University of Nevada, Reno

Using simulations to predict the genetic connectivity of the Mojave desert tortoise

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy in Geography

by

Derek Anthony Friend

Dr. Scott Bassett/Dissertation Advisor

May, 2022

THE GRADUATE SCHOOL

We recommend that the dissertation
prepared under our supervision by

Derek Anthony Friend

entitled

Using simulations to predict Mojave desert tortoise genetic connectivity

be accepted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

Scott Bassett, D. Des
Advisor

Todd Esque, Ph.D
Committee Member

Scott Kelley, Ph.D
Committee Member

Ken Nussear, Ph.D
Committee Member

Kevin Shoemaker, Ph.D
Graduate School Representative

David W. Zeh, Ph.D., Dean
Graduate School

May, 2022

ABSTRACT

The Mojave desert tortoise is a threatened species that is facing habitat fragmentation from human development. Understanding the impact of fragmentation on this species is critical for developing appropriate conservation actions, but the effects of habitat fragmentation are often delayed, making it difficult to assess the impacts of recent landscape change. One tool often used to predict the impacts of fragmentation are agent-based models, which simulate the behavior and life-history of individual "agents". Agent-based models allow researchers to investigate the impacts of habitat fragmentation under many scenarios, which is useful for guiding conservation actions. However, because agent-based models are computationally intense, they are often limited to small spatial extents and low numbers of agents – while performing these simulations at large scales could lead to important insights, this is often infeasible.

In this dissertation, I use a computationally efficient agent-based model to assess the impact of anthropogenic development on the range-wide genetic connectivity of the Mojave desert tortoise. In Chapter 1, I describe the *quadtree* R package, which implements the region quadtree data structure in C++ and makes it available to the R programming environment – using this data structure increases the speed of the agent-based model. In Chapter 2, I calibrate and validate an agent-based model for predicting desert tortoise genetic connectivity. In Chapter 3, I use the model to make range-wide projections of the influence of anthropogenic development on desert tortoise genetic connectivity.

ACKNOWLEDGEMENTS

First of all, thanks to Scott Bassett for guiding me through this entire process. In addition, I'd like to thank my committee for their feedback and guidance: Todd Esque, Scott Kelley, Kevin Shoemaker, and Ken Nussear.

I want to acknowledge the work of the people who were on the Clark County Project with me: Ken Nussear, Kirsten Dutcher, Scott Bassett, and Scott Wright. This entire dissertation was founded on the work that we did in that project and would be completely different if I hadn't been a part of that project.

A big thank-you to Ken for all the support he's given me through the years – he has answered my questions, provided me with data, and given me advice on my research. Also, he let me use his server for running my models – if I hadn't had access to that sort of computing power this research wouldn't have happened.

Kirsten let me use her data, gave me advice, and fielded my questions about genetics, for which I am very grateful.

Thanks to Bridgette Hagerty for letting me use her genetic data in Chapter 3.

Thanks to the members of the Albright-Heaton-Nussear lab group for their feedback on my work throughout my years here.

And finally, I'd like to thank my parents and the rest of my family for their constant and steadfast support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

A major conservation concern for many species is habitat loss and fragmentation – human development both removes suitable areas of habitat and isolates the remaining habitat (Fahrig 2003). Fragmentation causes a variety of negative conservation outcomes. As habitat fragmentation increases, population resilience decreases – in connected habitat, catastrophic population declines can be mitigated by the ”rescue effect”, in which population losses are recouped by immigration from neighboring populations (Brown & Kodric-Brown 1977). Fragmented habitats, however, may prevent immigration between populations, making individual populations more susceptible to catastrophic events (Fahrig 2002; Haddad et al. 2015). From a genetic perspective, fragmentation reduces gene flow by limiting movement between populations, which can lead to a loss of genetic diversity and inbreeding depression (Keyghobadi 2007; Mills & Allendorf 1996). Ultimately, habitat fragmentation can lead to increased extinction risk (Fischer & Lindenmayer 2007).

The Mojave desert tortoise is a species of conservation concern for which habitat fragmentation is a major threat (Averill-Murray et al. 2021; Dutcher et al. 2020). Human development like roads, urban areas, railroads, and renewable energy facilities are present across much of the tortoises’ historical range (Parker et al. 2018). The limited dispersal ability of desert tortoises makes many anthropogenic landscape features formidable barriers to movement. Maintaining habitat connectivity even as human

development increases is critical to the conservation of this species (U.S. Fish and Wildlife Service 2011).

Assessing the impact of fragmentation is a critical first step in identifying conservation actions to help preserve any at-risk species. As computers have become more powerful and more ubiquitous, computer simulations have become a common method of investigating the causes and consequences of habitat fragmentation (Landguth et al. 2010; Rebaudo et al. 2013). While simulations are by no means a replacement for empirical data, they offer a powerful way to explore the impacts of fragmentation (Pearson & Dawson 2005). The effects of fragmentation are often delayed – the negative consequences may not be detectable for years after fragmentation occurs (Blair et al. 2012; Tilman et al. 1994). Simulations allow for investigations into long-term effects that empirical data simply cannot address when habitat fragmentation is relatively recent (Landguth et al. 2010). In addition, simulations allow for the testing of assumptions – many simulations are used to assess the way that assumptions about individual behavior impact large-scale patterns (Landguth et al. 2010).

However, a major limitation for many simulations is their computational intensity. Computationally intense simulations are often limited in scope – while they may work for small study systems, their applicability to large-scale systems may be limited. Therefore, writing computationally efficient code can expand the capabilities and capacity of computer simulations, which can lead to insights at broader scales.

The primary goal of this dissertation is to develop a computationally efficient simulation that can enable insights into desert tortoise genetic connectivity at a range-wide scale. In Chapter 1, I describe the quadtree data structure, which is a data structure for landscape representation that can increase computational speed by using variable-sized grid cells. I present the *quadtree* R package, which implements a quadtree data structure in C++ and is made accessible to the R programming environment. In Chapter 2, I describe an agent-based model for assessing the genetic connectivity of desert tortoises. The model is computationally efficient and scalable, in part because of the use of the data structure discussed in Chapter 1. By comparing the model output to actual genetic data, I assess the performance of the model and select parameter values that reproduce observed patterns. In Chapter 3, I apply the agent-based model at a range-wide scale and assess the how human development has impacted genetic connectivity.

References

Averill-Murray, R. C., Esque, T. C., Allison, L. J., Bassett, S., Carter, S. K., Dutcher, K. E., … Nussear, K. E. (2021). *Connectivity of Mojave Desert tortoise populations—Management implications for maintaining a viable recovery network*. *Open-File Report*.

Blair, C., Weigel, D. E., Balazik, M., Keeley, A. T. H., Walker, F. M., Landguth, E., … Balkenhol, N. (2012). A simulation-based evaluation of methods for inferring linear barriers to gene flow. *Molecular Ecology Resources*, *12*, 822–833.

Brown, J. H., & Kodric-Brown, A. (1977). Turnover Rates in Insular Biogeography: Effect of Immigration on Extinction. *Ecology*, *58*(2), 445–449.

Dutcher, K. E., Vandergast, A. G., Esque, T. C., Mitelberg, A., Matocq, M. D., Heaton, J. S., & Nussear, K. E. (2020). Genes in space: what Mojave desert tortoise genetics can tell us about landscape connectivity. *Conservation Genetics*.

Fahrig, L. (2003). Effects of Habitat Fragmentation on Biodiversity. *Annual Review of Ecology, Evolution, and Systematics*, *34*, 487–515.

Fahrig, L. (2002). Effect of habitat fragmentation on the extinction threshold: A synthesis. *Ecological Applications*, *12*(2), 346–353.

Fischer, J., & Lindenmayer, D. B. (2007). Landscape modification and habitat fragmentation: a synthesis. *Global Ecology and Biogeography*, *16*, 265–280.

Haddad, N. M., Brudvig, L. A., Clobert, J., Davies, K. F., Gonzalez, A., Holt, R. D., … Townshend, J. R. (2015). Habitat fragmentation and its lasting impact on Earth's ecosystems. *Science Advances*, *1*(2), 1–10.

Keyghobadi, N. (2007). The genetic implications of habitat fragmentation for animals. *Canadian Journal of Zoology*, *85*(10), 1049–1064.

Landguth, E. L., Cushman, S. A., Schwartz, M. K., McKelvey, K. S., Murphy, M., & Luikart, G. (2010). Quantifying the lag time to detect barriers in landscape genetics. *Molecular Ecology*, *19*(19), 4179–4191.

Mills, L. S., & Allendorf, F. W. (1996). The One-Migrant-per-Generation Rule in Conservation and Management. *Conservation Biology*, *10*(6), 1509–1518. https://doi.org/10.1046/j.1523-1739.1996.10061509.x

Parker, S. S., Cohen, B. S., & Moore, J. (2018). Impact of solar and wind development on conservation values in the Mojave Desert. *PLoS ONE*, *13*(12), 1–16.

Pearson, R. G., & Dawson, T. P. (2005). Long-distance plant dispersal and habitat fragmentation: Identifying conservation targets for spatial landscape planning under climate change. *Biological Conservation*, *123*(3), 389–401.

Rebaudo, F., Le Rouzic, A., Dupas, S., Silvain, J.-F., Harry, M., & Dangles, O. (2013). SimAdapt: An individual-based genetic model for simulating landscape management impacts on populations. *Methods in Ecology and Evolution*, *4*, 595–600.

Tilman, D., May, R. M., Lehman, C. L., & Nowak, M. A. (1994). Habitat destruction and the extinction debt. *Nature*, *371*, 65–66.

U.S. Fish and Wildlife Service. (2011). *Revised Recovery Plan for the Mojave Population of the Desert Tortoise*. Sacramento, CA.

CHAPTER 1

quadtree: An R Package for Region Quadtrees

Introduction

An important consideration when collecting and analyzing data is the data structure used to represent the phenomena of interest. Different data structures have differing strengths and weaknesses, and the "best" data structure will vary depending on the situation. In a spatial context, data are often represented with the "field" data model, which represents the variation of a single variable across space. For example, climate variables like average temperature and precipitation are often modeled in this way. By far the most common data structure used for this type of data is a raster, which consists of a grid of equal-sized rectangular cells over a region, each of which contains the value of the variable for that particular area (Anselin & Getis 1992; Couclelis 1992).

However, a raster may not always be the best choice, and there are several other options for representing field data (Goodchild et al. 1992). One in particular is the region quadtree, a hierarchical data structure that can be used to store field data (Samet 1984). Unlike a raster, a quadtree can have variable sized cells, which allows for differing levels of detail across space. This is particularly useful for retaining fine-grain detail in heterogeneous areas while using larger cells for more homogeneous areas, and it can have more efficient storage than a raster data structure, which must have a uniform cell size (Samet 1984). In addition, variable cell sizes can allow for faster

computation of models and operations that rely on field data structures (Popinet 2011; van Bemmelen et al. 1993). While not as widely used as rasters, region quadtrees have still been used in a wide variety of contexts, including image compression (Gargantini 1982), fracture propagation modelling (Bittencourt et al. 1996; Ooi et al. 2015), flood modelling (Liang & Borthwick 2009), robot navigation (Kambhampati & Davis 1986; Chen et al. 1997), spatial sampling design (Minasny et al. 2007), and agent-based modelling (Tischendorf et al. 1998), among others.

In this paper, I present *quadtree*, a package written for the R environment (R Core Team 2022) that provides access to a quadtree data structure. For greater computational efficiency, the data structure is programmed in C++ and made accessible to R via the *Rcpp* package (Eddelbuettel & François 2011). In what follows, I describe the quadtree data structure in more depth and demonstrate how the *quadtree* package can be used. The source code is available online at github.com/dfriend21/quadtree, and the package can be downloaded through the Comprehensive R Archive Network (CRAN).

The quadtree data structure

As the name suggests, the quadtree data structure is a tree where each node has either zero or four children, and is well-suited for representing two-dimensional space. When used in this fashion, each node represents a rectangular cell, and the root of the tree represents a single cell that encompasses the entire area. If a node has children, they represent the four cells created when the parent cell is divided into quadrants. Each node is associated with a value for the variable being modelled (Hunter & Steiglitz

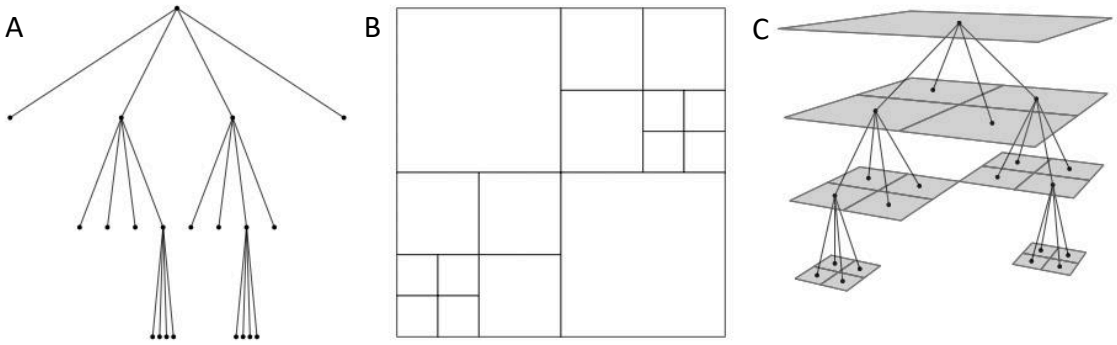

Figure 1 – Visual representations of the quadtree data structure. (A) A quadtree. (B) The 2-D representation of the quadtree presented in A. (C) A 3-D representation that shows how the data structure is represented spatially.

1979). This creates a hierarchical data structure where any given point in space may be contained within many different cells at different levels. To create a space-exhausting surface similar to a raster where each point is associated with a single value, the terminal nodes (leaves) can be treated as the only value for a given point, and the values at the higher levels can essentially be ignored. Unlike a raster, the hierarchical nature of a quadtree easily facilitates variable-sized cells. When a quadtree is used in this fashion it is referred to as a region quadtree. See Figure 1 for an illustration of the region quadtree data structure. It is worth noting that quadtrees have been used for other types of data as well, like points, lines, and curves (Samet 1984); however, only region quadtrees are implemented in this package.

The *quadtree* package

The *quadtree* R package is coded using R and C++. The quadtree data structure and other classes are defined in C++ and made available in R via the *Rcpp* package.

