BIODIVERSITY RESEARCH

Searching for non-indigenous species: rapidly delimiting the invasion boundary

Brian Leung[1,2]*, Oscar J. Cacho[3] and Daniel Spring[4]

[1]*Department of Biology, McGill University, Montreal, Quebec H3A 1B1, Canada,* [2]*School of Environment, McGill University, Montreal, Quebec H3A 2A7, Canada,* [3]*School of Business, Economics, and Public Policy, University of New England, Armidale, 2350, NSW, Australia,* [4]*School of Biological Sciences, Monash University, Clayton, VIC, 3800, Australia*

ABSTRACT

Aim At first detection, little information is typically known about an invader's characteristics, true arrival date or spatial extent. Yet, before management options such as control or eradication can be considered, we need to know where a nuisance species has already spread. This is particularly difficult because of stochastic processes. Here, we develop an approach that requires little *a priori* information, yet accurately delimits the range of a biological invader.

Location We used a simulated landscape, subjected to stochasticity inherent in establishment and spread, to test novel theory for delimiting locally spreading populations.

Methods We distinguish three stages to identify the boundary of an invasion, which we term Approach, Decline, Delimit (ADD). Our ADD algorithm uses general characteristics of the invasion pattern, obtained during a search for occupied sites, in combination with sampling and probability theory to delimit the invasion. We compare ADD against four naïve delimitation strategies, for long and normal dispersal kernels.

Results Our results illustrate the potential difficulty in delimiting invasions. Naïve strategies, such as stopping when the invader is absent, typically failed to properly delimit the invasion. In contrast, ADD operated relatively efficiently, and was robust to habitat heterogeneity and knowledge of the true epicentre, but was sensitive to the sparseness of the invasion. For long-distance dispersal kernels, ADD had 80% accurate delimitations when *c.* 5% or more of the cells were occupied within the invasion boundary; for normal dispersal kernels, ADD had 95% accurate delimitations when *c.* 2.5% or more of the cells were occupied.

Main conclusions There is virtually no existing theory for delimiting invasions. ADD is efficient and accurate, even with unknown time of invasion, unknown dispersal kernels, stochastic establishment dynamics and spatial heterogeneity, except for very low invasion densities.

Keywords
Bayesian, biological invasions, eradication, exotic species, simulation model.

*Correspondence: Dr. Brian Leung, Department of Biology, McGill University, Montreal, Quebec H3A 1B1, Canada. E-mail: brian.leung2@mcgill.ca

INTRODUCTION

Invasive species are increasingly being recognized as a driver of global environmental change, causing both economic and environmental damages (Mack *et al.*, 2000). These effects will likely be exacerbated as the rate of introduction of exotic species worldwide accelerates in conjunction with global trade (Hulme, 2009). Not surprisingly, there is consensus in the scientific community that society should invest in improved methods to deal with invaders, as these may pay dividends in terms of reduced impacts caused by the invader (Lodge *et al.*, 2006).

In an ideal world, we would identify and prevent the introduction of harmful exotic species. Failing that, we would be able to react quickly and eradicate an invader, or at least control its spread, and therefore the potential damages that it

1

might cause (Veitch & Clout, 2002). For many management options, the first step is to determine the extent of spread – to delimit the invasion. Failure to do so could allow the invader to continually spread to other new areas as we restrict management to one location (Panetta & Lawes, 2005). Yet, if we can accurately delimit the invasion, it may be possible to erect barriers to stop or slow spread (Sharov & Liebhold, 1998), or apply control/eradication strategies working from the boundaries inwards, such that dispersal occurs to managed zones or sites that are already occupied (Edwards & Leung, 2009).

The issue of delimiting an invasion, and therefore of control and eradication, can be conceptualized as a trade-off between the probability of escape, if the quarantined area is too small and do not encompass the range of the invader, and wasted effort, if the quarantined area is too large. If the barriers are set too large, search, eradication and control efforts will be conducted where the invader has not spread and/or the species will continue to spread until it reaches the barrier. Further, the speed at which we initiate our quarantine is also important. The population will continue to spread as we search, increasing the probability of escape, the extent of the invasion, and the ultimate effort required to manage the invasion. Thus, our objective is to delimit an invasion as quickly and as accurately as possible.

Unfortunately, delimitation is difficult. For a new invasion, a typical scenario would be as follows: an invader would be observed in one (or a few) location(s). Other than the initial observation, we might know little about the invasion. At first detection, we do not know when the species first arrived or potentially the location of the epicentre. Additionally, we typically cannot parameterize their dispersal characteristics, especially in a novel environment. Further, we should not simply keep searching until we fail to find the invader and assume that this is the boundary of the invasion (Mangel *et al.*, 1984). The invader may be absent from a given site because of the randomness inherent in the dispersal process (i.e. stochasticity), and therefore the invasion front should be sparse and heterogeneous. Thus, there is no clear rule about the distance at which one should stop searching (Fig. 1). In brief, given the typical lack of information about newly detected invasions and stochasticity in invasion dynamics, delimitation is currently a challenging unsolved problem.

In this manuscript, we develop a simple approach to delimit invasions for continuously spreading organisms. Our approach assumes no prior information – we do not know
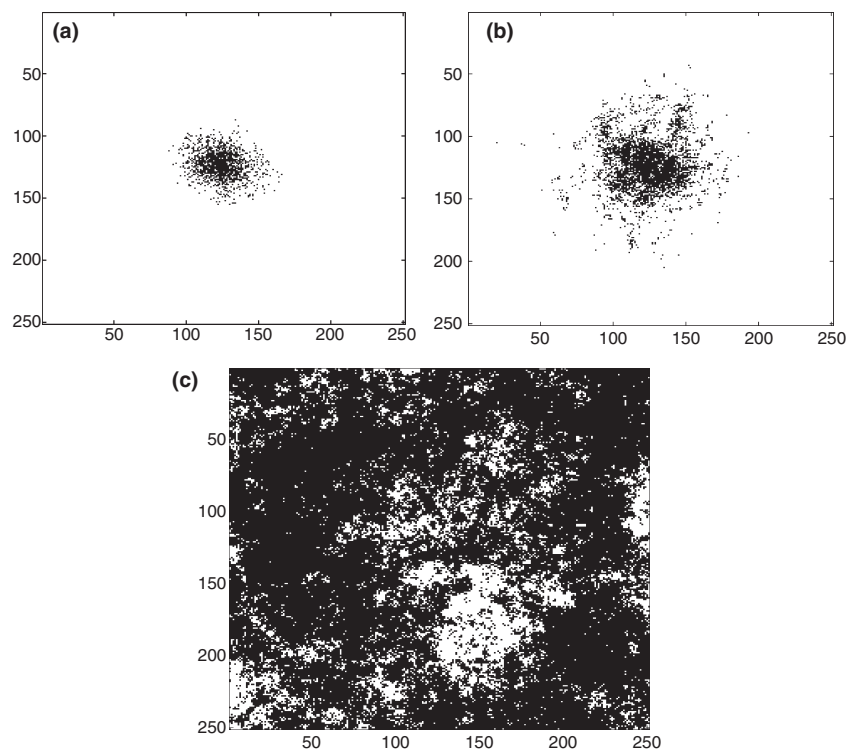

