Resource-Explicit Interaction Models for Spatial Populations

Bryan Chae

Project Overview

In recent years, population models that utilize continuous spatial frameworks have increasingly found use in lieu of traditional models that consider populations to be well-mixed. This is because the inclusion of spatiality can significantly alter simulation outcomes when modeling many population genetics processes, such as the evolution and spread of beneficial alleles. A fundamental requirement of spatial models is the determination of which individuals can interact with one another, and how strongly they interact. These calculations are used to regulate the population to its local capacity. However, these interactions represent a substantial computational workload, which can lead to prohibitively long runtimes for large populations. Here, we present a novel modeling method in which the resources available to a population are abstractly represented as an additional layer of the simulation. Instead of interacting directly with one another, individuals interact indirectly via this resource layer. We find that this method closely approximates interactions used in other spatial models, yet can increase the speed of the model by as much as an order of magnitude, allowing for the simulation of much larger populations. Additionally, structuring the model in this manner provides other desirable characteristics, including more realistic spatial dynamics near the edge of the simulated area, as well as an efficient route for implementing more complex heterogeneous landscapes and other features.

Introduction

Traditional population genetics models have generally considered groups of organisms to exist under the assumption of panmixia[1–3]. Panmictic population are considered to be fully intermixed, with no spatial structure and with no barriers to prevent any given individual from selecting any possible candidate as a mate. This abstraction is likely suitable for modeling laboratory populations, populations on small islands, and some others[4,5]. However, there are many other contexts in which the natural world cannot be accurately reflected without some representation of spatiality[2,6,7].

One option is to break up a population into an array or "lattice" of subpopulations linked by migration[2,8–10]. Each subpopulation is treated as a separate panmictic cell, while migration between cells allows for spatial variation. Increasingly, however, it has become apparent that the most accurate way to model populations that exist in real continuous geographical landscapes is to use simulated continuous landscapes[2,6,8]. Models structured in this way avoid artifacts produced by discretization in linked panmictic populations[2]. More importantly, there are a number of biological processes that play out very differently in continuous spatial models. One example is the evolution of beneficial mutations, which have been found to reach fixation at significantly different rates in continuous spatial models than predicted by non-spatial models[7,11,12].

While fully continuous spatial models are the best or only choice in many contexts, they are not without their pitfalls. Chief among these is the significant computational cost of spatial calculations. At each point in time throughout the simulation, it must be determined which individuals can mate with one another and which are in competition. Competition is represented by assigning fecundity or mortality as a function of local density in lieu of the global population regulation of a panmictic model[2]. To accomplish these tasks, each individual must ascertain whether or not it is interacting with each other individual; if two individuals interact, the strength of the interaction, usually a function of the distance between the individuals, must be computed. Superficially, it might appear that this process would require on the order of n^2 operations for a population of n individuals, but

the use of a space-partitioning data structure to store individual positions reduces the workload substantially, with the magnitude of the reduction inversely proportional to the density of the population[13,14]. Nonetheless, spatial calculations can occupy the majority of the runtime of such models.

Additional problems crop up near the boundaries of spatial models. If individuals are distributed approximately uniformly, as might be considered appropriate, individuals at the edges of the simulated area tend to interact with only half as many neighbors. Since survival is dependent on local density, this can result in the edge areas becoming overpopulated, with knock-on effects on density in the interior of the model. Such edge effects can be corrected by calculating the weight of the spatial interaction that has been cut off by the edge of the model and scaling competition accordingly[15]. However, this correction involves solving, or at least approximating, a double integral in polar coordinates for each individual whose interaction area falls partially outside the modeled area. This represents an additional computational burden, especially when using more complex interaction functions. The difficulty of this correction is substantially exacerbated if modeling an area with a naturally shaped border (e.g., an island's coastline) rather than a square or rectangular area.

In this study, we introduce a new continuous spatial modeling paradigm in which interactions can be evaluated with significantly greater speed, and which does not exhibit edge effects for which corrections need be applied. This is accomplished by directly simulating the resources available to the focal species. These resources are abstractly modeled via "resource node" entities which are uniformly distributed across the landscape at the center-point coordinates of a hexagonal tiling of the modeled area. Competing individuals collect resources from their foraging area which encompasses several such nodes, reducing the resources available to other individuals. In this way, local density is regulated by extrinsic resource availability, rather than local density being self-regulating. Thus, the spatial interactions used to model competition are interactions between individuals and resource nodes, removing the need to evaluate spatial interactions as between individuals of the focal species. Depending on the density of

individuals and the desired granularity of the tiling of resource nodes, one resource node might support as many as a dozen individuals, meaning evaluating interactions between individuals and nodes can be vastly less computationally intensive. Edge-behavior in a resource-explicit model is also preferable to that in a standard spatial model. Individuals at the edges of the modeled area have fewer local competitors, but they are also near fewer resource nodes (as there are neither resource nodes nor competitors outside of the modeled area). Consequently, local carrying capacity is the same near edge areas as it is in the interior of the model without the need for any time-consuming corrections.

The tiling of resource nodes results in minor spatial artifacts in interaction strength. However, quantitative assessments reveal that these artifacts are minor, and that the resource-explicit interaction has a high degree of similarity with standard spatial interactions. Performance comparisons reveal that models using this technique can run an order of magnitude faster than standard spatial models. Given this degree of runtime improvement, resource-explicit models could be used to simulate populations that were previously impractically large to simulate in full in spatial models. Additionally, these models are highly flexible, offering an elegant implementation path for a number of desirable features, including more realistic, heterogenous landscapes, irregular model boundaries (i.e., coastlines or otherwise impassable terrain), competition between multiple species, and other features.

Methods

The population models assessed in this manuscript are structured as continuous spatial models with overlapping generations in which competition impacts mortality (the resource-explicit approach can also be applied to models with non-overlapping generations and in which competition impacts fecundity; an example model structured in this way is available on the GitHub repository: https://github.com/MesserLab/ResourceExplicitModels). The models are written in

the SLiM individual-based forward-in-time population modeling framework (version 4.1; the models are not compatible with previous versions)[15].

The models share a common core, except as noted otherwise. An approximately square area with reprising boundaries was simulated, with the population starting at capacity (the area is slightly rectangular – see the Supplemental Methods regarding resource node placement for an explanation). The modeled area was specified in multiples of a basic "unit area," defined as the foraging area of an individual of the focal species. Capacity was calculated by multiplying the modeled area by a density parameter specifying the number of individuals supported by one unit area.

In each time-step, or "tick", of the model, the focal species first reproduces, and then experiences competition dependent mortality. Even in the resource-explicit models (with one exception), females search for mates using a standard spatial interaction, since mate-search interactions are generally less of a runtime burden than competition interactions because of the sex-segregated nature of the mate-search interaction. A fixed-strength interaction function is used for this purpose, resulting in an equal probability of picking any male within range. Each female who reproduces generates a number of offspring according to a Poisson distribution with an average of 8 (except as specified). This results in a post-reproduction (and pre-mortality) population size of five times the capacity of the system. Newly generated offspring are dispersed away from their mother by adjusting both the x and y coordinate of their mother by a draw from a normal distribution with a mean of zero and a standard deviation of $2\times\sqrt{\frac{1}{\pi}}$. This results in a distance that is twice the radius of the foraging area of the species, and was chosen such that offspring have a good chance to not compete strongly with their mother (and do not compete at all if they disperse more than one standard deviation away). After reproduction, competition for resources occurs. This is implemented differently in each model, as described below.

I. Standard Spatial Models

There are several functions that are commonly used to determine the interaction strength between individuals in a spatial model (Figure X). The simplest is a “fixed strength” interaction function: all individuals within range interact at strength 1 (100%). Individuals at the maximum interaction range compete as intensely as individuals who are right next to one another, which makes this function undesirable for simulating competition for resources and can lead to unrealistic clustering. However, this function is the most rapid to evaluate. A second option is a “linear” function: two individuals in the exact same location interact at strength 1, with interaction strength linearly declining to 0 at the maximum interaction distance. Another option is a “Gaussian” function, in which two individuals in the same location interact at strength 1, with strength declining according to:

$$strength = e^{\frac{-distance^2}{2\sigma^2}}$$

with distance as a fraction of the maximum interaction distance and where σ is a parameter which determines how quickly interaction strength decays as distance increases.

