelled susceptibility separately for each period, using only the corresponding subset of detection records to compute the dynamic variables ‘Propagule supply’, ‘Dispersal – flying animals’ and ‘Dispersal – water’ (Table 1). We evaluated the ability of our model to accurately predict the observed spatial progression of the *Cecropia* spp. invasion in North Queensland by validating mapped annual susceptibility against recorded detections in the following three periods (e.g. 2008-09 predictions validated against actual 2009-10, 2010-11 and 2011-12 detections). We included records from several fruiting periods to allow for delays in germination and detection. We used the Continuous Boyce Index (CBI) method, which was developed specifically for evaluation against presence-only observations (Hirzel et al. 2006) and has recently been successfully used to validate results of a Bayesian network habitat suitability model (Froese et al. 2017). A predicted-to-expected (P/E) ratio was computed as the (predicted) proportion of detection records relative to the (expected) proportion of background pixels on a continuous scale of susceptibility index values. The background was defined as the potential ‘dispersal area’ around all source infestations within a detection period. The CBI measures the Spearman rank correlation coefficient of P/E against susceptibility and varies from 1 (correct model, P/E steadily increases as susceptibility increases) to -1 (false model, P/E steadily decreases with increasing susceptibility). A CBI close to zero indicates a random prediction (Hirzel et al. 2006).

3. RESULTS

Mapped model predictions for *Cecropia* spp. suitability across the study area (encompassing the 2 million hectare ‘Wet Tropics’ bioregion; Figure 2a) are shown in Figure 2b. About half of Queensland’s Wet Tropics (47%) was predicted to be moderately suitable for *Cecropia* spp. invasion (index = 40-59), potentially supporting scattered individual trees. Another 35% was predicted as highly or very highly suitable habitat (index = 60-100), which may translate to widespread moderate or high densities of the invader. Only 18% of the study area was modelled as unsuitable (index = 0-39), supporting, at best, isolated individual trees that usually fail to persist. These overall results are unsurprising, given the high proportion of wet tropical rainforest habitat as well as favourable climatic and edaphic conditions in the study region.

An example of mapped model predictions for susceptibility to invasion from one *Cecropia* spp. source infestation detected during the 2012-13 period is shown in Figure 2c. Results from validating annual susceptibility maps for six consecutive detection periods against subsequently observed infestations (i.e. 2008-09 map vs. 2009-12 detections, ..., 2013-14 map vs. 2014-17 detections) are presented in Figure 3. In each year, modelled susceptibility was a good predictor of actual *Cecropia* spp. detections in subsequent years. CBI values ranged from 0.94 to 1, indicating near perfect predictions (i.e. proportionally more actual detections as predicted susceptibility increases). Maximum P/E ratios were typically above 15, indicating that the magnitude of discriminatory power at very high susceptibility values was very large. Despite these very good validation statistics, Figure 3 also shows that the model’s high discriminatory power was only evident at susceptibility indices above ~60 (highly and very highly susceptible habitat). Moderately susceptible habitat (index = 40-59) did not contain proportionally more actual *Cecropia* spp. detections in subsequent years. Furthermore, any detections of *Cecropia* spp. in locations outside the potential ‘dispersal area’ assumed by our susceptibility model (i.e. distance from nearest source infestation $\leq 5\text{km}$) due to human-mediated introduction or stochastic long-distance dispersal, was not adequately reflected in validation statistics. This was mostly because the number of such unanticipated detections remained relatively low. In summary, our model confidently predicted areas at high immediate invasion risk in close proximity to source infestations ($\leq 500\text{m}$) or in riparian habitats that may receive propagules via both bat/bird and (flood-)water dispersal. There is insufficient evidence for affirming or rejecting the model’s accuracy in predicting areas at moderate risk of *Cecropia* spp. invasion.