Figure 1 Examples of invasion patterns (simulated using equations 1–4). At first detection, we will only have information on the location of a single invaded point. For searches conducted on the ground, the pattern of invasion will not be visible and will be unknown, yet managers will still need to delimit the invasion (find the very outermost occurrence). This is particularly challenging because we expect that the invasion will become sparse near the invasion boundary with a few sites invaded and many empty sites. Panel (a) random invasion generated using a normal (negative exponential) dispersal kernel. Panel (b) random invasion generated using a long-distance (Cauchy) dispersal kernel. Both show substantial patchiness and low densities closer to the boundary. Panel (c) Example of habitat heterogeneity, with 10% unsuitable habitat (shown in white). For simulations using heterogeneous habitats, models were built with 5–20% unsuitable habitat, with different spatial configurations.

when or where the invasion first occurred or the form of the dispersal kernel of the invader. Only the site of first detection is known. Our approach consists of three useful stages for identifying the boundary of a species' spread (Approach–Decline–Delimit; ADD): (1) Approach towards the boundary, (2) as we move out towards the edge of the invasion, measure the Decline in density of occurrences; (3) use the rate of decline to Delimit the invasion. ADD is based on probability and sampling theory, and uses data collected during a search for invaded sites to draw inferences about the extent of an invasion.

METHODS

To theoretically test the ADD approach, we (1) simulate the 'real world' (an invasion pattern and subsequent first detection event), (2) develop and use ADD to delimit the extent of the invasion, and (3) evaluate the performance of our approach compared to other 'naïve' approaches, and examine the theoretic properties of our model.

Simulating the 'real' world

We model the following scenario: (1) an invader has been found; (2) the timing of invasion is unknown; (3) species characteristics, such as dispersal distance, are unknown; (4) invasion of local area (a cell) is a stochastic process, but depends upon propagule pressure, which is unknown; (5) the probabilistic relationship between propagule pressure and establishment within a local cell is unknown. These are realistic conditions that will occur for most new invaders. Because of the conditions 2 and 3, we do not know how far the invader has spread. Because of the conditions 4 and 5, we may not detect an invader, but that does not necessarily mean that we are at the boundary of the invasion. Additionally, we consider effects of knowledge of environmental suitability and knowledge of the invasion epicentre.

We have attempted to avoid unnecessary complexity in our simulations, keeping only the minimum elements required for analysis of the challenges we wish to address. A key requirement is to model the spread of an invader across a landscape. In the first step of our simulations, we create a landscape represented by an $n \times m$ matrix of cells. Each cell represents a local environment. Each cell can be invaded, depending on propagule pressure.

$$P_{i,j} = 1 - e^{-\alpha A_{i,j}} \qquad (1)$$

where $P_{i,j}$ is the probability that the cell at row i and column j becomes invaded, α is a coefficient, and A is the number of propagules reaching cell i,j from all other invaded cells.

Each invaded cell generates s propagules. For our simulations, we consider two general dispersal patterns, normal and long distance, by using a negative exponential and Cauchy dispersal kernel, respectively. The negative exponential is the standard means of generating normally distributed spread, whereas the Cauchy dispersal kernel has longer tails and is therefore often used to capture longer-distance dispersal events (Kot *et al.*, 1996).

$$A_{i,j} = s \sum_{k=-q}^{q} \sum_{l=-q}^{q} K_{ij} I_{k,l} \qquad (2)$$

where for the normal dispersal kernel,

$$K_{ij} = e^{-\beta d_{ij}} \qquad (3)$$

And for the long-distance dispersal kernel,

$$K_{ij} = \frac{1}{\pi \beta \left[1 + \left(\frac{d_{ij}}{\beta} \right)^2 \right]} \qquad (4)$$

where $I_{k,l}$ is zero if cell k,l is uninvaded and unity if cell k,l is invaded. Cell k,l is relative to the location of i,j; d is the distance between cell i,j and cell k,l, and β is a dispersal coefficient. We initiate the invasion at time $t = t_{ini}$, at a randomly chosen cell i_{ini}, j_{ini} in the landscape. From here, invasions progress according to equations 1 & 2 (Fig. 1), until subsequent discovery. We model a random discovery time (t_{begin}) after the invasion event to generate a variety of invasion ranges.

Additionally, we modelled both homogeneous and heterogeneous landscapes. Heterogeneous landscapes were comprised of a mix of suitable and unsuitable habitats, as is often the output from habitat suitability models. We allocated habitat suitability using a fractal algorithm – the mid-point displacement method (Saupe, 1988; Tyre *et al.*, 1999, Fig. 1c), which is a two-dimensional stochastic process defined by two parameters: variance, which determines the spread of values generated, and H, which determines the fractal dimension of the landscape. H can take values in the range of 0.0 (weak spatial autocorrelation) to 1.0 (strong spatial autocorrelation). The range of floating-point values generated by the algorithm was mapped to integer (0, 1) values. This approach allowed us to generate worlds that contained the same frequency distribution of suitable and unsuitable habitats but different spatial arrangements.