Creating a quadtree

In the *quadtree* package, quadtrees are constructed from rasters represented as matrices or as `RasterLayer` objects (Hijmans & van Etten 2021) using the `quadtree()` function. A top-down approach is used in which the raster is recursively divided into smaller and smaller quadrants – at each step a decision is made as to whether to divide a quadrant based on the values of the raster cells that fall within the quadrant. If the decision criteria are met, the quadrant is split into its four children and the same decision-making process is then applied recursively to each child quadrant. If, however, the decision criteria are not met, the quadrant is not split, and it is assigned a value based on the cells that fall within the quadrant. This procedure continues recursively until the decision criteria are not met or until the smallest possible cell size (i.e. the cell size of the input raster) has been reached.

The process begins at the highest level – at first, the entire raster is treated as a single "quadrant", and all of the raster cells are used in the split decision. If it is split, it is divided into four quadrants (so that each quadrant contains a quarter of the cells in the raster), and then the same process is applied to these child quadrants.

Before creating a quadtree, two sub-processes must be defined – the splitting criteria and the "combine" method. The splitting criteria uses the cell values of a quadrant to determine whether to split the quadrant. For example, the decision could be based on the standard deviation of all cell values within a quadrant – if the standard deviation is above a certain threshold, then the quadrant is split into its four children. The second process is the combine method, which is used to assign a single value to a quadrant based on the cell values it contains. For example, the mean or median could

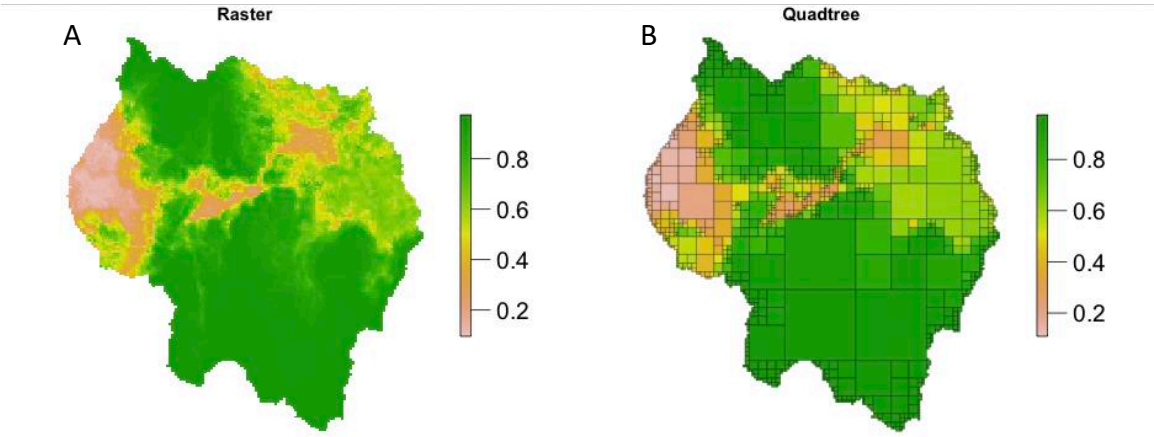

Figure 2 – A raster (A) and a quadtree (B) created using that raster.

be used since these functions aggregate multiple values into a single value. Figure 2 shows a raster and a quadtree created using that raster as input.

Split and combine functions

Three built-in methods for splitting a quadrant are provided. The "`range`" method calculates the difference between the minimum and maximum values within the quadrant, "`sd`" calculates the standard deviation, and "cv" calculates the coefficient of variation (Samet 1984; Wu et al. 1982). Regardless of which method is used, the resulting value is compared to the `split_threshold` parameter, and if it exceeds the threshold, the quadrant is split. In the following example and in all other examples, an example raster called `habitat` (a `RasterLayer` object which is provided with the *quadtree* package) is used as the input to the `quadtree()` function; the original raster can be seen in Figure 2.

```
R> data(habitat)
R> qt_sd <- quadtree(habitat, .1, "sd")
R> qt_range <- quadtree(habitat, .1, "range")
R> qt_cv <- quadtree(habitat, .1, "cv")
```

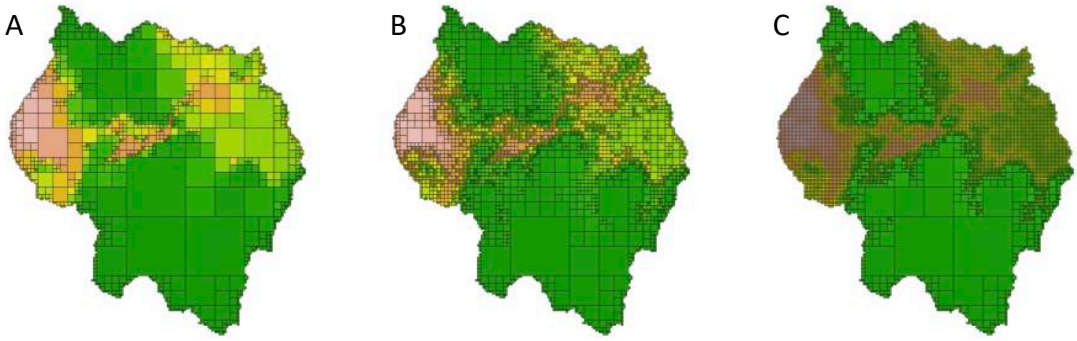

Figure 3 – Quadtrees created using different split methods. (A) A quadtree created using the 'standard deviation' method with a threshold of 0.1. (B) A quadtree created using the 'range' method with a threshold of 0.1. (C) A quadtree created using the custom splitting method described in the text.

See Figure 3 to see the result of using different split methods.

Four combine methods are provided – "mean", "median", "min", and "max" (Samet 1984; Wu et al. 1982). They aggregate values by calculating the mean, median, minimum or maximum, respectively, of all values in the quadrant.

```
R> qt_mean <- quadtree(habitat, .1, "sd", combine_method = "mean")
R> qt_median <- quadtree(habitat, .1, "sd", combine_method = "median")
R> qt_min <- quadtree(habitat, .1, "sd", combine_method = "min")
R> qt_max <- quadtree(habitat, .1, "sd", combine_method = "max")
```

See Figure 4 to see the results of using the different combine functions.

Users can also define custom split and combine functions, which allows for a high degree of flexibility in the quadtree creation process. These functions must take two parameters: `vals`, a numeric vector of the values of the cells within the current quadrant, and `args`, a named list that contains the arguments needed by the custom function. The `split_args` and `combine_args` parameters are passed to the `args` parameter of `split_fun` and `combine_fun`, respectively.

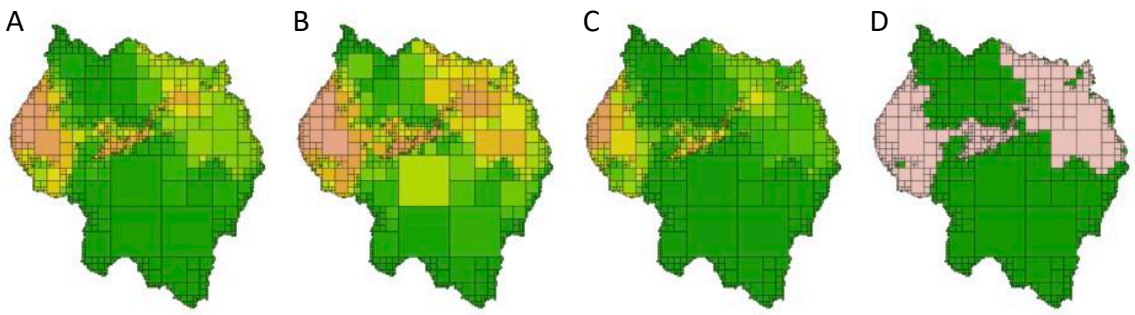

Figure 4 – Quadtrees created using the same splitting method but different combine functions. Note that the structure of each quadtree is identical – the only difference is the value of the cells. (A) mean (B) min (C) max (D) the custom combine function described in the text.

Custom split functions must return a boolean value, where TRUE indicates that the quadrant should be split. For example, we could create a simple function that splits a quadrant if any of the values fall below a certain threshold:

```
R> split_fun <- function(vals, args) {
+     return(any(vals < args$threshold))
+  }
R> qt_cust_split <- quadtree(habitat, split_method = "custom",
+     split_fun = split_fun, split_args = list(threshold = .8))
```

Custom combine functions must return a numeric value, which will be used as the value of the quadtree cell. For example, the following function returns either 0 or 1 based on whether the mean falls below a given threshold:

```
R> combine_fun <- function(vals, args) {
+     if(any(is.na(vals))) {
+           return(NA)
+     }
+     if(mean(vals) < args$threshold) {
+           return(0)
+     } else {
+           return(1)
+     }
+  }
R> qt_cust_combine <- quadtree(habitat, .1, "sd",
+     combine_method = "custom", combine_fun = combine_fun,
```

```
+     combine_args = list(threshold = .65))
```

The results from using the custom methods can be seen in Figure 3 and Figure 4.

Pre-creation dimension adjustment

The afore-mentioned method of creating quadtrees works most naturally on square rasters (same number of rows and columns) with the number of cells along each axis being a power of two. If the number of cells along an axis is not a power of two, there will be a point during the splitting process when a quadrant will have an odd number of cells along at least one of the dimensions, and therefore cannot be divided into four equal-sized child quadrants.

Forcing all input rasters to be square and have dimensions that are a power of two is a very restrictive requirement, so *quadtree* provides two methods for creating quadtrees from rasters that don't satisfy these conditions, specified by the `adj_type` parameter.

The first method, referred to as the "expand" method, adds `NA` rows and columns to both dimensions until the numbers of rows and columns are the same and are a power of two. For example, if a raster has 52 rows and 97 columns, `NA` rows and columns could be added to produce a raster with 128 rows and columns (which is a power of two). See Figure 5 for an example.

```
R> qt_expand <- quadtree(habitat, .1, "sd", adj_type = "expand")
```

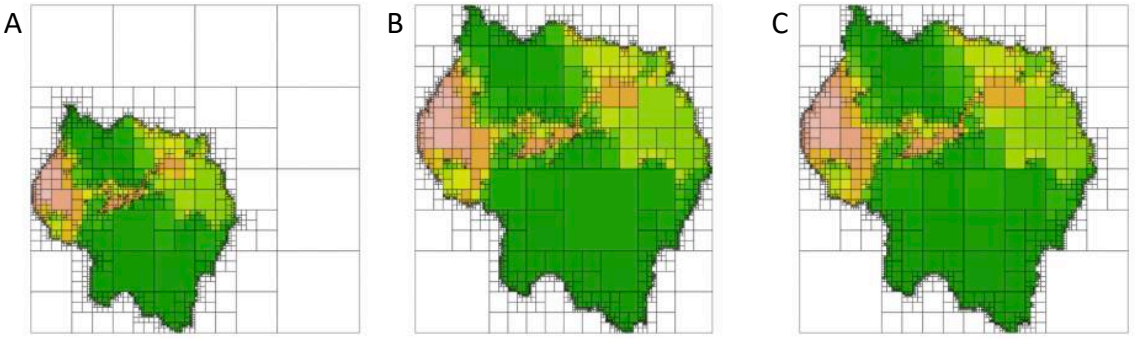

Figure 5 – Examples of quadtrees created using different pre-creation adjustment methods. (A) A quadtree created by padding both dimensions with NA rows and columns before constructing the quadtree. (B) A quadtree created by resampling the raster before constructing the quadtree so that the number of rows and columns are both a power of two. Note that the cells are not perfectly square. (C) A quadtree created by resampling the raster before constructing the quadtree so that the dimensions are a power of two, where the shorter dimension was padded with NA's before resampling so that the resulting quadtree has square cells, like the original raster.

The second method, called the "`resample`" method, resamples the raster so that the dimensions fulfill the requirement. This can be performed on a non-square raster – a raster with 52 rows and 97 columns could be resampled to have 128 rows and columns. However, doing so means that the quadtree cells will be differently proportioned than the original raster cells. For example, if the original raster cells are square, this will create a quadtree with rectangular but non-square cells. This may not be desirable, so an additional first step can be taken to prevent this. Before resampling, `NA` rows or columns can be added to the smaller dimension to make the raster square. For example, we can add `NA` rows to the raster with 52 rows and 97 columns to create a raster with 97 rows and columns that can then be resampled to have 128 rows and columns (or any other power of two). See Figure 5 for examples of both methods.

```
R> qt_resample_pad <- quadtree(habitat, .1, "sd",
+     adj_type = "resample", resample_n_side = 256,
+     resample_pad_nas = TRUE)
R> qt_resample_no_pad <- quadtree(habitat, .1, "sd",
+     adj_type = "resample", resample_n_side = 256,
```

```
+     resample_pad_nas = FALSE)
```

Using templates to create quadtrees

A quadtree can also be created using another quadtree as a template. In this case, the quadtree is created such that its structure is the exact same as the template quadtree, but the values of the quadtree are calculated from a new, user-provided raster. This is useful if the quadtree structure should be determined by a different variable than the variable represented in the input raster.

For example, a raster representing the presence or absence of roads (where zero indicates absence and one indicates presence) could be used to make a quadtree, where cells are split whenever a quadrant contains a one, resulting in the smallest possible cell size in areas that contain a road. First, we can define a splitting function for the situation:

```
R> split_if_one <- function(vals, args) {
+     if(any(vals == 1, na.rm = TRUE)) return(TRUE)
+     return(FALSE)
+  }
```

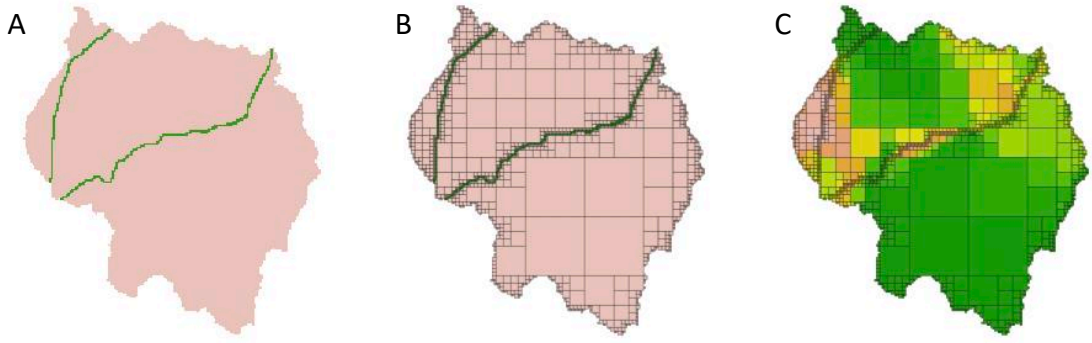

Figure 6 – An example of using a template to create a quadtree. (A) A raster representing the presence or absence of roads. (B) A quadtree created from the roads raster. (C) A quadtree created using the same raster as in previous examples but using the roads quadtree as a template.