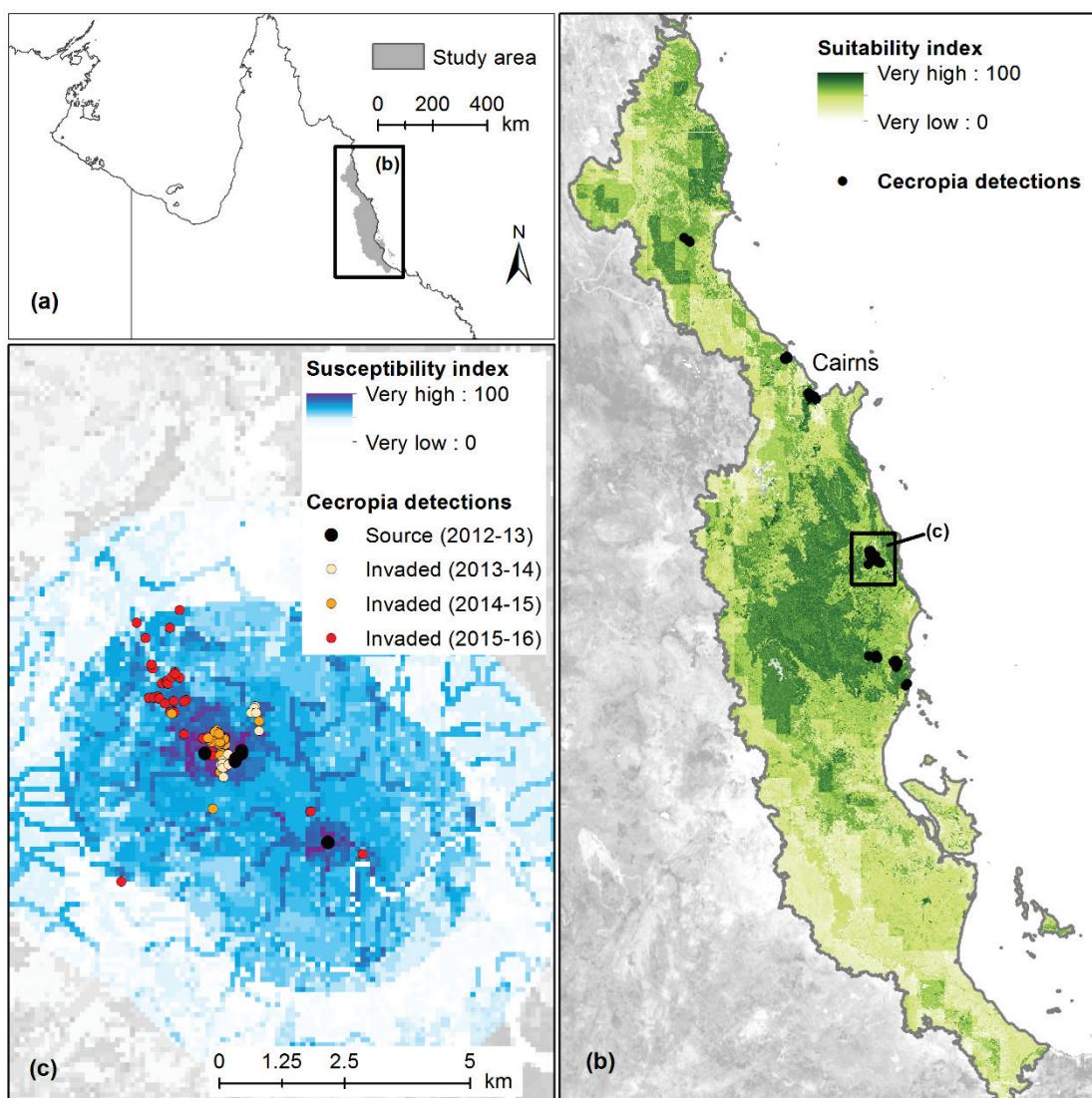


Figure 2. Panels show: (a) study area in North Queensland; (b) mapped environmental suitability for *Cecropia* spp., with actual detections during 2008–2016; and (c) mapped susceptibility to *Cecropia* spp. invasion from source infestations. Here, we show modelled invasion risk for one source infestation (location shown in panel b) during one detection period (2012–13, green dots), compared to actually invaded locations as recorded during the three subsequent detection periods (2013–16, red and orange dots).

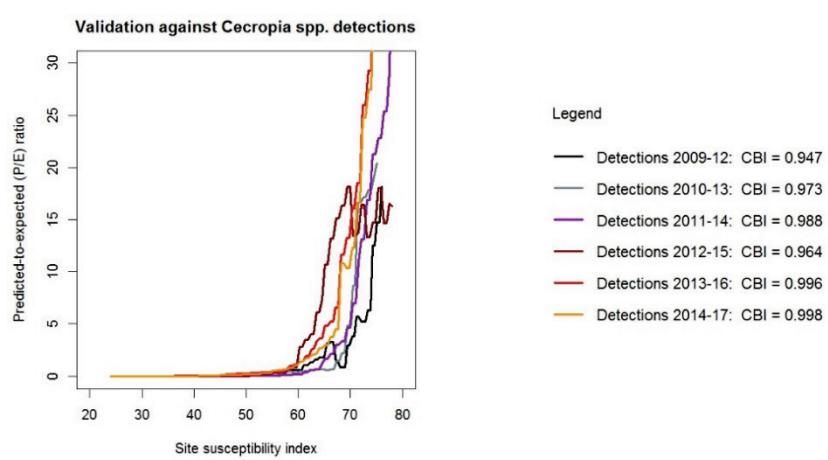


Figure 3. Validation plot for the *Cecropia* spp. susceptibility model. The predicted-to-expected ratio (y axis) measures the proportion of detection records relative to the proportion of background pixels (in this case the potential ‘dispersal area’ around all source infestations within the detection period) on a continuous scale of predicted susceptibility indices (x axis).

