The Impact of Landscape Sparsification on Modelling and Analysis of the Invasion Process*

Daniyah A. Aloqalaa[1], Jenny A. Hodgson[2], and Prudence W.H. Wong[1]

1 **Department of Computer Science, University of Liverpool, UK**
{d.a.aloqalaa, pwong}@liverpool.ac.uk
2 **Department of Evolution, Ecology and Behaviour, University of Liverpool, UK**
jenny.hodgson@liverpool.ac.uk

Abstract

Climate change is a major threat to species, unless their populations are able to invade and colonise new landscapes of more suitable environment. In this paper, we propose a new model of the invasion process using a tool of landscape network sparsification to efficiently estimate a duration of the process. More specifically, we aim to simplify the structure of large landscapes using the concept of sparsification in order to substantially decrease the time required to compute a good estimate of the invasion time in these landscapes. For this purpose, two different simulation methods have been compared: *full* and *R-local* simulations, which are based on the concept of dense and sparse networks, respectively. These two methods are applied to real heterogeneous landscapes in the United Kingdom to compute the total estimated time to invade landscapes. We examine how the duration of the invasion process is affected by different factors, such as dispersal coefficient, landscape quality and landscape size. Extensive evaluations have been carried out, showing that the R-local method approximates the duration of the invasion process to high accuracy using a substantially reduced computation time.

1998 ACM Subject Classification E. Data - Graphs and networks; G.3 Probability and Statistics; J.3 Computer Applications - Life and Medical Sciences - Biology and genetics

Keywords and phrases Landscape sparsification, invasion process, network sparsification, dense and sparse networks

Digital Object Identifier 10.4230/LIPIcs.SEA.2017.35

1 Introduction

Climate change and land use change are two threats that cause the extinction of numerous species [4, 7, 12, 17]. It was observed that species are responding to climate change by shifting their geographical area [1, 16], however the ability of their population to shift depends on the availability of suitable habitat to shift and colonise [4, 6, 12, 13]. Species become under a high risk of extinction if they are shifting very slowly or do not have the ability to shift [15]. Therefore, to maintain the functioning of an ecosystem and services in a changing climate, it becomes an important need to facilitate the adaptation of species, especially by enabling them to shift to new locations with more suitable climate [5]. It is an urgent need for policymakers and nature conservation organisations to find out whether and how they can facilitate range shifts [5]. A number of empirical and theoretical studies have shown that

* The source codes and data used are available at https://www.dropbox.com/sh/yu37cc8aftr5nox/AACW5Orvb8gCDSdOOJFiEDlua?dl=0

16th International Symposium on Experimental Algorithms (SEA 2017).
Editors: Costas S. Iliopoulos, Solon P. Pissis, Simon J. Puglisi, and Rajeev Raman; Article No. 35; pp. 35:1–35:16
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

spatial arrangement of habitats is an important factor that affects the speed of advance to new landscapes with more suitable climate [3, 5, 17]. Hodgson et al. [5] found the evidence of the benefits of using different tools such as habitat corridors and stepping stones to speed up shifting. However, it is still difficult for conservationists to make decisions that can facilitate range shifts in large landscapes, therefore there is a need for a tool efficiently computing the invasion time of the original and modified landscapes. Minimising computation time is especially important because the ultimate aim is for a decision-making tool that can tune the arrangement of the modified landscape to find scenarios where a small addition of habitat leads to a large decrease in invasion time. Running many scenarios with different permutations of habitat could require excessive computation times even with moderately-sized landscapes. Furthermore, planning for climate change requires the consideration of large landscapes (e.g., temperature isoclines are expected to shift at several km per decade).

In this study, our focus is to build a new sparse computational model for the invasion process using the network modelling approach. To model invasion process in a given landscape, we create a landscape network, where each vertex represents a patch of habitat (henceforth patch) in the landscape. We distinguish two sets of patches: the source patches represent initially populated patches in which species are located, and the target patches represent the target locations for the invasion process. The invasion process is to populate (some) target patches and we aim to estimate the time needed for achieving this. A stochastic model from [5] has been implemented in the simulation, which is based on the probability of a patch to be invaded expressed by a formula depending on various characteristics (distance, quality of patch, etc.) of all other patches in the network. Such simulation is computationally expensive, especially when the number of patches is large.

In this paper we propose to approximate the computed invasion time by exploiting network sparsification. We call the original approach in [5] the *full* invasion simulation method. In the protocol implementing *full* invasion method, in each round of the invasion process each populated patch tries to populate (independently) all other unpopulated patches in the landscape. On the other hand, we propose the *R-local* invasion method, and associated protocol, such that in each round, every patch only tries to populate other unpopulated patches within a local distance R. The *full* invasion protocol can be seen as an *R-local* invasion protocol with R equal to the diameter of the landscape. With a smaller value of R, the *R-local* invasion process is expected to take less time in each round of computation, while it may take more rounds to populate the target patches. If the local distance R is chosen properly, the *R-local* invasion protocol can compute a comparable (to the *full* method) duration of invasion process, while the computation time can be substantially reduced. That means our proposed method is much more efficient (less computational time) than the existing model. One important characteristics to consider when computing the total time of experiment is also the number of simulations needed for the (average) invasion time to stabilise on the outputted duration of invasion. In this paper, we illustrate how to determine the local distance R systematically to reach a good trade-off between quality of estimate and computation time.

In both *full* and *R-local* protocols, we investigate the effect of three factors: species' mean dispersal distance, landscape quality and landscape size, on the invasion duration. We apply the *full* and *R-local* protocols to real heterogeneous landscapes in the United Kingdom. An extensive experimental evaluation with real data illustrates the effectiveness and accurateness of the proposed *R-local* protocol in estimating the duration of invasion process in large landscapes in a relatively short time, with respect to the previously used *full* method.

Technically speaking, the work presented in this paper combines ideas from probability

Table 1 Aggregate classes [10].

Aggregate class number	Aggregate class	Aggregate class number	Aggregate class
1	Broadleaf woodland	6	Mountain, heath, bog
2	Coniferous woodland	7	Saltwater
3	Arable	8	Freshwater
4	Improved grassland	9	Coastal
5	Semi-natural grassland	10	Built-up areas and gardens

and random processes [2, 8, 9] with some use of network/graph foundations [11].