Using ADD to delimit the extent of the invasion

Our objective is to find the extent of the invasion as quickly and accurately as possible. We assert that we can use the spatial characteristics of invasion patterns to do so (Figs 1 & 2). In general, we expect more sites to be invaded closer to the epicentre. These sites have had the most time to become invaded. A smaller proportion of sites are expected to be invaded near the boundary. Because new sites become invaded from propagules generated from already invaded cells, we further expect the proportion of invaded sites to decline relatively smoothly from epicentre to edge. This is compatible with the concept of a travelling invasion wave (Neubert & Caswell, 2000), but our model accounts for the number of infested cells and produces a sparsely occupied, stochastic boundary with many gaps rather than a smooth wave front. Our ADD method consists of the following steps. First, we move close to the boundary (Approach). Second, we seek an efficient

method for collecting information that will characterize the rate of decline in the proportion of invaded sites (Decline). Third, we statistically estimate the rate of such decline to determine the likely boundary of the invasion (Delimit).

ADD seeks to sample efficiently and maximize the inferences possible based on the occurrence patterns observed from sampling. Importantly, ADD does not require additional sampling effort. Therefore, where insufficient information exists to construct the ADD algorithm, we would be no worse off than using current naïve approaches.
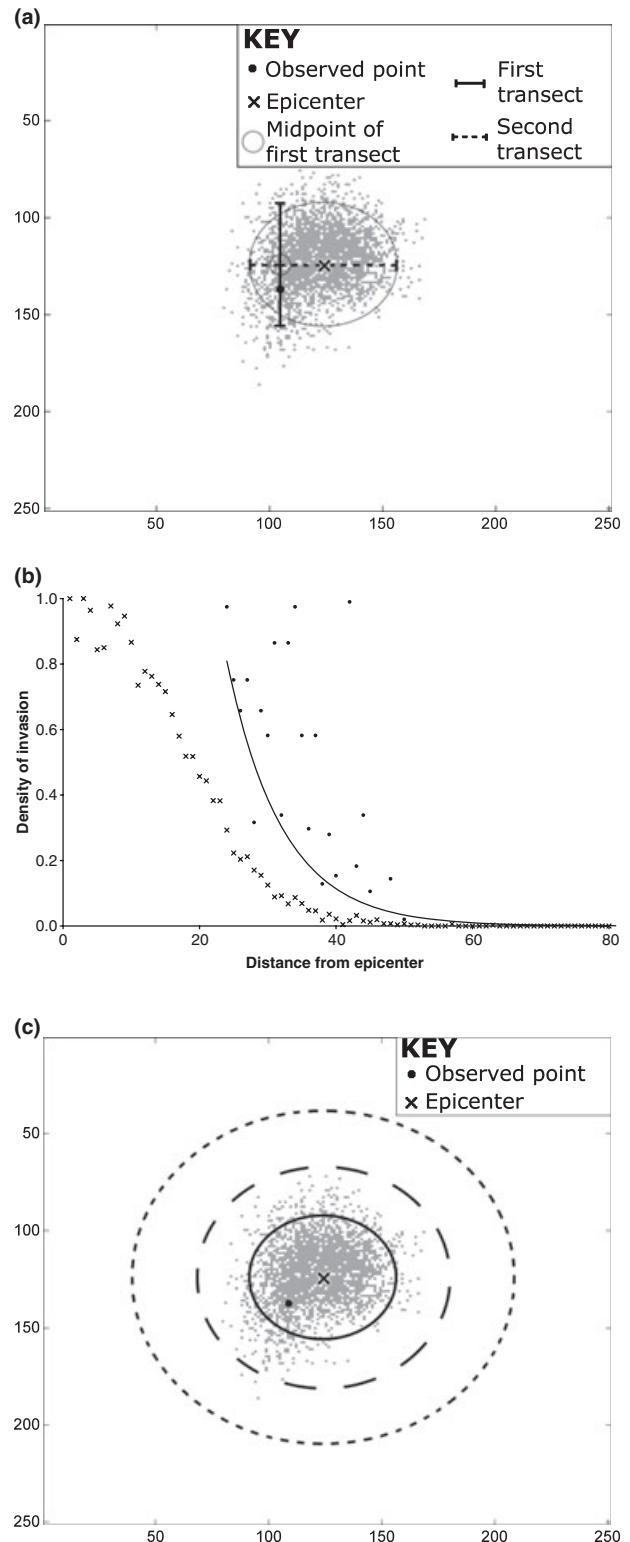
Approach: moving towards the boundary

The first stage is to use an algorithm to quickly move close to the boundary. The simplest method would be to search linearly, starting from the observed invaded site and moving in any given direction, one cell at a time, until cells cease to be invaded. However, if the extent of the invasion is unknown and large, this may be inefficient. Instead, searches can be carried out in an exponential manner – two cells away from the initial observed invasion, four cells, eight cells, etc. When an uninvaded cell is observed (e.g. the 16th cell), move back one step (e.g. the 8th cell from the 16th). This would have brought us to the point where the density of invaded cells is not asymptotically high (Fig. 2b), where we can begin to characterize the decline in densities using the next stage of ADD (see later).

This procedure is repeated in four directions (e.g. North, South, East, and West). If the invasion epicentre is unknown, we can further estimate the epicentre by taking the mid-point of one transect (e.g. North-South), and using this as the starting point for the next transect (e.g. East-West) (Fig. 2a). The mid-point of both transects is used as an estimate of the epicentre.

As a special note, if the invasion front is continuous and smooth, as is often assumed in ecological spread models (e.g. Neubert & Caswell, 2000), the following modified algorithm would be sufficient to delimit a species' extent: bisect the upper and lower bound. Specifically, let l be the lower bound, defined

Figure 2 The application of the ADD algorithm to a simulated invasion pattern. Panel (a) initial detection point (black dot) and transects conducted during stage 1 of ADD (Approach) are shown. The mid-point (open circle) of the first (vertical) transect (solid line) is used to identify where to sample the second (horizontal) transect (dashed line). The mid-point of both transects are used to estimate the epicentre of the invasion ($\times$). Panel (b) estimates of $\lambda(d)$ – the upper bound of the density of occurrences at each distance from the epicentre, estimated using equations 6 or 8 (shown as dots). These are generated during stage 2 of ADD (Decline). Note that, as expected, $\lambda(d)$ is higher than the true density at each distance from the epicentre (shown as $\times$). In stage 3 of ADD (Delimit), the best line is fit to the lower bound of dots (solid line) (equation 9). Final delimitation is determined by equation 10. Panel (c) the results of the three stages of ADD are shown. The inner circle represents the Approach and allows the densest region of invasion to be efficiently traversed (solid circle). The middle circle represents the distance traversed during stage 2 of ADD (Decline) (dashed circle). The outer circle represents the final delimitation from stage 3 (Delimit) (dotted circle). Note that stage 3 was necessary to capture the final few sparse occurrences. ADD, Approach–Decline–Delimit.