Once we make this quadtree, we can use it as a template to create a new quadtree that has the same structure as the template. Note that the new raster should have the same resolution and extent as the raster used to create the template.

```
R> data(habitat_roads)
R> qt_template <- quadtree(habitat_roads, split_method = "custom",
+     split_fun = split_if_one)
R> qt_roads <- quadtree(habitat, template_quadtree = qt_template)
```

See Figure 6 for plots of the quadtrees created from the sample code.

Other options for quadtree creation

There are other parameters that also control the construction of a quadtree. In particular, the user can specify a minimum and/or maximum cell size, thereby controlling the allowable cell size of the quadtree. When a maximum cell size is set, any quadrants with dimensions larger than the maximum size are forced to split. When a minimum size is set, a quadrant cannot be split if its children would be smaller than the minimum cell size. See Figure 7 for the plots of the quadtrees created by the following code.

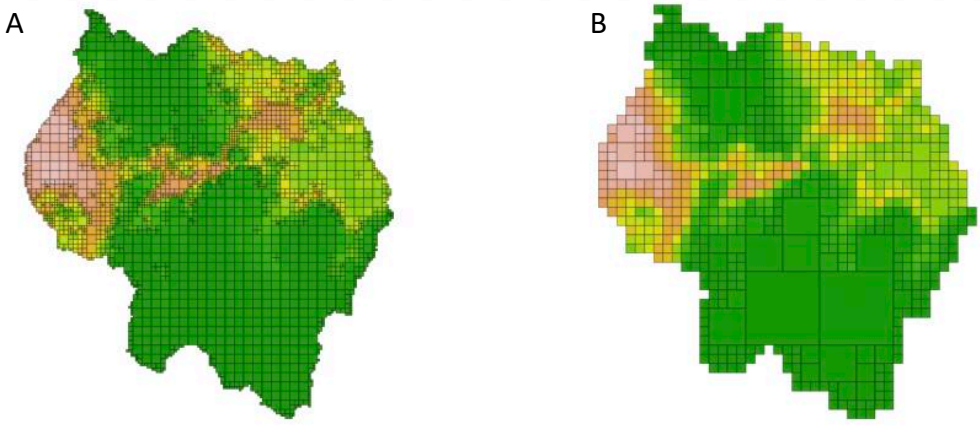

Figure 7 – Quadtrees created using (A) a maximum cell size of 1000 and (B) a minimum cell size of 1000.

```
R> qt_max_cell <- quadtree(habitat, .15, max_cell_length = 1000)
R> qt_min_cell <- quadtree(habitat, .15, min_cell_length = 1000)
```

In addition, users can specify how NA values should be treated during the creation process. By default, a quadrant is split if it has any NA values – this preserves the shape of a landscape if it is not square. This can be disabled, although it can sometimes produce unexpected results. In addition, by default quadrants that contain only NA cells are not split, but this can be disabled. In practice this causes all quadrants containing entirely NA cells to be split to their smallest possible cell values. Examples of quadtrees created using these options can be seen in Figure 8.

```
R> qt_any <- quadtree(habitat, .15, split_if_any_na = FALSE)
R> qt_all <- quadtree(habitat, .15, split_if_all_na = TRUE)
```

Working with quadtrees

Retrieving basic info about a quadtree

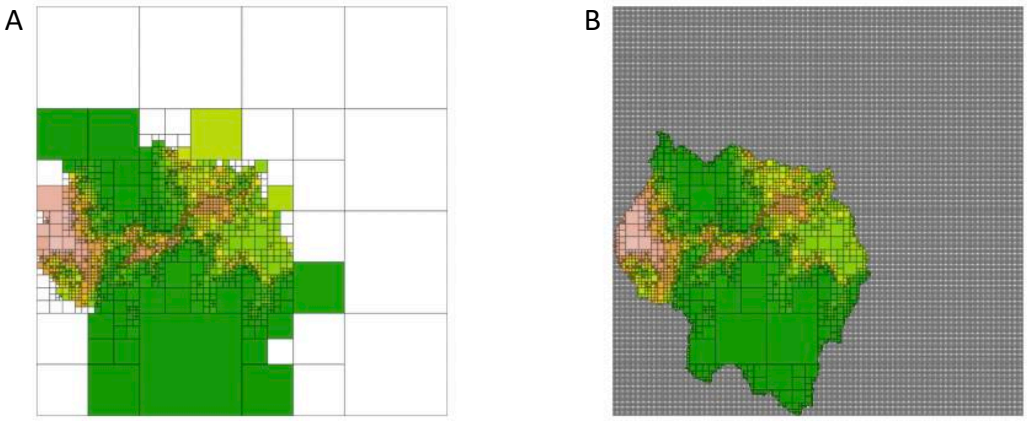

Figure 8 – Quadtrees created using different ways to handle NA values. (A) A quadtree created without forcing quadrants that contain NA values to split. (B) A quadtree created while forcing quadrants with only NA values to split.

Functions are provided for retrieving basic information about a quadtree. `summary()` can be used to print out basic information about a quadtree, while other functions like `projection()`, `extent()` and `n_cells()` can be a used to retrieve information about a quadtree.

```
R> qt <- quadtree(habitat, .2)
R> summary(qt)

class         : Quadtree
# of cells    : 3277
min cell size : 250
extent        : 0, 64000, 0, 64000 (xmin, xmax, ymin, ymax)
crs           : NA
values        : 0.109, 0.974 (min, max)

R> extent(qt)

class      : Extent
xmin       : 0
xmax       : 64000
ymin       : 0
ymax       : 64000

R> n_cells(qt)

[1] 3277
```

Retrieving cell-level data

Other functions can be used to retrieve information about quadtree cells. The `extract()` function allows users to retrieve the values at point locations. In addition, the `extents` parameter can be set to `TRUE` to return information on each cell (cell ID and extent) in addition to the cell value.

```
R> pts <- cbind(x = c(3959, 5609, 20161, 27662, 32763),
+               y = c(29586, 10835, 31836, 10834, 36337))
R> extract(qt, pts)
```

```
[1] 0.1451094     NaN 0.8269687 0.9556426 0.7241875

R> extract(qt, pts, extents = TRUE)

       id  xmin  xmax  ymin  ymax     value
[1,] 1454  2000  4000 28000 30000 0.1451094
[2,] 3223     0  8000  8000 16000       NaN
[3,] 2593 20000 22000 30000 32000 0.8269687
[4,] 3449 24000 32000  8000 16000 0.9556426
[5,] 1235 32000 33000 36000 37000 0.7241875
```

Another function, `get_neighbors()`, returns the cells that neighbor a given cell. This is especially useful for models of movement across a quadtree, as will be demonstrated in the "Example" section.

```
R> head(get_neighbors(qt, pts[1, ]))

       id xmin xmax  ymin  ymax     value
[1,] 1427 1750 2000 30000 30250 0.1840000
[2,] 1428 2000 4000 30000 32000 0.1364688
[3,] 1439 1000 2000 29000 30000 0.1478750
[4,] 1453 1000 2000 28000 29000 0.1163750
[5,] 1463 4000 5000 30000 31000 0.1403750
[6,] 1495 4000 5000 29000 30000 0.1693750
```

Copying a quadtree

As previously mentioned, the quadtree data structure is defined in C++. The R object accessible by the user consists of a pointer to the underlying C++ object. This means that copying a quadtree simply by assigning it to a new variable makes a shallow copy rather than a deep copy since only the pointer is copied rather than the underlying C++ object. The `copy()` function can be used to make a deep copy of a quadtree. This is

useful if the user wishes to make changes to a quadtree without modifying the original, as demonstrated in the next section.

Modifying cell values

Two functions are provided that modify the values of an existing quadtree. `transform_values()` applies a user-provided function to all cells. For example, it could be used to multiply all cell values by a constant or to take the square of all values.

```
R> qt_transform <- copy(qt)
R> transform_values(qt_transform, function(x) x * 2)
R> old_vals <- as_vector(qt)
R> new_vals <- as_vector(qt_transform)
R> old_vals[1000:1004]

[1] 0.5335 0.6510 0.7340 0.5120 0.7540

R> new_vals[1000:1004]

[1] 1.067 1.302 1.468 1.024 1.508
```

`set_values()` takes a set of points and values and changes the value of the cell each point falls in to its corresponding value.

```
R> qt_set_values <- copy(qt)
R> set_values(qt_set_values, pts[, 1:2], 1:5)
R> extract(qt_set_values, pts[, 1:2])

[1] 1 2 3 4 5
```

Reading and writing quadtrees

The *quadtree* package makes use of an external C++ library called *cereal* (Grant & Voorhies 2017) to facilitate reading and writing quadtrees to the file system, which can be done via the `read_quadtree()` and `write_quadtree()` functions.

```
R> write_quadtree("quadtree.qtree", qt)
R> qt_read <- read_quadtree("quadtree.qtree")
```

Converting quadtrees to other data types

Functions are also provided to convert the quadtree to other representations – `as_vector()` returns all cell values as a vector, `as_data_frame()` returns a data frame containing information on all cells, and `as_raster()` converts a quadtree to a `RasterLayer` (a class from the *raster* package).

```
R> head(as_vector(qt))

[1] NaN NaN NaN NaN NaN NaN

R> head(as_data_frame(qt)[,-9])

   id hasChildren level  xmin  xmax  ymin  ymax value parentID
3   2           0     2     0 16000 48000 64000   NaN        1
4   3           0     2 16000 32000 48000 64000   NaN        1
7   6           0     4     0  4000 44000 48000   NaN        5
9   8           0     5  4000  6000 46000 48000   NaN        7
10  9           0     5  6000  8000 46000 48000   NaN        7
11 10           0     5  4000  6000 44000 46000   NaN        7

R> as_raster(qt)

class      : RasterLayer
dimensions : 256, 256, 65536  (nrow, ncol, ncell)
resolution : 250, 250  (x, y)
extent     : 0, 64000, 0, 64000 (xmin, xmax, ymin, ymax)
crs        : NA
source     : memory
names      : layer
values     : 0.109, 0.974 (min, max)
```

Calculating least-cost paths

The quadtree package also includes functionality for calculating least-cost paths (LCPs) using the quadtree as a cost surface, which can be done by treating the quadtree as a network and then using a network-based shortest path algorithm (Huber & Church 1985; van Bemmelen et al. 1993). To interpret a quadtree as a network, each cell centroid is treated as a node, and two nodes are connected by an edge if they are adjacent to one another. Each edge has an assigned cost which is derived from the underlying surface. Because edges will always travel through two cells, the cost is determined by first dividing the edge into two so that each sub-segment falls in a single cell. The length of each sub-segment is multiplied by the cost of the corresponding cell, and then these two values are added together to determine the total cost of the edge. Once the network has been constructed, we can use Dijkstra's algorithm to find the LCP between two nodes.

Dijkstra's algorithm essentially builds a tree data structure where the starting node is the root of the tree. It iteratively builds the tree structure, and in each iteration it adds the node that is closest to the current tree – that is, it chooses the node which is easiest to reach. The result is that even if only one LCP is desired, LCPs to other nodes are also calculated in the process (Dijkstra 1959).

Once the tree has been constructed, LCPs can be found to any of the other nodes without further computation, allowing for efficient computation of multiple LCPs from a single point. The LCP functionality implemented in *quadtree* saves state – whenever an LCP is asked to be calculated, it first checks whether a path has been found

to that node already. If so, it simply returns the already-calculated path. If not, it builds out the existing tree until the desired node has been reached.

Two functions for calculating LCPs are provided that essentially differ in the way the algorithm stops running. `find_lcp()` finds the path between two points, so as soon as the path to the destination point is found, it stops and returns the path as a matrix.

```
R> qt_lcp <- quadtree(habitat, .05, "sd")
R> start_pt <- c(6989, 34007)
R> end_pt <- c(33015, 38162)
R> lcpf <- lcp_finder(qt_lcp, start_pt)
R> lcp <- find_lcp(lcpf, end_pt)
R> head(lcp)

        x     y cost_tot  dist_tot cost_cell
[1,] 6750 34250   0.0000    0.0000   0.47825
[2,] 6250 33750 301.4926  707.1068   0.37450
[3,] 5875 33375 489.4063 1237.4369   0.31400
[4,] 5625 33125 591.5832 1590.9903   0.26400
[5,] 5250 32750 718.5089 2121.3203   0.22700
[6,] 5250 32250 815.9464 2621.3203   0.16275
```

`find_lcps()`, on the other hand, uses a user-provided threshold and finds all paths whose cost-distance is less than the threshold. Alternatively, no threshold can be provided, in which case all possible LCPs from the given start point are found. Since the goal of `find_lcps()` is to find many paths rather than one specific path, it returns a summary matrix of all paths found rather than the path to any particular point. Individual paths can be extracted using `find_lcp()` with the `LcpFinder` object, which doesn't require any extra computation since the paths have already been found.

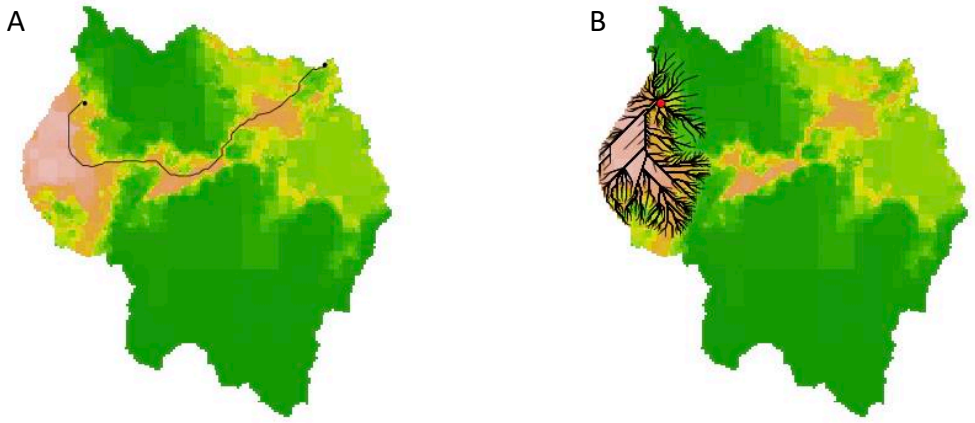

Figure 9 – Plots illustrating the least-cost path (LCP) functionality. (A) An LCP found between two points. (B) All LCPs from a starting point (red) that have a cost-distance of 4000 or less.

```
R> start_pt <- c(6989, 34007)
R> lcpf <- lcp_finder(qt_lcp, start_pt)
R> lcps <- find_lcps(lcpf, 4000)
R> head(lcps)

    id xmin xmax  ymin  ymax     value    area lcp_cost lcp_dist
1 2996 6500 7000 19500 20000 0.4382500  250000 3830.877 16544.39
2 2999 7000 8000 19000 20000 0.3544375 1000000 3774.968 16483.73
3 2774 3000 3250 20750 21000 0.3490000   62500 3747.612 15953.26
4 2775 3250 3500 20750 21000 0.3410000   62500 3784.512 16056.81
5 2778 3500 4000 20500 21000 0.4857500  250000 3957.449 16452.10
6 2757 2750 3000 21000 21250 0.4300000   62500 3621.361 15599.71
```

See Figure 9 for the output of the example code.

