Speeding up the simulation of population spread models

Mark A. Gilbert[*,1,2], **Steven M. White**[1,2], **James M. Bullock**[2] **and Eamonn A. Gaffney**[1]

[1]*Wolfson Centre for Mathematical Biology, Mathematical Institute, University of Oxford, Radcliffe Observatory Quarter, Woodstock Rd, Oxford, Oxfordshire OX2 6GG,UK; and* [2]*Centre for Ecology & Hydrology, MacLean Building, Benson Lane, Crowmarsh Gifford, Wallingford, Oxfordshire OX10 8BB, UK*

Summary

1. Simulating spatially explicit population models to predict population spread allows environmental managers to make better-informed decisions. Accurate simulation requires high spatial resolution, which, using existing techniques, can require prohibitively large amounts of computational resources (RAM, CPU, etc).

2. We developed and implemented a novel algorithm for the simulation of integro-difference equations (IDEs) modelling population spread, including stage structure, which uses *adaptive mesh refinement*.

3. We measured the accuracy of the adaptive algorithm by comparing the results of simulations using the adaptive and a standard non-adaptive algorithm. The relative error of the population's spatial extent was low (<0·05) for a range of parameter values. Comparing efficiency, we found that our algorithm used up to 10 times less CPU time and RAM than the non-adaptive algorithm.

4. Our approach provides large improvements in efficiency without significant loss of accuracy, so it enables faster simulation of IDEs and simulation at scales and at resolutions that have not been previously feasible. As an example, we simulate the spread of a hypothetical species over the UK at a resolution of 25 m. We provide our implementation of the algorithm as a user-friendly executable application.

Key-words: habitat fragmentation, IDE, integro-difference equation, invasive species

Introduction

Modelling changes in species' distributions enable environmental managers to make better-informed habitat management (Hulme 2006) and conservation (Guisan, Tingley & Baumgartner 2013) decisions. Predicting spread has two major applications: to the problem of invasive species (Shigesada & Kawasaki 1997; Williamson 1999) and to understanding population responses to the shifting of habitat under climate change (Zhou & Kot 2011; Bullock *et al.* 2012; Bennie *et al.* 2013), habitat loss (Fahrig 1997) and habitat fragmentation (Fahrig 2002). This provides motivation for spatially explicit population models which represent the relevant biological processes mathematically and enable the prediction of population dynamics. As population spread happens over large (Continental/National) spatial scales, large-scale models are required. One way of doing this is via *species distribution models* (e.g. Elith & Leathwick 2009), but these are static and predict only potential ranges which may not be realised (Zhu, Woodall & Clark 2012). They do not generally include the biological mechanisms of demography or dispersal which are likely to influence the rate and extent of spread (Schurr, Pagel & Cabral 2012) and have a comparatively weak underpinning in ecological theory (Thuiller *et al.* 2013). SDMs are trained only on data from locations within the range of the species. Predictions outside of these ranges may be unrealistic as they are beyond the scope of the data, making predictions at range limits or

scenario testing (e.g. adding another species or biological control) unreasonable. These problems can be overcome by using mechanistic models that explicitly include demography and dispersal. Mechanistic models take several forms, including *partial differential equations* (PDEs), *integro-difference equations* (IDEs), *individual-based models* (IBMs) and some *metapopulation models*, with the choice of model depending on the temporal and spatial scales underpinning the system and the form of the biological processes (e.g. whether dispersal can be characterised by diffusion or is more complex). Another option is hybrid models that use components of SDMs and mechanistic models, but these are vulnerable to over-fitting, caused by needing to estimate or assume large numbers of parameter values (Wisz, Pottier & Kissling 2013).

In many mechanistic population models, landscapes are assumed to be homogeneous, with the demographic and dispersal parameters spatially constant. This is because homogeneity (i) reduces the number of independent parameters in the model, making model specification easier and (ii) reduces the amount of computational resources required to determine the spread rate. In homogeneous landscapes, population growth can be characterised by a single scalar growth rate. Similarly, in spreading populations, one can calculate the spreading speed (the annual increment in the population's range) and how sensitive this is to life-history and dispersal parameters (Shigesada, Kawasaki & Teramoto 1986; Kot, Lewis & den Driessche 1996; Neubert & Caswell 2000). For landscapes where spatial variation is very slight (Gilbert *et al.* 2014a) or highly localised (Dewhirst & Lutscher 2009), the

results for homogeneous landscapes can be used as an approximation, with analogous results extending to some periodic (e.g. regular, repeating) landscapes (Gilbert *et al.* 2014b). However, landscape homogeneity is often a poor assumption and in many landscapes the dispersal and demographic parameters exhibit strong variation across large spatial scales, rendering this approximation inaccurate (Miller & Tenhumberg 2010; Svenning, Gravel & Holt 2014). Instead, to make predictions on complex landscapes, numerical simulations of the models are required. For these predictions to be accurate, both accurate ecological data and accurate simulation of the model are required. Accurate simulations require a large number of calculations (Scheffer *et al.* 1995). This means that intense computational power is needed to simulate spread in large two-dimensional landscapes. This is a particular problem for IBMs, such as RangeShifter (Bocedi *et al.* 2014), where a large number of individual mechanisms can be modelled, but where each organism must be explicitly represented in the numerical framework.

As predictions of species spread rely on computationally expensive simulations, the ability to predict spread is often limited by the model's scale and the available computational resources. This limitation inhibits the types of scientific questions and problems that can be investigated. For example, fitting spread models to data requires numerous model runs, so if each model run is computationally intensive, then the time to fit the model to data will be prohibitively long.

To be simulated numerically, most mechanistic population models must be discretised, with space and time represented as a mesh of points. The simulation's accuracy depends on the number of points, with the accuracy increasing with the number of points (Bocedi *et al.* 2012). Simulating spatially explicit models involves a large number of calculations at each time step. The computational complexity and computational resources (RAM, CPU time, etc) required to compute these calculations increase linearly or faster with the total number of points or individuals (Cooley & Tukey 1965; Solodovnikov 1985). If simulations over large landscapes are to be accurate, they need a high density of points and a very large total number of points. In addition, most models will require numerous iterations if the model is simulated over a long time, making simulations extremely computationally expensive. Large numbers of runs are also often required for parameter estimation and sensitivity studies, thus further motivating the need for efficient solvers (Csilléry *et al.* 2010). One way to improve efficiency has been developed for PDEs, using *adaptive mesh refinement* to allow the density of points (the resolution) to change in time and space (e.g. Davis & Flaherty 1982; Berger & Oliger 1984). The density of points is usually highest where the most accuracy is required (e.g. where the relative population change is greatest, such as in the wave-front). With only small reductions in accuracy (i.e. the accuracy of the simulations compared to the mathematical model), this approach allows limited computational resources to be focused on the spatial locations which have the greatest effect on the distribution as it changes. Adaptive mesh refinement has so far been restricted to PDEs. However, in cases where long distance

dispersal or annual cycles play an important role in the model, IDEs may be a more appropriate choice of modelling framework than PDEs, as they can incorporate these processes (Kot, Lewis & den Driessche 1996) and are relatively easy to parameterise, e.g. by using data from the COMPADRE database (Salguero-Gómez, Jones & Archer 2015) or population spread data (e.g. Gilbert *et al.* 2005), and thus, IDEs are extensively used in species modelling (Neubert & Parker 2004; Le Corff & Horvitz 2005; Bullock, Pywell & Coulson-Phillips 2008; Miller & Tenhumberg 2010; Zhou & Kot 2011; Bullock *et al.* 2012).