by the farthest known invaded cell (e.g. $x = 64$), and u be the upper bound (e.g. $x = 128$), defined by the closest known uninvaded cell. Sample a cell at the mid-point $c = l + (u - l)/2$ (e.g. $l = 64$, $u = 128$, $c = 64 + (128 - 64)/2 = 96$). If c is invaded, set the mid-point as the new upper bound ($u = c$); if it is uninvaded, set the mid-point as the new lower bound ($l = c$). Then proceed to resample a new mid-point until u and l converge. These steps would be repeated in all four directions. However, for the more general case, where invasions are stochastic and occurrences become sparse nearer the boundary, stages 2 and 3 of ADD may be required.

Decline: estimating the proportion of invaded cells

The problem of delimiting invasions becomes difficult because we expect many 'gaps' to be present as we approach the boundary and the frequency of invaded sites declines (Fig. 2). Thus, finding the last invaded site and delimiting the boundary will be difficult. To address this problem, we begin by estimating the proportion of cells invaded at a given distance from the epicentre by equal probability sampling of cells at that distance, without replacement.

For efficiency, we do not want to sample exhaustively to determine the proportion of cells invaded at a given distance, when the number of cells is large. Further, we want to sample only until we have observed at least one invaded cell. At this point, we know that we have not reached the invasion boundary and that we should sample at a farther distance. We can use the single observed invasion and the number of failed observations to estimate the upper bound of the proportion of invaded cells. Using Bayesian statistics:

$$P(p|N) = \frac{P(N|p)P(p)}{\int_0^1 P(N|p)P(p)\,\mathrm{d}p} \qquad (5)$$

where p is the probability that a cell is invaded (i.e. the proportion of invaded cells), and N is the number of cells sampled until the one invaded cell was found. In our case, the probability of observing N given p is the probability of observing $N - 1$ uninvaded cells until we find one invaded cell:

$$P(N|p) = p(1 - p)^{N-1} \qquad (6)$$

We can now use this to estimate the bounds (e.g. the 95% credible intervals) for p, assuming a uniform uninformative prior ($P(p)$ cancels out):

$$U(p < \lambda(d)|N) = \frac{\int_0^{\lambda(d)} p(1 - p)^{N-1}\,\mathrm{d}p}{\int_0^1 p(1 - p)^{N-1}\,\mathrm{d}p} \qquad (7)$$

$$= \frac{1/N(1 - (1 - \lambda(d))^N) - 1/(N+1)(1 - (1 - \lambda(d))^{N+1})}{1/N - 1/(N+1)} \qquad (8)$$

where U is the probability that p is less than or equal to the upper bound (λ) at a given distance from the epicentre (d). Numerically solve for λ, setting U to some preset level (e.g. $U = 0.95$).

If no invasions are found, then the upper bound of p would be

$$U(p < \lambda(d)|N) = \frac{\int_0^{\lambda(d)} (1 - p)^N\,\mathrm{d}p}{\int_0^1 (1 - p)^N\,\mathrm{d}p} \qquad (9)$$

$$\lambda(d) = 1 - (1 - U)^{1/N} \qquad (10)$$

We would continue to move out to farther distances, numerically solving for λ to correspond with a desired degree of confidence (U) (equations 7 & 8) each time an invaded site is observed, until λ reached a desired minimum value and no invasions were observed (equations 9 & 10), or we had enough points to estimate a curve (Stage 3) (Fig. 2b).

For the case where there is *a priori* knowledge of habitat suitability, the researcher should focus on sampling suitable habitats. In such a case, the following modification should be used:

$$\lambda(d)' = \lambda(d)\frac{A(d)}{A(d) + B(d)} \qquad (11)$$

where $\lambda(d)$ is calculated from equations 8 or 10, but sampling only suitable habitats, $A(d)$ and $B(d)$ are the number of suitable and unsuitable sites at distance d from the epicentre, and $\lambda(d)'$ is the modified estimate of $\lambda(d)$.

Delimit: fitting the rate of decline

Where the number of invaded sites is low, we may require an unfeasibly large effort to detect even a single invaded site. This typically may occur near the outside boundary of invasions. Our best estimate of the ultimate extent of an invasion, in a probabilistic sense, can be derived from the rate of decline in the proportion of invaded sites as the distance from the epicentre increases (estimated in stage 2). More specifically, the estimates of $\lambda(d)$ from stage 2 represent the upper bound that will typically exceed p (depending on the value of U set by the researcher). Thus, to estimate the rate of decline, we fit a curve to the lower bound of the $\lambda(d)$ data cloud (Fig. 2). The lower bound of $\lambda(d)$ can be determined with the following algorithm: (1) Move outwards towards the invasion boundary. (2) If $\lambda(d) < \lambda(d - 1)$, keep the estimate, otherwise discard the estimate. (3) Fit a curve to the $\lambda(d)$ estimates that have been kept.

To fit the rate of decline, we need to use some functional form. Common functional forms such as the negative exponential curve appear to fit the relationship between distance and proportion invaded quite well for simulations, starting at $P = 0.5$ or less (Fig. 2). We can estimate the relationship between $\lambda(d)$ and d using

$$\hat{\lambda}(d) = e^{-\gamma(d-\upsilon)} \qquad (12)$$

where $\hat{\lambda}(d)$ is the expected upper bound, γ is a fitted shape parameter and υ is a fitted distance offset. γ and υ are fit using the least squares difference between $\hat{\lambda}(d)$ and $\lambda(d)$ [i.e. $\underset{\gamma,\upsilon}{\text{MIN}}(\hat{\lambda}(d) - \lambda(d))^2$]. Once the best fit γ and υ are found, we place the boundary where the following condition is satisfied:

$$C(D)e^{-\gamma(D-\upsilon)}<1 \qquad (13)$$

where D is the final distance delimiting the invasion, C is the number of cells at that distance from the epicentre. This inequality indicates that we expect less than one cell to be invaded beyond D.

Where there are not enough occurrences to estimate a curve, exhaustive sampling may be required. We treat exhaustive sampling as the naïve alternative to ADD, and where such cases are needed, we cannot evaluate our algorithm.