It's important to note that the variable-sized cells of a quadtree will cause the least-cost path to differ from the path that would be found if the original raster data were used (van Bemmelen et al. 1993). Thus, it is important for anyone using the LCP functionality on a quadtree to evaluate whether this error is significant in the given context.

Example

To demonstrate how the quadtree can be used, I present a simple example of a situation in which a quadtree can result in significant performance gains over a raster – in this case I construct an agent-based model that simulates the movement of individual agents. This type of simulation is frequently used in ecological applications to model populations over time while taking the landscape into account (Gallien et al. 2010; Philips 2020). Note that the use of quadtrees is by no means limited to this type of situation – the example is simply meant to illustrate one situation where it can be useful.

The simulation presented here tests the speed at which agents can travel to the other side of a landscape when linear barriers are present. There are two objectives of the simulation. First, I will compare the rate at which agents can move northward using two landscapes – one with barriers, and one with no barriers. Secondly, I will run the barrier simulation twice, once using a raster and once using a quadtree, and then compare the run time of the two simulations.

The landscape

The original landscape is represented by a raster with 256 rows and columns where all cells have a value of one except for "barrier" cells, which have a value of `NA`. Therefore, the no-barrier landscape is simply a uniform raster where every cell is one. The barrier landscape, on the other hand, has several linear barriers running more-or-less horizontally through the landscape – see Figure 10.

Quadtrees are created from both landscapes – in the case of the uniform landscape, the result is a single-cell quadtree. When creating the quadtree from the barrier landscape, any quadrant that has any NA cells are split, resulting in small cells near the barriers.

As mentioned above, one of the goals is to compare the run-times of a simulation using a quadtree and a simulation using a raster. However, for the latter simulation, a quadtree that has been split down to the smallest possible cell size is used, which results in a quadtree that is equivalent to the raster (I refer to this as the "raster-like quadtree"). The reason for this is so that the effect of the larger cell sizes on the simulation speed can be assessed more directly. If using a raster, I would need to use least-cost path functionality that was suitable for a raster data structure, which would make it very difficult to determine whether speed differences are due to the quadtree data structure or simply due to differing implementations of the least-cost path algorithm. In fact, a simulation was originally run using a `RasterLayer` from the *raster* package, but the least-cost path functionality available for `RasterLayer` objects is not efficient for this specific scenario, so it ran significantly slower. Since there are other factors causing the speed disparity besides the inherent difference between rasters and quadtrees, comparing the two has little value with regards to the goal of this exercise. However, using two quadtrees means that the same least-cost path functionality will be used for both, making the comparison more meaningful.

Movement process

Each agent moves in an uncorrelated random walk – they move with a consistent step size and with no directional bias, and they travel for a pre-defined number of steps. However, they are not allowed to cross NA cells. In the case that the chosen point for a step is unreachable, a different, reachable point is chosen. Determining "reachability" is greatly facilitated by the quadtree data structure. If a chosen point falls in the same cell that the agent is currently in, it is guaranteed to be reachable, and the agent moves there. If the chosen point falls in a neighboring cell and that cell is not NA, then that cell is also guaranteed to be reachable. If the chosen point is not in a neighboring cell, the LCP functionality is used on a constrained area (so that an agent doesn't follow an unrealistically long path around a barrier) to check if the point is reachable.

In this way, agents moving in larger cells where the cell size is greater than the step size will never need to perform the more costly least-cost path calculations, which greatly reduces the amount of computation necessary while still maintaining the finer cell resolution near roads. When the raster-like quadtree is used, however, the LCP functionality is used at every step, greatly increasing the amount of computation.

Simulations

In each simulation, the agents start in the bottom quarter and the goal is to assess the agents' northward movement. To this end, the number of agents in the northern half is recorded after each iteration.

As mentioned above, three simulations were run – one using a quadtree of the no-barrier landscape, one using a quadtree of the barrier landscape, and one using a

raster-like quadtree of the barrier landscape. The two simulations using the barrier landscape used the same random seed and therefore produced identical results – the only difference between the two is runtime. Each simulation used 100 agents and ran for 50 iterations.

Results

The results are graphically displayed in Figure 10. Unsurprisingly, the barrier landscape showed slower northward movement than the no-barrier landscape – at the end of 50 iterations, the simulation using the no-barrier landscape had approximately twice as many agents in the northern half as the simulation using the barrier landscape. Clearly, the barriers significantly slowed the agents' northward movement.

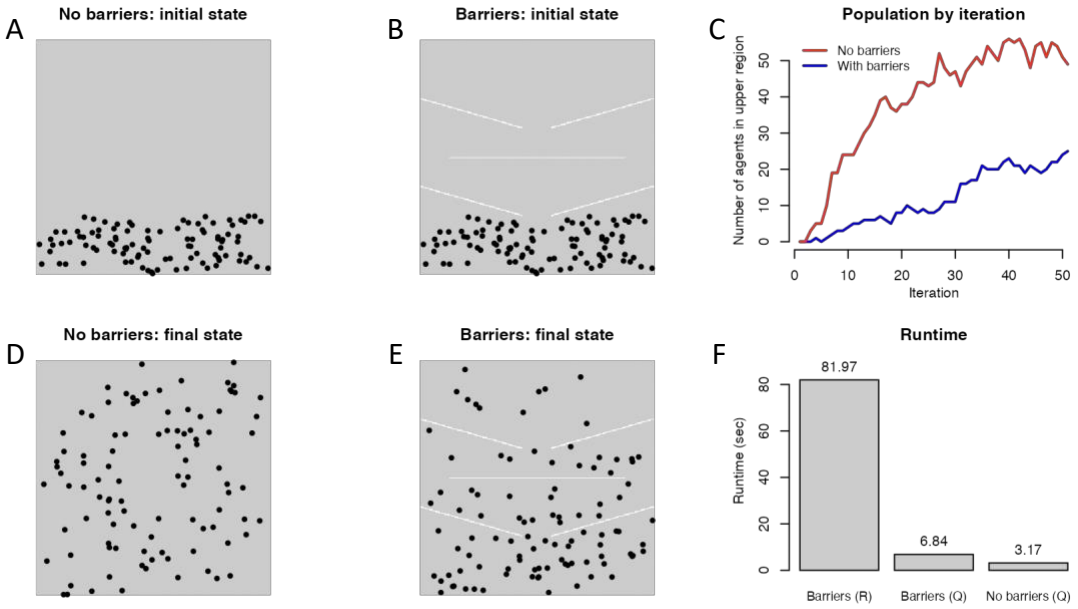

Figure 10 – The results of the simulation described in the text. (A & B) The initial distribution of agents on (A) the no-barrier landscape and (B) the barrier landscape. (C) The number of agents in the northern half of the landscape over time, for both landscapes. (D & E) The distribution of the agents after 50 iterations on (D) the no-barrier landscape and (E) the barrier landscape. (F) The runtime of each of the three simulations in seconds. "R" stands for "raster" and "Q" stands for "quadtree".

For the barrier landscape, the simulation using the raster-like quadtree performed about 12 times slower than the simulation using the variable cell-size quadtree (Figure 10) thus demonstrating the advantage of the quadtree data structure in this situation.

Conclusion

As illustrated by the example, region quadtrees can be used to speed up certain operations. However, the variable cell sizes of quadtrees can be useful for reasons other than computational speed. For example, they have been used to develop sampling schemes that have more sampling points in areas where a related variable has a high degree of variability, which can result in a relatively small number of sampling points that still cover the entire range of variability in the variable of interest (Minasny et al. 2007).

However, quadtrees have some disadvantages, and in many cases a raster will still be the best data structure to use. Quadtrees have the potential to save space in comparison with a raster, but they also have a considerable amount of overhead and can sometimes use more memory than a raster. While rasters need only store a cell's value, at minimum a quadtree must store the value as well as pointers to its children, meaning that a quadtree requires more memory per cell than a raster. If a quadtree is made to be identical to a raster by splitting it down to the smallest possible cell size, the quadtree will use more memory. It is only by aggregating enough cells that a quadtree can use less memory than a raster. In addition, the *quadtree* package stores additional

information about each cell, like the x and y limits as well as pointers to neighboring cells. While storing this information can make certain operations more efficient (like retrieving the neighbors of a given cell), it also makes it less space-efficient. Future versions of the package may implement a more efficient quadtree representation.

Another potential downside of using quadtrees is the arbitrary location of cell boundaries. This means that slight alterations to the dimensions of a raster will likely lead to a different quadtree structure – for example, if a raster is trimmed to a smaller extent, a quadtree created on the smaller raster will not simply be a trimmed version of the quadtree created on the original raster. Changing the extent changes the boundaries where a cell will be split, thereby changing the quadtree structure.

The *quadtree* package offers a data structure for region quadtrees, giving users another option for representing spatial data. In many cases a raster will still be the best data structure to use; as the example illustrates, however, region quadtrees can provide distinct advantages over rasters in certain contexts.

References

Anselin, L., & Getis, A. (1992). Spatial statistical analysis and geographic information systems. *The Annals of Regional Science*, *26*, 19–33.

Bittencourt, T. N., Wawrzynek, P. A., Ingraffea, A. R., & Sousa, J. L. (1996). Quasi-automatic simulation of crack propagation for 2D LEFM problems. *Engineering Fracture Mechanics*, *55*(2), 321–334.

Chen, D. Z., Szczerba, R. J., & Uhran, J. J. (1997). A framed-quadtree approach for determining euclidean shortest paths in a 2-D environment. *IEEE Transactions on Robotics and Automation*, *13*(5), 668–681.

Couclelis, H. (1992). People manipulate objects (but cultivate fields): Beyond the raster-vector debate in GIS. In A. U. Frank, I. Campari, & U. Formentini (Eds.), *Theories and*

Methods of Spatio-Temporal Reasoning in Geographic Space (pp. 65–77). Berlin, Heidelberg: Springer Berlin Heidelberg.

Dijkstra, E. W. (1959). A note on two problems in connexion with graphs. *Numerische Mathematik*, *1*, 269–271.

Eddelbuettel, D., & François, R. (2011). Rcpp: Seamless R and C++ Integration. *Journal of Statistical Software*, *40*(8), 1–18.

Gallien, L., Münkemüller, T., Albert, C. H., Boulangeat, I., & Thuiller, W. (2010). Predicting potential distributions of invasive species: where to go from here? *Diversity and Distributions*, *16*(3), 331–342.

Gargantini, I. (1982). An effective way to represent quadtrees. *Communications of the ACM*, *25*(12), 905–910.

Goodchild, M., Haining, R., & Wise, S. (1992). Integrating GIS and spatial data analysis: Problems and possibilities. *International Journal of Geographical Information Systems*, *6*(5), 407–423.

Grant, S. W., & Voorhies, R. (2017). cereal – A C++ Library for Serialization. Retrieved from http://uscilab.github.io/cereal/

Hijmans, R. J., & van Etten, J. (2021). raster: Geographic Data Analysis and Modeling. Retrieved from https://rspatial.org/raster

Huber, D. L., & Church, R. L. (1985). Transmission corridor location modeling. *Jounral of Transportation Engineering*, *111*(2), 114–130.

Hunter, G. M., & Steiglitz, K. (1979). Operations on Images Using Quad Trees. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *PAMI-1*(2), 145–153.

Kambhampati, S., & Davis, L. S. (1986). Multiresolution path planning for mobile robots. *IEEE Journal on Robotics and Automation*, *2*(3), 135–145.

Liang, Q., & Borthwick, A. G. L. (2009). Adaptive quadtree simulation of shallow flows with wet-dry fronts over complex topography. *Computers and Fluids*, *38*(2), 221–234.

Minasny, B., McBratney, A. B., & Walvoort, D. J. J. (2007). The variance quadtree algorithm: Use for spatial sampling design. *Computers and Geosciences*, *33*, 383–392.

Ooi, E. T., Man, H., Natarajan, S., & Song, C. (2015). Adaptation of quadtree meshes in the scaled boundary finite element method for crack propagation modelling. *Engineering Fracture Mechanics*, *144*, 101–117.

Philips, I. (2020). An agent based model to estimate lynx dispersal if re-Introduced to Scotland. *Applied Spatial Analysis and Policy*, *13*, 161–185.

Popinet, S. (2011). Quadtree-adaptive tsunami modelling. *Ocean Dynamics*, *61*(9), 1261–1285.

R Core Team. (2022). R: A Language and Environment for Statistical Computing. Vienna, Austria.

Samet, H. (1984). The quadtree and related hierarchical data structures. *ACM Computing Surveys (CSUR)*, *16*(2), 187–260.

Tischendorf, L., Irmler, U., & Hingst, R. (1998). A simulation experiment on the potential of hedgerows as movement corridors for forest carabids. *Ecological Modelling*, *106*(2–3), 107–118.

van Bemmelen, J., Quak, W., van Hekken, M., & van Oosterom, P. (1993). Vector vs. raster-based algorithms for cross country movement planning. *Proceedings Auto-Carto*, *11*.

Wu, A. Y., Hong, T. H., & Rosenfeld, A. (1982). Threshold Selection Using Quadtrees. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *PAMI-4*(1), 90–94.

CHAPTER 2

Calibrating and validating a landscape-genetic agent-based model for the Mojave desert tortoise

Introduction

Before humans started large-scale development in the Mojave Desert, Mojave desert tortoise habitat was largely contiguous (Averill-Murray et al. 2013). However, anthropogenic development, which began in the 19th century and has increased rapidly in recent years (Parker et al. 2018; Theobald et al. 2013), has caused this habitat to become fragmented. Human-built structures like railroads, interstates, solar facilities, and urban development have reduced both the quantity and connectivity of available habitat (Carter et al. 2020; Dutcher et al. 2020; Hagerty et al. 2011). Indeed, habitat fragmentation is a major threat to the species, and fragmentation of tortoise habitat is unlikely to slow. Las Vegas has been growing rapidly, and the recent emphasis on renewable energy has resulted in many large-scale solar facilities being built in tortoise habitat (Farnsworth et al. 2015). Previous studies have shown that anthropogenic development has already caused detectable genetic differentiation between populations (Dutcher et al. 2020). Therefore, finding ways to maintain tortoise connectivity in the face of development is a major conservation concern (Averill-Murray et al. 2013). Tools that can assess the impact of anthropogenic development on tortoise connectivity, and thus inform building plans and conservation actions, are important for this species' conservation (Averill-Murray et al. 2021).