2 Materials and methods

2.1 Notation and technical preliminaries

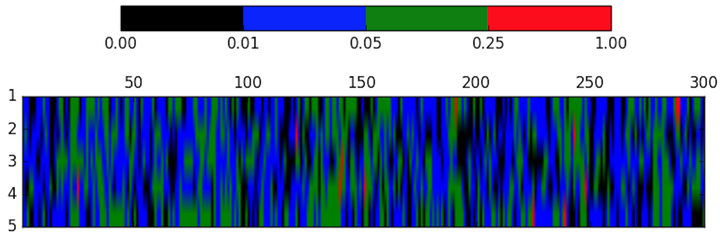
We are given a 2-dimensional rectangular grid landscape as an input, which we represent as a landscape graph $G = (V, E)$. Denote by $n = |V|$ the number of vertices/patches in the landscape graph, and by $m = |E|$ the total number of edges between patches. Let $q(v)$ denote the quality of a patch v, where the quality is a number between zero and one given as input. We distinguish two sets of patches, S and T, where S denotes the set of populated source patches that are non-zeros in quality and T denotes the set of unpopulated target patches that are also non-zeros in quality. In addition, we define the maximum and minimum quality as $q_{max} = \max\{q(v) : v \in V\}$ and $q_{min} = \min\{q(v) > 0 : v \in V\}$, respectively. We define the height H as the difference in y-coordinates of any top and bottom patch in the landscape graph G and the width W as the difference in x-coordinates of any left patch in the source set and right patch in the target set. We also define *R-local* graph $Loc(G, R) = (V', E')$ as a subgraph of G that contains all the vertices of the landscape graph G (i.e., $V' = V$) and a set of edges $E' \subseteq E$ such that for any $(u, v) \in E$, $(u, v) \in E'$ if the Euclidean distance between vertices v and u is at most R. For a defined *R-local* graph $Loc(G, R)$, we define d_{min} as the minimum distance d such that $Loc(G, d)$ connects S to T, i.e., d_{min} is the minimum distance d such that every node in T is reachable from some node in S in graph $Loc(G, d)$.

2.2 The studied landscape

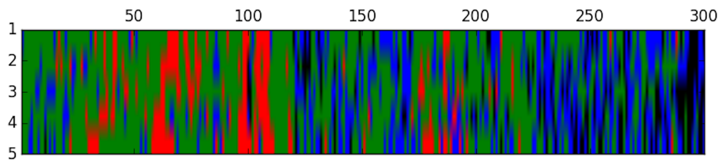
For evaluation purposes, the dataset of the 1km resolution raster version of the Land Cover Map 2007 (LCM2007) for Great Britain [14] is used. To examine the proposed methods, different sized landscapes from different maps of the aggregate classes are extracted. The aggregate classes data contain one tiff file for each land use class as in Table 1. Each map consists of 1300 rows/height (pixels) and 700 columns/width (pixels) and each 1km pixel provides the percentage cover of a particular land cover at LCM2007 Class level [10]. We consider the percentage cover at each patch in such an extracted landscape as the quality of each patch. For examination purposes, from percentage values we formed three groups of landscape qualities, namely: low quality, medium quality and high quality to represent the quality of the extracted landscape. If the average of all patches qualities in such an extracted landscape is between 0% and 5%, 5% and 25%, 25% and 100%, then the extracted landscape is of low quality, medium quality, and high quality, respectively. For each quality type, we extract a rectangular landscape that consists of 5 rows and 300 columns. Landscapes of low and medium qualities were extracted from semi-natural grassland UK map [10], while the one of high quality from an improved grassland UK map [10]. On these extracted rectangular landscapes, we assume that all patches at the first column of each landscape are occupied and the goal is to populate patches at the target columns (col. 10, 20, 30, etc.). Each landscape is extracted according to the following criteria:

1. The qualities of all source occupied patches are non-zeros.
2. At each target column, at least one of the patches is non-zero in quality.

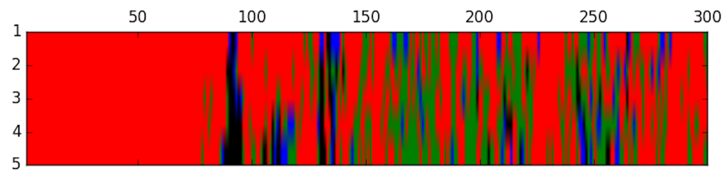
Figure 1 shows real landscapes used which are extracted from two LCM2007 UK (aggregate classes) maps.



(a) Landscape of size 5×300 and of low quality.



(b) Landscape of size 5×300 and of medium quality.



(c) Landscape of size 5×300 and of high quality.

Figure 1 The studied landscapes; three landscapes of size 5×300 and of low, medium and high quality extracted from LCM2007 UK (aggregate classes) maps. In each landscapes, the colour corresponds to the quality of each patch; black, blue, green, and red corresponds to zero, low (0-0.5), medium (0.5-0.25), and high quality (0.25-1), respectively.

2.3 New formulas estimating duration of the invasion process

For a given landscape graph G, we use the formula of colonisation probability proposed by Hodgson et al. [5] to define the *transition probability* between patches v and u as $p(v,u) = q(v) \cdot \frac{\exp(-\alpha d(v,u))}{\left(\frac{2\pi}{\alpha^2}\right)-1}$, where $\alpha > 0$ is the dispersal coefficient assumed to be the same for all patches and $d(v,u)$ is the Euclidean distance between patches v and u.

For a given landscape graph G, source S and target T, we define "all non-zero target patches" as the total number of target patches that are non-zero in quality, and "majority of all non-zero target patches" as the number of patches being more than half of the total number of non-zero quality target patches. In this work, we consider three types of invasions, namely: *first* success, *majority* success and *all* successes invasion. The *first* success invasion measures the estimated time to populate any of the non-zero target patches. Therefore, the *first* success is defined as the estimated time needed until the first non-zero target patch becomes populated. The *majority* success is the estimated time to populate the majority (more than 50%) of all non-zero target patches. Finally, *all* successes is defined as the estimated time to populate all non-zero target patches.

Based on the formula of the *transition probability*, we propose three new estimating times of invasions in such a landscape. For a given landscape graph G, in order to hop from source S to target T, we need to find a path that contains at most $\frac{H+W}{d_{min}(G)}$ hops. According to the formula of *transition probability* between patches, the probability of a single hop in the landscape graph G is at least $q_{min}(G) \cdot \frac{\exp(-\alpha \cdot d_{min}(G))}{\left(\frac{2\pi}{\alpha^2}\right)-1}$. Thus, the expected time for each hop is the inverse of its probability, i.e., $\frac{\left(\frac{2\pi}{\alpha^2}\right)-1}{q_{min}(G)\cdot\exp(-\alpha\cdot d_{min}(G))}$. Since the number of hops that are needed in such a path to connect source S with target T is $\frac{H+W}{d_{min}(G)}$, the total expected time of invasion for the *first* success is at most

$$\frac{H+W}{d_{min}(G)} \cdot \frac{\left(\frac{2\pi}{\alpha^2}\right)-1}{q_{min}(G)\cdot\exp\left(-\alpha\cdot d_{min}(G)\right)} \cdot c\ ,$$

where c is a small constant to be determined by simulation. Consequently, the total expected time of invasion for the *majority* success is at most