Evaluating the performance of ADD

We tested the theoretic behaviour of our ADD heuristic. Most importantly, we examined the success/failure rate of ADD for accurately delimiting the boundary. We defined successful delimitation as having encompassed each and every occurrence of the invader. We compared ADD to four naïve alternatives: (1) we searched along north, south, east and west transects, and delimited the invasion when an uninvaded cell was observed. We termed this approach 'Transect'. (2) We searched at a given radius from the epicentre until we observed an occurrence, at which point we moved farther from the epicentre and repeated the process. If no occurrences were found after an exhaustive search at a given radius, we delimited the invasion. We termed this approach 'Circumference'. (3) We repeated the Circumference algorithm, but instead of stopping when invasions were absent at a given radius, we continued searching farther from the epicentre, until the search effort was equivalent to the effort required by ADD. In this way, we controlled for search effort in our comparison with ADD. We term this approach 'Equivalent effort'. (4) We used adaptive cluster sampling (ACS) – we searched a five-cell radius around the epicentre, and searched an additional five-cell radius around any additional detections resulting from this search, after eliminating overlap. We repeated this process until no more occurrences were found. While ACS has been applied to general searches (Turk & Borkowski, 2005), it has not been previously applied to delimitation.

Although accurate delimitation is the primary objective, we also examined sampling efficiency – the fraction of cells that needed to be sampled using ADD compared to the actual number of cells within the true boundary (i.e. taking the radius from the farthest invaded cell and counting the number of cells within that radius). This provided a conservative metric of efficiency – without ADD, there would be no obvious stopping rule, and the search would need to extend beyond the true boundary.

Finally, we examined the degree of overestimation of the boundary. There will be a balance between setting conservative (large) boundaries and failure rates. Ideally, ADD will overestimate the true boundary by the smallest amount possible, while having a reasonable success rate for delimiting the invasion (e.g. 80%). We measured overestimation as the proportional difference in predicted versus true radius ($\frac{r_p - r_t}{r_t}$), where r_p is the predicted radius of invasion and r_t is the true radius. We only considered overestimation and effort when delimitation was successful so that we did not overestimate the efficacy of ADD.

To test the behaviour of the model, we simulated the 'real' world 4000 times. We applied ADD and the naïve algorithms to each real world, such that each algorithm was applied to identical sets of invasion patterns. Additionally, we limited the analysis to scenarios where there was sufficient information to apply the ADD algorithm, such that results for different algorithms were for the same set of scenarios and were comparable; for scenarios where ADD could not be constructed, we would be limited to using some naïve approach and comparison would therefore not be meaningful.

To generate a range of invasion patterns and identify the behaviour of ADD across a range of conditions, we varied parameter values simulating our 'real world' (equations 1–4). Specifically, for the normal dispersal kernel (equation 3), we randomly chose β values ranging from 0.05 to 0.4, α values ranging from 0.001 to 0.01 and initial detection time t_{begin} at random between 4 and 20 time intervals. For the long-distance dispersal kernel (equation 4), we randomly chose β values ranging from 0.3 to 1.5, α values ranging from 0.006 to 0.06 and initial detection time t_{begin} at random between 4 and 10 time intervals. We arbitrarily set s, the number of propagules to 100, and did not consider s further, because variation in probability of establishment could be modelled entirely through variation in α, as both s and α are scalars multiplied together (equations 1 & 2).

To further examine the sensitivity of ADD, for each set of 4000 simulations, we also considered the scenario where the epicentre was known and where it was not. When the epicentre was not known, we chose the initial discovery point randomly from all sites that were invaded at t_{begin} and we proceeded with the fitting techniques described earlier. Where the epicentre was known, we forced the epicentre used in ADD to cell (i_{ini}, j_{ini}). Finally, for both long-distance and normal dispersal kernels, we ran addition sets of 4000 simulations to generate 'real worlds' with habitat heterogeneity (Fig. 1, Panel c), such that the amount of unsuitable habitat ranged from 5% to 20% of the landscape. We considered the behaviour of ADD, assuming no knowledge of habitat suitability, and sampling based on knowledge of habitat suitability (equation 11).

RESULTS

The ability to characterize the decline in occurrences (stage 2) and construct the ADD algorithm was dependent on the number of invaded cells (Fig. 3). When 60–80 cells were
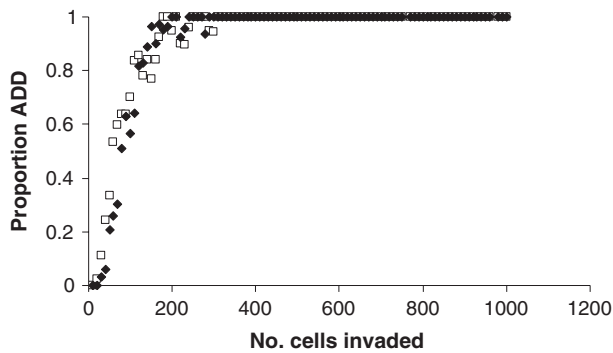
Figure 3 Relationship between the number of invaded cells and the proportion of simulations where there was sufficient data to construct the ADD algorithm. Empty squares = normal dispersal kernel; solid diamonds = long-distance dispersal kernel. ADD, Approach–Decline–Delimit.

invaded, we were able to construct ADD 50% of the time; when 100–120 cells were invaded, ADD was applicable in 80% of simulations and reached an asymptote of 100% for higher invasions (Fig. 3).

When ADD was analysable, the ability to successfully delimit invasions was dependent on the density of invaded cells, but offered substantial improvements compared to naïve approaches (Fig. 4). For normal dispersal, ADD achieved *c.* 80% successful delimitations (0.2 error rate) when invasion density was as little as 0.01 (i.e. 1% of the cells invaded) (calculated as the number of cells occupied divided by the number of cells within the true invasion boundary). Error rates declined to an asymptote of *c.* 0.05 (95% successful delimitations) at an invasion density above 0.025 (Fig. 4a). For long-distance dispersal, ADD achieved *c.* 80% successful delimitations at an invasion density of 0.05, reflecting the greater difficulty in delimiting long-distance invasions (Fig. 4b).

