ECOGRAPHY

Software note

samc: an R package for connectivity modeling with spatial absorbing Markov chains

Andrew J. Marx, Chao Wang, Jorge A. Sefair, Miguel A. Acevedo and Robert J. Fletcher Jr.

A. J. Marx (https://orcid.org/0000-0002-7456-1631) ✉ (andrewjmarx@ufl.edu), M. A. Acevedo (https://orcid.org/0000-0002-8289-1497) and R. J. Fletcher, Jr. (https://orcid.org/0000-0003-1717-5707), Dept of Wildlife Ecology and Conservation, PO Box 110430, 110 Newins-Ziegler Hall, Univ. of Florida, Gainesville, FL 32611-0430, USA. – C. Wang (https://orcid.org/0000-0003-2706-5122) and J. A. Sefair (https://orcid.org/0000-0002-5887-8938), School of Computing, Informatics, and Decision Systems Engineering, Arizona State Univ., Tempe, AZ, USA.

Quantifying landscape connectivity is fundamental to better understand and predict how populations respond to environmental change. Currently, popular methods to quantify landscape connectivity emphasize how landscape features provide resistance to movement. While many tools are available to quantify landscape resistance, these do not discern between two fundamentally different sources of resistance: movement behavior and mortality. To address this issue, we developed the samc R package that quantifies landscape connectivity using absorbing Markov chain theory. Within this mathematical framework, movements are represented as transient states in the Markov chain, while mortality is represented by transitions to absorbing states. Not only does this framework explicitly account for these different issues, it provides a probabilistic approach that can incorporate both short-term and long-term dynamics, as well as species distribution and abundance. The package includes functions to quantify life expectancy, long-term visitation rates, and various spatially and temporally explicit measures of mortality and movement at the local and landscape scales. These functions in samc have been optimized to find computationally practical solutions in landscapes comprised of $> 2 \times 10^6$ cells. Here, we illustrate the workflow of the samc package with publicly available movement and mortality data on the endangered Florida panther *Puma concolor coryi*. This analysis showed that movement and mortality are generally correlated except for locations near roads (areas of high mortality risk) that are within the dispersal range of source locations. This pattern would have been undetectable with current methods that quantify movement resistance. Overall, the samc package provides a means for implementing spatial absorbing Markov chains that can distinguish between movement behavior and mortality resulting in more reliable landscape connectivity measures.

Keywords: circuit theory, dispersal, landscape connectivity, Markov chain, mortality, movement ecology

Introduction

Understanding landscape connectivity, or the degree to which the landscape alters movements among resources (Taylor et al. 1993), is essential for ecology, evolution and conservation (Slatkin 1993, Hanski 1999, Fletcher and Fortin 2018), and it lies at the heart of the emerging sub-discipline of movement ecology (Nathan et al. 2008). Nonetheless, accurately predicting and mapping landscape connectivity remains challenging (Fletcher et al. 2011, Sawyer et al. 2011).

Currently, most methods for mapping connectivity rely on a spatially explicit perspective of matrix or landscape resistance. The rationale is that aspects of the matrix (e.g. land-use, topography) can alter movement routes of individuals dispersing across landscapes, such that incorporating 'matrix resistance' (or conversely, 'matrix permeability') may be crucial for interpreting connectivity (Zeller et al. 2012). Yet resistance can emerge for two fundamentally different reasons: 1) organisms may be less likely to move through a location (Elliot et al. 2014); or 2) organisms may suffer mortality at a location (Nowakowski et al. 2015). Because the demographic effects of mortality when moving through the landscape are likely greater than those resulting from changes in movement routes, isolating these issues may help guide more effective conservation actions (Vasudev et al. 2015).

Recently, we introduced a framework for connectivity by advancing random walk theory with absorbing Markov chains that honor the idea that the matrix can influence both movement behavior and mortality risk (Fletcher et al. 2019). This framework provides probabilistic accounting of both movement behavior and mortality risk of dispersers across landscapes, provides both short- and long-term predictions of connectivity, it can directly incorporate population distribution and abundance into predictions of connectivity, and it can quantify demographic parameters related to successful dispersal and connectivity. This framework, what we term the 'spatial absorbing Markov chain' (SAMC), differs from other connectivity mapping approaches (Calabrese and Fagan 2004, Rudnick et al. 2012, Fletcher et al. 2016) in several ways. Unlike individual-based modeling of connectivity (Bocedi et al. 2014), the SAMC is an analytical framework like least-cost analysis (Etherington 2016), randomized shortest paths (Saerens et al. 2009), and circuit theory (McRae et al. 2008). Overall, it is most similar to circuit theory: SAMC and circuit theory are both rooted in Markov chain theory. Yet, the SAMC describes the general case that accounts for time-specific movement and mortality while circuit theory can be thought of a special case with no explicit absorbing states. Fletcher et al. (2019) evaluated this modeling framework relative to least-cost and circuit theory approaches, finding that the SAMC outperformed other approaches for predicting observed movements of an insect herbivore across 15 experimental landscapes undergoing habitat destruction by accounting explicitly for mortality risk.

Here we introduce the samc package for modeling connectivity within the SAMC framework using the R programming language. First, we begin by providing a brief overview of the SAMC framework and how it is implemented in the samc package, including the methods and features of the package. Second, we illustrate the workflow of using the SAMC framework for quantifying connectivity with the samc package by providing an example for the endangered Florida panther *Puma concolor coryi*.