In this study we provide a solution to large-scale spread problems by developing an adaptive mesh refinement algorithm to simulate IDEs. At each iteration, the algorithm divides the landscape up into square *regions* of equal area that are either *high resolution* (containing a large number of mesh points) or *low resolution* (containing only one mesh point). We present the algorithm for stage-structured populations with homogeneous dispersal and heterogeneous demography in Section 'Materials and methods'. In Section 'Results', we demonstrate that the algorithm is both highly accurate and uses fewer computational resources (RAM, CPU time) than non-adaptive algorithms for a given problem, as well as enabling us to tackle problems that were previously impossible due to insufficient computational resources. Finally, we discuss extending the algorithm to landscapes with heterogeneous dispersal in the Discussion (Section 'Discussion'). We provide an executable application with a worked example via GitHub (Gilbert *et al.* 2016) so that others may use this algorithm to simulate the spread of species under a variety of scenarios.

Materials and methods

In this section, we introduce single species stage-structured IDEs with constant dispersal and show how to simulate them with adaptive meshes and achieve improvements in computational efficiency compared to non-adaptive meshes. We restrict ourselves to IDEs on two-dimensional landscapes, with the method easily restricted to one spatial dimension. The technical details of the algorithm are presented in full in Appendix S2.

Integro-difference equations are discrete time, spatially continuous population models that have been used to model many different organisms including annual (Bullock, Pywell & Coulson-Phillips 2008) and perennial herbs (Le Corff & Horvitz 2005), shrubs (Neubert & Parker 2004), trees (Bullock *et al.* 2012), weevils (Miller & Tenhumberg 2010) and butterflies (Zhou & Kot 2011). Their strength lies in the trade-off between simplicity, flexibility and ease of use in terms of parametrisation and implementation; they can incorporate a range of dispersal mechanisms and demographic structures, including stage-structured populations which represent each life-history stage explicitly (Neubert & Caswell 2000), but with few parameters. The general stage-structured two-dimensional IDE on a landscape Ω relates the stage-structured vector population density $\mathbf{u}^{t+1}(\mathbf{x})$ at each location $\mathbf{x}$ for generation $t + 1$ with the population for generation t via

$$\mathbf{u}^{t+1}(\mathbf{x}) = \int_{\Omega} \left[\mathbf{K}(\mathbf{x} - \mathbf{y}, \mathbf{y}) \circ \mathbf{B}(\mathbf{u}^t(\mathbf{y}), \mathbf{y}) \right] \mathbf{u}^t(\mathbf{y}) d\mathbf{y}, \qquad \text{eqn 1}$$

where $\circ$ denotes the Hadamard (element-wise) product of two matrices. Here, the population undergoes two consecutive phases, growth and dispersal. The growth phase incorporates the processes of birth,

maturation and death that occur during a generation. Growth is local and depends only on the location **x** and on the population $\mathbf{u}^t(\mathbf{x})$ at **x** at time t. Locations may differ in terms of habitat, meteorological conditions, soil quality, etc., impacting growth rates and carrying capacities. The current population may affect demographic rates through intraspecific competition or via Allee effects. Population growth is determined by the population projection matrix $\mathbf{B}(\mathbf{u}^t(\mathbf{y}),\ \mathbf{y})$, where the (i,j)-th element is the ratio of the number of individuals in stage j after the growth phase and the number of individuals in stage i at time t (at location **y**), e.g. the survival probability between juveniles and adults. The dispersal phase is the process by which individuals (or propagules) move between locations. In stage-structured IDEs, it is characterised by the matrix of dispersal kernels $\mathbf{K}(\mathbf{x} - \mathbf{y}, \mathbf{y})$, where the (i,j)-th element $K_{i,j}(\mathbf{x} - \mathbf{y},\ \mathbf{y})$ gives the density of population that has transitioned from stage i to j and has transferred between locations **y** and **x**, e.g. dispersal of seeds from adult plants. Each pair of demographic stages is considered separately as individuals' dispersal depends on both their current stage and their stage prior to the growth phase.

In general, the dispersal kernel depends on the disperser's starting location **y** and the displacement $\mathbf{x} - \mathbf{y}$ (as in Fig. 1e). However, in this study, we consider only models where the dispersal kernel is spatially constant and has no explicit **y** dependence (see Fig. 1d), which is a useful simplifying assumption. Prospects for relaxing this assumption are reviewed in the Discussion.

With the exception of homogeneous landscapes (see section 'Analytical Results' in Appendix S1, Supporting Information), quantifying spread requires numerical simulation, with the landscape typically discretised into a square mesh (Powell, White & McMillen 1998). If dispersal is homogeneous, then the integral in (1) can be computed efficiently using fast Fourier transforms (FFTs) (Andersen 1991). If dispersal varies spatially, then the discretised spatial distributions at t and $t + 1$ are related by a matrix multiplication. Considering either case, the computational resources required increase linearly or faster with the number of mesh points, leading to very large computational costs for simulating the model accurately over large and complex landscapes (see Appendix S1). To reduce these computational requirements, we propose a simulation algorithm with adaptive mesh refinement.