$$\left[\frac{H+W}{d_{min}(G)} \cdot \frac{\left(\frac{2\pi}{\alpha^2}\right)-1}{q_{min}(G)\cdot\exp\left(-\alpha\cdot d_{min}(G)\right)} + \left(\frac{H}{2}-1\right) \cdot \frac{\left(\frac{2\pi}{\alpha^2}\right)-1}{q_{min}(G)\cdot\exp\left(-\alpha\cdot d_{min}(G)\right)}\right] \cdot c\ ,$$

where $\left(\frac{H}{2}-1\right)$ is an upper bound on the total number of *majority* of target patches decreased by one and c is again a small constant to be interpolated by simulation. Therefore, the total expected time of invasion for *all* successes is at most

$$\left[\frac{H+W}{d_{min}(G)} \cdot \frac{\left(\frac{2\pi}{\alpha^2}\right)-1}{q_{min}(G)\cdot\exp\left(-\alpha\cdot d_{min}(G)\right)} + (H-1) \cdot \frac{\left(\frac{2\pi}{\alpha^2}\right)-1}{q_{min}(G)\cdot\exp\left(-\alpha\cdot d_{min}(G)\right)}\right] \cdot c\ ,$$

where $(H-1)$ is an upper bound on the total number of *all* target patches decreased by one and c is a small constant to be interpolated by simulation.

2.4 The R-*local* simulation method

We simulate the behaviour of the invasion process by building a simulator that uses the *R-local* method to compute the number of rounds needed for invasion. The inputs to the simulator are: a two-dimensional array that represents a given real landscape and stores qualities of patches, a source vector containing indices of populated patches, a target vector containing indices of unpopulated target patches, and dispersal coefficient α (in Algorithm 1 initialised to 0.25). For a given landscape, the simulator constructs a two-dimensional array of a size equal to the one of the given landscape. Each cell in the constructed array corresponds to a patch in the landscape and can take only two values, zero or one, where zero means the cell is unpopulated while one means it is populated. At the beginning of the invasion process, only source populated patches take value of one and others take value of zero. The simulator returns an estimated duration needed (number of rounds) to invade targets by the use of real probabilities for each pair of patches v, u, in which v is populated and u is not. We use transition probabilities between patches v and u to decide whether patch v populates patch u or not.

More formal description of Algorithm 1 provides the structure of the *R-local* method. The generic structure of the *R-local* method contains inputs (as mentioned above), outputs, and COUNT ROUNDS function. The COUNT ROUNDS function counts the number of rounds required for *first*, *majority* and *all* successes and to compute the real time execution for each simulation. The function includes nested loops of three levels. The main loop (lines 12-37) counts the number of rounds to populate target patches. The second level loop (lines 14-36) is for all populated patches that are trying to populate unpopulated patches. The inner level loop (lines 17-33) is for all unpopulated patches. Each unpopulated patch becomes populated if the *transition probability* between the populated and unpopulated patches is greater than a random generated number between zero and one (lines 22-25). We consider only populating a patch with non-zero quality. Each time when an unpopulated target patch becomes populated, the algorithm checks if the total number of non-zero patches at target is equal to one or *majority* or *all* targets' number, and the number of rounds is recorded accordingly. The COUNT ROUNDS function terminates when all non-zero target patches become populated and returns the number of rounds needed for each type of successful invasions as well as the execution time of simulation.

Recall that in the *full* invasion method, each populated patch in the landscape tries to populate every other unpopulated patch in the whole landscape, while in the *R-local* invasion method each populated patch in the landscape only tries to populate every other unpopulated patch within local distance R around the populated patch. Thus, parameter R in the *full* method takes the whole size (diameter) of the landscape (i.e., $R = H + W$), while in the *R-local* method we propose an equation to compute the local distance R. In addition, the inner level loop in COUNT ROUNDS function runs for all unpopulated patches in the whole landscape (*from 0 to* $H + W$) in the *full* method, while runs only for unpopulated patches that are of distance at most R from the populated patch (*from populated patch to* R) in the *R-local* method (lines 17 and 18 in Algorithm 1).

It is expected that the number of rounds required for the *R-local* method is larger. In our simulation, we aim to find the local distance R that allows the following accuracy:

$$\frac{\text{FULL}}{\text{R-LOCAL}} = \frac{\text{average number of rounds using the } \textit{full} \text{ method}}{\text{average number of rounds using the } \textit{R-local} \text{ method}} \geq 90\% \ .$$

As a starting point, we run simulation using both *full* and *R-local* methods with some expected local distances R. It has been observed that the required local distance R for the FULL/R-LOCAL ratio to be at least 90% in the *first* success is greater than or equal to the needed local distance R for *majority* and *all* successes. Based on this observation, we propose an equation that computes the local distance R for a given landscape graph G based on the total expected time of invasion for the *first* success. Observe that the probability of a single hop in a given landscape graph G is less than the inverse of the total expected time of the invasion process for the *first* success:

$$q_{min}(G) \cdot \frac{\exp\left(-\alpha \cdot d_{min}(G)\right)}{\left(\frac{2\pi}{\alpha^2}\right) - 1} \ll \frac{1}{\text{total expected time of invasion for } \textit{first} \text{ success}} \ .$$

Therefore, the following holds:

$$q_{max}(G) \cdot \frac{\exp\left(-\alpha \cdot R(G)\right)}{\left(\frac{2\pi}{\alpha^2}\right) - 1} \ll \frac{d_{min}(G)}{H + W} \cdot q_{min}(G) \cdot \frac{\exp\left(-\alpha \cdot d_{min}(G)\right)}{\left(\frac{2\pi}{\alpha^2}\right) - 1}$$

$$-\alpha\left(R(G) - d_{min}(G)\right) < \ln\left(\frac{d_{min}(G) \cdot \bar{q}(G)}{H + W}\right) \ .$$

Finally, we get the best (smallest) local distance R needed to create a landscape subgraph $Loc(G, R)$, for a given landscape graph G, and guarantees the FULL/R-LOCAL accuracy is:

$$R(G) \approx \left[\frac{1}{\alpha} \cdot \ln\left(\frac{H+W}{d_{min}(G) \cdot \bar{q}(G)}\right) + d_{min}(G)\right] \cdot c\ , \tag{1}$$

where $\bar{q}(G) = \frac{q_{min}(G)}{q_{max}(G)}$ and c is a small constant to be determined by simulation.

2.4.1 Interpolating constant c in Equation (1)

We use the following three objective functions in order to interpolate constant c in Equation (1). We use terminology "approx R" and "opt R" to express the local distance R using Equation (1) and simulation such that it is the smallest distance satisfies the FULL/R-LOCAL accuracy, respectively. We call it opt R because this value of R fulfills our goal of accuracy.