In order to implement conservation actions to improve genetic connectivity, an understanding of the current genetic pattern on the landscape as well as the features inhibiting tortoise movement connectivity is critical. An important step towards this goal is collecting genetic data from tortoises across the Mojave and using it to assess current patterns of genetic connectivity while identifying features that may inhibit gene flow (Dutcher et al. 2020; Hagerty et al. 2011; Latch et al. 2011; Sánchez-Ramírez et al. 2018). Genetic data have limitations, however. One of the most significant limitations is the lag time between landscape changes and their subsequent appearance as a signal in the genetic data (Gregory & Beier 2014; Landguth et al. 2010; Safner et al. 2011). This makes it difficult to assess the genetic effects of recent landscape changes, particularly in long-lived species like the desert tortoise – their long generation time means that current genetic data likely reflect the past landscape more than the current landscape. Therefore, genetic data can provide little information on the impact of recent landscape change, making it difficult to predict the future impacts of landscape changes whose effects are not yet present in the genetic signal.

While not a replacement for genetic data, simulations can help fill some of these gaps. In particular, agent-based modelling (also referred to as individual-based modelling) can be used to model individual agents while taking into account the landscape and the species' life history (Shirk et al. 2012). By allowing the researcher to run the model with various landscape scenarios, these models can provide inference on the possible genetic effects of landscape change (Aben et al. 2016; Gauffre et al. 2008;

Thatte et al. 2018), which is especially valuable for conservation purposes as the models can be used to predict how genetic connectivity might change under different development scenarios.

However, an obvious challenge in using agent-based models is the potential for unrealistic results, which is of particular concern when the model will influence management decisions – incorrect inferences drawn from a simulation model could lead to ineffective or, at worst, harmful conservation actions. It is therefore critical that models be evaluated to ensure that they lead to reliable inferences.

Model performance can be evaluated by running the model on a real-world system and then comparing the results to empirical data, allowing researchers to adjust the model to produce results in line with observed data (Augusiak et al. 2014; Shirk et al. 2012). While a model that reproduces real-world patterns is not guaranteed to be structurally correct, it is still an important step in the modelling process, especially if the model will be used to make predictions into the future. When applied to agent-based modelling, this approach is called pattern-oriented modelling, a framework in which complex models are refined by tuning parameters and mechanisms to produce a model that closely reproduces real-world patterns (Grimm et al. 2005; Grimm & Railsback 2012). In addition to testing model performance, running a model with a variety of parameters and mechanisms helps the researcher to better understand how certain factors influence the model output.

In this chapter I describe an agent-based model for predicting genetic connectivity of the Mojave desert tortoise in the Ivanpah Valley area. If the model's

performance is considered acceptable, the model can be run on a larger scale to assess range-wide genetic connectivity (see Chapter 3). Therefore, the goal of this chapter is to assess the performance of the model and inform the selection of parameter values. I use previously published genetic data for this area (Dutcher et al. 2020) to assess how well the model predicts observed genetic patterns on the landscape. In addition, I use a range of parameter values to assess the sensitivity of the model and to find a parameter set that produces accurate results.

Review of tortoise movement

Before describing the methods used, it is valuable to briefly discuss Mojave desert tortoise movement and how it impacts the species' landscape genetics. The spatial genetic pattern of tortoises is influenced by both small-scale and large-scale movements. Small-scale tortoise movement (e.g. at a daily time scale) reflect behaviors like feeding and mating (Bulova 1994; Ruby et al. 1994). These movements typically take place within an established home range that remains largely constant from year to year (Freilich et al. 2000; Harless et al. 2009; Lovich et al. 2018; Nussear et al. 2012) and allow tortoises with overlapping home ranges to encounter and mate with each other, resulting in genetic exchange between these nearby tortoises. In addition to shorter movements within a home range, tortoises are also known to occasionally make longer-distance dispersal movements (Berry 1986; Nussear et al. 2012). Longer-distance dispersal movements also contribute to genetic connectivity and can result in more rapid genetic interchange.

Relative to the overall range of the desert tortoise, even long-distance dispersal movements of desert tortoises can only cover a small portion of the range, meaning that direct gene flow is limited to populations within the tortoises' maximum dispersal distance. At a landscape level, these types of movements can lead to genetic connectivity that emerges not from direct genetic exchange between populations but rather at a multi-generational time scale (Boulanger et al. 2020; Dutcher et al. 2020). Genetic exchange limited by distance leads to a pattern called isolation-by-distance where nearby tortoises tend to be more genetically similar than distant tortoises, and previous genetic studies have indeed detected a pattern of isolation-by-distance for the Mojave desert tortoise (Dutcher et al. 2020; Hagerty et al. 2011; Murphy et al. 2007; Sánchez-Ramírez et al. 2018). When habitat is largely continuous, this results in a gradient of genetic differentiation across the landscape with no clear boundaries. In addition to distance, landscape structure can also shape spatial genetic patterns. Tortoise movement can be restricted by natural features like mountains and rivers as well as human features like roads and urban areas (Dutcher et al. 2020). Restricted movement leads to reduced gene flow, and two populations with little to no gene flow are likely to become more genetically different over time (Landguth et al. 2010).

The goal of the simulation presented here is to accurately simulate desert tortoise movement, as well as other processes like reproduction and mortality, to produce realistic predictions of the spatial genetic pattern. Comparing the simulated genetic pattern to the observed genetic pattern allows for evaluation of how realistic the simulation results are.

Methods

Study area and data collection

Genetic data were collected at eleven sites located in the central Mojave desert. Most of the sites were located in the Ivanpah Valley, with other sites located in the Mesquite and Piute-Eldorado Valleys (see Figure 11). A total of 299 genetic samples were collected, with varying numbers of samples per site. For each sample, 20 variable microsatellite loci were amplified (Edwards et al. 2003; Hagerty et al. 2008; Schwartz et al. 2003). See Dutcher et al. 2020 for more details.

Landscape representation

Movement resistance

An updated version of the habitat model described in a previous study was used to represent the landscape (Nussear et al. 2009). The habitat model uses variables related to topography, soil characteristics, climate, and plant cover to produce a raster of habitat suitability values ranging from zero to one, where zero indicates poor habitat and one indicates high quality habitat. I cropped the surface down to an extent that included all 11 plots using a set of contiguous watershed polygons (U.S. Geological Survey 2015). The resulting raster was scaled up to a 1km cell size from the original resolution of 250m. Initial results showed that simulations on the 1km surface produced

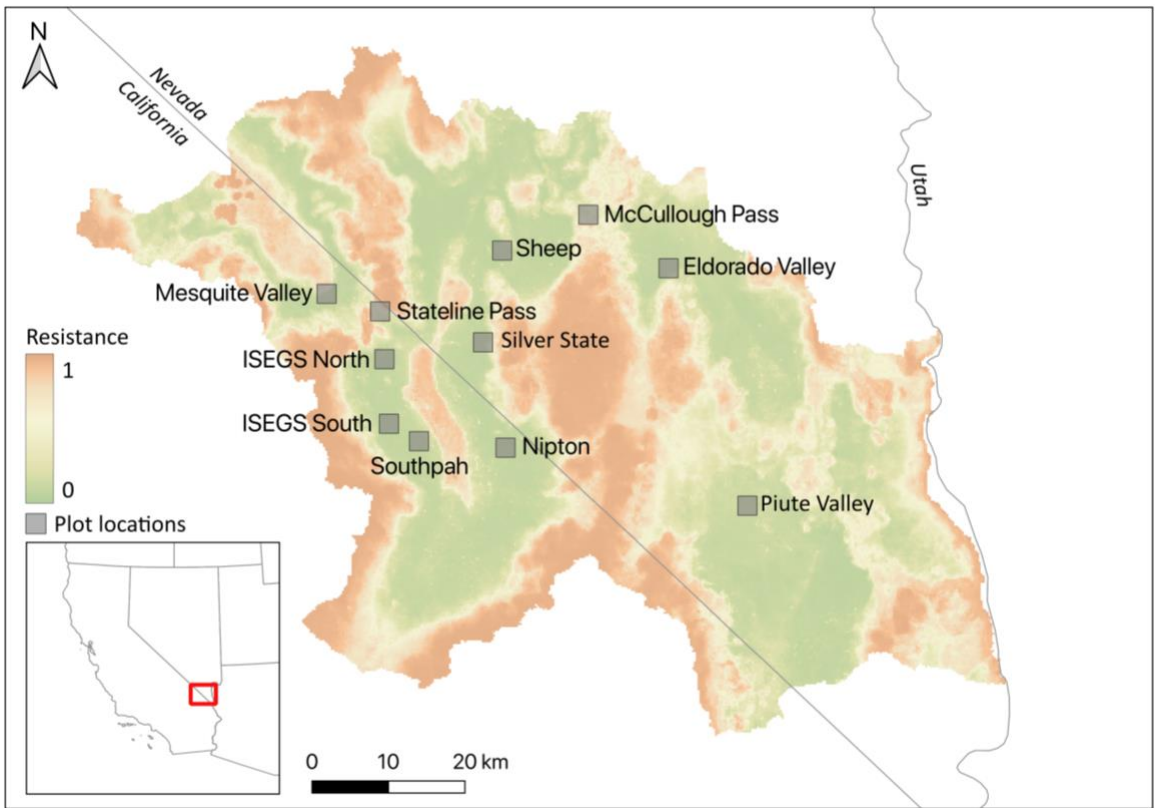

Figure 11 – The location of the 11 sites overlaid on the resistance surface used for the simulation.

very similar results as those performed on the 250m surface while running considerably faster, so the 1km surface was used.

The movement algorithm used in the simulation requires that the raster represent resistance to movement rather than habitat suitability. One approach towards creating movement resistance rasters is by taking the inverse of a habitat suitability raster (Beier et al. 2008; Chetkiewicz et al. 2006; Hagerty et al. 2011) – following this approach, I inverted the habitat suitability raster by subtracting the cell values from one, resulting in values between zero and one, where zero represents low resistance to movement and one represents high resistance to movement.

It is important to note, however, that this approach makes some important assumptions that have been criticized. Habitat suitability models represent the likelihood of occurrence, and the variables that influence occurrence may not be the same as those that influence movement (Keeley et al. 2017; Mateo-Sánchez et al. 2015; Zeller et al. 2012). For example, many animals are able to easily travel over areas that are not habitable, which violates the assumption inherent in the approach used here that areas of bad habitat are more difficult to travel through.

That being said, the validity of this approach depends to a large degree on the characteristics of the species in question. The desert tortoise is considered to be a "corridor dweller" – due to its limited dispersal capabilities it generally can only move through areas that also offer suitable habitat for longer-term survival (Averill-Murray et al. 2013; Beier & Loe 1992). Because of these characteristics, it is more likely that the inverse of habitat suitability represents movement resistance than it would in other, more dispersal-capable species (Keeley et al. 2017). A recent study found the same habitat model used here to be a significant predictor of fine-scale tortoise movement (Hromada et al. 2020), suggesting that the habitat model at least somewhat corresponds with movement suitability.

Quadtree representation

To improve processing time, the movement resistance surface was converted to a quadtree representation (see Figure 12 and Chapter 1). The quadtree data structure allows for variable cell sizes, and the quadtree was constructed by dividing quadrants

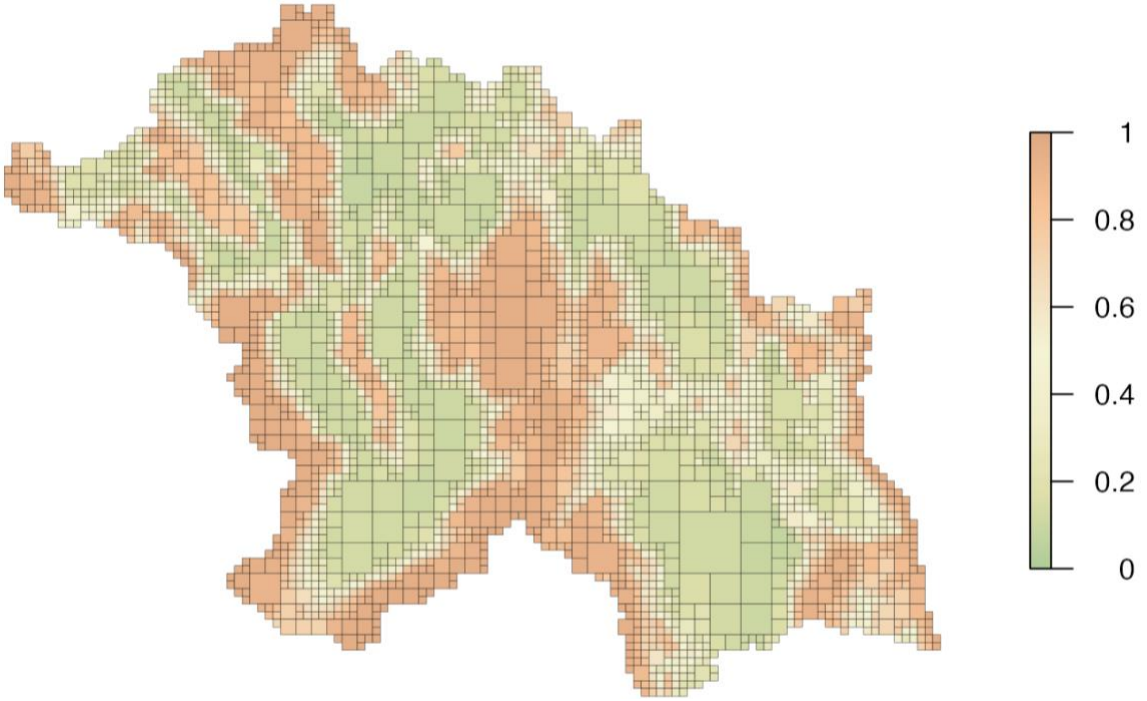

Figure 12 – The quadtree created from the original landscape raster. Cell sizes range from 1km to 8km.

into smaller cells when the range between the minimum and maximum values exceeded 0.1. If the values of all cells within a quadrant fell within a range of 0.1, the cells were combined into a single, larger cell and the mean was used as the value for that cell. The threshold of 0.1 was chosen because the resulting quadtree had some areas with larger cells while still maintaining the landscape structure present in the original raster. Because the original raster used had a resolution of 1km, quadtree cells ranged from 1km to 8km. Using variable cell sizes allows for faster computation of movement by limiting small cell sizes (which result in higher computation times) to areas with a higher level of heterogeneity (see Chapter 1).