Methods and features

Using absorbing Markov chain theory (Kemeny and Snell 1976, Ross 2010, Fletcher et al. 2019) to model connectivity, the samc package can distinguish between movement behavior and mortality risk ('absorption') on the likelihood of movement across landscapes. It is also a temporally and spatially explicit probabilistic framework, which provides a means to calculate several relevant metrics of connectivity across landscapes. This framework assumes that landscapes are considered as discrete representations of the environment via the use of raster maps. It also treats time as a discrete variable, which matches the typical discretization of empirical movement data. Next, we briefly summarize key aspects of the SAMC framework described in Fletcher et al. (2019) and how they relate to the samc package.

The spatial absorbing Markov chain

For organisms dispersing across complex landscapes, for each time step an individual can either survive and stay at the same location (i.e. site fidelity), survive and move to a nearby site, or suffer mortality. The SAMC framework honors this simple idea by considering 'transient' states of fidelity and movement, and an 'absorbing', permanent state of mortality.

To capture this idea and apply it to the problem of predicting and mapping connectivity, there are two general aspects of the SAMC framework. The first aspect is the quantification of a probability matrix, $\mathbf{P}$, that contains both transition probabilities between transient states (i.e. landscape cells) and an absorbing state representing death. For a landscape divided into C cells, matrix $\mathbf{P}$ can be written as:

$$\begin{pmatrix} \mathbf{Q} & \mathbf{R} \\ \mathbf{0} & 1 \end{pmatrix} \tag{1}$$

where $\mathbf{Q}$ is a sparse, $C \times C$ transition matrix reflecting transitions between transient states, $\mathbf{R}$ is a $C \times 1$ vector containing transition probabilities from the transient states to the absorbing state, and $\mathbf{0}$ is a $1 \times C$ vector of zeros. The elements p_{ij} describe the probability of transitioning from state i to j in one time step. The i-th element of $\mathbf{R}$ $p_{i,C+1}$ describes the probability of death in one time step for an individual located in cell i. Note that $p_{C+1,C+1} = 1$, since a dead individual must remain dead. $\mathbf{Q}$ reflects the permeability of the landscape to movement; we assume that movement transitions occur

locally over short time steps (e.g. via a 4- or 8-neighbor rule), although the framework is flexible and this assumption could be relaxed. Overall, **P** results in a biased random walk model for movement driven by spatial heterogeneity. In the samc package, there are several ways of parameterizing **P** through the use of the samc() function. In its simplest form, the samc package requires two maps (either as matrices or as rasters) as inputs: one that describes resistance or cost to movement and a second that describes mortality risk. When fidelity is of interest, a map of potential site fidelity could be used. The second component of the SAMC is the derivation and quantification of connectivity-related metrics based on the **P** matrix (Table 1). These metrics can be broadly classified in two ways: 1) short-term versus long-term metrics; and 2) metrics that quantify movement or demography.

Short-term metrics can be derived from **P** in two general ways. First, time-specific predictions are possible, which can be helpful for interpreting problems of changes in distribution, such as range expansion and spread of species, across complex landscapes (Hudgins et al. 2017). For instance, the probability of being in location j after t steps if starting at location i is the (i, j)-th element of $\mathbf{Q}^t$. Using the distribution() function, $\mathbf{Q}^t$ can be calculated, as well as a variant that accounts for the initial distribution or abundance of individuals, $\mathbf{\Psi}$ (i.e. a $C \times 1$ vector that describes the probability of occurrence of an individual or the abundance of individuals at time $t = 0$). Second, cumulative predictions across given time periods can be made, such as asking whether a location might be colonized over the next year. Cumulative predictions include quantifying: the probability of ever visiting location j if an individual starts at location i within t or fewer steps, $\tilde{\mathbf{D}}_{jt}$ and the probability of experiencing mortality at location j within t or fewer steps if starting in location i, $\tilde{\mathbf{B}}_t$. These metrics can be calculated with the dispersal() and mortality() functions, respectively, when passing the amount of time being considered into the time parameter of the function (i.e. the number of time steps).

Long-term metrics can also be calculated, which are direct extensions of the time-specific metrics above. Summing $\mathbf{Q}^t$ over all t gives the 'fundamental matrix', **F**:

$$\mathbf{F} = (\mathbf{I} - \mathbf{Q})^{-1} \qquad (2)$$

where **I** is the identity matrix. f_{ij} is the expected number of times an individual that starts in location i uses location j before it dies. Therefore, **F** provides a time-focused description of long-term visitation rates across complex landscapes, which can be calculated with the visitation() function.

Time-specific predictions of $\tilde{\mathbf{D}}$ can be scaled to long-term predictions of movement and dispersal. The probability that location j is ever visited when starting from location i is the (i, j)-th element of matrix:

$$\mathbf{D} = (\mathbf{F} - \mathbf{I}) \operatorname{diag}(\mathbf{F})^{-1} \qquad (3)$$

where diag(**F**) is a matrix with diagonal elements from **F** and zeros otherwise (Kemeny and Snell 1976). The unconditional probability of ever visiting location j, taking into account the probability of each initial species distribution, is given by the j-th element of $\mathbf{\Psi}^T\mathbf{D}$. Consequently, **D** provides a long-term approximation of the probability of movement between locations and can be calculated with the dispersal() function.

Time-specific predictions of $\tilde{\mathbf{B}}$ can also be scaled to long-term predictions of mortality likelihood across landscapes. The probability of suffering mortality in location j if starting in location i is the (i, j)-th element of **B**:

$$\mathbf{B} = \mathbf{F}\tilde{\mathbf{R}} \qquad (4)$$

where $\tilde{\mathbf{R}}$ is a $C \times C$ matrix with diagonal elements equal to the mortality probabilities ($\tilde{R}_{jj} = R_j$ for all j) and off-diagonal

Table 1. Connectivity metrics derived from SAMC and fit with the samc package.