In comparison, the naïve Transect approach and ACS failed to properly delimit invasions almost 100% of the time (Fig. 4). Exhaustive searches at a given distance from the epicentre (Circumference) typically failed more than 80% of the time. The naïve 'Equivalent effort' approach (searching exhaustively, but searching the same number of cells as the ADD algorithm) still performed considerably worse than ADD. For the long-distance dispersal kernel, above invasion densities of 0.025, error rates were 2–3 times higher using the Equivalent effort approach. For the normal dispersal kernel, above invasion densities of 0.01, error rates were 4 to 10 times higher using Equivalent effort compared to ADD (Fig. 4).

ADD was fairly robust to the presence and knowledge of habitat heterogeneity and knowledge of the invasion epicentre (Fig. 5). For both normal- and long-distance dispersal kernels, error rates were comparable, although lack of knowledge of the invasion epicentre resulted in higher variability (Fig. 5a,b). Similarly, the effort expended to delimit the invasion was fairly robust to habitat heterogeneity and knowledge, although it was strongly affected by the invasion density (Fig. 5c,d). At invasion densities of 0.01 (1% cells invaded), *c.* 50% of the
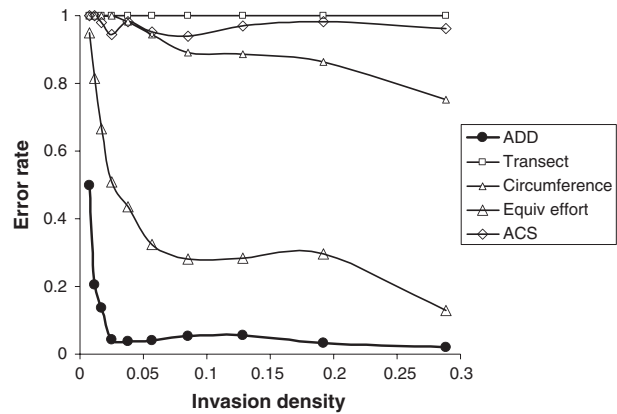
Panel A
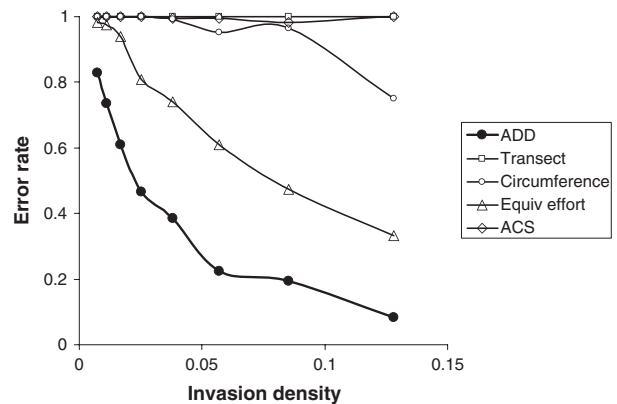
Panel B

Figure 4 Relationship between invasion density (the fraction of cells invaded within the true invasion boundary) and the error rate (fraction of simulations where the delimited boundary failed to encapsulate all occurrences). Here, results from the ADD algorithm are compared against four naïve alternatives, Transect, Circumference, Equivalent effort and adaptive cluster sampling (ACS), which are described in the methods section, for the normal dispersal kernel (Panel a), and the long-distance dispersal kernel (Panel b). ADD, Approach–Decline–Delimit.

cells needed to be searched; at invasion density of 0.12, *c.* 10% of the cells needed to be searched. Interestingly, the fraction of cells searched was not greatly different when sampling was based on known and unknown habitat suitability, and differed by only *c.* 2.5% across simulations. However, this translated to *c.* 11% fewer cells that needed to be sampled if knowledge of habitat suitability were incorporated (as expected given that between 5% and 20% unsuitable habitats were generated).

In contrast to error rates and sampling effort, the magnitude of overestimation generally was not strongly affected by invasion density (Fig. 5e,f). Over estimation ranged from 20% to 40% overestimation when the epicentre was known and was not affected by habitat heterogeneity. However, knowledge of the epicentre strongly affected the degree of overestimation; when we did not know the location of the epicentre, overestimation was sometimes as high as 85%.
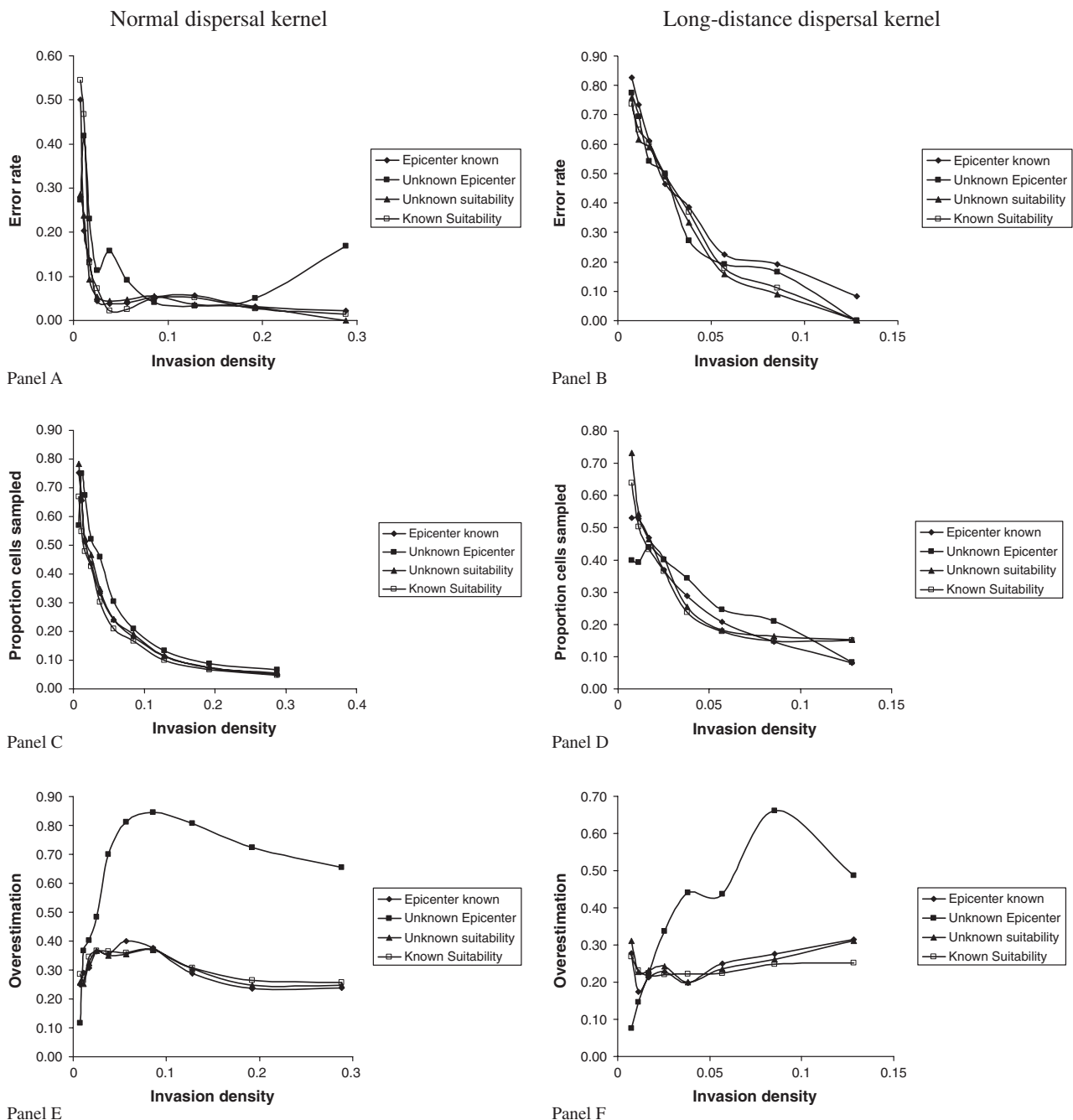
Normal dispersal kernel