Each of these three interaction kernels is presented as a separate model in this manuscript. For the Gaussian interaction, a σ of 1/3 of the maximum interaction distance was used such that the interaction strength declines to a suitably negligible value (of 0.011) at the maximum interaction distance.

An additional function that could be used for calculating interaction strength is the less frequently used “circular intersection” interaction function. This function treats each individual as foraging from a circular area, with the interaction strength between two individuals equal to the portion of their foraging areas that intersects (Figures X and 1, panels a and d). The radius of the foraging area in this function is half of the maximum interaction range. The resulting function has a strength of 1 for individuals who are in the exact same location, with strength decreasing to 0 at the maximum interaction range, as described by:

$$strength = \frac{2(distance) - 2 \times distance\sqrt{1 - distance^2}}{\pi}$$

with distance as a fraction of the maximum interaction distance. This function is not commonly used in spatial population models (and was not implemented in the population models used in this manuscript), likely due to the amount of computation required. However, it is a nicely continuous function based on biologically plausible assumptions, and is presented here as a useful point of comparison with the resourced-explicit models.

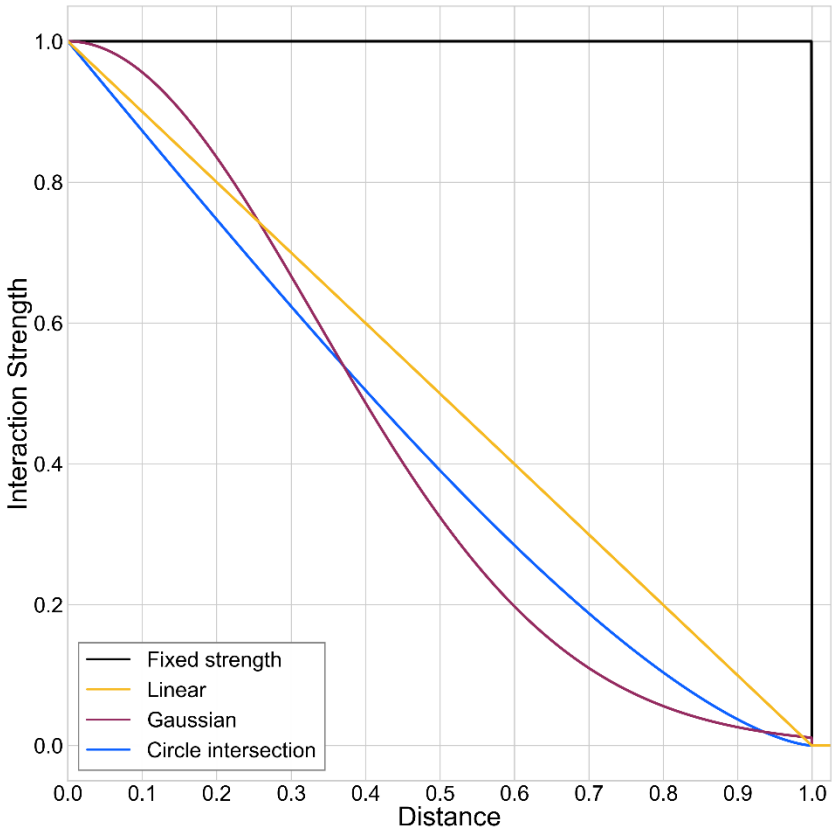

Figure X. Interaction functions. Individuals within a continuous space model interact with one another according to an interaction strength function. A number of potential function choices are depicted.

II. Resource-Explicit Models

The resource-explicit models are analogous the “circular intersection” interaction function, as individuals in these models are considered to forage from an approximately circular area. At the outset of the simulation, the simulated area is tiled with resource nodes. These nodes are placed according to a hexagonal tiling of the modeled area. The dimensions of the hexagons are determined by the desired granularity of the tiling relative to the foraging area of the modeled species. In the models presented below, the granularity of the tiling is chosen such that the

area of one hexagon is 1/37 of a unit area (with the exception of variants where one hexagon is 1/19 of a unit area). Two individuals in the exact same location forage from the same resource nodes, and could thus be considered to have an interaction strength of 1, with interaction strength decreasing as further apart individuals share fewer resource nodes.

The value 37 is chosen from the sequence of "hexagonal numbers" and represents an area consisting of a central hexagon and all other hexagons within three "steps" (similarly, 19 represents a central hexagon and all others within two steps). Although the hexagons in the model are considered to have an "area" and are depicted as having edges, the actual resource node entities in the model consist of single-point entities located at the center-points of the hexagonal grid. Resource nodes are then assigned a value representing the amount of resources available at that node as determined by the density parameter. E.g., if the focal species has a density of 37 individuals per unit area, then the number of resources available at each node is 1.

See the Supplemental Methods for a detailed description of the construction of the grid, along with the rationale for choosing a hexagonal tiling. In the SLiM models assessed in this manuscript, the resource node entities are implemented as a second species for which no life cycle is defined (i.e., no mobility, mortality, or reproduction) which spatially interacts with the focal species. Similar implementations are possible within other modeling frameworks or in from-scratch models that allow for inter-species interactions.