Function	Metric	Description
dispersal()	$\tilde{\mathbf{D}}_{jt}$	Probability of an individual visiting a location, if starting at any other location, before or at time t
	$\mathbf{\Psi}^T\tilde{\mathbf{D}}_{jt}$	Probability of an individual visiting a location, before or at time t, accounting for initial location
	D	Probability of an individual visiting a location
	$\mathbf{\Psi}^T\mathbf{D}$	Probability of an individual visiting a location, accounting for initial location
distribution()	$\mathbf{Q}^t$	Probability of an individual being at a location at time t
	$\mathbf{\Psi}^T\mathbf{Q}^t$	Probability of an individual being at a location at time t, accounting for initial location
mortality()	$\tilde{\mathbf{B}}_t$	Probability of an individual experiencing mortality at a location before or at time t
	$\mathbf{\Psi}^T\tilde{\mathbf{B}}_t$	Probability of an individual experiencing mortality at a location, before or at time t, accounting for initial location
	B	Probability of an individual experiencing mortality at a location
	$\mathbf{\Psi}^T\mathbf{B}$	Probability of an individual experiencing mortality at a location, accounting for initial location
survival()	**z**	Expected life expectancy of an individual
	$\mathbf{\Psi}^T\mathbf{z}$	Overall life expectancy accounting for initial location
visitation()	**F**	Expected number of times an individual will visit a location

elements equal to 0. The i-th element of $\mathbf{\Psi}^T\mathbf{B}$ represents the unconditional probability of suffering mortality in location i, taking into account the probability of each initial species distribution, which can be calculated with the `mortality()` function.

Finally, the life expectancy of individuals starting at different locations can be derived from $\mathbf{F}$ as a demographic metric to describe survival across landscapes. This metric was initially introduced by Acevedo et al. (2015) for patch-based networks (see also Sefair et al. 2017). The life expectancy if starting from location i is defined as the i-th element of vector

$$\mathbf{z} = \left(\mathbf{I} - \mathbf{Q}\right)^{-1}\mathbf{1} = \mathbf{F} \times \mathbf{1} \qquad (5)$$

where $\mathbf{1}$ is a $C \times 1$ vector of ones. Species initial distribution or abundance can also be incorporated (Table 1). $\mathbf{z}$ and its variants can be calculated with the `survival()` function.

Features of the `samc` package

The `samc` package makes extensive use of the S4 object system in R to create an efficient framework that enforces correct usage and minimizes accidental misuse. It currently requires the `gdistance` package (van Etten 2012) for constructing transition matrices, the `raster` package (Hijmans 2017) for processing raster GIS data, and the `Matrix` (Bates and Maechler 2017) package for sparse matrix manipulation and calculations. It also requires the `Rcpp` (Eddelbuettel and Balamuta 2017) and `RcppEigen` (Bates and Eddelbuettel 2013) packages, along with C++ development tools, when building manually from source. It contains functions for calculating all metrics described in Fletcher et al. (2019), as well as utility functions for setting up the framework, checking input data, and mapping results. Finally, it contains the example data used in Fletcher et al. (2019).

To calculate the connectivity metrics, the user starts by creating a `samc-class` object using the `samc()` function. This object contains the probability matrix and other pieces of data used to perform error checking and ensure correct usage of the framework. Actual details about the internal components of this object are not provided here because they are expected to change in the future as new features and optimizations are added, and it is generally expected that users will not be directly accessing or modifying these components.

All the metrics described in Fletcher et al. (2019) are implemented into a set of analytical functions provided by the package. For conciseness, related equations are grouped together under the same function name (Table 1). The package uses parameter matching to select the correct underlying calculation. For example, passing a value for the `time` parameter of a function will result in the calculation of a short-term metric (e.g. for the `mortality()` function, using the `time` parameter results in calculating $\tilde{\mathbf{B}}_t$).

Additional details about the various components of the package are provided in the 'Overview' vignette included in the package and online documentation. This includes how to use the built-in example data, important requirements for creating the `samc-class` object, and details about how to use the analytical functions to calculate different metrics.

Performance

Two major issues that can affect performance (time to perform an analysis) in connectivity modeling are the number of sites or locations from which dispersers may occur and the number of cells (or pixels) in the landscape being considered. For the first issue, many of the SAMC metrics only slightly reduce in performance when more sites are considered because calculations are not done iteratively (unlike some approaches such as circuit theoryrandomized shortest paths; Saerens et al. 2009, Panzacchi et al. 2016). In terms of the number of cells, there are performance concerns when working with maps with large raster grids, because such maps result in large transition matrices. The size of the matrices increases quadratically as the number of cells in an analysis increase. While the transition matrix can be stored efficiently in a sparse format, many of the metrics in Fletcher et al. (2019) perform calculations that involve large dense matrices that can become impractical for modern consumer computers at scales sometimes used by ecologists. However, these dense matrices represent pairwise data for every pair of cells in the landscape, most of which will typically be superfluous information for practical applications. Instead, ecologists will generally only be interested in information about specific regions or cells of the landscape. By focusing on calculating these more useful partial solutions, we have managed to find computationally practical solutions for every metric for landscapes with > 1 000 000 cells (e.g. a 1000×1000 raster), and up to 3–4 M cells for some metrics (e.g. mortality probability accounting for initial locations, $\mathbf{\Psi}^T\mathbf{B}$). With the special exception of $\mathbf{\Psi}^T\mathbf{D}$, these metrics can be computed in a manner of seconds or minutes on an average modern consumer desktop. We also emphasize that because of the probabilistic, absorbing nature of the model, expected values for metrics decline with distance from source locations (Fletcher et al. 2019), such that landscapes with > 4 M cells could be tiled into smaller areas or users could buffer around source locations based on maximum known dispersal distances (i.e. modeling only within specified distances of source locations) without loss of information in mapping. Most of these metrics appear to be similar or more computationally efficient than the related randomized shortest paths analysis in R (Saerens et al. 2009, Panzacchi et al. 2016) for > 2×10^6 cells, but metrics are still less efficient than using circuit theory and least-cost metrics using the `gdistance` package (van Etten 2012). For smaller landscapes, or for users with access to exceptional hardware resources, the full metrics can still be calculated. The optimized functions are enabled by default to ensure that users do not accidentally run the more demanding calculations that can cause issues for computers with limited resources. This default behavior can be disabled by explicitly enabling the override option when creating a `samc-class` object.