In spreading populations, there is usually little change in the population density in the *wave-back* (or within range), where the population has reached carrying capacity, or in the *far-field*, where the population is very small (see Fig. 1a). Behaviour in these regions has only a small effect on the wave-front's behaviour, so computational resources can be focused on the wave-front (or leading edge) with a loss of resolution in the wave-back and far-field. For our novel algorithm, we subdivide the landscape into non-overlapping square *regions* of equal width Dx, which are further divided into an equal number of *cells* of width dx. Regions can be either *High Resolution* (HR) or *Low Resolution* (LR). High-resolution regions
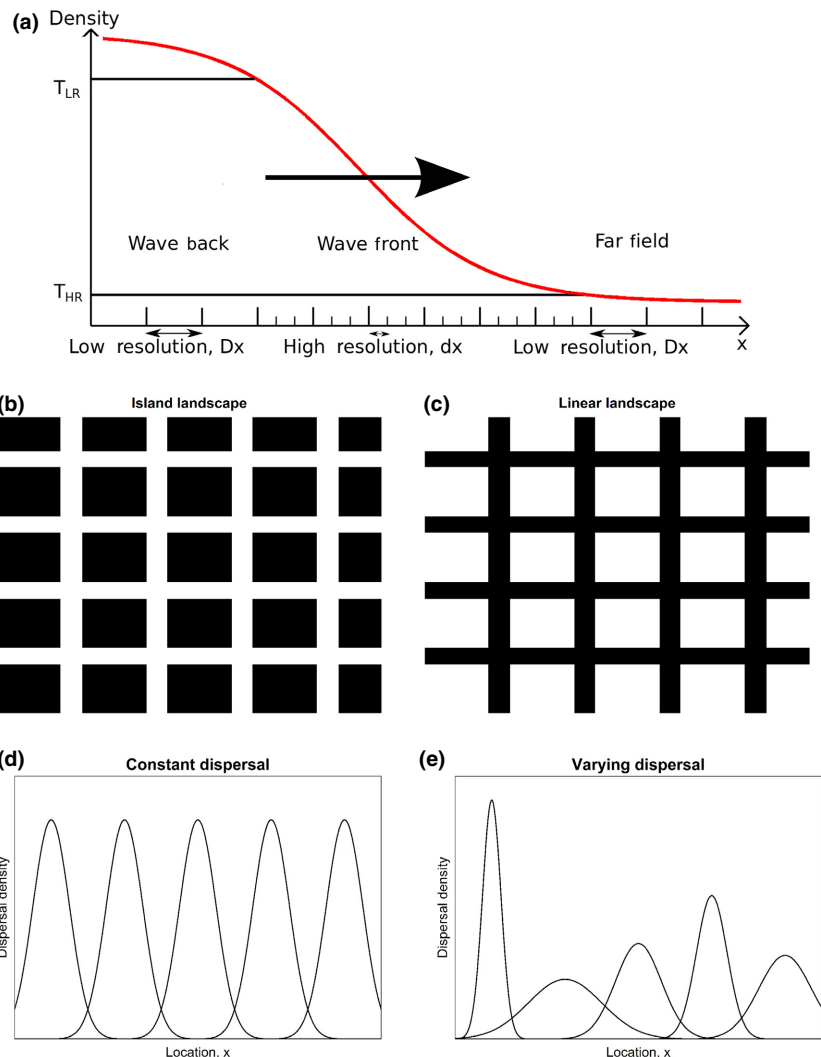

Fig. 1. (a) Schematic diagram of a travelling wave showing the algorithm parameters: the upper threshold T_{LR} where the population returns to low resolution *LR*, the lower threshold T_{HR} where the population becomes *HR*, the distances between each cell dx and between each region Dx in relation to the spreading population (red online). (b, c) Heterogeneous Landscape Scenarios used to test the accuracy and efficiency of the model, with suitable habitat shown in black and unsuitable habitat in white. The suitable habitat in (b) is a regular grid of unconnected square islands separated by unsuitable habitat. The suitable habitat in (c) is a connected linear habitat with islands of unsuitable habitat. (d, e) One-dimensional dispersal kernels at select points with (d) constant dispersal parameters, and (e) varying dispersal parameters (for comparison).

have a mesh point for each cell, but low-resolution regions are represented by a single mesh point.

The algorithm requires the following inputs: maps (provided as raster, i.e. gridded data at the desired resolution) of the initial population distribution, demographic parameters and corresponding growth functions over the entire landscape and the (spatially homogeneous) dispersal kernel. Initially, any region in which any mesh point is in the *wave-front* is high resolution (see Fig. 1a). The wave-front is any location where the juvenile (or lowest) demographic stage, $u_1^t(\mathbf{x})$, has population density between the thresholds T_{LR} and T_{HR}. The upper threshold T_{LR} is taken as a proportion of the carrying capacity and T_{HR} is set as small as required to allow small populations to be resolved (e.g. $T_{HR} = 10^{-13}$ in the examples in Sections 'Materials and methods' and 'Results'), with suitable values provided for these parameters in the worked example.

In standard simulations of IDEs there are two sequential processes during each generation: population growth and dispersal. However, the adaptive algorithm also incorporates the additional process of mesh refinement.

1. *Population Growth* at each cell depends on the cell's intrinsic demographic rates and carrying capacity, which are supplied by the user. For high-resolution regions, population growth is calculated by applying each cell's population projection matrix $\mathbf{B}(\bullet, \mathbf{y})$ to its stage-structured population density $\mathbf{u}^t(\mathbf{y})$. For low-resolution regions, each demographic stage's population density is represented by a single scalar, so the growth phase for the whole region is calculated using the mean of the population projection matrix over the region (see Appendix S2, equation (B2)).

2. *Dispersal*. The assumption of homogeneous dispersal reduces the integral for each pair of demographic stages in (1) to a convolution and allows dispersal within a regular square mesh to be calculated using FFTs (Andersen 1991). The key idea with our method is to divide dispersal into two categories of *dispersal resolution*.

(a) *High-resolution dispersal* is used to denote the dispersal processes which require more computational resources for simulations to be accurate. High resolution is required for dispersal that (i) is relatively short distance and (ii) has its destination in the wave-front. In practice, this is dispersal to a high-resolution region (a region in the wave-front) from any neighbouring region (short distance). For a region containing high-resolution information, we take the pre-dispersal population density of each cell in the region and its neighbours (for low-resolution neighbouring regions, all cells are assumed to have population density equal to the region's population density) and take the discrete convolution of this distribution with the *HR dispersal kernel* (a high-resolution discretisation of the dispersal kernel) to get the post-dispersal population density distribution.

(b) *Low-resolution dispersal* is used to denote the dispersal processes which can be simulated with fewer computational resources without significantly affecting the simulation's accuracy. This is dispersal between regions which are not both neighbours and in the wave-front. This is justified by the observation that all realistic dispersal kernels eventually decay after sufficient distance from the origin. For each pair of demographic stages, the dispersal between non-neighbouring regions is calculated by taking the convolution of the array of mean pre-dispersal population densities for each region with the *long-distance LR dispersal kernel*, a low-resolution discretisation of the dispersal kernel restricted to dispersal between non-neighbouring regions. For dispersal to low-resolution regions from neighbouring regions, we take the mean pre-dispersal population densities for the region and its neighbours and take the discrete convolution of this distribution with the *local LR dispersal kernel*, a low-resolution discretisation of the dispersal kernel

restricted to dispersal between neighbouring regions. Finally, the contributions of low-resolution dispersal are added to the post–high-resolution dispersal population distribution. For low-resolution regions, this involves the addition of two scalars (for each pair of demographic stages). For high-resolution regions, the contribution of long distance dispersal to a region is approximated as equal for all cells, so the region's global dispersal density is added to each cell's population density.

In summary, picking and choosing which level of dispersal resolution to use, based on the change in population and distance between locations, allows gains in efficiency with little reduction in accuracy (See Appendix S2 for full details).

1. *Mesh refinement* is where the resolution of each region is changed to ensure that while the population distribution and the wave-front (the area with population density between the lower T_{HR} and upper T_{LR} thresholds) change, the mesh always has high resolution where needed. Individuals in the far field may drive population expansion (Lewis 2016), so the choice of T_{HR} will depend on the IDE model being studied and is in general set as low as required to ensure no significant change in model prediction. The mesh refinement phase enables regions which enter the wave-front to become high resolution, and regions which leave to return to low resolution. Regions in the far-field of the wave-front where the juvenile stage's minimum population density has crossed the T_{HR} threshold become high resolution. Regions in the wave-back which have either a juvenile population that has crossed the T_{LR} threshold, or densities which have changed less than a further parameter $T_{LR,2}$, become low resolution. As regions with only zero intrinsic population growth rates will always have zero population, they will always remain low resolution (making the algorithm very efficient for sparse landscapes).