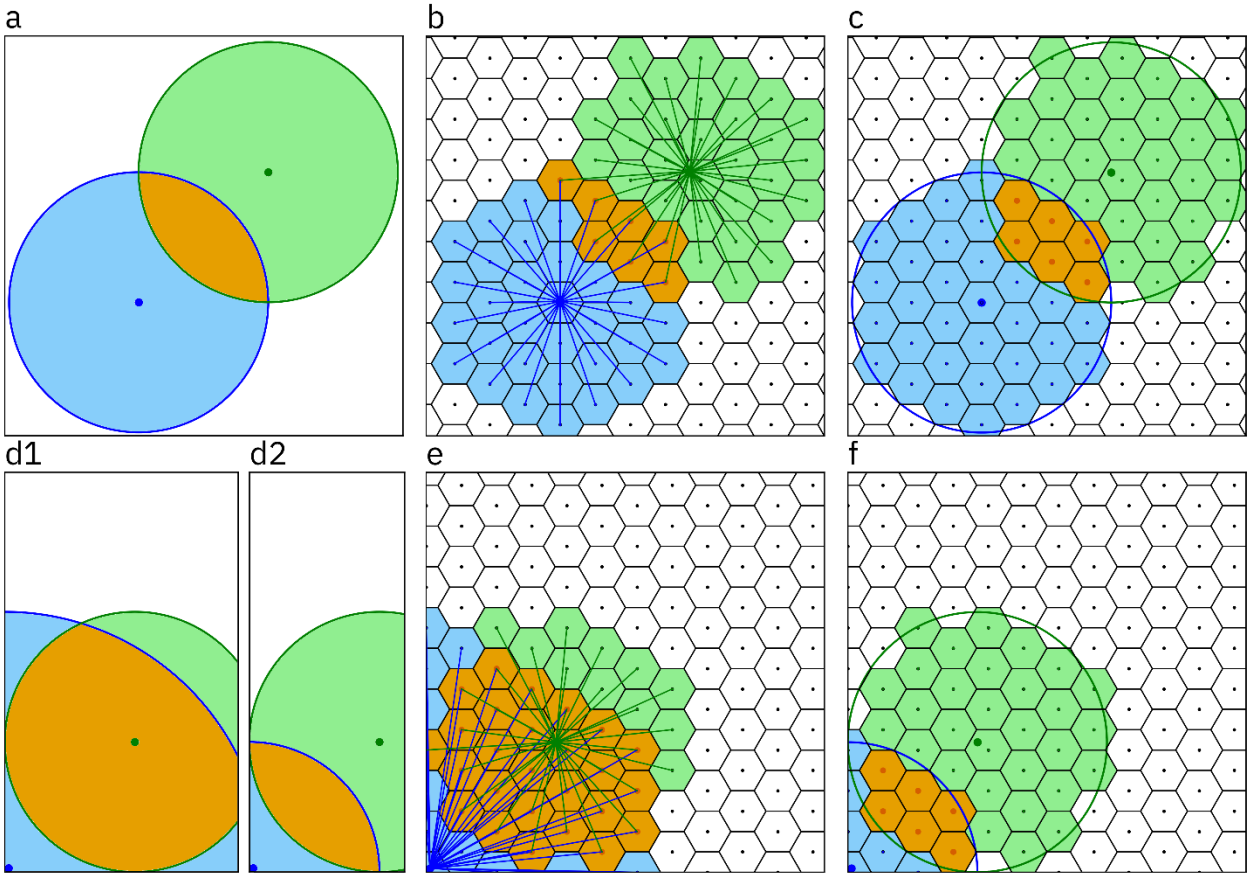

Figure 1. Visualization of resource-explicit interaction functions. Panels a, b, and c show the interaction strength between two individuals ("blue" and "green"), as determined by the portion of the foraging area of the two individuals that overlaps. The foraging areas of the two individuals are represented by blue and green shading; the shared area is shaded orange. Panel a: a circular intersection function; b: the elastic resource-explicit model wherein each individual forages from the nearest 37 nodes (rays are drawn between the focal individual and the nodes they forage at); c: the inelastic resource-explicit model wherein each individual forages from resource nodes within their foraging radius. In the inelastic model, the overlapping area is smaller by one resource node, which lies just outside of green's foraging radius (because of green's exact position, green can only access 36 nodes in the inelastic model). That resource node is, nonetheless, one of the 37 closest nodes to green, which is why green forages from that node in the elastic model. Panels d, e, and f show the interaction dynamics when blue is in the corner of the modeled area. Panel d1 shows the circular intersection function under the assumption of the modeled species having an elastic range; panel d2 shows the same under the assumption of an inelastic range; panels e and f show the elastic and inelastic implementations of the resource-explicit model.

II.A. Models with an Elastic Foraging Area

In the elastic models (Figure 1, panels b and e), individuals forage from the nearest 37 resource nodes, though only from nodes that are within radius, R, with R being chosen as larger than the radius of the foraging area, r. Individuals near the edges of the modeled area are thus able to forage from further away than an inland individual would, resulting in an elastic foraging area. R can be chosen as only somewhat larger than r, producing slightly elastic ranges, or can be chosen as much larger than r, resulting in fully elastic ranges in which individuals always have access to a full 37 resource nodes. Models in this manuscript use an R of $2 \times r$, which is sufficient for fully elastic ranges given a square shaped landscape. The nearest 37 resource nodes are chosen from a continuous spatial interaction based on the individual's exact position – thus, two individuals in different locations within the same hexagon may forage from a different set of nodes (though most of the nodes foraged at by two such individuals will overlap).

In this model, competition is determined as follows. First, each individual is iterated through, and the nearest 37 resource nodes to each are recorded. In the process, every time a resource node is recorded, a variable belonging to the node that tracks "demand" is incremented by 1. After this, individuals are iterated through once again. During the second iteration, each individual receives resources from each of their nodes according to the amount at the node divided by the demand on that node. After receiving resources, individuals are assigned a likelihood of survival equal to the amount received.

E.g., if each resource node has enough resources for 1 individual, and the population size is five times the capacity of the system, then each of the 37 resource nodes provides an average of $1/(37 \times 5)$, resulting an average survival rate of 1/5.

In addition to this default implementation, we include three additional variants. Any combination of the changes in these variants could be included in a single model if desirable.

Non-migratory variant. An improvement in runtime can be achieved in a model in which individuals only disperse a single time (an initial dispersal of offspring from their parents) by finding the nearest resource nodes to each individual only a single time, and then caching and reusing this information. Individuals do not disperse in any of the models presented here (beyond their initial dispersal), but this optimization has not been made in the default model in order to provide an accurate reflection of the runtime that could be expected in a model where individuals move during each tick. This optimization is not relevant in a model with non-overlapping generations, as the cached locations would never be used.

Preferential-foraging variant. In this version of the model, individuals preferentially forage from resource nodes at which fewer competitors are present. If all resource nodes are crowded, this has basically no effect. However, if the landscape has not been fully exploited, preferential foraging may prove relevant. For example, if an invasive species is spreading across a newly invaded landscape, individuals at the forefront of the invading wave can fully meet their needs by using the unexploited resources in front of the wave, and thus do not compete with individuals immediately behind them (even if their foraging ranges overlap); those individuals in turn can forage from the relatively underutilized areas at the front of the wave, and so on. The result is less "back-pressure" on the wave, potentially resulting in a faster spread across the landscape.

Nineteen hexagon variant. In this version of the elastic model, individuals only forage from the nearest 19 nodes. Accordingly, the landscape is tiled more loosely, such that the area of 19 nodes is equal to one unit area.

II.B. Models with an Inelastic Foraging Area

In the inelastic models (Figure 1, panels c and f), individuals forage from any resource nodes that are within their foraging radius (i.e., within a circle of one unit area). Unlike in the elastic model, individuals are not guaranteed access to a full 37

resource nodes. Individuals in the corners of the simulation might have access to only about a quarter this many. Even inland individuals, depending on their exact spatial position within the grid, may have access to somewhat more or somewhat fewer resource nodes.

Competition in the elastic model is calculated by first iterating through each resource node. Each node counts local demand as the number of individuals within r distance and then sends each such individual an amount of resources equal to the amount available at that node divided by the local demand. Then, individuals are assigned a likelihood of survival equal to the amount of resources received.

Thus, the number of nodes in range of an individual directly impacts that individual's survival rate. Individuals near the edge of the area are near fewer nodes, though those nodes are likely to be used by fewer individuals, which may result in only small net effects. Inland individuals in range of fewer nodes have proportionately lower rates of survival, and individuals in range of more nodes have proportionately higher rates of survival.

In addition to this default implementation, we include three additional variants. As with the elastic model, these variants are compatible with one another, and could be combined if desired.

Fair variant. This version of the model corrects for the spatial unfairness of the inelastic model such that the survival rate of an individual should be unimpacted by having access to more or fewer resource nodes. This is achieved by performing a preliminary iteration through the nodes and incrementing a variable for each individual that tracks the number of nodes foraged from, n. The amount of resources each individual receives from each resource node is determined in a second iteration through the resource nodes. The amount received from a resource node by an individual i, where f is the amount of resources available at the node and C is the set of all individuals receiving resources from the node (the "customers") is given as:

$$received = \frac{\frac{f}{i.n}}{\sum_{j}^{C} \frac{1}{j.n}}$$

This variant could be seen as tracking the amount of time or energy individuals spend at each of the resource nodes within their foraging range. Individuals with fewer nodes in range compete more intensely at those nodes, while individuals with more than 37 nodes within range draw proportionately less from those nodes.

Variant with a spatial approximation for reproduction. This variant replaces the mate-search interaction with a spatial approximation that uses the resource node entities, thus replacing all interactions between members of the focal species with interactions between individuals and resource nodes. Before reproduction takes place, each node caches a list of males who are within reproduction distance. Then, rather than females directly searching for and choosing a nearby male, they instead select a mate from the list of males cached at the nearest resource node.

This spatial approximation means that all females within a given hexagon have access to the same list of potential mates, regardless of the female's position within a hexagon. However, due to the fairly granular tiling of the area, the spatial approximation is not large. With 37 nodes per unit area, the nearest resource node is guaranteed to be within 1/5 of the radius of the foraging area (or about 1/4 with 19 nodes per unit area). E.g., if the modeled species has a foraging area of 1 hectare, the nearest resource node to an individual is guaranteed to be within 10.2 meters (and is usually much closer).

Nineteen hexagon variant. As with the elastic model, this implementation differs from the default inelastic model only in that the landscape is tiled more loosely, such that the area of 19 nodes is equal to one unit area.

III. Panmictic Model

In order to provide a baseline understanding of what portion of the runtime of the other models is spent performing spatial computations versus other necessary functions (e.g., offspring generation), a panmictic version of the simulation was also implemented. In the panmictic model, a global survival rate is imposed, which is calculated as the capacity of the system divided by the current population. Instead of choosing a nearby mate, females in the panmictic model randomly sample a male from the population. Other than these changes, and individuals not being assigned spatial coordinates, the panmictic model is identical to the other models.