1. The Euclidean distance objective function (ED) chooses the constant c such that it minimises the sum over all prefixes z of the square difference between the approx R and opt R:

$$c_{ED} = \underset{c \in \mathbb{R}^+}{\operatorname{argmin}} \left\{ \sum_{j=1}^{z} (approxR_j \cdot c - optR_j)^2 \right\}.$$

2. The absolute objective function (AB) chooses the constant c such that it minimises the sum over all prefixes z of the absolute difference between the approx R and opt R:

$$c_{AB} = \underset{c \in \mathbb{R}^+}{\operatorname{argmin}} \left\{ \sum_{j=1}^{z} |approxR_j \cdot c - optR_j| \right\}.$$

3. The min-max (MM) objective function chooses the constant c such that it minimises the approx R to be greater than or equal to opt R for all prefixes z:

$$c_{MM} = \underset{c \in \mathbb{R}^+}{\operatorname{argmin}} \left\{ \max_{1 \le j \le z} (approxR_j \cdot c - optR_j) \ge 0 \right\}.$$

2.5 Simulation

2.5.1 Main simulation

The simulation is directed at four goals. The first is to monitor the behaviour of *full* and *R-local* methods and compare results obtained by each method. The second is to investigate what values of local distance R would allow the FULL/R-LOCAL accuracy. Based on the results obtained from the simulation, the third goal is to predict an equation for the local distance R, depending on landscape size, dispersal coefficient α and landscape quality. Finally, we aim to combine results from simulation and the predicted equation to compare them and conduct an independent validation based on the value obtained from the proposed equation on the local distance R.

For each prefix $5 \times 10, 5 \times 20, 5 \times 30, ..., 5 \times 300$ in each of the extracted rectangular landscapes in Figure 1, we run the *full* simulation and, simultaneously, the *R-local* simulation with some predicted distances; ideally, one of these R gives the FULL/R-LOCAL accuracy, while the local distance R decreased by one does not satisfy it. We consider four values

Algorithm 1 Modelling Invasion Process using *R-local* Method

```
Inputs:
    2-dimensional array G, which stores qualities of patches in a real landscape
    Source: vector S contains indices of initial populated patches
    Target: vector T contains indices of unpopulated target patches
    α ← 0.25
Outputs:
    Number of rounds needed for first, majority and all successful invasions
    Execution time of simulation
function COUNT ROUNDS(G, S, T, α)
    SimulationStartTime ← record start time of simulation
    R ← Use Equation (1)
    Create 2-dimensional array B having size equal to array G ← 0
    for each index of populated patch in vector S do
        B(index) ← 1
    end for
    Counter ← 0                    ▷ counter to count rounds for successful invasions
    StartTime ← Record start time of rounds
    while any of target patch in T is unpopulated do
        Counter ← Counter+1
        for i ← 0 to number of rows in array G do         ▷ loop for all populated patches (1)
            for j ← 0 to number of columns in array G do
                if patch B(i, j) is populated then
                    for k ← i − R to i + R do             ▷ loop for all unpopulated patches (0)
                        for l ← j − R to j + R do
                            if patch B(k, l) is unpopulated and its quality ≠ 0 then
                                distance ← Euclidean distance between B(i, j) and B(k, l)
                                if 0 < distance ≤ R then
                                    p ← Transition probability between B(i, j) and B(k, l)
                                    w ← Generate a random number between 0 and 1
                                    if w < p then
                                        Populate patch B(k, l)
                                        if populated patch B(k, l) is at target column then
                                            Check the invasion type first or majority or all
                                        end if
                                    end if
                                end if
                            end if
                        end for
                    end for
                end if
            end for
        end for
    end while
    SimulationEndTime ← Record end time of simulation
    SimulationTimeExecution ← SimulationEndTime - SimulationStartTime
    return Counter, SimulationTimeExecution
end function
```

for the dispersal coefficient α: 0.25, 0.5, 1 and 2. For each prefix and for each dispersal coefficient α, we run *full* and *R-local* simulation 100 times and compute the average number of rounds for *first*, *majority*, and *all* successes.

2.5.2 Validation of the R-*local* simulation method

We are interested in the most convenient time for stopping or cutting simulation and determine the stabilisation time (ST) of simulation. For that purpose we define the stabilisation time (ST) for a given landscape as the time t such that the change in average number of rounds for *all* successes (AFAS) between t and $2t$ is less than or equal to 2%: $\forall\tau \in (t, 2t]$, $|AFAS(\tau) - AFAS(t)| \leq 0.02 \cdot AFAS(t)$, where $AFAS(\tau) = \frac{\sum_{j=1}^{\tau} AS(j)}{\tau}$ and $AS(j)$ is the number of rounds needed for the *all* successes at iteration j.

To test the robustness of our method, we have used different sizes from the sizes used in deriving R. We apply it to four different sizes of landscapes: 5×50, 10×50, 15×50 and 20×50, as in the following steps.

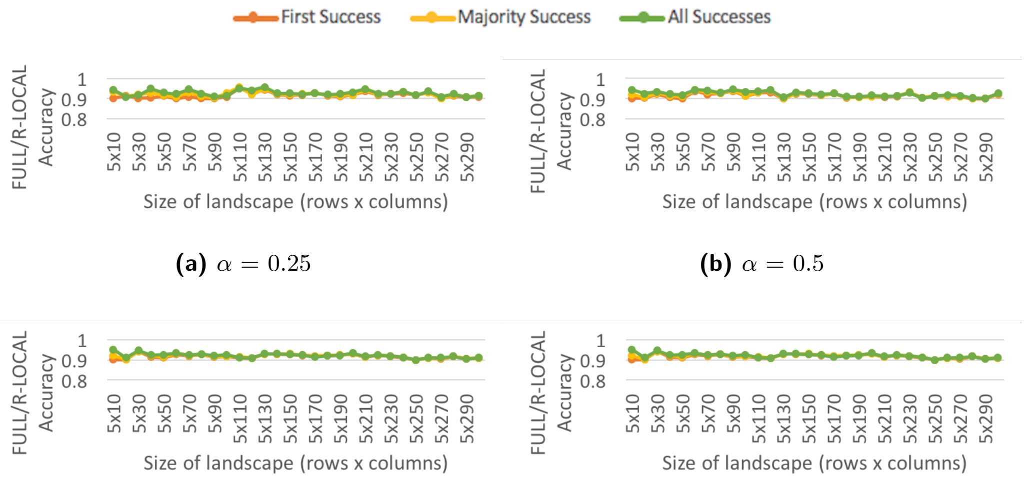
1. Extract randomly three landscapes of each size from different LCM2007 UK (aggregate classes) maps for each quality low, medium, and high.
2. On each extracted nine landscapes and for each considered value of the dispersal coefficient α:
   a. Run *full* simulation, stop simulation at the ST and record the results.
   b. Compute the average number of rounds for *first*, *majority* and *all* successes, and the average of the execution times of *full* simulation.
   c. Compute the minimum quality q_{min}, the maximum quality q_{max} and the minimum distance d_{min}.
   d. Compute three local distances R using Equation (1) with the constants in Table 3 which are calculated as in Section 2.5.1 with the data in Section 2.4.1 (see Section 3).
   e. Run R-*local* simulation with each of the three computed local distances R (based on c_{ED}, c_{AB}, and c_{MM} constants), stop simulation at the ST and record the results.
   f. Compute the average number of rounds for *first*, *majority* and *all* successes, and the average of the execution times of R-*local* simulation.
   g. For all three types of successful invasions and all local distances R computed in (d), compute the FULL/R-LOCAL accuracy. Then, check all accuracies if they are guaranteed.
   h. Specify which of the computed local distance R is the opt, where the opt one is the smallest distance that gives the FULL/R-LOCAL accuracy to be at least 90%.
   i. Compute the ratio between the average of the execution times (AETS) of *full* and R-*local* simulation methods.