Additional details about the different performance aspects of the package are included in the 'Performance' vignette and online documentation. This includes a more thorough breakdown of performance into the separate issues of computation time and memory requirements. It also goes into more detail and illustrates the effect of landscape size. Finally, it includes crude benchmarks illustrating the relative performance of each the various metrics, which will be kept up-to date as future optimizations are implemented.

Workflow: an example with Florida panthers

We illustrate the workflow of the `samc` package (Fig. 1) with data on Florida panthers, an endangered carnivore that is confined to south Florida. This case study is useful because large carnivores are often the target of connectivity conservation (Carroll et al. 2001, Dixon et al. 2006), decades of data have been collected on panther movement, dispersal and mortality, and connectivity models have been applied to this species previously (Maehr et al. 2002, Kautz et al. 2006, van de Kerk et al. 2019). Furthermore, a key criterion for delisting this species is dispersal and population establishment of panthers north of the Caloosahatchee River (Fig. 2).

Step 1: data preparation

At a minimum, the package requires absorption (or mortality) and landscape resistance data on movement behavior. Additionally, fidelity and occupancy (or abundance) data can also be used. As in other applications that interpret 'resistance' for landscape connectivity, parameterizing the transition matrix that is used to interpret movement behavior in the SAMC (**Q**) can come from behavioral experiments, telemetry, habitat use or expert opinion (Zeller et al. 2012). In these ways, parameterizing **Q** of the SAMC is similar to that of parameterizing least-cost and circuit theory approaches.

We parameterize resistance for Florida panthers based on the analysis in Kautz et al. (2006), using land-cover data and resistance coding illustrated in Fletcher and Fortin (2018). In this example, data from telemetered Florida panthers were used to estimate habitat preferences based on resource selection functions, which were converted to a rank-based measure of resistance to movement (Fig. 2a). We consider a raster-based, land-cover map, aggregated to a 500 m resolution (reflecting the approximate grain of telemetry error; post-aggregation, 274 664 cells on the map).

Information on mortality risk is less common but can be included in the SAMC as simply a constant low rate, or it
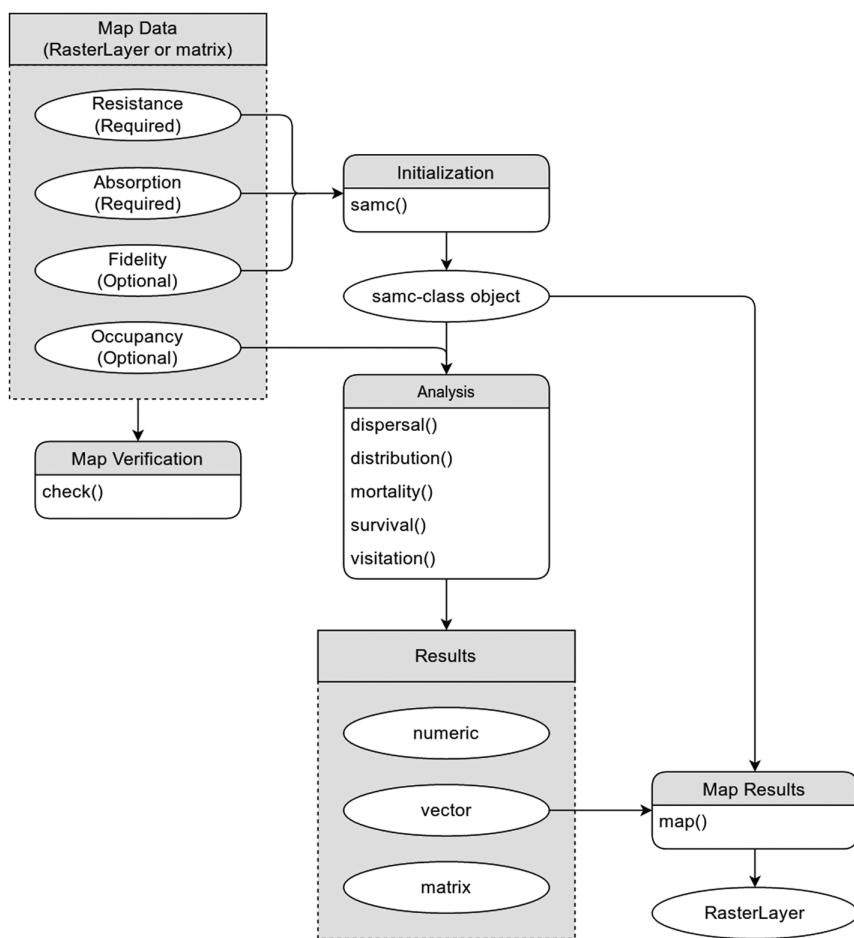