The algorithm was implemented in C++. An executable which runs on Windows operating systems, is downloadable from GitHub (Gilbert *et al.* 2016); source code for non-Windows machines is available upon request.

We measured the algorithm's accuracy in simulating a non–stage-structured spreading population compared to a non-adaptive simulation algorithm (Section 'Accuracy'), its efficiency in CPU time and RAM usage (Section 'Efficiency') and tested it by simulating hypothetical biological invasions into coniferous woodland with (i) a single demographic stage across Great Britain, and (ii) two demographic stages into the New Forest (Hampshire, UK). In both cases, we used fine-scale mapping data from the CEH Land Cover Map 2007 (LCM2007) in Section 'Examples'. We also demonstrated the accuracy of our algorithm for stage-structured populations in Appendix S5.

Throughout the following section, we simulate IDEs of the form (1) with a single demographic stage and the commonly used 2D Laplace (exponential) dispersal kernel (Carrasco *et al.* 2010) with spatially constant dispersal parameter α, given by

$$k(\mathbf{x} - \mathbf{y}, \mathbf{y}) = \frac{1}{2\pi\alpha^2} \exp\left(\frac{-\|\mathbf{x} - \mathbf{y}\|}{\alpha}\right), \qquad \text{eqn 2}$$

where $\|\mathbf{x} - \mathbf{y}\|$ is the distance between $\mathbf{y}$ and $\mathbf{x}$. We choose a piece-wise linear scalar population growth function due to its simplicity. Other stage-structured population projection matrices and single-stage growth rates, including those with Allee effects, can also be used. We have tested the algorithm for a growth function with an Allee effect and found that the simulations were accurate for selected examples. For the piece-wise linear growth function, the population growth rate is constant below the carrying capacity C, with the population after the growth phase at $\mathbf{x}$ given by

$$r(\mathbf{u}^t(\mathbf{x}), \mathbf{x})\mathbf{u}^t(\mathbf{x}) = \min(r_0(\mathbf{x})\mathbf{u}^t(\mathbf{x}), C) \qquad \text{eqn 3}$$

where $r_0(\mathbf{x})$ is the intrinsic population growth rate at $\mathbf{x}$. Without loss of generality, we take the carrying capacity $C = 1$ throughout this section. In Sections 'Accuracy' and 'Efficiency' we take $\alpha = 1$, but in Section 'Examples' we simulate over a large real landscape, and take $\alpha = 50$ m to allow wider dispersal.

When measuring the algorithm's accuracy (Section 'Accuracy') and efficiency (Section 'Efficiency'), we take the initial distribution to be along one side of the square domain. For the example of spread of a specialist into a fragmented coniferous habitat in the UK as given by the LCM 2007 (Section 'Examples'), we took the initial distribution to be localised at a single cell.

Results

ACCURACY

We tested the accuracy of the adaptive algorithm presented in Section 'Materials and methods' by comparing its output with a high-resolution non-adaptive algorithm across a range of parameters and landscape types. We varied four user-defined tolerance parameters to investigate the trade-off between accuracy and speed, as well as two model parameters to determine how the demography and landscape pattern affect the algorithm's accuracy. For a spatially homogeneous landscape and the spatially periodic linear and island landscapes presented in Fig. 1b,c, we simulated both algorithms while varying six different parameters to explore the relationship among the parameter value, the landscape type and the algorithm's accuracy. We varied: the density threshold at which a region returns to low resolution once in the wave-back T_{LR}, the density threshold at which a region in the wave-front becomes high resolution T_{HR}, the cell size dx, the region size Dx, the growth rate in suitable habitat r and the landscape period p (the spatial distance taken for landscape features to repeat).

We measured the accuracy using the area occupied (Appendix S3) and the furthest distance reached by the population (Appendix S4), and found similar responses to the different parameters in both cases. For standard and default parameter values, the algorithm gave accurate results for the scenarios considered. For all three landscape types, simulation errors were minimised for smaller values of dx, larger values of T_{LR} and r and intermediate values of Dx. The relationship between T_{HR} and error differed between landscape scenarios and there was no straightforward relationship between p and simulation error. To show that our algorithm gives accurate results for stage-structured populations, we tested it against a stage-structured example and present our results in Appendix S5.

EFFICIENCY

The adaptive algorithm was tested for efficiency against a non-adaptive algorithm. We varied the parameters that have the greatest effect on the algorithm's computational complexity as these demonstrated the differences in performance between the adaptive and non-adaptive algorithms. Different values to the

ones used to test the accuracy were used as performance gains are only possible for larger simulations (see Table 1). We varied the landscape/domain length, L, the number of generations, T, and the ratio M of the region width Dx to the cell width dx, $M = \mathrm{D}x/\mathrm{d}x$ (Fig. 2).

For each of the three parameters varied, 24 values were investigated under the three landscape scenarios introduced in Section 'Accuracy'. For each parameter value and landscape scenario, the average of 10 runs was taken. The following performance measures were used:

1. *CPU time.* The total time used by the processor on the simulation. This ignores any interruptions from other tasks, so is generally less than the actual time taken for the simulation, and was measured via the UNIX command `time()`.

2. *Maximum RAM* is measured by the maximum *resident set size* over the programme's run that is the amount of main memory (RAM) occupied by the process.

In Fig. 2, we plot the adaptive algorithm's performance against the performance of the non-adaptive algorithm for the three different landscape scenarios and for the three different parameters. In all cases, the adaptive algorithm was generally faster than the non-adaptive algorithm, particularly as the domain length L, total number of generations, T, and the width in cells of each region M increased, with the adaptive algorithm using up to 10 times fewer resources than the non-adaptive algorithm (Fig. 2a,c,e). The adaptive algorithm also achieved a lower maximum RAM usage than the non-adaptive algorithm, especially as the domain length L and the ratio of region width to cell width $M = \mathrm{D}x/\mathrm{d}x$ increased.