Simulation

An individual-based model was used to simulate individual tortoise life-histories and genetics. The model was coded in C++, and analysis of model output was done in R (R Core Team 2022). The model runs on a yearly basis and consists of three main processes: movement, reproduction, and mortality (Figure 13).

Movement

The movement algorithm operates by moving a tortoise some amount of cost-distance. When a segment is contained within a single cell, the cost-distance of that segment is defined as:

$$distance + distance \times cost$$

When a path passes through multiple cells, the distance traveled through each cell is calculated. The formula is used to calculate the cost-distance for each segment that passes through a different cell, and the results are summed to get the total cost-distance of the path.

Every iteration (year) a tortoise has the possibility of moving to a new location. For each tortoise, a random distance is drawn from a pre-defined distribution, which is considered to be the maximum cumulative cost-distance that the tortoise can travel. Next, a random angle is selected and is used with the chosen distance to define the "destination point". If the cell containing the destination point has a value (i.e. is not NA), a path from the initial point to the destination point is calculated, with three situations possible:

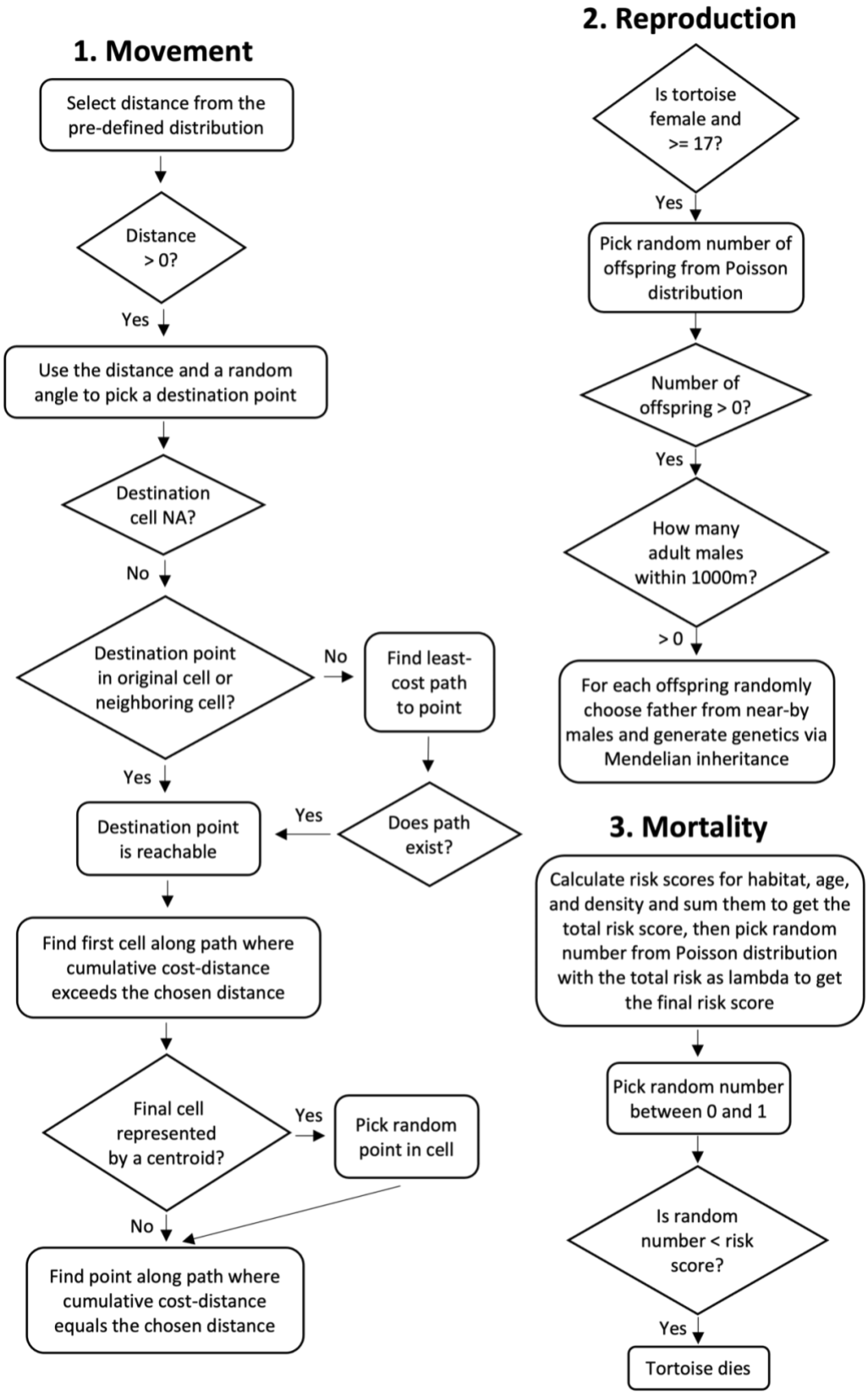

Figure 13 – Flow charts depicting the movement, reproduction, and mortality processes.

1. The destination point falls within the same cell as the initial point. In this case, the path simply consists of these two points. The point along this path where the maximum cost-distance is reached is used as the new location of the tortoise (see Figure 14).
2. The destination point falls in a neighboring cell. If a straight line between the two points would pass through a third cell, the path is adjusted so that it travels through the closest corner at which the cells meet, thereby creating the shortest path that doesn't travel through a third cell (see Figure 14). After the path has been determined, the point along this path where the cumulative cost distance is reached is used as the new location of the tortoise.
3. The destination point is not in the initial cell or a neighboring cell. In this case, the least-cost path algorithm is used on the quadtree (see Chapter 1) to find the least-cost path from the tortoise's current location to the destination point. The algorithm uses a single point to represent each cell – the initial point and the destination point are used to represent their respective cells, while all other cells are represented by their cell centroid. After the path has been calculated using the least-cost path algorithm, the first point along the path where the maximum cost-distance is exceeded is found. If this point is not the destination point, then it is a cell centroid; to avoid agents congregating along a straight line between

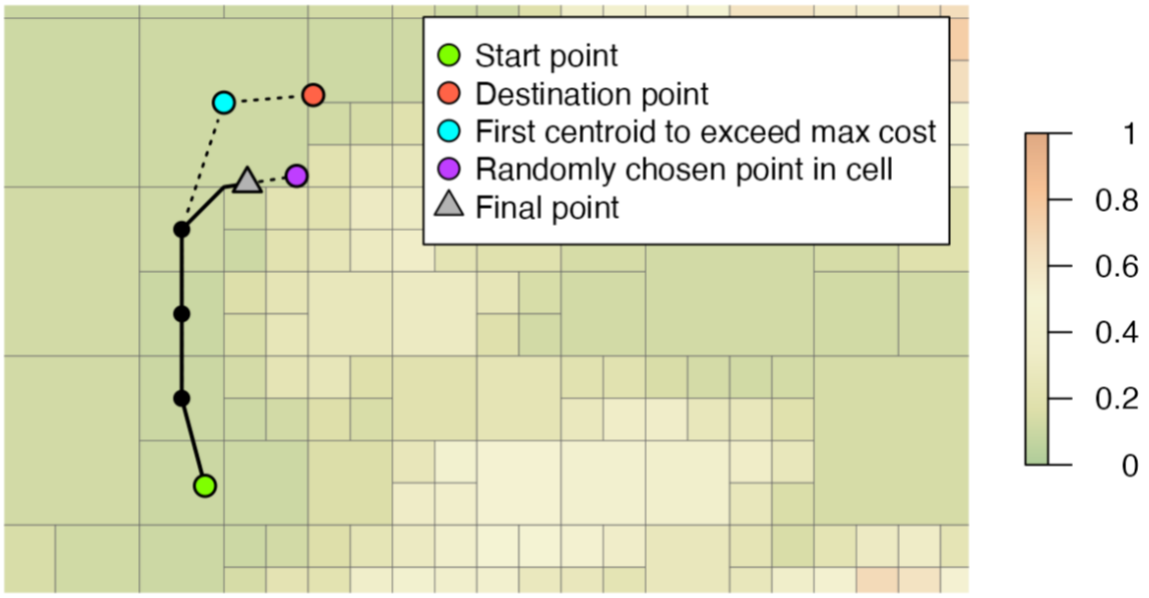

Figure 14 – An example of the movement algorithm. The last segment of the path (the segment that contains the final point) demonstrates the method used to adjust a path so it only travels through two cells – this method is used in both the second and third situations described in the text. It also demonstrates the process of selecting the point on the segment at which the maximum allowed cost-distance is reached – this method is used in all three situations.

centroids, a random point is selected within this cell. The path between this point and the previous point is found. If the path passes through a third cell, the same procedure described in situation 2 is used to adjust the path, and then the point along the path where the maximum cost-distance is reached is calculated and used as the tortoise's new location. If the randomly chosen point is closer to the previous point than the centroid, it is possible that the new point may be beyond the chosen point. Very rarely this can result in the final point being in a different cell – in this case, the randomly chosen point is used as the tortoise's new location. See Figure 14 for an illustration of the method.

The movement algorithm is split into these three cases because finding the least-cost path is a computationally expensive operation – only using the algorithm when the destination is more than one cell away (situation 3) helps speed up model computation.

Reproduction

Every year adult female tortoises have the possibility of reproducing. Adults are defined as tortoises that are 17 years or older, which falls within the range of predicted sexual maturity reported by previous studies (Curtin et al. 2009; Medica et al. 2013). When one or more adult males are within 1000m of a female, that female reproduces. A random number of offspring is drawn from a Poisson distribution with a lambda of six, which approximates the number of eggs laid per year by desert tortoises (Mitchell et al. 2021). Desert tortoises are known to have multiple paternity (Davy et al. 2011), so the father of each offspring is chosen randomly (with replacement) from the set of males within the mating distance of the female. The genetics of each offspring are created by randomly selecting one allele from each parent for each of the 20 loci.

In some simulations, a mutation rate is specified. When a tortoise is born, the mutation rate is applied to each locus – if a locus is chosen for mutation, one of the two alleles is chosen. Mutation is modelled in a stepwise fashion, such that one is added to or subtracted from the current value for an allele.

Mortality

At the end of the year, each tortoise has some probability of dying that is based on three factors: age, local tortoise density, and habitat value. For each of these three

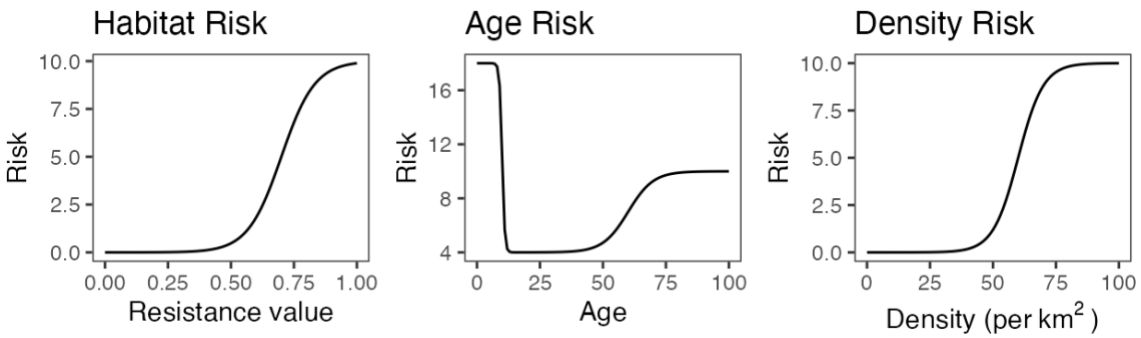

Figure 15 – The functions used to determine risk for habitat, age, and density.

factors a sigmoidal function is defined for calculating the amount of "risk" assigned to each tortoise (Figure 15 and Table 1).

Mortality risk for tortoises is not equal across age classes. Juvenile tortoises are known to have high mortality rates (Nagy et al. 2015), so tortoises less than 10 years old are assigned high risk. The risk decreases for adult tortoises, and then starts increasing once the tortoises are approximately 50 years old.

Habitat-related variables have been shown to be correlated with tortoise density and mortality rates (Andersen et al. 2000; Berry et al. 2013; Esque et al. 2010; Longshore et al. 2003; Nafus et al. 2017). Therefore, the value of the habitat model at a tortoise's location impacts its probability of dying. Tortoises in high quality habitat (represented by low resistance values) have a lower probability of dying while those in low quality habitat (high resistance values) have a higher probability of dying.

Finally, to prevent exponential growth, tortoises in areas with a higher local tortoise density are considered to be at a greater risk of dying. A grid of 1000m cells is overlaid over the area and the number of tortoises in each cell is calculated. The

number of tortoises in a cell is used to calculate the density risk for each tortoise within that cell.

The risk determined by these three functions is then summed for each tortoise to get the total mortality risk. The final risk score for a tortoise is determined by using a random Poisson draw with the risk score as lambda – the resulting number is divided by 100 and treated as the tortoise's probability of dying. A random number between zero and one is drawn from a uniform distribution, and if it is less than the final risk score the tortoise is considered to have died.

Table 1 – Parameter values that were constant for all simulations.

Parameter	Value
Number of years	25,000
Number of replications	10
Mating age	17
Number of offspring (lambda of Poisson distribution)	6
Mating distance	1000m
Maximum age	100
Age risk score functions	juveniles: $14 / (1 + e^{2 * (x - 10)})$ adults: $6 / (1 + e^{0.2 * (x - 60)})$
Habitat risk score function	$10 / (1 + e^{15 * ((1 - x) - 0.3)})$
Density risk score function	$10 / (1 + e^{0.2 * (x - 60)})$

Parameter calibration

In order to test the sensitivity of the model, I adjusted the values of certain parameters to observe how this affected the model output. Ideally, I would use a full factorial design, in which all possible parameter combinations are tested (Thiele et al. 2015), but because of the long run time of the model, this is not realistic. Therefore, I

chose to test three parameters that are likely to affect the model output: movement distance, the resistance surface, and mutation.

Movement distance

One of the key factors influencing genetic patterns on a landscape is the dispersal behavior of an organism. In order to explore the influence of dispersal distance on the output of the model, I ran the simulation with different distributions used for the distance parameter.

The original distribution of dispersal distances was based on the yearly home range shift of resident radio-telemetered tortoises tracked for multiple years at eight sites in the Mojave Desert (previously reported in Nussear et al. 2012; Drake et al. 2012, 2015; Sah et al. 2016; Hromada et al. 2020). The data was grouped into 100m bins and the frequency of occurrence for each bin was used as the probability of that distance being selected for a given tortoise.

In order to test different distance distributions, I modified the base distribution by multiplying the binned distance by some constant while keeping the probabilities the same. The values were then interpolated at 100m increments and the maximum distance was cut off at 3100m, the same as in the original data. Finally, the probabilities were scaled to sum to one. The five different distributions used can be found in Figure 16.