Figure 1. Workflow diagram illustrating the relationship between data inputs, outputs and functions of the `samc` package.
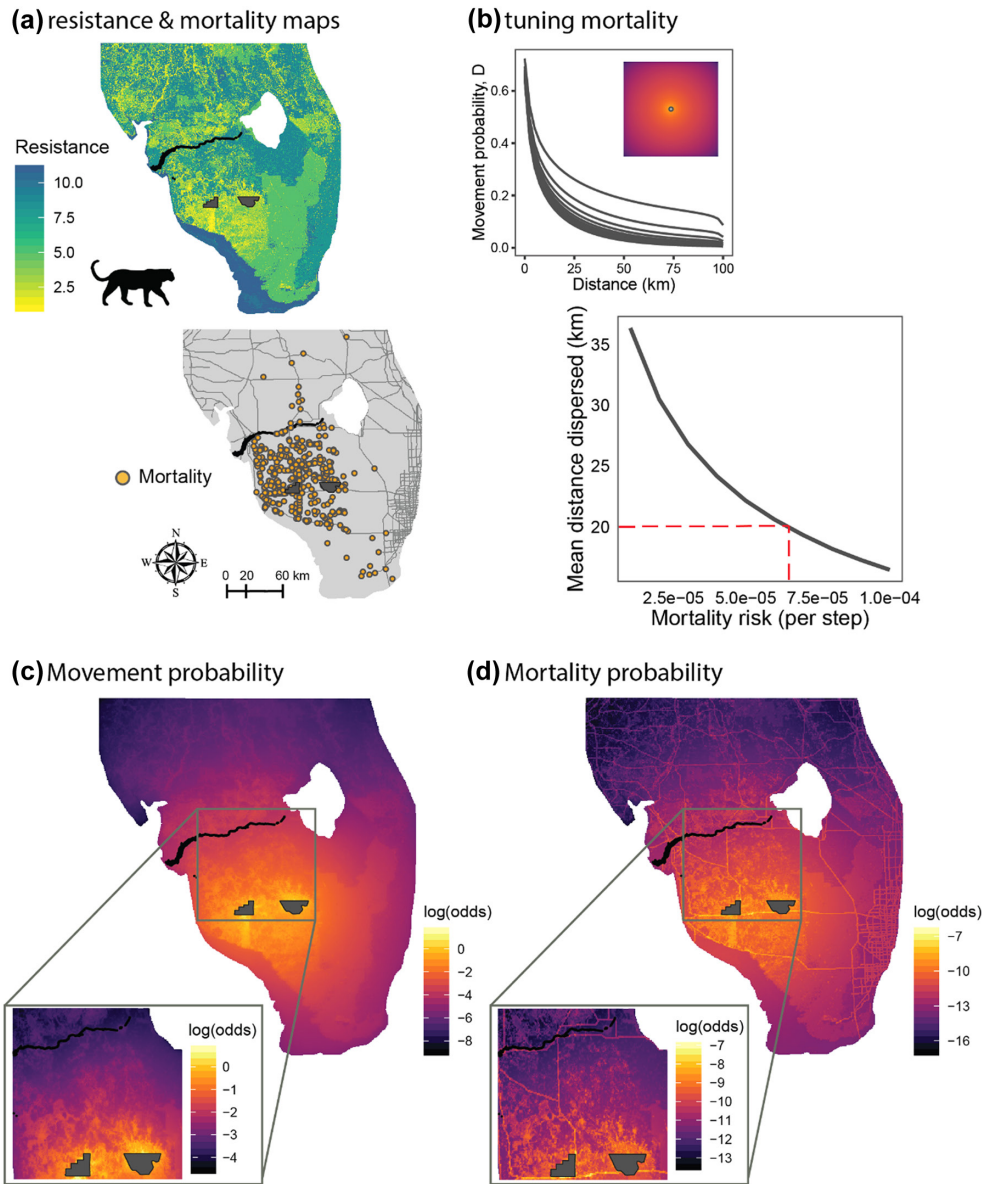
Figure 2. Connectivity and predicted movement by the Florida panther using the spatial absorbing Markov chain approach. (a) Resistance map based on resource selection functions created by telemetered panthers (Kautz et al. 2006) and data on mortality (including with respect to major roads in southern Florida). (b) Tuning mortality rates using a homogeneous grid of constant mortality to determine predicted movement distances and relating predictions to known dispersal distances. (c) Mapping predicted movement (D). (d) Mapping predicted mortality risk (B). In (a, c, d) black lines show the Caloosahatchee River, dark grey source areas considered by Kautz et al. 2006 (Florida Panther Wildlife Refuge and the northern portion of Big Cypress Preserve).

could be spatially variable across the landscape. Mortality rates, **R**, can be approximated using several approaches, such as using field experiments (Fletcher et al. 2019), habitat-specific survival estimates (Low et al. 2010), or mortality risk data (e.g. roadkill data; Santos et al. 2013, Grilo et al. 2018). For instance, if annual survival rates are available, such rates could be scaled to reflect expected mortality rate per time step (where a 'time step' reflects the average time expected for movement between adjacent cells). In addition, if the mean dispersal distance for a species is known, **R** could be tuned to

derive dispersal distance estimates from the SAMC to equal known dispersal distances, as we show below. The SAMC assumes non-zero mortality risk, because if mortality risk is 0, it would predict that all locations are eventually reached by mobile organisms.