EXAMPLES

To run the algorithm, the landscape map must be saved in the same directory as the executable, and the model and algorithm parameters (which determine the speed and

Table 1. Table of default parameter values (used when the parameter is not being varied)

| Parameter	Description	Default value (Accuracy)	Default value (Efficiency)
T_{LR}	HR to LR density threshold	0·5	0·5
$T_{LR,2}$	HR to LR density change threshold	10^{-10}	10^{-10}
T_{HR}	LR to HR density threshold	10^{-13}	10^{-13}
dx	cell size	0·1	0·4
Dx	Region size	10	40
r	Growth rate (suitable habitat)	3	3
p	Landscape period	3	3
L_1	Habitat length	160	800
L_2	Habitat width	80	400
α	Dispersal parameter	1	2
T	Number of generations	40	40
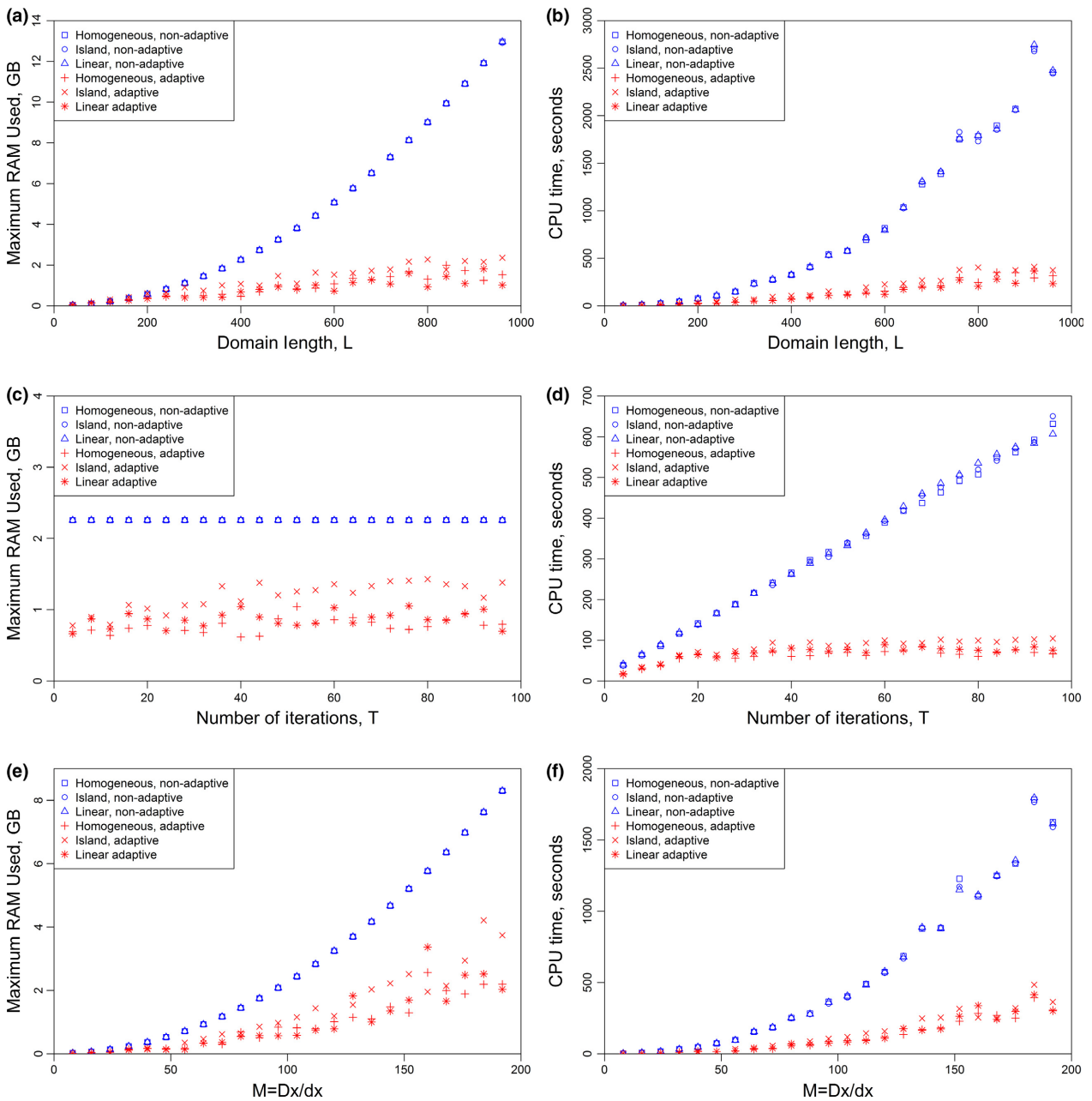
Fig. 2. Effect of varying parameters on the average CPU time (a, c, e) and the maximum RAM (b, d, f) used by 10 simulations. We vary (a, b) the domain length L, (c, d) the number of generations T and (e, f) the width of each region in cells $M = Dx/dx$. For each parameter varied, we show the results for all three landscape scenarios for both the adaptive and non-adaptive algorithms.

accuracy of the algorithm) must be set in the parameter text file. The parameter file includes the option of specifying the location of a point release for the spreading population, or of defining an initial population through another map file. The executable and user guide are available from GitHub (Gilbert *et al.* 2016).

UK land cover map

To assess our algorithm's performance on high-resolution maps of large landscapes, we simulated an invasion of a hypothetical specialist non–stage-structured species that can only inhabit coniferous woodland habitat using data from the 25-m

resolution CEH Land Cover Map, LCM2007, covering the whole of Britain (Morton *et al.* 2011). LCM2007 has $1 \cdot 456 \times 10^9$ individual cells and our ability to simulate the model over this scale demonstrates the usefulness of the algorithm (the maximum RAM used by the adaptive algorithm is seven times less than the non-adaptive).

We present the results of this simulation in Fig. 3, with the parameter values: growth rate $r^* = 200$ in coniferous habitat and zero elsewhere (non-coniferous terrain and sea/ocean), dispersal parameter $\alpha = 50$ m, region size $Dx = 2 \cdot 5$ km, landscape length $L_1 = 1300$ km and landscape width $L_2 = 700$ km. All other model parameters took the values tabulated under default accuracy values in Table 1.