Long-distance dispersal kernel

Figure 5 Relationship between invasion density (the fraction of cells invaded within the true invasion boundary) and error rate (fraction of simulations where the delimited boundary failed to encapsulate all occurrences) (Panels a, b), proportion of cells sampled within the true invasion boundary (Panels c, d), and degree of overestimation of the boundary (Panels e, f). Comparisons are made for known 'true' epicentres and unknown epicentres estimated in Stage 1 of ADD, and also for heterogeneous environments, where sampling does not take into account environmental suitability (unknown suitability) and where sampling is only performed in suitable habitats (known suitability). Relations are shown for the normal dispersal kernel and the long-distance dispersal kernel. ADD, Approach–Decline–Delimit.

DISCUSSION

One of the central requirements in invasion biology is to delimit the range of the invasion (Panetta & Lawes, 2005), once a new invader has been detected as well as for expanding invaders. With proper delimitation, we have the possibility of erecting barriers (e.g. Emerald Ash Borer, Muirhead *et al.*, 2006) or applying strategies to slow spread (e.g. Gypsy Moth, Sharov & Liebhold, 1998). Unfortunately, there has been little option but to base delimitation simply on the farthest observed location of an invader, even though, logically we should expect to have missed some occurrences near the boundary, where

there may be only a few invaded sites across a large geographical area. As illustrated in this manuscript, it can be very difficult to find the true boundary of an invasion, and such naïve approaches should often fail. Without an appropriate delimitation strategy, we will constantly be chasing the invasion as it expands into new areas.

Surprisingly, we know of few attempts to develop theory for delimitation, despite its obvious relevance (but see Mangel et al., 1984; Panetta & Lawes, 2005). The work presented here, therefore, represents an early attempt to generate useful ideas for delimitation, given a general lack of data that characterize most invasions. The conceptual logic is relatively straightforward and should be extensible. For Stage 1 (Approach), we simply suggest that we need not search linearly, especially in regions where most sites are invaded and little information exists to infer the location of the boundary. For Stage 2 (Decline), we theorize that densities of occurrences should generally decrease as one moves towards the boundary, and that often it may be possible to sample only a small fraction of cells to estimate density. For Stage 3 (Delimit), we assert that the last observed occurrence may not be the true boundary and sampling the absolute farthest occurrence will be difficult. Simply put, fitting a curve and statistically estimating the boundary based on the observed invasion pattern should be more feasible.

Each Stage of ADD can be useful under different scenarios. Where the invasion progresses as a smooth front as in many ecological spread models, Stage 1 may be sufficient to delimit an invasion. Where stochasticity exists, Stage 1 will only allow us to move nearer to the boundary. The importance of stochasticity is illustrated by the 100% failure of the naïve 'Transect' approach, which would have worked if invasions had a smooth front. Stage 1 may be particularly advantageous where invaders are cryptic and search effort is more expensive than travel effort (e.g. invertebrate invaders such as red imported fire ants (Solenopsis invicta) are difficult to detect without intensive search effort, Forbes et al., 2000). Where invaders can easily be detected at a distance, this step may not be necessary. Nevertheless, for completeness and generality, we present these ideas. Where Stage 1 is not needed, Stages 2 and 3 may remain useful.

We should sample as efficiently as possible and maximize the inferences possible; Stage 2 does this using Bayesian logic. However, where density data already exist (e.g. S. invicta in Brisbane Australia, Biosecurity Queensland Control Center, unpublished data) or when sampling cannot be adaptive (e.g. traps are more feasible to place many at once rather than adaptively based on information from previously placed traps), researchers may decide to proceed directly to Stage 3 (Delimit). Regardless, the ideas underlying ADD can be used in concert or separately, should be generalizable, and can be modified to suit the idiosyncrasies of different systems.

Although ADD worked well, many challenges remain: (1) alternatives need to be found when there is insufficient data to build the ADD algorithm. Currently, in the absence of ADD, we are limited to naïve approaches, which were shown to perform poorly. Of course, one could sample exhaustively, but the stopping rules and effort required remain unclear. Nevertheless, there is no additional sampling effort required to construct the ADD algorithm; thus, there is no disadvantage to consideration of ADD. 2) When densities of invasions are low, ADD performs sub-optimally. For instance, if less than 1% or 5% of sites are occupied, for normal- and long-distance dispersal kernels, respectively, the delimitation success rate of ADD drops below 80%. Researchers should work to derive alternative solutions. (3) Knowledge of the epicentre was shown to be important for the degree of overestimation of the boundary. In some cases, ADD could be integrated with other approaches to estimate invasion epicentres, such as tree ring analysis, which was used to reconstruct an emerald ash borer invasion in south-eastern Michigan (Siegert et al., 2008). (4) ADD provides a means to delimit an invasion. However, it should be integrated with other management strategies and other bodies of theory in the invasion biology literature. For instance, once an invasion has been delimited, managers could employ search theory to find the invaders (Cacho et al., 2006), and management options could be employed – for instance treatment zones could be initiated for eradication (Sharov & Liebhold, 1998; Edwards & Leung, 2009). Alternatively, although ACS was not useful for delimitation, after a barrier had been set using ADD, ACS could be useful to find the maximum number of invaded sites with the minimum effort (Thompson, 1990; Turk & Borkowski, 2005). (5) We did not consider the relationship between local population size and detectability. These should logically be correlated and if the data existed, would provide useful additional information. As such, other important extensions could include consideration of probabilities of non-detection within sites (Delaney & Leung, in press).