To parameterize spatial variation mortality risk for Florida panthers, we focus on mortality risk from roads, which is a major cause of mortality in panthers (Schwab and Zandbergen 2011, van de Kerk et al. 2019). Data on mortality locations have been collected over the past several decades (1972–2018)

(<https://geodata.myfwc.com/datasets/florida-panther-mortality>) (Fig. 2a). Using these locations, we determined the proportion of mortality locations within 500 m (the map resolution) of major roads relative to what would be expected based on available habitat. Mortality locations (n = 488) were 3.7× more likely to be within 500 m of roads than > 500 m based on these data. We use this information to adjust relative expected mortality rates near roads. We then tune absolute rates of mortality risk based on known dispersal distances in Florida panthers, where the mean dispersal distance has been estimated for females as 20.3 km (Maehr et al. 2002). We focus on mapping potential dispersal by females, as males are known to be wide ranging, but females are key to breeding range expansion north of the Caloosahatchee River. To do so, we first consider different baseline rates of mortality across a homogeneous 400 × 400 cell (160 K) grid with the cell resolution considered in mapping (500 m; Fig. 2b). The samc package can be used to calculate the probability of visiting locations as a function of distance with the dispersal() function (D, Eq. 3) to generate potential dispersal kernels based on known mortality risk (Fig. 2b). We quantified probabilities from the source location to locations as a function of distance to derive expected mean movement distances based on the parameters of an exponential dispersal kernel. In this case, baseline mortality risk would be 6.5×10^{-6} per cell to generate mean movement distances of approximately 20 km (Fig. 2b). We then increased this base mortality probability 3.7× near roads, given mortality locations to create a mortality risk map for the SAMC.

Information on species distribution and abundance can also be included as a map, which is necessary for some (but not all) metrics. In most cases, we expect that such information will be passed into the model to capture source-destination locations (e.g. protected areas) (Cushman et al. 2009), local populations (Larkin et al. 2004), or individual dispersers (Fletcher et al. 2014). This information can either reflect the probability that an individual occurs at a location, in which output generally reflects probabilities of individual movement or mortality, or it can reflect the number of individuals, in which output generally reflects the number of individuals moving or suffering mortality. Including information on distribution and abundance facilitates computation efficiency. If such information is not included, users are implicitly assuming that individuals start from all locations, akin to some graph theory applications to resistance modeling (Carroll et al. 2012) and loosely similar to factorial least-cost analysis (Elliot et al. 2014).

For the Florida panther, we set the Florida Panther National Wildlife Refuge and the northern portion of the Big Cypress Reserve as source locations (as in Kautz et al. 2006; Fig. 1) for individual dispersers by creating a raster map where the centroids of these two locations are set to 1 and all other locations on the raster are set to 0 (protected area data come from: <www.fnai.org/conservationlands.cfm>).

All sets of data must be stored either as matrices or raster maps. Mixing matrices and rasters is not supported due to underlying differences in the way they store data and potential issues that can arise when users are not aware of these differences. Every set of data must have the same dimensions (number of rows and columns), and NA data is supported as long as it occurs in the same locations in every set of data. If the data sets are stored as rasters, then they must have the same geographic coordinate extents and coordinate reference system (CRS). The check() function is provided to simplify the process of checking that all these conditions are met.

Step 2: samc-class object creation

The resistance, absorption and (optional) fidelity data are used to create an samc-class object via the samc() function. If the data sets are in a raster format, then the latlon parameter is mandatory. A transition function must always be provided for constructing the transition matrix. Typically, transition functions use the mean of resistance values between pairs of cells (tr_fun = function(x) 1/mean(x)), but other alternative functions can be considered.

For the Florida panther example, we take the resistance (cost_panther) and mortality (mort_panther) raster maps (Fig. 2a) to create the samc object using:

```
Pmatrix <- samc(resistance=cost_panther,
absorption=mort_panther,          tr_fun=
function(x) 1/mean(x), latlon=TRUE))
```

See Supplementary material Appendix 1 for all relevant code for this example.

Step 3: analysis

Once the samc-class object has been created, it can be used to calculate the various metrics described in Fletcher et al. (2019) using the functions listed in Table 1. Each function supports a variety of inputs, including occupancy/abundance data, start locations, end locations and time steps. Depending on the inputs used, the functions can return a single value, a vector of values, or a matrix. The various input options and interpretation of results are explained in detail in the help documentation for each function.

For Florida panthers, we use the dispersal() function to calculate the probability of movement across the landscape, given these source locations (by providing a raster map of source locations identified as 1 and 0 otherwise) and the mortality() function to map the probability of mortality, given these source (source) locations. We also calculated life expectancy of dispersers from each of these locations with the survival() function. For each, we pass the initial distribution map of source locations as:

```
D_panther   <-   dispersal(samc=Pmatrix,
occ=source)
B_panther   <-   mortality(samc=Pmatrix,
occ=source)
Z_panther   <-    survival(samc=Pmatrix,
occ=source)
```

In the above function calls for dispersal and mortality, if `time=100` was added, the function would calculate $\tilde{\mathbf{D}}$ and $\tilde{\mathbf{B}}$ for 100 time steps, which in the case for panthers would capture a maximum possible distance of 50 km given the cell resolution. The absolute amount of time that this value reflects will vary based on application, but it is useful to think about it as the number of local transition steps considered. Given the biased random walk nature of the model, users may want the time considered to be much larger than the number of cells in the longest linear dimension for a given extent or Euclidean distance of interest.

In this example (274 664 cell map), on a desktop with 32 GB RAM and an Intel Core i7-3770 CPU @3.40 GHz processer, `dispersal()` function took 2.1 h to run, whereas the `mortality()` and `survival()` functions to < 4 s. If interest is in quantifying movement probabilities between pairs of patches with the `dispersal()` function (Fletcher et al. 2019), users can provide source and destination locations with the `origin` and `dest` parameters, which substantially increases computational efficiency. For instance, in the panther example, calculating $\mathbf{D}_{ij}$ between the two source locations took < 4 s.

Step 4: visualization and interpretation

Vector results returned by analyses correspond to cells in the landscape data and can be visualized accordingly. The `map()` function is provided to simplify this process; it ensures that the values in the vector results are mapped to the landscape in the correct order, and properly handles results based on datasets with `NA` values. The result is a raster that can be plotted directly using `plot()` without any additional processing.