3 Results

3.1 Main simulation

The estimated time of invasion (i.e., average number of rounds over 100 independent experiments) has been computed for each prefix 5×10, 5×20, 5×30, ..., 5×300 in each landscape of low, medium, and high quality using *full* and R-*local* simulation methods. In order to investigate the local distance R that allows FULL/R-LOCAL accuracy in each prefix in each of the extracted landscapes we use some predicted local distances R to run R-*local* simulation; ideally, one of the predicted distances is the opt that guarantees the

FULL/R-LOCAL accuracy, while decreasing the opt local distance R by one does not satisfy it. The simulation results show that the FULL/R-LOCAL accuracy of at least 90% is satisfied in all scenarios, as shown in Figures 2-4. On these landscapes, the opt local distances R used for simulations and guaranteeing the FULL/R-LOCAL accuracy are provided in Figure 5a, while Figure 5b shows the approx local distances R using Equation (1). From the opt local distances R in Figure 5a, we observe that the dispersal coefficient α is the most important parameter in both simulation methods. In all qualities, it has been investigated that a larger local distance R is required when the dispersal coefficient α equals to 0.25, while for 0.5, 1 and 2 a smaller local distance R is sufficient. Therefore, the local distance R that ensures the FULL/R-LOCAL accuracy depends on the dispersal coefficient α, and hence with decreasing mean dispersal distance: R decreases with the increase in α. Furthermore, a logarithmic growth has been observed in the local distance R with the growth of landscape size. On the other hand, the difference in the landscape quality has not caused a significant difference in the local distance R.



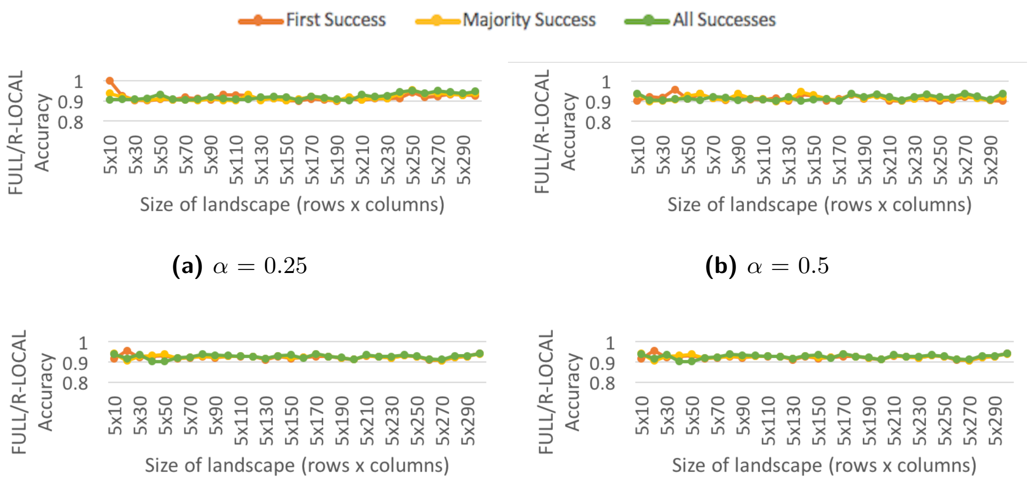
(a) $\alpha = 0.25$ **(b)** $\alpha = 0.5$

(c) $\alpha = 1$ **(d)** $\alpha = 2$

Figure 2 The FULL/R-LOCAL accuracy for each prefix in landscape of size 5×300 and of **low** quality when the dispersal coefficient α takes values of 0.25, 0.5, 1 and 2.

Comparison of the obtained results based on the proposed equation with simulation results demonstrates that the proposed equation (Equation (1)) gives a good estimate of the local distance R. We define the error rate between the approx and opt local distances R as: $\frac{\sum_{j=1}^{z}(approxR_j - optR_j)}{\sum_{j=1}^{z} optR_j}$. Table 2 gives the error rates between the approx and opt local distances R for each 5×300 landscape of low, medium, and high quality. All error rates in Table 2 are high, which means that we need to find the best constant c such that it minimises the error for various dispersal coefficients α and different qualities. Table 3 presents the interpolated constants c based on the objective functions ED, AB and MM for each 5×300 landscape of low, medium, and high quality. These constants are affected by the opt local distances R for all prefixes in each landscape of different quality.

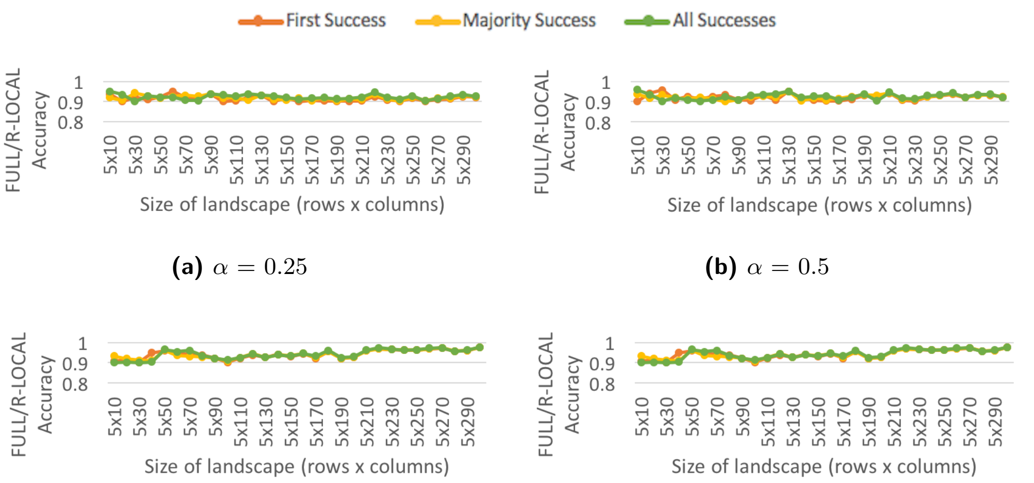
In addition, we define the following three error rates, which correspond to constant c produced by each objective function:



(a) $\alpha = 0.25$ **(b)** $\alpha = 0.5$

(c) $\alpha = 1$ **(d)** $\alpha = 2$

Figure 3 The FULL/R-LOCAL accuracy for each prefix in landscape of size 5×300 and of **medium** quality when the dispersal coefficient α takes values of 0.25, 0.5, 1 and 2.