There is one aspect of the movement algorithm that is worth noting since it affects the interpretation of the movement distance parameter. The distance parameter

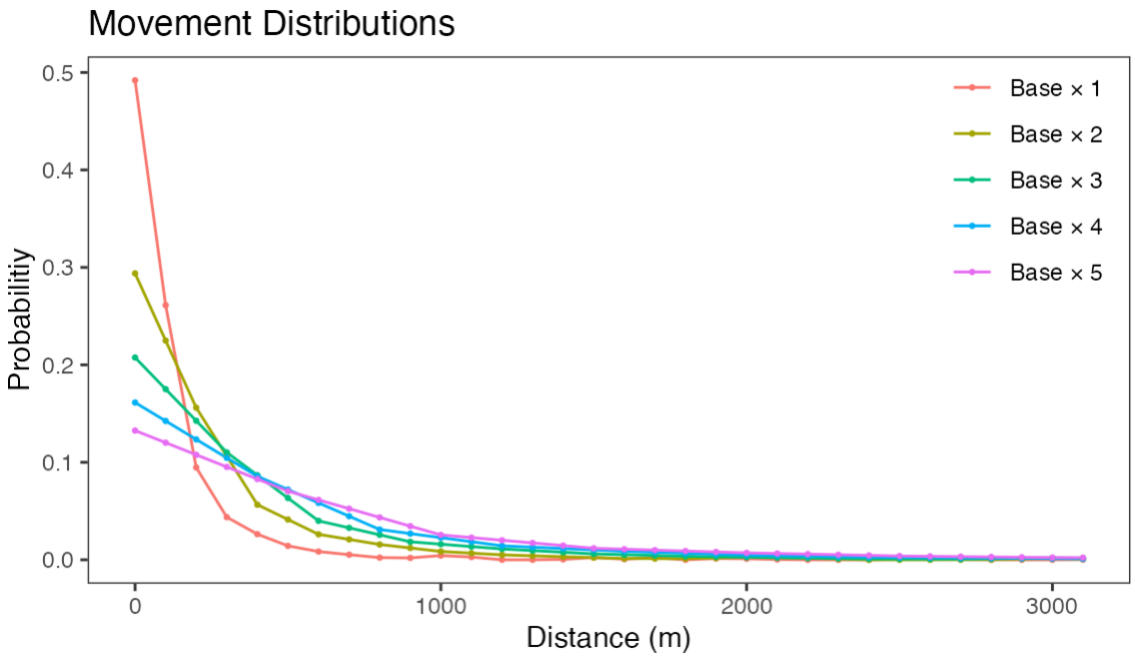

Figure 16 – The five movement distributions used in the simulations.

is actually interpreted as "cost-distance" rather than as distance alone. For example, if a distance of 2000m is selected from the provided distribution, the tortoise will move 2000 cost-distance units as defined in the formula previously discussed. Unless the path passes only through cells with a resistance of 0, the tortoise will travel less than 2000m. Therefore, on the habitat surface, the resulting distribution of simulated movement distances will always have shorter distances than the distribution used as the input parameter.

Landscape

In landscape genetics it is common to test for a genetic pattern of isolation-by-distance (IBD), which occurs when the only factor influencing genetic connectivity is the distance between populations (i.e. landscape structure has no influence) (Wright 1943). In order to explore the outcome of the IBD hypothesis on this landscape, I conducted

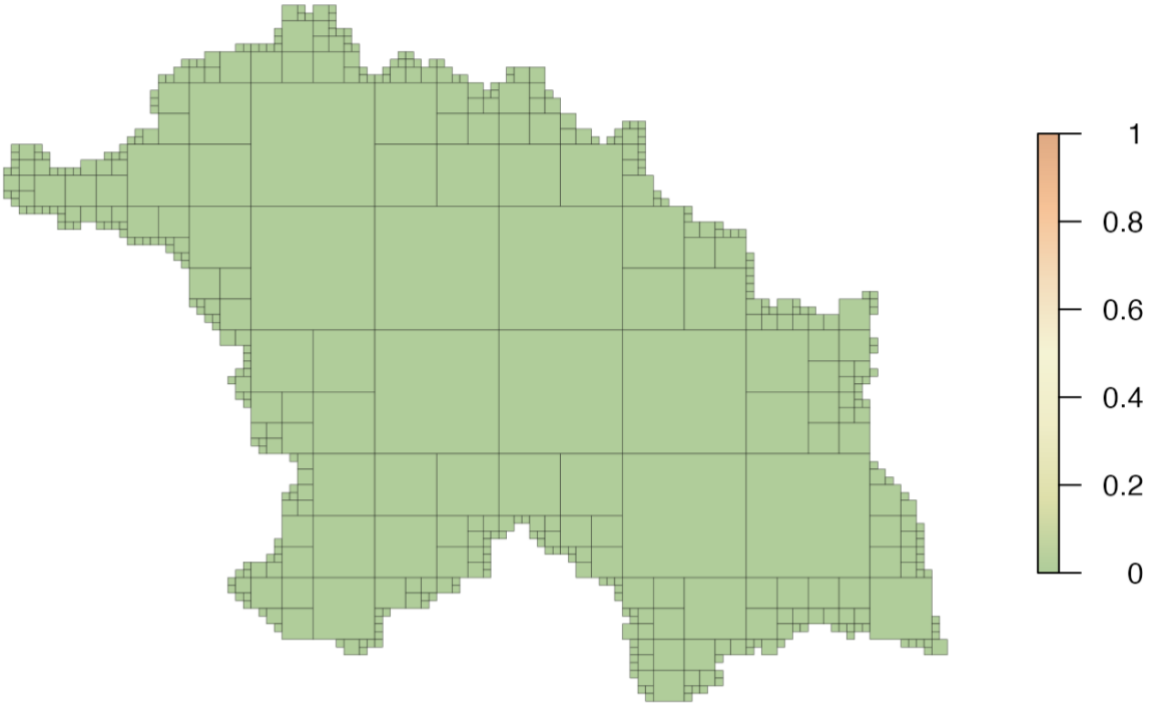

Figure 17 – The neutral landscape.

simulations on a neutral surface in addition to the habitat surface described above. This surface had the same extent and all cells had a resistance of zero, indicating unrestricted movement across the landscape (Figure 17).

Mutation

To assess the impact of mutation on the results, I varied the mutation rate while keeping the other parameters constant. The habitat surface and the "base $\times$ 4" distance parameter were used in conjunction with mutation rates of 0.0001, 0.0003, and 0.0005. The mutation rate of 0.0001 reflects the upper limit of the range of estimated mutation rates for desert tortoises (0.00001 – 0.0001, Edwards et al. 2015) and uses 0.0005 as the maximum mutation rate since this has been used in previous simulations of desert tortoises (Dutcher et al. 2019).

Table 2 – The parameter sets used for the simulations. “Base” refers to the original distances calculated from the home-range shift data.

Landscape	Distance	Mutation rate
Habitat	Base × 1	0
Habitat	Base × 2	0
Habitat	Base × 3	0
Habitat	Base × 4	0
Habitat	Base × 4	0.0001
Habitat	Base × 4	0.0003
Habitat	Base × 4	0.0005
Habitat	Base × 5	0
Neutral	Base × 1	0
Neutral	Base × 2	0
Neutral	Base × 3	0

Initialization and simulation runs

A burn-in step was used to create the set of initial agents used in each simulation. First, 200,000 tortoises were randomly placed on the landscape. A preliminary simulation was then run with the chosen parameters until the population size stabilized. The final set of agents was used as the initial state for the simulation.

After the initial set of tortoises had been generated, randomly generated genetic data were assigned to each tortoise. To initialize the genetics, the allele frequencies for each of the 20 loci from the real data were calculated, and then these frequencies were used to randomly assign alleles to each of the tortoises, resulting in a panmictic population.

Each simulation was run for 25,000 years with 10 repetitions.

Output analysis

Basic metrics

Basic metrics like the total number of tortoises present in the simulation were tracked and then compared among simulations, and tortoise density of the final year of the simulation was calculated using a quadrat count with 1km $\times$ 1km cells. These quadrat counts were performed for all tortoises regardless of age as well as for adults only (those 17 years or older).

Genetic metrics

To evaluate genetic distance between the 11 sites, I calculated the Weir and Cockerham pairwise F_{ST} for all pairs of sites (Weir & Cockerham 1984). In addition, I calculated the overall F_{ST}, the expected and observed heterozygosity, and the number of alleles present per locus. These metrics were calculated for both the real data and the simulated data; for the simulated data, these metrics were calculated every 100 years.

Comparison with empirical data

To assess how well the simulated genetic pattern matched the actual genetic pattern, I calculated the correlation between the empirical pairwise F_{ST} values and the simulated pairwise F_{ST} values. In addition, I used Mantel tests to test for a significant relationship between the simulated pairwise values and the empirical values and reported the resulting *p*-values (Shirk et al. 2012).

Correlation can identify a relationship between two variables but does not check for similarity in the magnitude of the values. To evaluate how closely the values from the simulated data matched the actual data, I calculated the root mean squared error (RMSE) between the simulated pairwise F_{ST} and the observed pairwise F_{ST}. I also

compared the empirical and simulated values of the overall F_{ST} and the expected and observed heterozygosity (H_e and H_o, respectively).

Simulation code and computation

The simulation was written in C++ and run on an 80-core Linux machine with 377GB of RAM. Multiple simulation runs were run in parallel, and computation time for a single run ranged from 26-51 hours. All data analysis was performed in R (R Core Team 2022).

Results

Simulation output

Most simulations resulted in similar demographic results – Figure 18 shows results for a single repetition of one parameter set. Simulations performed on the neutral surface had significantly larger populations, which is as expected given that the entire landscape consists of suitable habitat.

In the habitat simulations, population size was around 175,000 tortoises, with variation between years. The vast majority of the tortoises were juvenile – in all simulations, adults (defined as those 17 years or older) made up about 10% of the population. When all tortoises were considered, tortoise density ranged from 0 to 110 tortoises per km^2, with most densities falling between 30 and 60 tortoises per km^2. When only adults were considered, densities ranged from about 0 to 13 tortoises per

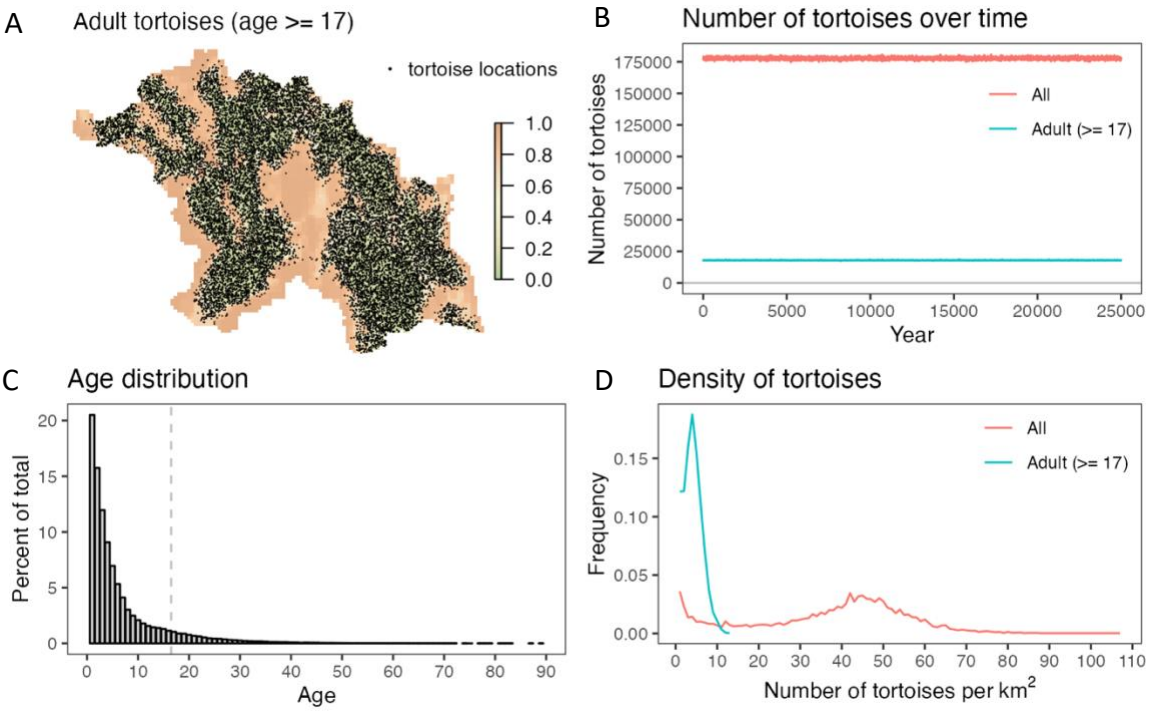

Figure 18 – Summary of demographic information from the "base × 3" simulation with no mutation at year 25,000. (A) The adult tortoises on the landscape. (B) Number of tortoises over time. Note that the number of adults is not recorded for each year and is instead estimated using the percentage of adults in the final year. (C) The age distribution of year 25,000. The vertical dashed line is at age 17, which is when tortoises are considered "adults" in this simulation.(D) The density of tortoises. Density was calculated using a quadrat count with 1km x 1km cells. The plot shows the frequency of each number of tortoises among the quadrat cells. Cells with a count of 0 were excluded from the plot.

km^2; when zero densities were excluded, about 4 adult tortoises per km^2 was the most common density value.