In summary, where typically few analyses have existed, we argue that the approach developed in this manuscript can set a benchmark for expectations for delimitations, from which extensions and improvements can be made – currently, ADD is theoretically operational with 1–5% area invaded depending on the dispersal kernel, and achieve >80% successful delimitations, using virtually no a priori information. ADD is relatively simple conceptually and can be adapted depending on the data that are available. It provides a starting point and many opportunities for synthesis with other existing theories. In combination with other branches of invasion biology research, this work should help address the problem of biological invasions.

ACKNOWLEDGEMENTS

REFERENCES

Cacho, O.J., Spring, D., Pheloung, P. & Hester, S. (2006) Evaluating the feasibility of eradicating an invasion. *Biological Invasions*, **8**, 903–917.

Delaney, D. & Leung, B. (in press) An empirical probability model of detecting species at low densities. *Ecological Applications*.

Edwards, P.K. & Leung, B. (2009) Re-evaluating eradication of nuisance species: invasion of the tunicate, *Ciona intestinalis*. *Frontiers in Ecology and the Environment*, **7**, 326–332.

Forbes, A.R., Mueller, J.M., Mitchell, R.B., Dabbert, C.B. & Wester, D.B. (2000) Accuracy of red imported fire ant mound density estimates. *Southwestern Entomologist*, **25**, 109–112.

Hulme, P.E. (2009) Trade, transport and trouble: managing invasive species pathways in an era of globalization. *Journal of Animal Ecology*, **46**, 10–18.

Kot, M., Lewis, M.A. & van den Driessche, P. (1996) Dispersal data and the spread of invading organisms. *Ecology*, **77**, 2027–2042.

Lodge, D.M., Williams, S.L., MacIsaac, H., Hayes, K., Leung, B., Reichard, S., Mack, R.N., Moyle, P.B., Smith, M., Andow, D.A., Carlton, J.T. & McMichael, A. (2006) Biological invasions: recommendations for U.S. policy and management. *Ecological Applications*, **16**, 2035–2054.

Mack, R.N., Simberloff, D., Lonsdale, W.M., Evans, H., Clout, M. & Bazzaz, F.A. (2000) Biotic invasions: causes, epidemiology, global consequences, and control. *Ecological Applications*, **10**, 689–710.

Mangel, M., Plant, R.E. & Carey, J.R. (1984) Rapid delimiting of pest infestations: a case study of the Mediterranean fruit fly. *Journal of Applied Ecology*, **21**, 563–579.

Muirhead, J.R., Leung, B., van Overdijk, C., Kelly, D.W., Nandakumar, K., Marchant, K.R. & MacIsaac, H.J. (2006) Modelling local and long-distance dispersal of invasive emerald ash borer *Agrilus planipennis* (Coleoptera) in North America. *Diversity and Distributions*, **12**, 71–79.

Neubert, M.G. & Caswell, H. (2000) Demography and dispersal: calculation and sensitivity analysis of invasion speed for structured populations. *Ecology*, **81**, 1613–1628.

Panetta, F.D. & Lawes, R. (2005) Evaluation of weed eradication programs: the delimitation of extent. *Diversity and Distributions*, **11**, 435–442.

Saupe, D. (1988) Algorithms for random fractals. *The science of fractal images* (ed. by H.O. Peitgen and D. Saupe), pp. 71–113, Springer-Verlag New York Inc, New York.

Sharov, A.A. & Liebhold, A.M. (1998) Bioeconomics of managing the spread of exotic pest species with barrier zones. *Ecological Applications*, **8**, 833–845.

Siegert, N.W., McCullough, D.G., Liebhold, A.M. & Telewski, F.W. (2008) Reconstruction of the establishment and spread of emerald ash borer through dendrochronological analysis. *Proceedings of 2008 USDA Research Forum on Invasive Species* (ed. by K.A. McManus and K.W. Gottschalk), p. 70, Gen. Tech. Rep. NRS-P-36, Annapolis, MD.

Thompson, S.K. (1990) Adaptive cluster sampling. *Journal of American Statistical Association*, **85**, 1050–1059.

Turk, P. & Borkowski, J.J. (2005) A review of adaptive cluster sampling: 1990–2003. *Environmental and Ecological Statistics*, **12**, 55–94.

Tyre, A.J., Possingham, H.P. & Lindenmayer, D.B. (1999) Modelling dispersal behaviour on a fractal landscape. *Environmental Modelling and Software*, **14**, 103–111.

Veitch, C. & Clout, M. (2002) Turning the tide: the eradication of invasive species. *Proceedings of the international conference on eradication of island invasives; Occasional Paper of the IUCN Species Survival Commission No. 27* (ed. by C. Veitch and M. Clout), pp. 1–3, International Union for the Conservation of Nature and Natural Resources, Gland, Switzerland.

BIOSKETCHES

Brian Leung's research interests include ecological forecasting of invasive species (which species will arrive, where they will occur, and what effects will they have), and using decision theory and bioeconomics to structure their management.

Oscar Cacho's research interests centre on the application of bioeconomics to invasive species, forestry and carbon sequestration, and management of grazing systems among others.

Daniel Spring's research interests include prediction and management of invasive species, with a particular interest in surveillance and the optimal integration of surveillance and treatment.

Author contributions: BL, OC and DS conceptualized the problem. BL derived the ADD solution. OC built the model. OC performed simulation analyses with contribution from DS. BL wrote the manuscript with contributions from OC and DS.

Editor: David Richardson