Results

I. Dynamics of the Resource-Explicit Models

The resource-explicit models take place in a continuous spatial framework but replace direct interactions between individuals by instead having the individuals compete for resources at discretized nodes that are tiled across the landscape. The resultant indirect interaction function differs from commonly used interaction functions in a number of qualitative ways. First, the indirect interaction is a discontinuous function. Two individuals who are right next to one another share 37 of 37 nodes, and thus could be considered to interact with a strength of 1. As the distance between the two individuals increases, the number of shared nodes decreases, resulting in the interaction strength decreasing in increments of 1/37. Second, the indirect function is not perfectly radially symmetric: given a fixed position of a focal individual and a fixed distance between two individuals, the strength of the interaction may vary slightly depending on the angle between the two individuals. Finally, the strength of the indirect interaction is also dependent on the precise location of individuals within the hexagonal grid: given a fixed distance and angle between two individuals, the strength of the interaction may vary slightly depending on their coordinates.

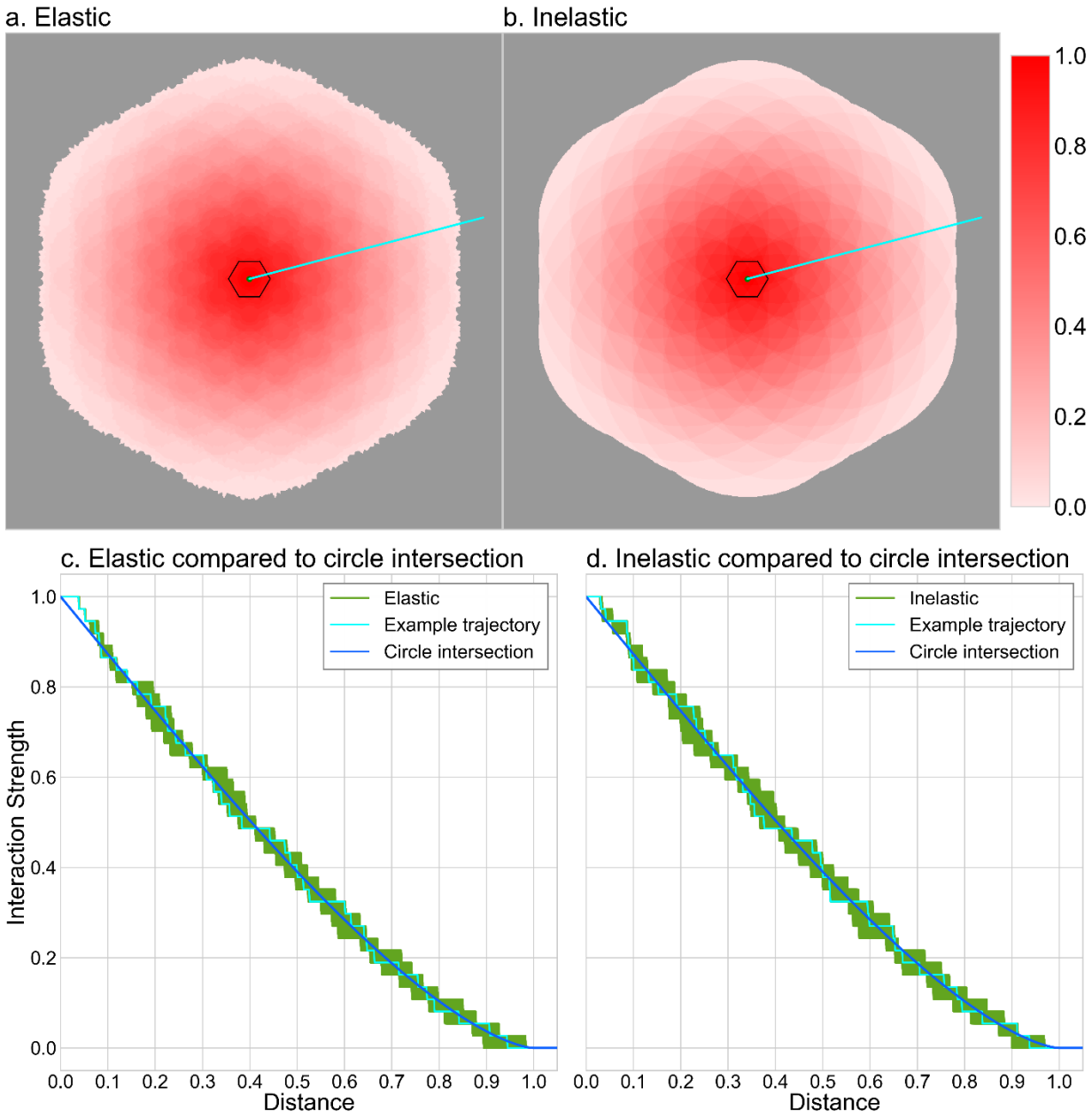

Figure 2. Visualization of the indirect interaction between individuals competing for resources. These panels depict the shape of the interaction and the interaction strength between a focal individual (located in the center of the depicted hexagon) and a competitor in any other spatial position within a grid of 1000 by 1000 points for the elastic (panel a) and inelastic (panel b) model. Panels c and d depict the comparison between the circle overlap function and the interaction strength in the resource-explicit functions as between individuals in every possible direction. The cyan line overlaid on panels a and b is shown as a specific example trajectory on panels c and d.

However, despite these seemingly quite substantial qualitative differences, a quantitative assessment reveals that the resource-explicit models are a very close

match to the circular intersection function which they seek to emulate. To evaluate the functions and measure the differences between them, an approximately square area consisting of 418 hexagons was subdivided into heatmap grid with about 1000 by 1000 points. For each function, interaction strength was measured between a focal individual located in a central hexagon and every other point in the grid (Figure 2). The average difference between two functions was calculated by measuring the absolute value of the difference between the two functions as evaluated at each point where at least one of the functions had a non-zero value. This average difference was recalculated with the focal individual located in every point in the heatmap that fell within the central hexagon, totaling 2543 points, and the resulting 2543 average differences were then averaged again to give an average interaction strength as averaged across all possible coordinates.

For the elastic model with 37 nodes per unit area, when the focal individual was located at the exact center point of the hexagon, the average absolute difference between the resource-explicit interaction and the circular intersection interaction was 0.0128 (Figure 2, panels a and c, Supplemental Figure 1, panel a). The average absolute difference as averaged across all 2543 focal points within the central hexagon was 0.0170. At the difference maximizing point, the average absolute difference was 0.0239 (Supplemental Figure 1, panel c).

For the inelastic model with 37 nodes per unit area, the average absolute difference between the resource-explicit interaction and the circular intersection interaction given a focal individual in the center of a hexagon was 0.0138 (Figure 2, panels b and d, Supplemental Figure 1, panel b). The average absolute difference as averaged across 2543 points was 0.0185. At the difference maximizing point, the average absolute difference was 0.0253 (Supplemental Figure 1, panel d).

See the Supplemental Results for similar measurements for the models with 19 hexagons per unit area.

Unequal Resource Availability in the Inelastic Model. In the inelastic implementation of the model, individuals are not guaranteed to forage from

exactly 37 nodes. Depending on their position within the array of nodes, they may forage from as few as 35 and as many as 40 nodes (Supplemental Figures 3 and 4), not including individuals near the edge of the model who tend to forage from substantially fewer. Except in the "fair" variant of the model, foraging from fewer nodes directly corresponds with a higher mortality rate, while foraging from more corresponds with a lower mortality rate. This gradient in the mortality rate between individuals does not impact the average mortality of the system as a whole, as the average number of individuals who survive the mortality phase of the model is still dependent on the amount of resources available across the landscape.

The degree to which an unequal mortality rate could be considered undesirable likely depends on the system being modeled. The tessellation of varying resource node availability is repeated in each hexagon of the model, and thus by definition has a very small spatial scale compared to the range of the focal species. For most species, it can likely be assumed that individuals disperse at least as far as they travel to forage. Given the large difference in spatial scales between the pattern of resource node availability and dispersal, the number of resource nodes available to an individual given their exact coordinate in the grid could be considered an entirely random property. Similarly, after competition, even though more individuals are removed from areas with access to fewer nodes and vice-versa in areas with access to more nodes, this should not have a measurable impact on dispersal patterns in the following generation, again given the much larger spatial scale of dispersal compared to this repeating pattern. Thus, the total impact of this phenomenon is that some individuals, at random, have a higher likelihood of survival, while others, at random, have a lower likelihood, without any change to the overall survival rate. Considering that survival is already stochastic, this should cause no noticeable change in most circumstances. However, this phenomenon may be more of an issue in very sparse populations with small litter sizes. For such cases, the "fair" variant may be considered desirable. In the fair variant, the unequal resource node availability is counteracted such that mortality rates are not affected.