For Florida panthers, we mapped both $\mathbf{D}$ and $\mathbf{B}$ with the `map()` function. Mapping dispersal from two source areas provides a means to understand the probability of dispersers visiting any location on the landscape (Fig. 2c; plotted on the log(odds) scale), illustrating that dispersal is expected to decrease with distance from the source locations. We note that using the `dispersal()` function for each source location separately can isolate roles of source habitats on potential connectivity. We also map the expected probability of mortality of individuals dispersing from these two source areas (Fig. 2d; plotted on the log(odds) scale). Expected mortality locations are generally similar to movement locations, except that mortality risk spikes near roads that are within the dispersal range from the source locations. Finally, we found that life expectancy (number of expected time steps before mortality, z) was estimated to be 2% greater for individuals dispersing from Big Cypress Reserve (z = 13 734 steps) than from Florida Panther National Wildlife Refuge (z = 13 465 steps). Other types of interpretation and visualization are possible, such as demarcation of potential corridors (Pinto and Keitt 2009), using models to parameterize network (or patch-based graph) links in connectivity models (Saura and Rubio 2010), and identifying the extent to which protected areas may contribute dispersers to larger metapopulations (Runge et al. 2006); see Fletcher et al. (2019) for discussion.

We note that at least three sources of uncertainty occur in predictions and interpretation of this framework: 1) uncertainty in the model framework (e.g. assumption of biased random walks); 2) uncertainty in inputs (e.g. resistance, mortality uncertainty as a function of land use); and 3) uncertainty in the metrics themselves (e.g. the variance of $\mathbf{D}$). The first source can be evaluated by comparing the SAMC to other modeling frameworks (Fletchet et al. 2019), the second can be addressed when resistance and mortality inputs have estimates of precision (e.g. SEs), whereas the third requires new model derivations (e.g. see Ross 2010 for an example for estimating the variance in life expectancy with an absorbing Markov chain).

Future extensions

We expect improved performance with continued development. At least one metric ($\mathbf{\Psi}^T\mathbf{D}$) is an ideal candidate for parallel computing (dividing the calculation into smaller tasks that can be executed simultaneously across multiple cores of a computer processor). Additionally, several metrics would benefit substantially from caching matrix decompositions (at the expense of memory). New convenience features would focus on expanding the input options to the current metrics. Examples include alternative methods for specifying locations more easily, calculating multiple time steps for short-term metrics using a vector input of time steps, and automating the calculation of metrics for pairwise start and end locations. Finally, fundamentally new features that could be implemented include supporting multiple absorption layers (which can represent different sources of mortality such as natural mortality, parasitism or predation), quantifying uncertainty in metrics, and alternatives to matrix/raster data for generating the transition matrix (e.g. patch-based networks, resource selection functions and related statistical models).

Conclusions

The spatial absorbing Markov chain framework delivers several advances in connectivity modeling (Fletcher et al. 2019) by providing a probabilistic approach for modeling and mapping movement, mortality and visitation rates of dispersers over short and long time scales. The `samc` package provides a means for implementing those advances. We expect that the `samc` package will catalyze addressing new questions in connectivity science and serve as a helpful tool for more reliable connectivity conservation.

Data availability statement

Data are available from the Dryad Digital Repository: <http://dx.doi.org/10.5061/dryad.0k6djh9wk> (Marx et al. 2019).

Software availability

The `samc` package is distributed under the GNU Public License (GPL) ver. 3 or later. It is available directly from

the Comprehensive R Archive Network (CRAN) <https://CRAN.R-project.org/package=samc> and can be installed using `install.packages('samc')`. The package source and documentation are hosted on GitHub <https://andrewmarx.github.io/samc>.

To cite samc or acknowledge its use, cite this Software note as follows, substituting the version of the application that you used for 'version 0':

Marx, A. J., Wang, C., Sefair, J. A., Acevedo, M. A. and Fletcher, Jr., R. J. 2019. samc: an R package for connectivity modeling with spatial absorbing Markov chains. – Ecography 42: 000–000 (ver. 0).

References

Acevedo, M. A. et al. 2015. Conservation under uncertainty: optimal network protection strategies for worst-case disturbance events. – J. Appl. Ecol. 52: 1588–1597.

Bates, D. and Eddelbuettel, D. 2013. Fast and elegant numerical linear algebra using the RcppEigen package. – J. Stat. Softw. 52: 1–24.

Bates, D. and Maechler, M. 2017. Matrix: sparse and dense matrix classes and methods. – R package ver. 1.2-12. <http://matrix.r-forge.r-project.org/>

Bocedi, G. et al. 2014. RangeShifter: a platform for modelling spatial eco-evolutionary dynamics and species' responses to environmental changes. – Methods Ecol. Evol. 5: 388–396.

Calabrese, J. M. and Fagan, W. F. 2004. A comparison-shopper's guide to connectivity metrics. – Front. Ecol. Environ. 2: 529–536.

Carroll, C. et al. 2001. Carnivores as focal species for conservation planning in the Rocky Mountain region. – Ecol. Appl. 11: 961–980.

Carroll, C. et al. 2012. Use of linkage mapping and centrality analysis across habitat gradients to conserve connectivity of gray wolf populations in western North America. – Conserv. Biol. 26: 78–87.

Cushman, S. A. et al. 2009. Use of empirically derived source-destination models to map regional conservation corridors. – Conserv. Biol. 23: 368–376.

Dixon, J. D. et al. 2006. Effectiveness of a regional corridor in connecting two Florida black bear populations. – Conserv. Biol. 20: 155–162.

Eddelbuettel, D. and Balamuta, J. J. 2017. Extending R with C++: a brief introduction to Rcpp. – PeerJ 5: e3188v1.

Elliot, N. B. et al. 2014. The devil is in the dispersers: predictions of landscape connectivity change with demography. – J. Appl. Ecol. 51: 1169–1178.