Stage-structured example

To demonstrate the opportunities our algorithm offers for modelling stage-structured spread that our algorithm offers, we simulated the spread of a population with two demographic stages (juvenile and adult) across part of the New Forest (Hampshire, UK). The hypothetical species has the same demographic and dispersal parameters as the example in Appendix S5, except its population projection matrix is given by
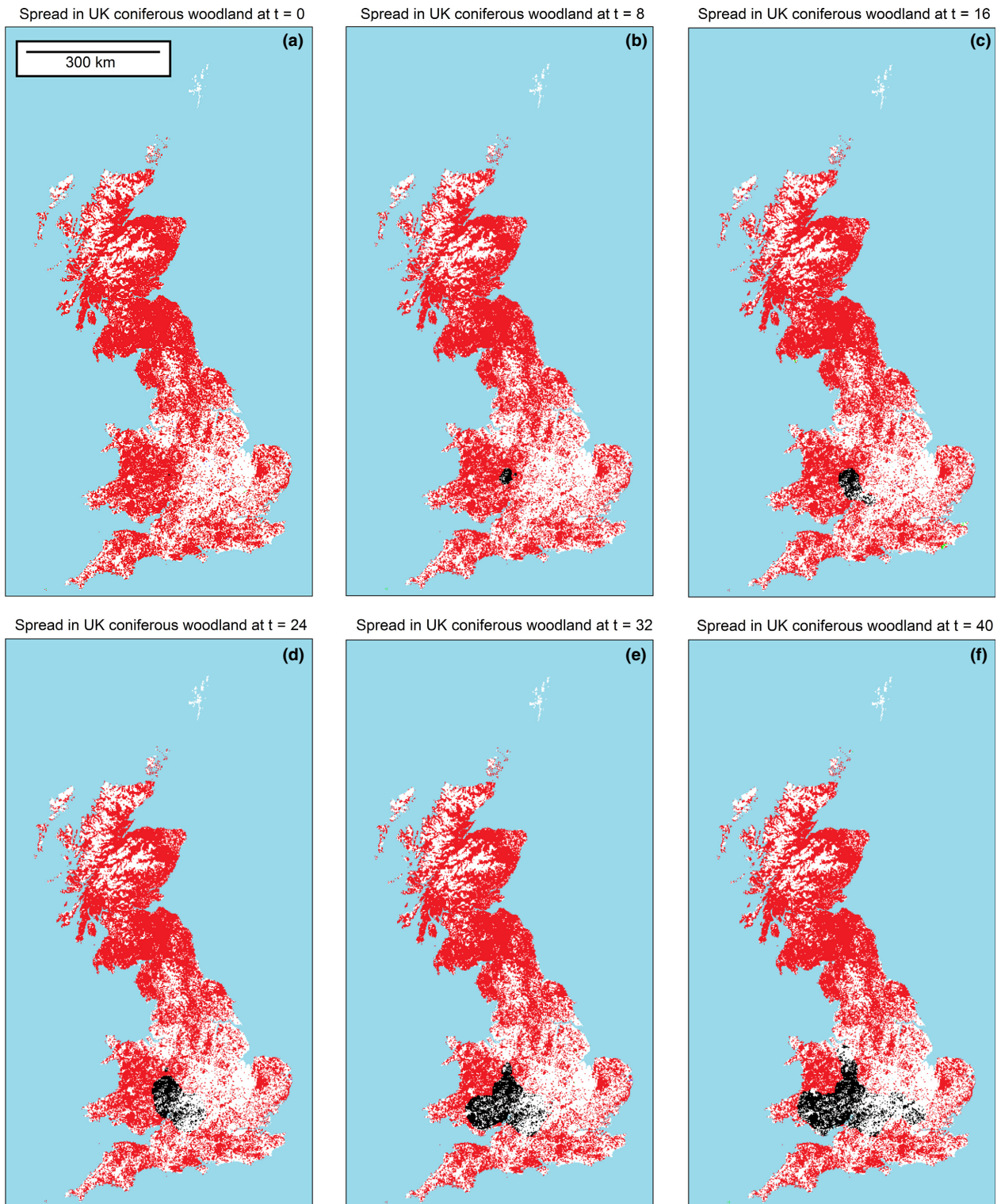

Fig. 3. Spread of a hypothetical invasive species (black online) into UK coniferous woodland (green online) shown at (a) $t = 0$, (b) $t = 8$, (c) $t = 16$, (d) $t = 24$, (e) $t = 32$, (f) $t = 40$. The population growth rate in the coniferous habitat $r = 200$ and the dispersal parameter $\alpha = 2$. The blue region represents salt water (oceans and seas) which together with terrestrial non-habitat has a zero population growth rate.

$$\mathbf{A} = \begin{bmatrix} 0{\cdot}2 & 30 \\ 0{\cdot}7 & 0{\cdot}5 \end{bmatrix} \qquad\qquad \text{eqn 4}$$

in coniferous habitat, and zero elsewhere. The other (algorithm) parameters are given by the default values in Table 1. We present the results of this simulation in Fig. 4, which show the areas occupied by juveniles (Fig. 4a) and adults (Fig. 4b) at each time step.

Discussion

Spatially explicit, mechanistic population models facilitate a quantitative understanding of changing population distributions that incorporate dispersal and demographic behaviour. These models enable predictions of population spread for any species and landscape fitting the assumptions of the model. Under a limiting set of assumptions, spread rates of spatial models can be calculated analytically (Shigesada, Kawasaki & Teramoto 1986; Kot, Lewis & den Driessche 1996; Neubert & Caswell 2000), but in general, numerical simulation must be used to predict the dynamics of changing population distributions in realistic scenarios, particularly when landscapes are either inhomogeneous or aperiodic. For all spatially explicit models, the accuracy of results obtained through simulation depends on the quality of the data used as input to the model and on the resolution of the discretisation of the landscape, with more grid points providing greater accuracy, but requiring more computational resources.

In this study we have developed an adaptive mesh method which facilitates the simulation of IDE models of spatial spread efficiently and without significant losses in accuracy. This method allows ecologists to simulate spatial dynamics for cases that have previously been too computationally demanding even on computer clusters, and thus permits previously unachievable research questions to be addressed, and allows researchers to get more out of limited computational resources. In particular, we show that it is possible to simulate spread straightforwardly over very large areas with appropriate resolution. This adaptive algorithm has the additional cost of a mesh refinement phase, but this is significantly smaller than the savings in computational resources in the growth and dispersal phases. We implemented this algorithm in C++ and have made an executable available on GitHub (Gilbert *et al.* 2016).

For a range of parameter values and landscape types, we found that the adaptive algorithm was accurate and that for IDEs on large and/or high-resolution domains, in the tests (Section 'Accuracy' and 'Efficiency'), the adaptive algorithm used up to 10 times less CPU time and maximum RAM than the non-adaptive algorithm. The algorithm was also applied to stage-structured IDE (Appendix S5), and was found to be highly accurate.

The loss of resolution in the far-field and wave-back of the invading population introduces some error. In particular, if T_{LR} is too low, the errors in the wave-back may affect the spreading behaviour and if T_{HR} is too high, large errors will appear in the wave-front that may significantly affect spread. These effects are further exacerbated by landscape heterogeneity, which affects the spreading population. As with most numerical solvers (Berger & Oliger 1984), there is a trade-off between model accuracy and efficiency, and the simulation parameters must be chosen to reduce both computational resources and the errors added by the numerical scheme. The acceptable level of error will depend on the application, but when calculating wave speeds an error of a few per cent will usually be acceptable. We suggest that users use the parameter values used in our examples (Table 1), and then lower the error tolerance and see if the solution significantly changes. If a significant change occurs, then a lower error tolerance is probably required.

To further demonstrate the algorithm's efficiency, we simulated a hypothetical invasive species' spread across coniferous woodland in Great Britain using the CEH LCM2007 habitat map of the UK. Using a non-adaptive algorithm to simulate over the whole of Great Britain at a resolution of 25×25 m^2 would require around 300 GB of RAM, which is not presently