Edge Effects in the Resource-Explicit Models. The dynamics at the edge of the modeled area in the resource-explicit models are not completely identical to those in the interior of the model. When offspring are dispersed away from parents located near the edge of the area, those offspring have a smaller area into which they can disperse, since they cannot disperse off the edge of the area. Thus, before mortality is imposed, density tends to be higher near the edge of the area, leading to an accompanying increase in mortality rates, despite local capacity near the edge of the model being equal to local capacity in further inland areas. This could be considered realistic, depending on the species under investigation. As an alternative, a larger dispersal distance could be used for individuals near coastal areas, as such individuals might invest extra effort to avoid this type of clustering.

In the inelastic implementation, though not quantitatively assessed, a slightly reduced density was observed near the very edges of the simulated area. This is likely due competitive pressure from inland individuals on coastal individuals. Since individuals near the edge of the model have access to very few resource nodes, they are less likely to survive than the slightly further inland individuals who can forage from the coastal nodes but also from further inland nodes. This was not observed in the "fair" variant of the inelastic model, since coastal individuals in that variant who forage from fewer nodes have a commensurate increase in the amount of resources they receive from those nodes.

Wave Advance Speed in the Resource-Explicit Models. In some contexts, such as considering the spread of an invasive species, it can be important to measure the rate at which a population can spread into an unoccupied habitat. In one of the resource-explicit models, individuals forage preferentially from resource nodes that are less crowded, which was expected to potentially increase the speed at which a wave could spread across the landscape.

To assess potential differences in wave speed between the models, a series of simulations was performed in which a population of individuals with varying

density and litter size parameters was released at one end of a rectangular area, and the time required for the population to spread across the landscape was recorded (see Supplemental Results). With most parameters, the wave advance speed of all of the models was indistinguishable, as the advancing wave was able to fully exploit the landscape, with the population able to spread to the full extent that migration would allow during each tick. Only when the density and litter size parameters were very low did differences in the models appear. At low density and small litter sizes, the wave advance speed of the preferential foraging model was fastest. Second fastest was the standard model with a fixed strength interaction function. All of the other models were slightly slower still (Supplemental Figure 5). The differences between the models were not extreme, but differences of this nature may be worth consideration in some systems.

II. Computational Performance of the Resource-Explicit Models

To assess the runtime performance of each of the models, a series of simulations with a varying landscape size was run at three different densities – a low density (20 individuals per unit area), a medium density (100 individuals per unit area), and a high density (200 individuals per unit area). For each density, the landscape size was varied such that population size after reproduction and before mortality (i.e., the number of individuals involved in the spatial interaction calculating competition) was increased in increments of 1 million, up to a maximum of 10 million individuals. The time elapsed during each tick was recorded, starting with the third tick in order to allow the simulation time to equilibrate. Simulations were performed on a desktop computer using an Intel i9-9900K processor, on which only a single simulation was performed at a time in order to keep conditions across simulations as equal as possible.

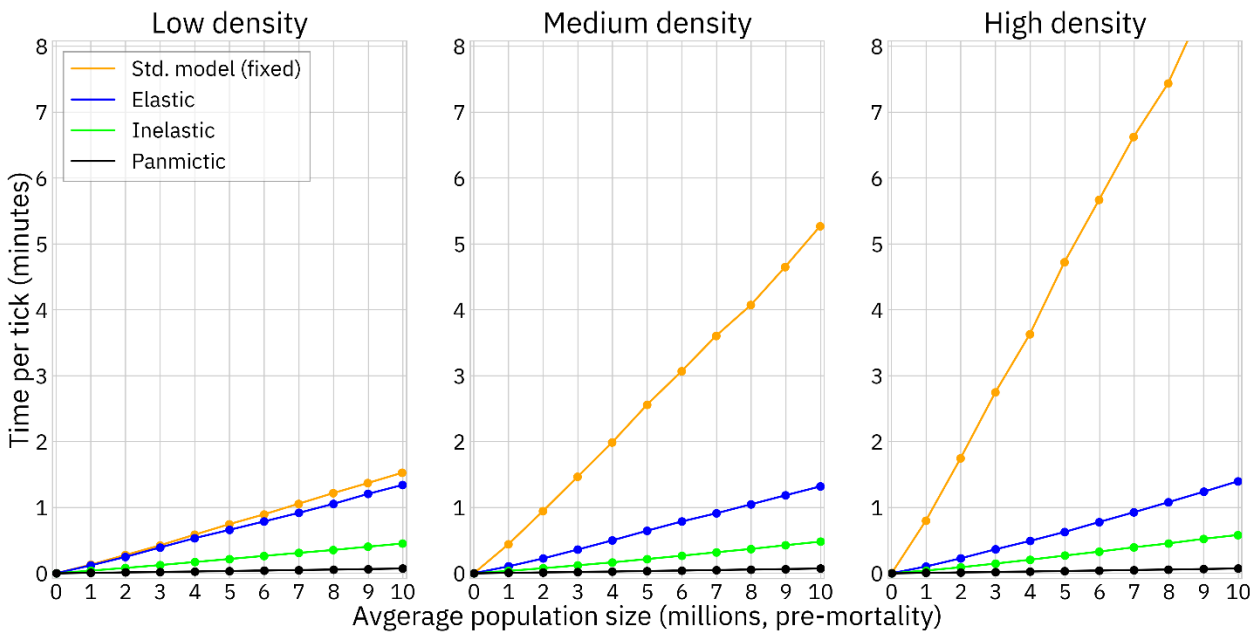

Figure 4. Runtime comparison across model types. A series of simulations with a varying landscape size was run at three different densities (20, 100 and 200 individuals per unit area) with the size of the landscape increasing such that the population size (after reproduction and before mortality) increased in steps of 1 million. After the model equilibrated for 3 ticks, the runtime duration of the next ten ticks was recorded and averaged. Note that figures 4-6 have different y-axis scales.

An initial comparison between the default implementations of the resource explicit models and the standard spatial model reveals an increase in runtime speed across all population sizes at all densities (Figure 4). This is in spite of the standard spatial model using a fixed strength interaction, the fastest of the standard spatial functions (though the linear function is only slightly slower; Supplemental Figure 6).

A comparison between the variants of the elastic model (Figure 5) shows that the implementation with preferential foraging is by far the slowest of the resource-explicit models. This model's purpose is to add a novel capability, rather than being intended as a runtime optimization, so this is not unexpected. This version of the elastic model is the only resource-explicit model that runs more slowly than any of the standard spatial models (though it only ran more slowly at low density). The optimization for populations where migration is only performed by newborn offspring represents a small runtime improvement over the default

elastic model. This optimization would certainly yield more improvement in a model with longer lived individuals – individuals in this model only had a survival rate of 20% per tick, and the caching of spatial information only saves time when individuals actually survive to the next tick to use that data. The implementation with 19 nodes per unit area cuts the runtime by even more, though this implementation represents somewhat more of a spatial approximation compared to the default 37 nodes (see the Supplemental Results). The runtimes of the elastic models are almost unaffected by the density of the population, unlike the standard spatial models, which are highly sensitive to increases in density.