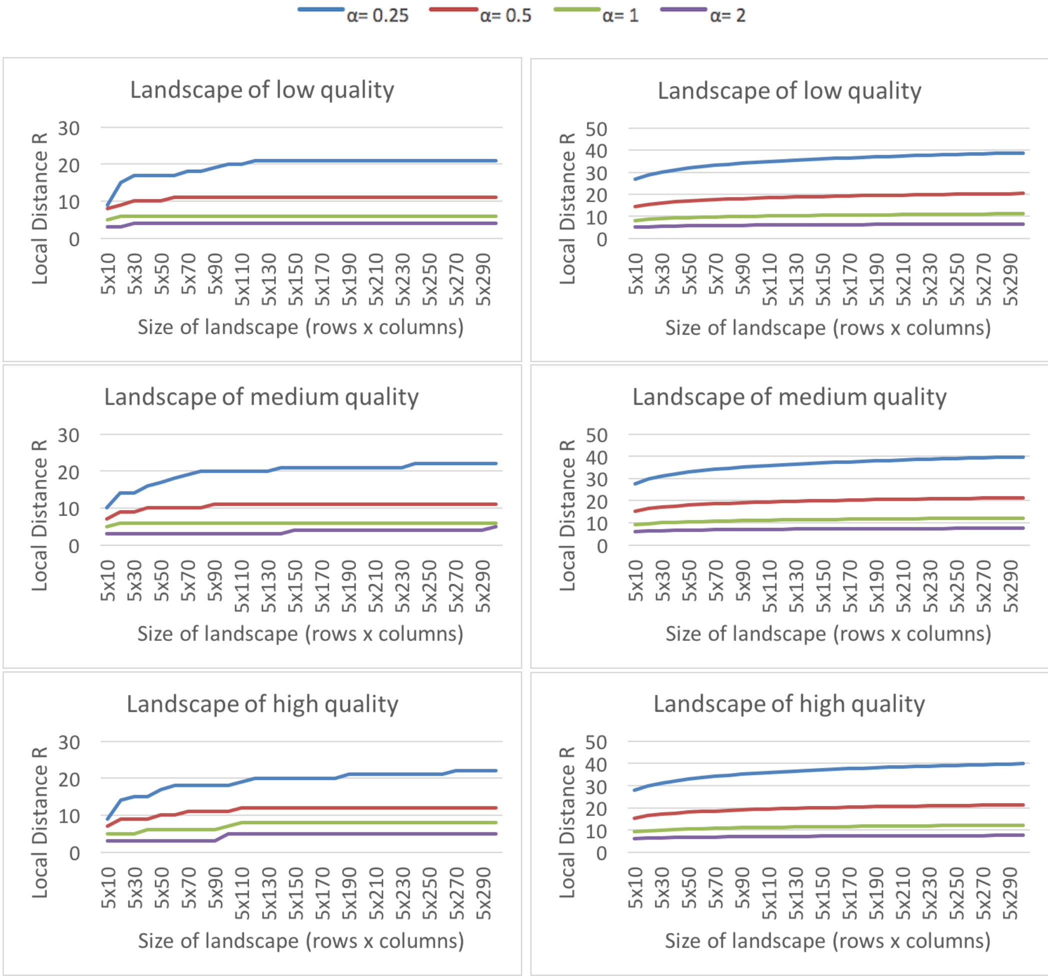
(a) $\alpha = 0.25$ **(b)** $\alpha = 0.5$

(c) $\alpha = 1$ **(d)** $\alpha = 2$

Figure 4 The FULL/R-LOCAL accuracy for each prefix in landscape of size 5×300 and of **high** quality when the dispersal coefficient α takes values of 0.25, 0.5, 1 and 2.

Table 2 Error rate between approx (when constant $c = 1$) and opt local distances R for each 5×300 landscape of low, medium and high quality when the dispersal coefficient $\alpha = 0.25, 0.5, 1, 2$.

Landscape quality	α=0.25	α=0.5	α=1	α=2
Low quality	0.81	0.74	0.73	0.57
Medium quality	0.83	0.86	0.90	1.01
High quality	0.89	0.75	0.56	0.63



(a) Opt local distances R computed by simulations

(b) Approx local distances R computed by Equation (1)

Figure 5 The local distance R for each prefix in landscape of size 5×300 and of low, medium, and high quality when $\alpha = 0.25, 0.5, 1, 2$. **(a)**The opt local distances R that allows the FULL/R-LOCAL accuracy and computed by simulations. **(b)**The computed local distances R by the proposed formula (Equation (1) when constant $c = 1$).

Table 3 The computed constant c by the objective functions ED, AB and MM for each 5×300 landscape of low, medium and high quality when the dispersal coefficient $\alpha = 0.25, 0.5, 1, 2$.

Landscape quality	Parameters	Constant c	α=0.25	α=0.5	α=1	α=2
Low quality	$q_{min} = 0.01$	c_{ED}	0.55	0.55	0.6	0.65
	$q_{max} = 0.64$	c_{AB}	0.55	0.55	0.55	0.65
	$d_{min} = 2$	c_{MM}	0.6	0.65	0.7	0.75
Medium quality	$q_{min} = 0.01$	c_{ED}	0.55	0.55	0.55	0.5
	$q_{max} = 0.96$	c_{AB}	0.55	0.55	0.5	0.5
	$d_{min} = 3$	c_{MM}	0.6	0.6	0.65	0.7
High quality	$q_{min} = 0.01$	c_{ED}	0.55	0.55	0.65	0.6
	$q_{max} = 0.99$	c_{AB}	0.55	0.6	0.65	0.65
	$d_{min} = 3$	c_{MM}	0.6	0.65	0.75	0.75

Table 4 The error rate between approx and opt local distances R, when $c = c_{ED}, c_{AB}, c_{MM}$, for each 5×300 landscape of low, medium and high quality when α = 0.25, 0.5, 1, 2.