(a) Spread of juvenlies
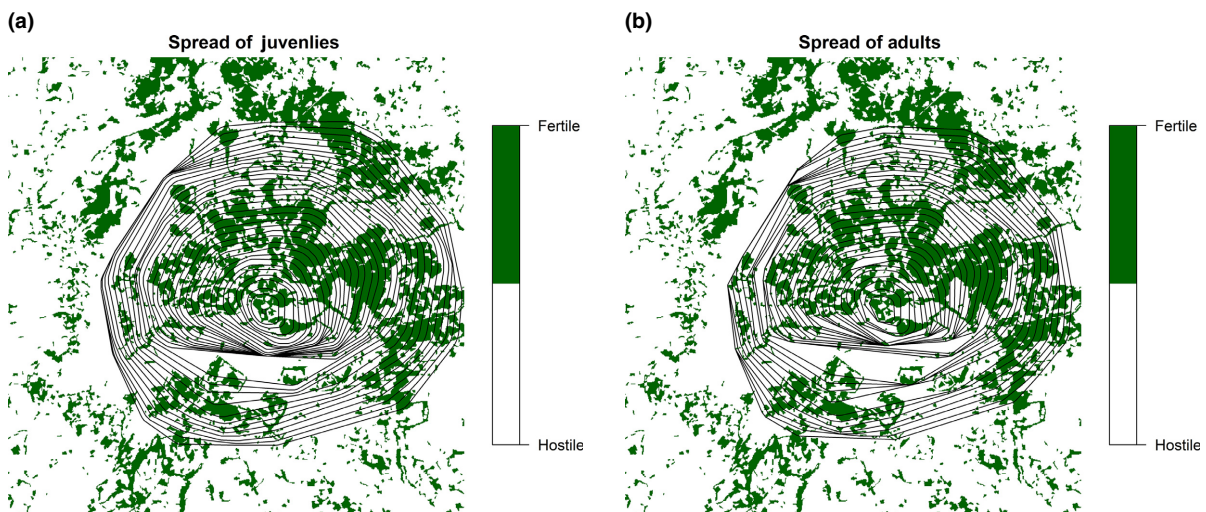
(b) Spread of adults

Fig. 4. Spread of a hypothetical stage-structured invasive species into an area of the New Forest (green online) with the areas occupied by (a) the juvenile and (b) the adult stage shown at 1-year intervals (black boundary).

possible on a regular desktop computer. For the same problem, the adaptive algorithm used a maximum of 40 GB RAM and therefore enables the running of simulations that would have been impossible with non-adaptive algorithms.

In particular, the adaptive algorithm enables the simulation of the spread of species over large landscapes with variation in demographic parameters or complex fragmentation patterns. Spatial variation in demography is commonly observed (Oostermeijer *et al.* 1996), with environmental variation having a strong relationship to parameter variation (Bullock, Silvertown & Hill 1996). Fine-scale variation has been identified as a strong determinant of spread and persistence (Carrasco *et al.* 2010; Bennie *et al.* 2013). Therefore, incorporating high-resolution spatial variation into large-scale models will enable more accurate predictions as well as greater insight into the effectiveness of fine-scale ecological interventions such as landscape buffers and wildlife corridors in preventing and facilitating spread. Indeed, models of spread along corridors, stepping stones, etc. in realistic landscapes have, to date, been limited with simplifying assumptions about spatial extent, structure and dynamics (Moilanen 2011; Fennell *et al.* 2012; Hodgson *et al.* 2012). Our method will facilitate more complex modelling over scales and resolutions appropriate for conservation policy and planning.

A natural extension to the method presented in this study is the inclusion of dispersal heterogeneity (Fig. 1e) and density-dependent dispersal, which is required to simulate organisms which settle based on habitat quality or population density as well as distance (e.g. Bonte, Van Dyck & Bullock 2012). This is more computationally intensive to simulate, and simulating these IDEs adaptively would provide even greater relative improvements in speed and efficiency than for homogeneous IDEs. Without dispersal homogeneity, the integral in (1) cannot be calculated using FFTs. Instead dispersal is normally computed by multiplying the vector of discretised population densities by an $N \times N$ matrix, where N is the number of points in the discretisation, requiring $O(N^2)$ floating point operations. An adaptive algorithm for simulating IDEs with heterogeneous dispersal would reduce the size of the matrix to $\hat{N}^2$, where $\hat{N}$ is the number of cells in regions with high-resolution added to the number of regions with low-resolution information, leading to even larger reductions in computational resources than for the homogeneous cases presented in this study. Extending the algorithm to heterogeneous and density-dependent dispersal would vastly extend the types of models that could be simulated, and is for future development.

Authors' contributions

M.G., E.G., S.W. and J.B. conceived the ideas and structure of the Adaptive algorithm; M.G. led the coding and implementation of the algorithm. All authors contributed to the manuscript and gave final approval for publication.

Data accessibility

An executable of our implementation of the algorithm which runs on Windows operating systems, together with the source code, is downloadable from GitHub (Gilbert *et al.* 2016, http://dx.doi.org/10.5281/zenodo.45509). All mapping data used in this study are from the CEH Land Cover Map, LCM2007 (Morton *et al.* 2011). The Land Cover Map has not been archived as the authors do not have permission to do so, but it is freely available to academics and can be requested from http://www.ceh.ac.uk/data-request-form?ref_nid = 16315&node_title = Land%20Cover%20Map%202007. To replicate the examples presented here, users would need to request the *Land Cover Map 2007 – 25 m raster* data set from the Data Product list on the request form and should request the following co-ordinates: $X_{min} = 0$, $Y_{min} = 0$, $X_{max} = 28\,000$ and $Y_{max} = 52\,000$.

References

Andersen, M. (1991) Properties of some density-dependent integrodifference equation population models. *Mathematical Biosciences*, **104**, 135–157.

Bennie, J., Hodgson, J.A., Lawson, C.R., Holloway C.T.R., Roy D.B., Brereton, T., Thomas, C.D. & Wilson, R.J. (2013) Range expansion through fragmented landscapes under a variable climate. *Ecology Letters*, **16**, 921–929.

Berger, M.J. & Oliger, J. (1984) Adaptive mesh refinement for hyperbolic partial differential equations. *Journal of Computational Physics*, **53**, 484–512.

Bocedi, G., Pe'er, G., Heikkinen, R.K., Matsinos, Y. & Travis, J.M. (2012) Projecting species' range expansion dynamics: sources of systematic biases when scaling up patterns and processes. *Methods in Ecology and Evolution*, **3**, 1008–1018.

Bocedi, G., Palmer, S.C., Pe'er, G., Heikkinen, R.K., Matsinos, Y.G., Watts, K. & Travis, J.M. (2014) Rangeshifter: a platform for modelling spatial eco-evolutionary dynamics and species' responses to environmental changes. *Methods in Ecology and Evolution*, **5**, 388–396.

Bonte, D., Van Dyck, H., Bullock, J.M., *et al.* (2012) Costs of dispersal. *Biological Reviews*, **87**, 290–312.

Bullock, J.M., Silvertown, J. & Hill, B.C. (1996) Plant demographic responses to environmental variation: distinguishing between effects on age structure and effects on age-specific vital rates. *Journal of Ecology*, **84**, 733–743.

Bullock, J.M., Pywell, R.F. & Coulson-Phillips, S.J. (2008) Managing plant population spread: prediction and analysis using a simple model. *Ecological Applications*, **18**, 945–953.

Bullock, J.M., White, S.M., Prudhomme, C., Tansey C., Perea, R. & Hooftman, D.A.P (2012) Modelling spread of British wind-dispersed plants under future wind speeds in a changing climate. *Journal of Ecology*, **100**, 104–115.

Carrasco, L., Harwood, T., Toepfer, S., MacLeod, A., Levay N., Kiss, J., Baker, R., Mumford, J. & Knight, J. (2010) Dispersal kernels of the invasive alien western corn rootworm and the effectiveness of buffer zones in eradication programmes in europe. *Annals of Applied Biology*, **156**, 63–77.