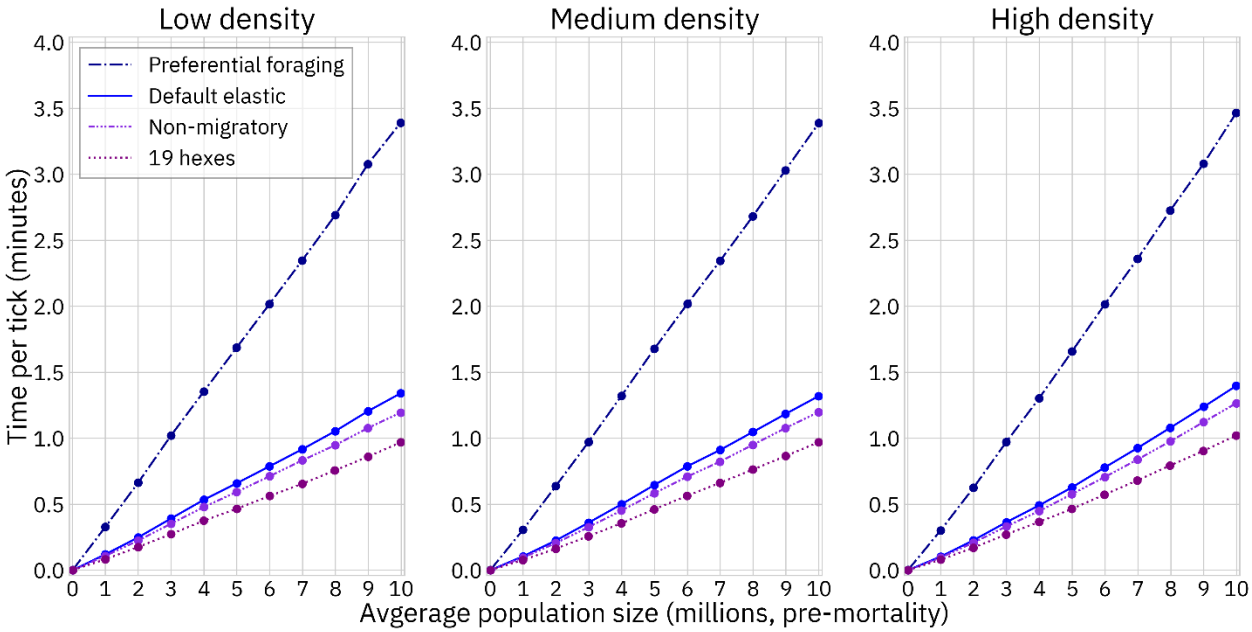

Figure 5. Runtime comparison of elastic models. A series of simulations with a varying landscape size was run at three different densities (20, 100 and 200 individuals per unit area) with the size of the landscape increasing such that the population size (after reproduction and before mortality) increased in steps of 1 million. Note that figures 4-6 have different y-axis scales.

The inelastic models, across all variations, perform extremely quickly (Figure 6). Even the “fair” variant, which is slower than the other inelastic models, performs faster than any of the elastic models. In general, the inelastic models slowed down very slightly as density increased. However, the variant using a spatial approximation for reproduction showed a reverse trend, and was faster at medium density than at low density, and was approximately as fast at high density as at medium density. The variant with 19 nodes per unit area is the fastest of the

spatial models, achieving the notable feat of requiring only about 5 times as much runtime as the panmictic model.

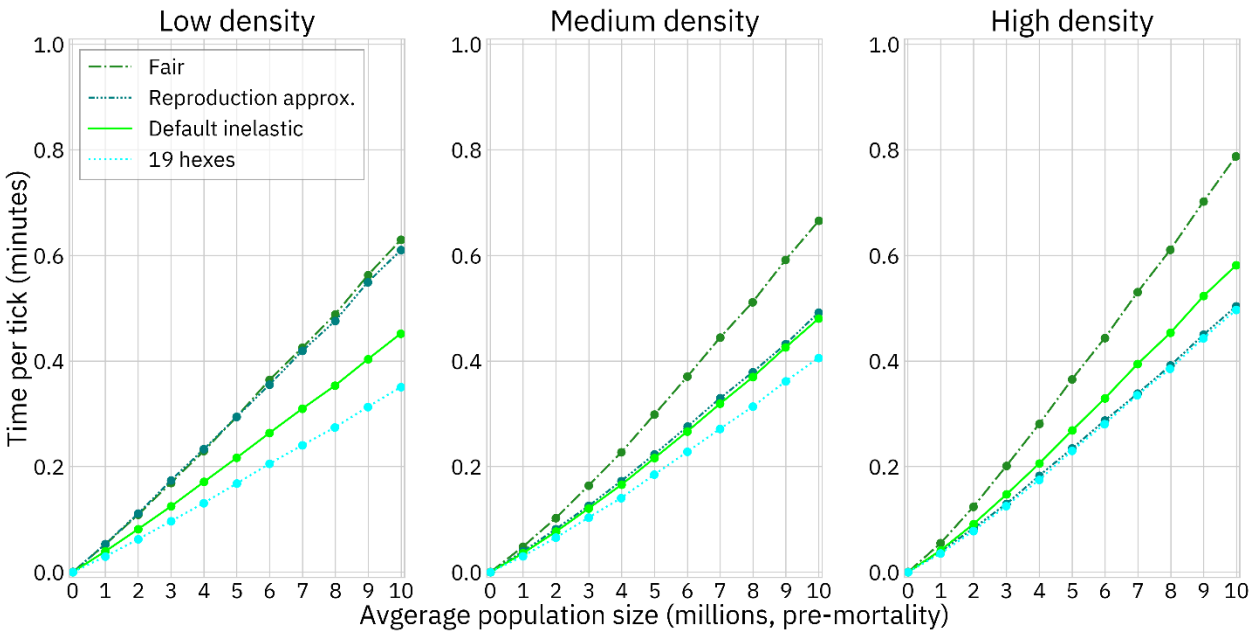

Figure 6. Runtime comparison of inelastic models. A series of simulations with a varying landscape size was run at three different densities (20, 100 and 200 individuals per unit area) with the size of the landscape increasing such that the population size (after reproduction and before mortality) increased in steps of 1 million. Note that figures 4-6 have different y-axis scales.

An additional set of simulations was used to continue exploring the relationship between density and runtime in the default implementations of the elastic and inelastic model versus the standard spatial model (Figure 7). When modeling higher density populations, given a certain population capacity, there are fewer resource nodes (each of which distributes a greater amount of resources). This offsets the reduced efficiency of the spatial data structure at higher densities (indeed, density within the data structure is no longer dependent on the density of the population). At high densities, the elastic model had performance of up to about 8 times faster than the standard spatial model, while the inelastic model performed up to 20 times as faster. At decreasing densities, the resource-explicit models are eventually outperformed by the standard spatial model. The elastic model performs more slowly than the standard spatial model at densities below about 19 individuals per unit area, while the inelastic model is slower only when the density is below about 7 individuals per unit area. Note that the actual density

of individuals taking part in the spatial interaction is five times greater due to this being a mortality regulated population rather than a fecundity regulated population: each female has a full litter with an average size of 8, and the population is then regulated back to its capacity size. Thus, it can be anticipated that in a fecundity regulated model, the resource-explicit models would perform more slowly than the standard spatial model at densities under about 95 individuals per unit area for the elastic model and 35 for the inelastic model.