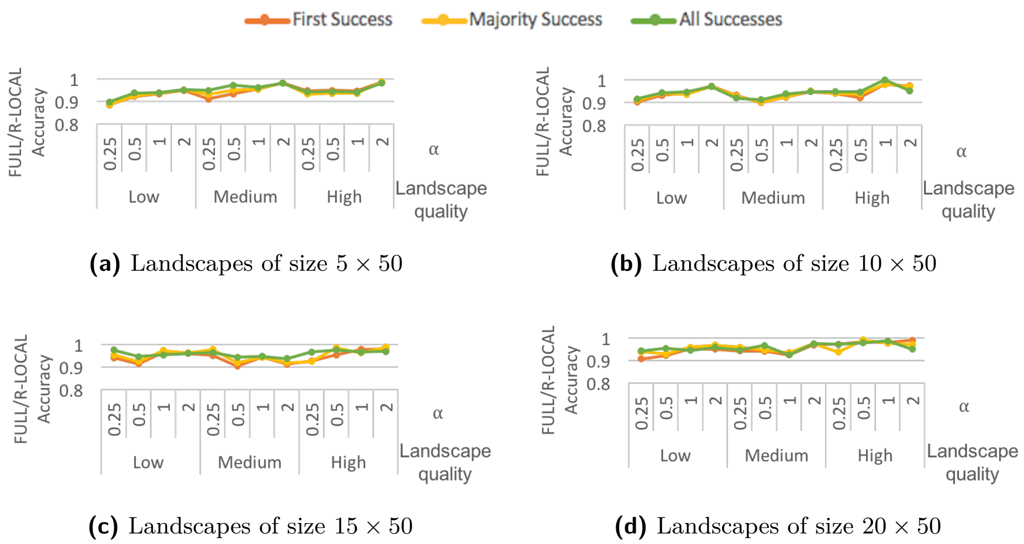
Landscape quality	Parameters	Error rate	α=0.25	α=0.5	α=1	α=2
Low quality	$q_{min} = 0.01$	Error rate (c_{ED})	0.01	0.01	0.01	0.01
	$q_{max} = 0.64$	Error rate (c_{AB})	0.04	0.05	0.06	0.05
	$d_{min} = 2$	Error rate (c_{MM})	0.08	0.12	0.21	0.17
Medium quality	$q_{min} = 0.01$	Error rate (c_{ED})	0.01	0.01	0.01	0.02
	$q_{max} = 0.96$	Error rate (c_{AB})	0.04	0.04	0.05	0.11
	$d_{min} = 3$	Error rate (c_{MM})	0.1	0.11	0.23	0.40
High quality	$q_{min} = 0.01$	Error rate (c_{ED})	0.01	0.01	0.02	0.03
	$q_{max} = 0.99$	Error rate (c_{AB})	0.05	0.05	0.08	0.12
	$d_{min} = 3$	Error rate (c_{MM})	0.13	0.13	0.17	0.22

1. error rate$(c_{ED}) = \dfrac{\sqrt{\sum_{j=1}^{z}(approxR_j \cdot c_{ED} - optR_j)^2}}{\sum_{j=1}^{z} optR_j}$,
2. error rate$(c_{AB}) = \dfrac{\sum_{j=1}^{z}|approxR_j \cdot c_{AB} - optR_j|}{\sum_{j=1}^{z} optR_j}$, and
3. error rate$(c_{MM}) = \dfrac{\sum_{j=1}^{z}(approxR_j \cdot c_{MM} - optR_j)}{\sum_{j=1}^{z} optR_j}$.

Observe that the error rate gives a measure of how well the constant c interpolated by the corresponding objective function minimises the error rate between the approx and opt local distances R, for a given landscape. When the approx local distance R is very close or equal to the opt local distance R, the error will be small or zero. Table 4 provides the computed error rates between the approx and opt local distances R for each 5×300 landscape of low, medium and high quality, when the constant c is equal to the interpolated c_{ED}, c_{AB}, and c_{MM} (constants in Table 3). As can be seen in Table 4, all error rates are between 0.01 and 0.4. While c_{ED} and c_{AB} produce error rates smaller than c_{MM}, constant c_{MM} is the best among the three. One of the reasons could be that constant c_{MM} in all landscapes reduces the approx local distance R to be greater than or equal to the opt local distance R. On the other hand, in some cases c_{ED} and c_{AB} decrease the approx local distance R to be less than the opt local distance R, and that means they give an estimated local distance R which does not allow the sought FULL/R-LOCAL accuracy.

3.2 Validation

We performed validation on 36 different landscapes. The 36 landscapes are divided into groups of nine landscapes and the four groups each has size: 5×50, 10×50, 15×50, and 20×50, respectively. For each landscape size, the nine landscapes are further divided into subgroups of three and each group has associated low, medium and high quality, respectively. All landscapes are extracted randomly from different LCM2007 UK (aggregate classes) maps [14] (i.e., aggregate classes in Table 1). On those landscapes, we run *full* and *R-local* simulations independently as described in Section 2.5.2 in order to get the averages of the estimated time of invasion for *first*, *majority* and *all* successes as well as the averages of the execution times of simulations. Running *full* and *R-local* simulations on those landscapes, using constants in Table 3, gives a good result as the FULL/R-LOCAL accuracy has been achieved as presented in Figure 6. It has been illustrated by the validation experiments that the MM objective function is the best function to be used because it gives constant c_{MM} such that it reduces the approx local distance R to be greater than or equal to the opt for all 36 landscapes.



(a) Landscapes of size 5×50

(b) Landscapes of size 10×50

(c) Landscapes of size 15×50

(d) Landscapes of size 20×50

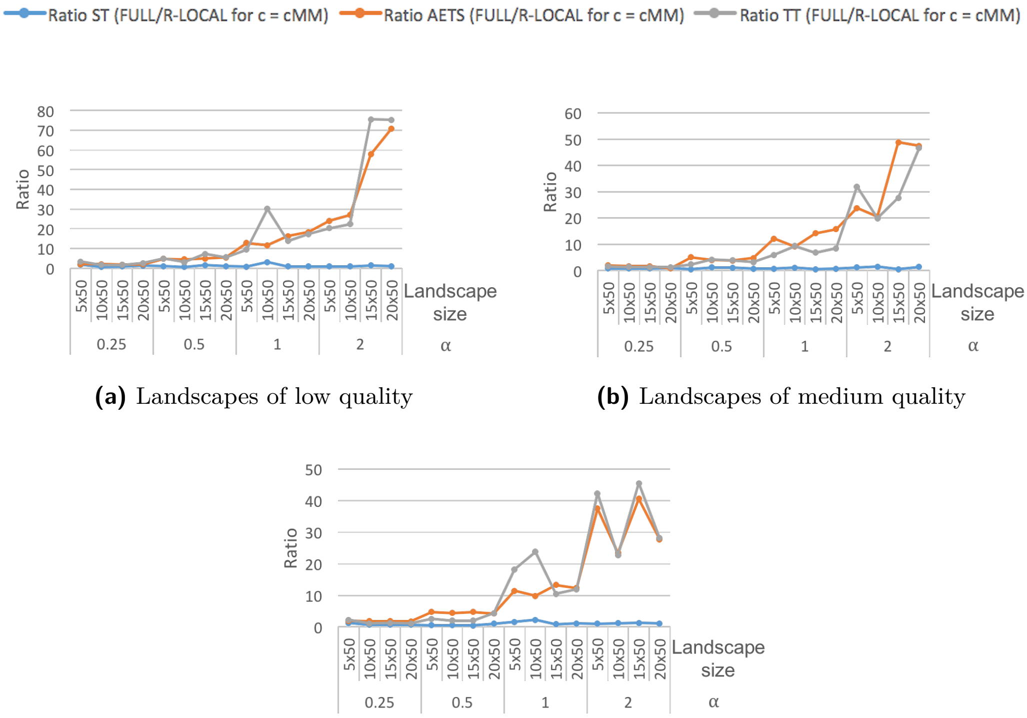
Figure 6 The average of FULL/R-LOCAL accuracy over the three landscapes in each subgroup with same size and same landscape quality; the R-LOCAL is with $R = approxR$ for $c = c_{MM}$. This is done for different values of dispersal coefficients α.