Cooley J.W. & Tukey J.W. (1965) An algorithm for the machine calculation of complex fourier series. *Mathematics of Computation*, **19**, 297–301.

Csilléry K., Blum, M.G., Gaggiotti, O.E. & François, O. (2010) Approximate bayesian computation (abc) in practice. *Trends in Ecology & Evolution*, **25**, 410–418.

Davis, S.F. & Flaherty J.E. (1982) An adaptive finite element method for initial-boundary value problems for partial differential equations. *SIAM Journal on Scientific and Statistical Computing*, **3**, 6–27.

Dewhirst, S. & Lutscher, F. (2009) Dispersal in heterogeneous habitats: thresholds, spatial scales, and approximate rates of spread. *Ecology*, **90**, 1338–1345.

Elith, J. & Leathwick, J.R. (2009) Species distribution models: ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics*, **40**, 677.

Fahrig, L. (1997) Relative effects of habitat loss and fragmentation on population extinction. *The Journal of Wildlife Management*, **61**, 603–610.

Fahrig, L. (2002) Effect of habitat fragmentation on the extinction threshold: a synthesis*. *Ecological Applications*, **12**, 346–353.

Fennell, M., Murphy J.E., Armstrong, C., Gallagher, T. & Osborne, B. (2012) Plant spread simulator: a model for simulating large-scale directed dispersal processes across heterogeneous environments. *Ecological Modelling*, **230**, 1–10.

Gilbert, M., Guichard, S., Freise, J., Grégoire, J.C., Heitland, W., Straw, N., Tilbury, C. & Augustin, S. (2005) Forecasting cameraria ohridella invasion dynamics in recently invaded countries: from validation to prediction. *Journal of Applied Ecology*, **42**, 805–813.

Gilbert, M.A., Gaffney E.A., Bullock, J.M. & White, S.M. (2014a) Spreading speeds for plant populations in landscapes with low environmental variation. *Journal of Theoretical Biology*, **363**, 436–451.

Gilbert, M.A., White, S.M., Bullock, J.M. & Gaffney, E.A. (2014b) Spreading speeds for stage structured plant populations in fragmented landscapes. *Journal of Theoretical Biology*, **349**, 135–149.

Gilbert, M.A., White, S.M., Bullock, J.M. & Gaffney, E.A. (2016) Data from: Adaptiveide: Adaptiveide version1. *Dryad Digital Repository*, http://dx.doi.org/10.5281/zenodo.45509

Guisan, A., Tingley R., Baumgartner, J.B., *et al.* (2013) Predicting species distributions for conservation decisions. *Ecology Letters*, **16**, 1424–1435.

Hodgson, J.A., Thomas, C.D., Dytham, C., Travis, J.M.J. & Cornell, S.J. (2012) The speed of range shifts in fragmented landscapes. *PLoS One*, **7**, e47141.

Hulme, P.E. (2006) Beyond control: wider implications for the management of biological invasions. *Journal of Applied Ecology*, **43**, 835–847.

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

Le Corff, J. & Horvitz, C.C. (2005) Population growth versus population spread of an ant-dispersed neotrop- ical herb with a mixed reproductive strategy. *Ecological Modelling*, **188**, 41–51.

Lewis, M.A. (2016) Finding the sweet spot for invasion theory. *Proceedings of the National Academy of Sciences*, **113**, 6819–6820.

Miller, T.E. & Tenhumberg, B. (2010) Contributions of demography and dispersal parameters to the spatial spread of a stage-structured insect invasion. *Ecological Applications*, **20**, 620–633.

Moilanen, A. (2011) On the limitations of graph-theoretic connectivity in spatial ecology and conservation. *Journal of Applied Ecology*, **48**, 1543–1547.

Morton, D., Rowland, C., Wood, C., Meek, L., Marston, C., Smith, G., Wadsworth, R.. & Simpson, I. (2011) Final Report for LCM2007-the new UK land cover map. Countryside Survey Technical Report No 11/07. NERC/Centre for Ecology & Hydrology.

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

Neubert, M.G. & Parker, I.M. (2004) Projecting rates of spread for invasive species. *Risk Analysis*, **24**, 817–831.

Oostermeijer, J.G.B., Brugman, M.L., De Boer, E.R. & Den Nijs, H.C. (1996) Temporal and spatial variation in the demography of gentiana pneumonanthe, a rare perennial herb. *Journal of Ecology*, **184**, 153–166.

Powell, J.A., White, P & McMillen, T. (1998) Connecting a chemotactic model for mass attack to a rapid integro-difference emulation strategy. *SIAM Journal on Applied Mathematics*, **59**, 547–572.

Salguero-Gómez, R., Jones, O.R., Archer, C.R., *et al.* (2015) The compadre plant matrix database: an open online repository for plant demography. *Journal of Ecology*, **103**, 202–218.

Scheffer, M., Baveco, J., DeAngelis, D., Rose, K. & Van Nes, E. (1995) Super-individuals a simple solution for modelling large populations on an individual basis. *Ecological Modelling*, **80**, 161–170.

Schurr, F.M., Pagel, J., Cabral, J.S., *et al.* (2012) How to understand species' niches and range dynamics: a demo- graphic research agenda for biogeography. *Journal of Biogeography*, **39**, 2146–2162.

Shigesada, N. & Kawasaki, K. (1997) *Biological Invasions: Theory and Practice Oxford Series in Ecology and Evolution*, 1st edn. Oxford University Press, Oxford, UK.

Shigesada, N., Kawasaki, K. & Teramoto, E. (1986) Traveling periodic waves in heterogeneous environments. *Theoretical Population Biology*, **30**, 143–160.

Solodovnikov, V. (1985) Upper bounds on the complexity of solving systems of linear equations. *Journal of Soviet Mathematics*, **29**, 1482–1501.

Svenning, J.C., Gravel, D., Holt, R.D., *et al.* (2014) The influence of interspecific interactions on species range expansion rates. *Ecography*, **37**, 1198–1209.

Thuiller, W., Münkemüller, T., Lavergne, S., Mouillot, D., Mouquet, N., Schiffers, K. & Gravel, D. (2013) A road map for integrating eco-evolutionary processes into biodiversity models. *Ecology letters* **16**, 94–105.

Williamson, M. (1999) Invasions. *Ecography*, **22**, 5–12.

Wisz, M.S., Pottier, J., Kissling, W.D., *et al.* (2013) The role of biotic interactions in shaping distributions and realised assemblages of species: implications for species distribution modelling. *Biological Reviews*, **88**, 15–30.

Zhou, Y. & Kot, M. (2011) Discrete-time growth-dispersal models with shifting species ranges. *Theoretical Ecology*, **4**, 13–25.

Zhu, K., Woodall, C.W. & Clark, J.S. (2012) Failure to migrate: lack of tree range expansion in response to climate change. *Global Change Biology*, **18**, 1042–1052.

Supporting Information

Additional Supporting Information may be found online in the supporting information tab for this article:

Appendix S1. Existing approaches.

Appendix S2. Method.

Appendix S3. Accuracy.

Appendix S4. Furthest distances.

Appendix S5. Stage structured example.