Etherington, T. R. 2016. Least-cost modelling and landscape ecology: concepts, applications and opportunities. – Curr. Landscape Ecol. Rep. 1: 40–53.

Fletcher Jr., R. J. and Fortin, M. J. 2018. Spatial ecology and conservation modeling: applications with R. – Springer.

Fletcher Jr., R. J. et al. 2011. Social network models predict movement and connectivity in ecological landscapes. – Proc. Natl Acad. Sci. USA 108: 19282–19287.

Fletcher, R. J. et al. 2014. The matrix alters the role of path redundancy on patch colonization rates. – Ecology 95: 1444–1450.

Fletcher, R. J. et al. 2016. Divergent perspectives on landscape connectivity reveal consistent effects from genes to communities. – Curr. Landscape Ecol. Rep. 1: 67–79.

Fletcher Jr., R. J. et al. 2019. Towards a unified framework for connectivity that disentangles movement and mortality in space and time. – Ecol. Lett. 22: 1680–1689.

Grilo, C. et al. 2018. BRAZIL ROAD-KILL: a data set of wildlife terrestrial vertebrate road-kills. – Ecology 99: 2625–2625.

Hanski, I. 1999. Metapopulation ecology. – Oxford Univ. Press.

Hijmans, R. J. 2017. raster: geographic data analysis and modeling. – R package ver. 2.6-7. <https://CRAN.R-project.org/package=raster>

Hudgins, E. J. et al. 2017. Predicting the spread of all invasive forest pests in the United States. – Ecol. Lett. 20: 426–435.

Kautz, R. et al. 2006. How much is enough? Landscape-scale conservation for the Florida panther. – Biol. Conserv. 130: 118–133.

Kemeny, J. G. and Snell, J. L. 1976. Finite Markov chains. – Springer.

Larkin, J. L. et al. 2004. Landscape linkages and conservation planning for the black bear in west-central Florida. – Anim. Conserv. 7: 23–34.

Low, M. et al. 2010. Habitat-specific differences in adult survival rates and its links to parental workload and on-nest predation. – J. Anim. Ecol. 79: 214–224.

Maehr, D. S. et al. 2002. Florida panther dispersal and conservation. – Biol. Conserv. 106: 187–197.

Marx, A. et al. 2019. Data from: samc: an R package for connectivity modeling with spatial absorbing Markov chains. – Dryad Digital Repository, <http://dx.doi.org/10.5061/dryad.0k6djh9wk>.

McRae, B. H. et al. 2008. Using circuit theory to model connectivity in ecology, evolution and conservation. – Ecology 89: 2712–2724.

Nathan, R. et al. 2008. A movement ecology paradigm for unifying organismal movement research. – Proc. Natl Acad. Sci. USA 105: 19052–19059.

Nowakowski, A. J. et al. 2015. Evaluating connectivity for tropical amphibians using empirically derived resistance surfaces. – Ecol. Appl. 25: 928–942.

Panzacchi, M. et al. 2016. Predicting the continuum between corridors and barriers to animal movements using step selection functions and randomized shortest paths. – J. Anim. Ecol. 85: 32–42.

Pinto, N. and Keitt, T. H. 2009. Beyond the least-cost path: evaluating corridor redundancy using a graph–theoretic approach. – Landscape Ecol. 24: 253–266.

Ross, S. M. 2010. Introduction to probability models, 10th edn. – Academic Press.

Rudnick, D. A. et al. 2012. The role of landscape connectivity in planning and implementing conservation and restoration priorities. – Issues Ecol. 16: 1–20.

Runge, J. P. et al. 2006. The role of local populations within a landscape context: defining and classifying sources and sinks. – Am. Nat. 167: 925–938.

Saerens, M. et al. 2009. Randomized shortest-path problems: two related models. – Neural Comput. 21: 2363–2404.

Santos, S. M. et al. 2013. Relative effects of road risk, habitat suitability and connectivity on wildlife roadkills: the case of tawny owls (*Strix aluco*). – PLoS One 8(11): e79967.

Saura, S. and Rubio, L. 2010. A common currency for the different ways in which patches and links can contribute to

habitat availability and connectivity in the landscape. – Ecography 33: 523–537.

Sawyer, S. C. et al. 2011. Placing linkages among fragmented habitats: do least-cost models reflect how animals use landscapes? – J. Appl. Ecol. 48: 668–678.

Schwab, A. C. and Zandbergen, P. A. 2011. Vehicle-related mortality and road crossing behavior of the Florida panther. – Appl. Geogr. 31: 859–870.

Sefair, J. A. et al. 2017. A defender–attacker model and algorithm for maximizing weighted expected hitting time with application to conservation planning. – IISE Trans. 49: 1112–1128.

Slatkin, M. 1993. Isolation by distance in equilibrium and nonequilibrium populations. – Evolution 47: 264–279.

Taylor, P. D. et al. 1993. Connectivity is a vital element of landscape structure. – Oikos 68: 571–573.

van de Kerk, M. et al. 2019. Dynamics, persistence and genetic management of the endangered Florida panther population. – Wildl. Monogr. 203: 3–35.

van Etten, J. 2012. gdistance: distances and routes on geographical grids. – R package ver. 1.1-4, <http://CRAN.R-project.org/package=gdistance>.

Vasudev, D. et al. 2015. From dispersal constraints to landscape connectivity: lessons from species distribution modeling. – Ecography 38: 967–978.

Zeller, K. A. et al. 2012. Estimating landscape resistance to movement: a review. – Landscape Ecol. 27: 777–797.

Supplementary material (available online as Appendix ecog-04891 at <www.ecography.org/appendix/ecog-04891>). Appendix 1.