Furthermore, the validation experiments demonstrate that the total time (TT) of simulation execution, where TT = ST · AETS, needed to compute the estimated duration of the invasion process is substantially reduced by the *R-local* simulation method for all landscapes of different qualities. Figure 7 illustrates how much the *R-local* method is faster than the *full* method for all three qualities. For many cases, the *full* method takes 5-10 times longer to compute and this ratio can become as high as 75 for low quality landscape. We note that in general, for a given landscape size, the speedup of the *R-local* method increases as the dispersal coefficient α increases. On the other hand, in most cases, when the dispersal coefficient α is fixed, the speedup increases as the size of landscape increases.

In more details, in the landscape of size 20×50 and of low quality, the average execution time of *full* simulation is 176.9 seconds while only 2.5 seconds in the *R-local* simulation. That means the *full* method takes around 70 times longer to compute. We could envisage that when we are running *full* simulation in very large landscapes e.g., landscape of size 500×500, the computation time will be substantially reduced from maybe weeks/days to hours.

4 Conclusion

This study was prompted by a desire to construct the *R-local* model that visualises the invasion process based on the landscape network sparsification tool to efficiently estimate a duration of the process. The capability of our model is to reduce the time needed to compute the estimated duration of the invasion process on large landscapes while maintaining a comparable duration of the invasion.



(a) Landscapes of low quality

(b) Landscapes of medium quality

(c) Landscapes of high quality

Figure 7 In each landscape quality (low, medium, high), the ratio between the stabilisation time (ST) of *full* and R-*local* simulations; the ratio between the average of the execution times (AETS) of *full* and R-*local* simulations; the ratio between the total time (TT) of *full* and R-*local* simulations in four different sizes of landscapes 5×50, 10×50, 15×50 and 20×50, when $R = approxR$ for $c = c_{MM}$, and dispersal coefficient $\alpha = 0.25, 0.5, 1, 2$.

The simulations demonstrate how the local distance R depends on two factors: the dispersal coefficient α and the landscape quality. A small dispersal coefficient α requires a large local distance R in all types of quality, while a small R is sufficient for a large α. The difference in landscape quality does not cause a significant difference in the needed value of the local distance R. Indeed, the tool of landscape network sparsification proves its efficiency in computing the invasion time especially for large landscapes. Even when the size of landscape is increased, the local distance R does not grow significantly (see Figure 5). This implies a sparser landscape networks for large landscapes, and therefore the time needed to compute the invasion duration decreases substantially.

As for future work, we aim to study how to improve/spoil the invasion process, e.g., to increase/decrease the speed of invasion by modifying landscapes. (Note that in some applications decreasing the speed of invasion could be more desirable, e.g., in epidemics.) This would require the computation of the invasion duration many times to verify the effectiveness of landscape modification, and therefore our work improving the speed up of the computation would be beneficial.

Acknowledgments. The authors would like to thank Dariusz Kowalski (University of Liverpool) for fruitful discussions and comments on the content of this paper. Jenny A. Hodgson acknowledges support from NERC grant ref NE/L002787/1.

References

1 I Chen, JK Hill, R Ohlemüller, DB Roy, and CD Thomas. Rapid range shifts of species associated with high levels of climate warming. *Science*, 333:1024–1026, 2011.

2 G Grimmett and D Stirzaker. *Probability and random processes*. Oxford university press, 2001.

3 JA Hodgson, A Moilanen, BA Wintle, and CD Thomas. Habitat area, quality and connectivity: striking the balance for efficient conservation. *Journal of Applied Ecology*, 48(1):148–152, 2011.

4 JA Hodgson, CD Thomas, S Cinderby, H Cambridge, P Evans, and J K Hill. Habitat recreation strategies for promoting adaptation of species to climate change. *Conservation Letters*, 4:289–297, 2011.

5 JA Hodgson, CD Thomas, C Dytham, JMJ Travis, and SJ Cornell. The speed of range shifts in fragmented landscapes. *PLoS ONE*, 7:e47141, 2012.

6 O Honnay, K Verheyen, J Butaye, H Jacquemyn, B Bossuyt, and M Hermy. Possible effects of habitat fragmentation and climate change on the range of forest plant species. *Ecol Lett*, 5:525–530, 2002.

7 B Huntley, YC Collingham, SG Willis, and RE Green. Potential impacts of climatic change on European breeding birds. *PLoS ONE*, 3:e1439, 2008.

8 VG Kulkarni. *Modeling and analysis of stochastic systems*. CRC Press, 2016.

9 M Mitzenmacher and E Upfal. *Probability and computing: Randomized algorithms and probabilistic analysis*. Cambridge university press, 2005.

10 D Morton, C Rowland, C Wood, L Meek, C Marston, G Smith, R Wadsworth, and I Simpson. Final report for lcm2007-the new uk land cover map. *Countryside Survey Technical Report No 11/07*, 2011.

11 M Newman. *Networks: An Introduction*. Oxford University Press, 2010.

12 SJ Phillips, P Williams, G Midgley, and A Archer. Optimizing dispersal corridors for the cape proteaceae using network flow. *Ecological Applications*, 18:1200–1211, 2008.

13 F Skov and JC Svenning. Potential impact of climatic change on the distribution of forest herbs in europe. *Ecography*, 27:366–380, 2004.

14 The Centre of Ecology and Hydrology Information Gateway. Land cover map 2007 (1km percentage aggregate class, gb) v1.2.

15 CD Thomas, A Cameron, and RE Green et al. Extinction risk from climate change. *Nature*, 427:145–148, 2004.

16 GR Walther, E Post, P Convey, A Menzel, C Parmesan, TJ Beebee, JM Fromentin, O Hoegh-Guldberg, and F Bairlein. Ecological responses to recent climate change. *Nature*, 416(6879):389–395, 2002.

17 MS Warren, JK Hill, and JA Thomas et al. Rapid responses of british butterflies to opposing forces of climate and habitat change. *Nature*, 414:65–69, 2001.