4. CONCLUSIONS

Our spatial risk model accurately predicted the observed progression of the *Cecropia* spp. invasion in North Queensland over the past decade. Annual susceptibility maps may be used for allocating surveillance and containment resources around existing infestations. Validation results suggest that efforts should be confidently focused on highly susceptible areas, as these were indeed preferentially invaded in the existing data. It remains uncertain whether moderate susceptibility also translates into increased invasion. Low-frequency long-distance dispersal and human-mediated translocations may be important to containment but were not captured in the model. In conclusion, we adapted an existing conceptual framework of plant invasions, and an established modelling methodology for implementing this framework, to the needs of rapid incursion response for emerging weeds. This meant (a) limiting data/knowledge acquisition to desktop research in line with common approaches to non-spatial pest risk assessments, and (b) simplifying model calibration by using adaptable mathematical weighting functions. Further automation of risk factor / spatial proxy selection and data pre-processing may make our rapid methodology accessible to biosecurity agencies that typically have significant field knowledge and GIS capacities but may have limited expertise in applying complex modelling methods.

ACKNOWLEDGEMENTS

We thank Moya Calvert and Joe Scanlan from Biosecurity Queensland for assisting with case study selection and access to spatial data on weed detections.

REFERENCES

- Adams VM, Petty AM, Douglas MM et al. (2015). Distribution, demography and dispersal model of spatial spread of invasive plant populations with limited data. *Methods in Ecology and Evolution* 6(7): 782-794.
- Boets P, Landuyt D, Everaert G et al. (2015). Evaluation and comparison of data-driven and knowledge-supported Bayesian belief networks to assess the habitat suitability for alien macroinvertebrates. *Environmental Modelling & Software* 74: 92-103.
- Csurhes S (2016). *Invasive plant risk assessment: Mexican bean tree (Cecropia spp.)*. State of Queensland.
- Department of Agriculture and Fisheries (2016). *Mexican bean tree (Cecropia spp.)*. State of Queensland.
- Fenton N and Neil M (2013). *Risk assessment and decision analysis with Bayesian networks*. CRC Press.
- Fenton NE, Neil M and Caballero JG (2007). Using ranked nodes to model qualitative judgments in Bayesian networks. *IEEE Transactions on Knowledge and Data Engineering* 19(10): 1420-1432.
- Fletcher CS and Westcott DA (2013). Dispersal and the design of effective management strategies for plant invasions: matching scales for success. *Ecological Applications* 23(8): 1881-1892.
- Froese JG, Smith CS, Durr PA et al. (2017). Modelling seasonal habitat suitability for wide-ranging species: invasive wild pigs in northern Australia. *PLOS ONE* 12(5): e0177018.
- Hastings A. (2014). Persistence and management of spatially distributed populations. *Population Ecology* 56(1): 21-26.
- Hirzel AH, Le Lay G, Helfer V et al. (2006). Evaluating the ability of habitat suitability models to predict species presences. *Ecological Modelling* 199(2): 142-152.
- O'Reilly-Nugent A, Palit R, Lopez-Aldana A et al. (2016). Landscape effects on the spread of invasive species. *Current Landscape Ecology Reports* 1(3): 107-114.
- Panetta FD and Lawes R (2005). Evaluation of weed eradication programs: the delimitation of extent. *Diversity and Distributions* 11(5): 435-442.
- Parry H, Sadler R and Kriticos D (2013). Practical guidelines for modelling post-entry spread in invasion ecology. *NeoBiota* 18: 41-66.
- Savage D and Renton M (2014). Requirements, design and implementation of a general model of biological invasion. *Ecological Modelling* 272: 394-409.
- Smith CS, van Klinken RD, Seabrook L and McAlpine C (2012). Estimating the influence of land management change on weed invasion potential using expert knowledge. *Diversity and Distributions* 18(8): 818-831.
- Wilson JR, Panetta FD and Lindgren C (2017). *Detecting and responding to alien plant incursions*. Cambridge University Press, New York.