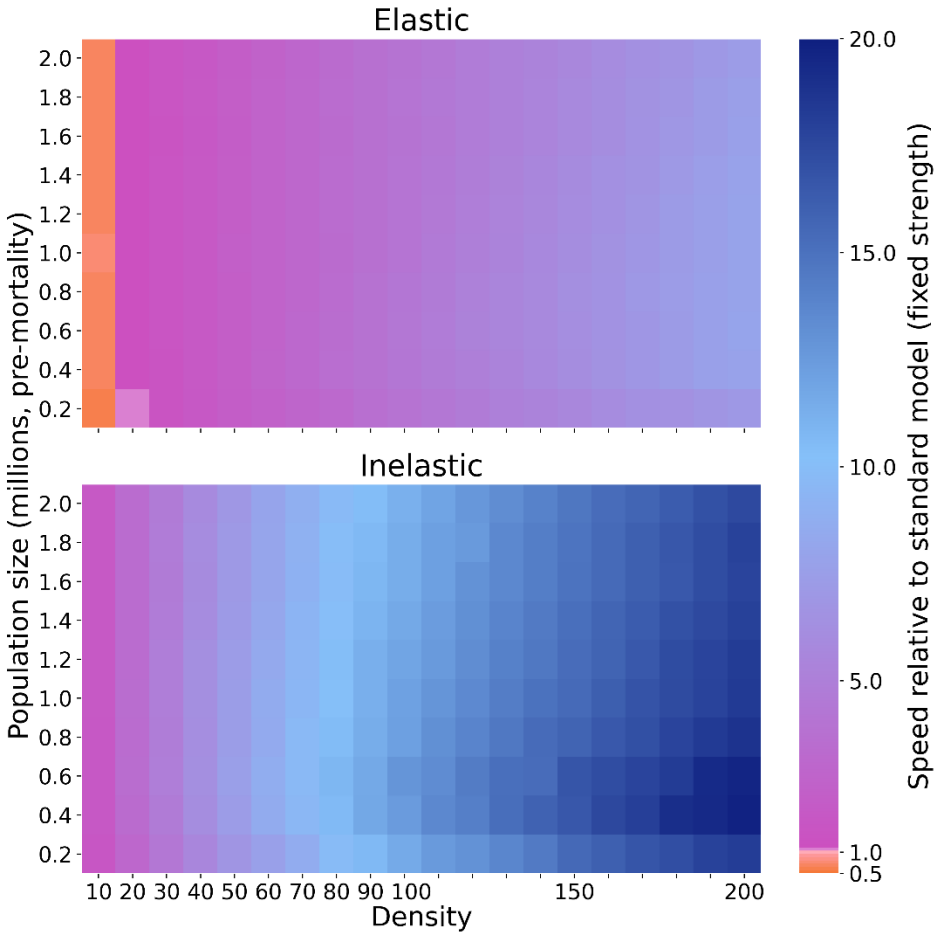

Figure 7. Runtime comparison with varying density and population size. Density was varied in increments of 10 individuals per unit area, and the modeled area was varied such that the population size increased in steps of 200,000. The colors displayed on the two heatmaps indicates the relative runtime speed of the resource-explicit models as compared to a standard spatial model using a fixed strength interaction function.

One potential drawback of the resource-explicit models is that they require more memory than standard spatial models, as the resource node entities represent a large array of objects that must be accounted for. In a simulation with a

density of 100 individuals per unit area and an area of 2000 (resulting in a post-reproduction population size of one million individuals), the standard models had an average of 540 Mb peak memory usage. The elastic models required substantially more, with a 1650 Mb peak usage for most variants (slightly less for the variant with 19 hexagons per unit area), while the inelastic models required only somewhat more than the standard spatial models.

The panmictic model serves to give a baseline understanding how much of the runtime of the various other models is spent doing work unrelated to spatial interactions (Fig 6). However, this should also serve as a caveat to the rest of the results: as the runtime of the panmictic model demonstrates, very little is going on in these models other than spatial interactions. But in models with a complex genetic component, spatial calculations may occupy a substantially smaller portion of the overall runtime of the model. For such models, switching to a resource-explicit modeling approach cannot be expected to yield the same relative increases in speed observed here. Additionally, runtimes in this manuscript were measured on a highly performant desktop CPU which was only performing a single simulation at a time. Runtimes may be much slower on a cluster CPU performing many other tasks in parallel. Nonetheless, the relative runtime improvements measured in this manuscript should approximately hold in such contexts.

]

Discussion

Before consideration of the runtime improvements and other potential advantages of the resource-explicit approach, the threshold question must be answered: are the interaction dynamics within the model fit for purpose? We assert that they are in the majority of contexts.

The circle overlap interaction function is inspired by the biological reality of resource foraging behavior, and is likely as reasonable an interaction function as any of the more commonly used functions. The magnitudes by which indirect interaction strengths in the resource-explicit frameworks differs from this function are quite small. While the interaction strengths are not perfectly radially

symmetric, the small random variation in pairwise interaction strengths is negligible within a model with so many other stochastic elements, amounting to little more than random seafoam atop what is already a random wave.

The unequal availability of resource nodes in the inelastic model (Supplemental Figures 3 and 4) should perhaps not be dismissed quite as readily. In a species with a smaller population or with a lower birthrate, a model in which individuals might occasionally perish just because they happened to step on the wrong spot in the grid is not desirable. For such species, the elastic model or the "fair" variant of the inelastic model should be preferred. On the other hand, for species with large populations and high birthrates, randomly assigning some individuals a somewhat higher mortality rate while randomly assigning others a somewhat lower mortality rate should not disturb model outcomes.

This modeling approach could be seen as a relative of approaches used in several other fields and which also utilize spatial approximations or discretizations to increase performance. In computational fluid dynamics fluids are treated as a set of discrete cells, as the simulation of individual molecules would not be tractable for most applications[16]. In computational astrophysics, the Barnes-Hut method reduces the complexity of n-body simulations from $O(n^2)$ to $O(n \log n)$ by approximating the interaction forces between distant objects by treating multiple distant objects as having a single interaction force emanating from a spatial cell's center of mass[17].

The resource-explicit approach to modeling comes with a number of exciting advantages and possible applications. The runtime speed of the models is among the most exciting of these advantages. Studies involving continuous spatial models have tended to contemplate populations of between 10000 and 100000 individuals. In addition to allowing simulations of such populations to be conducted more quickly and efficiently, the resource-explicit approach will allow for the simulation of hitherto impractically large populations. Many large populations which are currently studied by applying a scaling multiplier to mutation rates, or which are not modeled in continuous space at all, can now be modeled in full in spatial models.

In addition to the runtime advantages demonstrated in this manuscript, there are a number of features that can be implemented with much greater efficiency in a resource-explicit model than in a standard spatial model, either in terms of implementation effort, runtime, or both. The following is an incomplete list of such features.

Multiple classes of resources. Many species require different classes of resources at different stages of life, or need a specific resource in order to enable reproduction. For example, monarch butterflies can feed on nectar from many plants, but only lay their eggs on milkweed; mosquitos feed on nectar or other plant juices, but female mosquitos only reproduce after taking a blood meal from their selected host, after which their eggs can only be laid in water sources. In a resource-explicit model, resource nodes can be parameterized with a separate value for each relevant class of resources, allowing for the simulation of complex life histories while adding almost no computational overhead to the model.

Interspecies competition. In a standard model with multiple competing species, an interaction function between each species and each other species must be evaluated, in addition to an intraspecies interaction for each species. E.g., in a model considering three competing species, A, B, and C, a total of six types of interactions must be evaluated: three intraspecies interactions (one for each of A, B, and C) and interspecies interactions between species A and B, between A and C, and between B and C.

The resource-explicit modeling approach is significantly more efficient in this context. Instead of six types of interaction, only three interactions are necessary – the interaction between each of the three species with resource node entities. There may be some complications in setting up such a model if the foraging areas of the species involved have very different scales. Otherwise, each species can be parameterized with a different foraging radius, and the exact tiling density of the resource nodes can be selected such that the species with the smallest foraging area has enough resource nodes within foraging range that the interaction is as artifact-free as desired (perhaps 37 or 19 nodes per foraging area as in this manuscript, though 7 might suffice in some modeling contexts). The

number of hexagons corresponding to the interaction area of an individual need not be selected from the sequence of "hexagonal numbers," though not doing so may result in a slight increase in spatial artifacts.

This efficiency increase only applies to the simulation of interspecies resource competition. Other interspecies interactions, such as a predator-prey relationship, are neither more nor less efficient in this model framework, as the interaction between the predator and prey species must still be evaluated.

Heterogenous landscapes and irregular area boundaries. The resource-explicit approach elegantly lends itself to the simulation of realistic heterogenous geography. To accomplish this, nodes can simply be parameterized with different resource amounts depending on the local habitat quality in different areas of the model. Heterogenous landscapes can be simulated in a standard spatial model as well, but doing so requires additional computational overhead, as each individual must ascertain the quality of the local habitat (e.g., by referencing their position against the color of the corresponding pixel in an image of the landscape) every time they compete for resources. By contrast, no such additional overhead is necessary in a resource-explicit model after the initial setup of the model.

An irregular area boundary, such as a coastline, is another feature that is substantially easier to implement in a resource-explicit model. With an irregular boundary, standard techniques for avoiding edge-effects cannot be applied. In a standard spatial model, the best methods to simulate such a boundary would probably require substantial programmer effort and pre-calculation of interaction strength multipliers near the boundary. In resource-explicit models, no such effort is necessary (aside from preventing individuals from dispersing into the sea, which must be done in either type of model).

Resource variability as a function of time. Similarly, resource-explicit models are amenable to simulating varying resource availability as a function of time, whether in the form of periodic seasonal variation, longer term periodic or random variation (such as tree masting, a phenomenon in which, during some years, certain species of trees produce abnormally large crops of seeds in order to

overwhelm foragers and ensure that some seeds go uneaten), or long-term trajectories (such as degradation of habitat due to climate change). To accomplish this, after each tick of the model, resource nodes can be replenished according to an appropriate function of time, rather than being replenished to the starting resource value of the node. Implementing this feature would not likely result in a noticeable increase in the runtime of the model.

Focal species induced habitat change. In addition to competing for resources, many species also have a direct impact on the landscape on which they live. For example, many species of bacteria and yeast, such as the historically beloved *Saccharomyces cerevisiae* (brewer's yeast), ingest sugars or other organic molecules, and excrete ethanol[18]. As the concentration of alcohol increases and the availability of sugars decreases, the capacity of such a system decreases. Systems of this nature are usually modeled as panmictic systems via differential equations, but the resource-explicit modeling approach can allow for the spatial modeling of these processes by explicitly tracking the amount of sugar and alcohol at each resource node. While panmictic models are likely suitably descriptive of the brewing process, there are many similar contexts where a spatial model may be desirable (e.g., for the simulation of petri dish experiments).

Imperfect resource regeneration rate. In many natural systems, groups of animals locally extract resources much faster than those resources can replenish themselves. For example, herds of grazing animals might eat grass in a local area much faster than it can grow, but the animals survive by migrating across the landscape, leaving previously exploited resources ample time to recover. Dynamics of this type are omitted from standard spatial models, but are quite straightforward to implement in a resource-explicit model by simply parameterizing resource nodes with a resource maximum along with a regeneration rate or function.

Simulation of other landscape features. Resource availability isn't the only landscape feature that can have a profound impact on animal behavior. Natural and artificial barriers, such as rivers and highways respectively, might severely curtail migration between otherwise adjacent areas[19]. Dispersal behavior of

human-commensal species may be entirely different in urban environments than in any other context (e.g., a mouse may disperse over a kilometer in the wild, while a mouse in a well-provisioned kitchen might be quite content to never travel more than a room away)[20,21].

The resource-explicit modeling approach provides at least a basic framework in which such features could be implemented, some more easily than others. Implementing different dispersal characteristics in urban areas, for example, would be quite easily accomplished by adding an additional property to each resource node denoting the local human population density, and altering the dispersal of individuals based on that property. To model dispersal barriers such as roads, hexes through which such infrastructure passes could be given an additional property, which could then impact the dispersal characteristics of individuals within interaction range of those hexes.

Adding a large number of such features to the simulated landscape might have undesirable consequences for model runtime. Other simulation frameworks, such as HexSim, may be more appropriate for simulating landscape features such as dispersal barriers if they are central to a research question[22]. HexSim treats landscapes as consisting of a grid of hexagonal cells rather than a continuous space. However, the ranges exploited by individuals and groups within the framework can span multiple hexes, which likely allows the avoidance of many artifacts otherwise found in models consisting of linked panmictic cells. In this framework, hexagon edges are an explicit part of the model, and collections of edges can be used to define migration barriers with greater ease and with less computational overhead than is likely necessary in a continuous space model.

There are a number of contexts in which the resource-explicit approach might not be the preferred choice. This framework is designed to emulate competition to exploit resources. If direct or “interference” competition is more dispositive to system outcomes, it makes more sense to directly measure interaction strengths between individuals by using a standard spatial framework.

Or in systems with very low densities or small populations (particular relevant for the modeling of endangered or threatened species), a resource-explicit model may actually run more slowly than a standard spatial model, which might warrant the selection of the latter (though the flexibility of the resource-explicit approach for modeling heterogenous landscapes or interspecies competition may still make it a preferable choice).

In the majority of modeling contexts, the resource-explicit method promises superior runtime speeds and excellent flexibility compared to standard spatial models. Adoption of this method will allow for the construction of continuous-space models of populations that were previously intractably large. Not only that, but these populations can be simulated on realistic landscapes with realistic boundaries. The flexibility of this method will also allow for increased realism in modeling numerous factors that impact resource availability which are prohibitively computationally intensive to include in current models.

Data and Code Availability

The SLiM code for the models presented in this paper, including each of the variations, along with the data used in this manuscript, are available at https://github.com/MesserLab/ResourceExplicitModels. The SLiM simulation software used in the project is available at https://messerlab.org/slim/ or at https://github.com/MesserLab/SLiM.

Acknowledgements

This study was supported by funding from the National Institutes of Health awards R01GM127418 to P.W.M.

References

1. Wright, S. Evolution in Mendelian Populations. *Genetics* **16**, 97 (1931).
2. Battey, C. J., Ralph, P. L. & Kern, A. D. Space is the Place: Effects of Continuous Spatial Structure on Analysis of Population Genetic Data. *Genetics* **215**, 193 (2020).
3. Kingman, J. F. C. The coalescent. *Stoch Process Their Appl* **13**, 235–248 (1982).
4. Pujolar, J. M. Conclusive evidence for panmixia in the American eel. *Mol Ecol* **22**, 1761–1762 (2013).
5. Beveridge, M. & Simmons, L. W. Panmixia: an example from Dawson's burrowing bee (Amegilla dawsoni) (Hymenoptera: Anthophorini). *Mol Ecol* **15**, 951–957 (2006).
6. Champer, J., Kim, I. K., Champer, S. E., Clark, A. G. & Messer, P. W. Suppression gene drive in continuous space can result in unstable persistence of both drive and wild-type alleles. *Mol Ecol* **30**, 1086–1101 (2021).
7. Muktupavela, R. A. *et al.* Modeling the spatiotemporal spread of beneficial alleles using ancient genomes. *Elife* **11**, (2022).
8. Lundgren, E. & Ralph, P. L. Are populations like a circuit? Comparing isolation by resistance to a new coalescent-based method. *Mol Ecol Resour* **19**, 1388–1406 (2019).
9. Epperson, B. K. Geographical genetics. 356 (2003).
10. Birch, C. P. D., Oom, S. P. & Beecham, J. A. Rectangular and hexagonal grids used for observation, experiment and simulation in ecology. *Ecol Modell* **206**, 347–359 (2007).
11. Garud, N. R., Messer, P. W., Buzbas, E. O. & Petrov, D. A. Recent Selective Sweeps in North American Drosophila melanogaster Show Signatures of Soft Sweeps. *PLoS Genet* **11**, e1005004 (2015).
12. Min, J., Gupta, M., Desai, M. M. & Weissman, D. B. Spatial structure alters the site frequency spectrum produced by hitchhiking. *Genetics* **222**, (2022).
13. Bentley, J. L. Multidimensional binary search trees used for associative searching. *Commun ACM* **18**, 509–517 (1975).

14. Haller, B. C. & Messer, P. W. SLiM 3: Forward Genetic Simulations Beyond the Wright–Fisher Model. *Mol Biol Evol* **36**, 632–637 (2019).

15. Haller, B. C. & Messer, P. W. SLiM 4: Multispecies eco-evolutionary modeling. *The American Society of Naturalists* (2023) doi:10.1086/723601.

16. Leer, B. van & Powell, K. G. Introduction to Computational Fluid Dynamics. *Encyclopedia of Aerospace Engineering* (2010) doi:10.1002/9780470686652.EAE048.

17. Barnes, J. & Hut, P. A hierarchical O(N log N) force-calculation algorithm. *Nature* **324**, 446–449 (1986).

18. Shopska, V., Denkova-Kostova, R. & Kostov, G. Modeling in Brewing—A Review. *Processes 2022, Vol. 10, Page 267* **10**, 267 (2022).

19. Byers, K. A., Lee, M. J., Patrick, D. M. & Himsworth, C. G. Rats About Town: A Systematic Review of Rat Movement in Urban Ecosystems. *Front Ecol Evol* **7**, 13 (2019).

20. Sol, D., Lapiedra, O. & González-Lagos, C. Behavioural adjustments for a life in the city. *Anim Behav* **85**, 1101–1112 (2013).

21. Dammhahn, M., Mazza, V., Schirmer, A., Göttsche, C. & Eccard, J. A. Of city and village mice: behavioural adjustments of striped field mice to urban environments. *Scientific Reports 2020 10:1* **10**, 1–12 (2020).

22. Schumaker, N. H. & Brookes, A. HexSim: a modeling environment for ecology and conservation. *Landsc Ecol* **33**, 197–211 (2018).

23. Grünbaum, B. & Shephard, G. C. Tilings by Regular Polygons. *Mathematics Magazine* **50**, 227 (1977).

24. Thompson, K. & Dray, T. Taxicab Angles and Trigonometry. *Pi Mu Epsilon Journal* **11**, 87–96 (2000).