Genetic metrics

All simulations showed an increase in the overall F_{ST}, although the rate of increase varied by simulation. In addition, the observed and expected heterozygosity declined over time in the simulations without mutation (Figure 19). In the simulations with mutation, however, the observed and expected heterozygosity either remained steady or increased (Figure 20). A plot showing the variability in metrics for a single

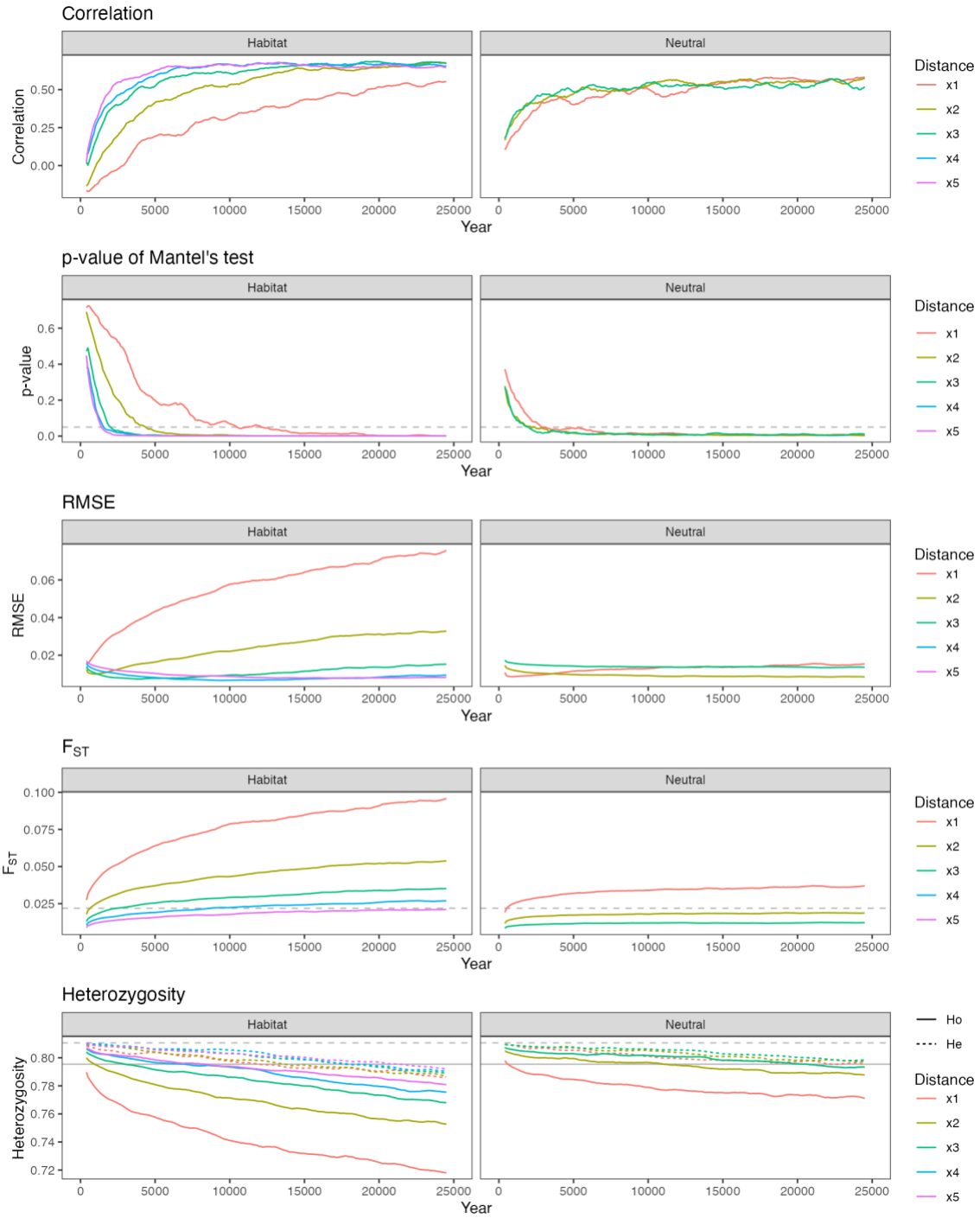

Figure 19 – Results for the simulation runs, comparing between movement distances and landscape surfaces. All simulations displayed here had no mutation. For each statistic, the mean of the 10 replications was calculated for each year and then a 10 data point (1000 year) rolling average was applied to smooth the mean. In the F_{ST} plot and the heterozygosity plot, the horizontal lines represent the empirical value. In the heterozygosity plot, the dashed line represents H_e (expected heterozygosity) while the solid line represents H_o (observed heterozygosity).

parameter set can be found in Figure 21.

Comparison with empirical data

All models showed a correlation near zero at the beginning of the simulation, which is as expected, given that the genetics were randomly initialized with no pre-existing pattern on the landscape. Over time, all simulations showed a pattern in which the correlation initially increased rapidly and then leveled off at a stable value. However, the rate at which the simulations reached a stable value varied across the different scenarios. All models had a correlation between 0.55 and 0.65 by the end of the run.

By year 25,000, all simulations had a significant *p*-value from the Mantel test (at the 0.05 significance level). However, as with the correlation, the rate of convergence varied between simulations.

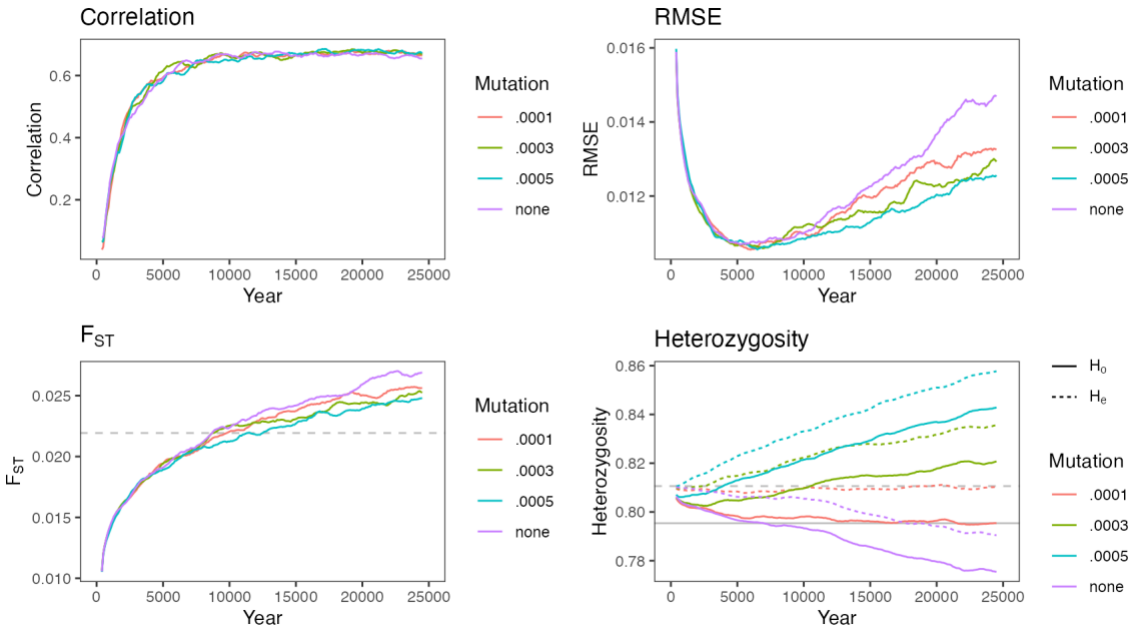

Figure 20 – Plots comparing the results for different mutation rates. For each statistic, the mean of the 10 replications was calculated for each year and then a 10 data point (1000 year) rolling average was applied to smooth the mean. The simulations were performed on the habitat surface with the "base × 3" distance parameter – the only parameter changed was the mutation rate. The horizontal lines in the F_{ST} and heterozygosity plots represent the empirical values.

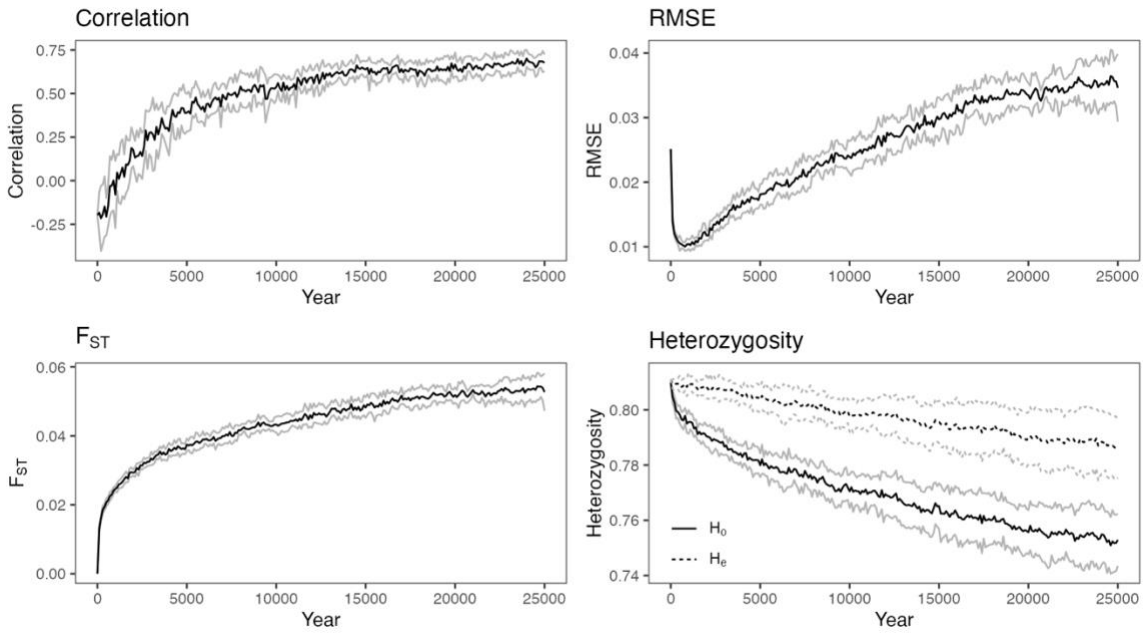

Figure 21 – Plots showing the variation in the correlation, RMSE, F_{ST}, and heterozygosity for a single parameter set (distance of base × 2, a mutation rate of 0, and using the habitat surface). The black lines represent the means across all 10 repetitions while the gray lines indicate +/- one standard deviation from the mean.

While all models had high correlations and significant Mantel tests, the RMSE varied significantly by scenario. The RMSE increases when a simulation either overestimates or underestimates the observed values.

Effects of specific parameters

Effect of the distance parameter

The movement distance appears to have little influence on the correlation with the observed F_{ST} values. Simulations using the same resistance surface but different distance parameters all converged on a similar correlation value. However, the movement distance affected the time it took for the correlation values to converge – scenarios with shorter movements took longer to reach a stable value. For the

simulations performed on the habitat surface, the simulations with the longest dispersal distance appear to converge by year 5000, while the simulations using the shortest dispersal distance had still not converged by year 25,000. A similar pattern occurs in the Mantel tests, where simulations with shorter movement distances reach a significant *p*-value more quickly (Figure 19).

However, while the final value of the correlation is similar across all distance parameters used, the genetic connectivity is greatly influenced by the distance parameter. This trend is clearly evidenced in several of the metrics. In particular, the F_{ST} shows a clear pattern – simulations with shorter movement distances have higher F_{ST} values, indicating more genetic differentiation (Figure 19, Figure 22). In addition, the heterozygosity showed a relationship with movement distance – while all simulations without mutation showed a decrease in both observed and expected heterozygosity, simulations with shorter movement distances showed a more pronounced decline.

The effect of movement on genetic connectivity can also be observed in the RMSE – because simulations with shorter distances had higher pairwise F_{ST} values, they

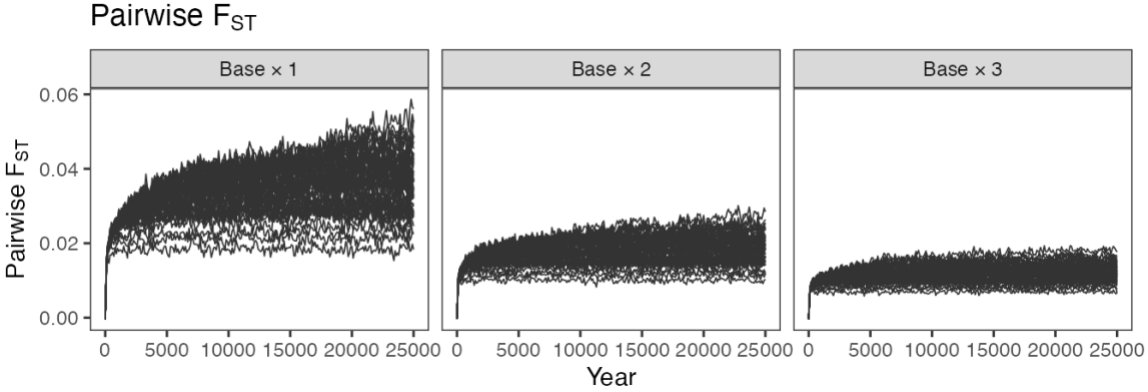

Figure 22 – The pairwise F_{ST} values over time for three different distance parameters, all on the neutral surface. Each line represents the pairwise F_{ST} between one pair of plots

greatly exceed the empirical values, resulting in a large RMSE. Longer distances resulted in lower F_{ST} values – as the movement distance increased, the simulations tended to underestimate the empirical values, which also led to higher RMSE values. For example, among the simulations on the neutral landscape, "base × 1" and "base × 3" both had higher RMSE values than "base × 2" (Figure 19).

Effect of the resistance surface

The resistance surface used appears to have minimal effect on the correlation with the real data. That being said, simulations using the habitat surface appear to have slightly, but not significantly, higher correlations than simulations using the neutral landscape (Figure 19). In addition, the correlations for simulations on the neutral landscape converge more quickly than their counterparts on the habitat landscape. A similar pattern holds for the Mantel tests, where the neutral landscape simulations reach significance more quickly than the habitat simulations.

Overall F_{ST} and H_o both show signs of increased connectivity when compared to the habitat landscape – F_{ST} is lower and H_o is higher than simulations run on the habitat landscape, a trend that is also reflected in the RMSE. For example, for "base × 1", the RMSE of the neutral landscape simulation is lower than the RMSE for the habitat simulation (Figure 19).

Effect of mutation

The mutation rate had no effect on the correlation (Figure 20). However, the F_{ST} and RMSE appear to show slight differences between the different mutation rates, with

higher mutation rates having lower F_{ST} and RMSE. Of the metrics tracked here, mutation appeared to have the most noticeable influence on heterozygosity. In the absence of mutation, both expected and observed heterozygosity showed a steady downward trend. As the mutation rate increased, this effect was reversed – a mutation rate of 0.0001 resulted in steady values for the heterozygosity, while 0.0003 and 0.0005 showed an increase over time.

Discussion

Impact of movement distance on genetic patterns

In general, shorter dispersal distances can cause both a decrease in heterozygosity and an increase in genetic differentiation between populations (Furstenau & Cartwright 2016; Jaquiéry et al. 2011). In addition, studies have found that the rate of approach to an equilibrium is longer when movements distances are short (Blair et al. 2012; Epps & Keyghobadi 2015; Landguth et al. 2010; Wang 1997). These trends are consistent with the results of the simulations presented here – longer movement distances resulted in a decrease in H_o, an increase in pairwise F_{ST}, and a slower approach to equilibrium (Figure 19, Figure 22).

Impact of mutation rate on genetic patterns

In the absence of mutation, genetic diversity is expected to decrease due to genetic drift (Allendorf et al. 2012; Varvio et al. 1986), resulting in a loss of heterozygosity. As mutation rates increase, heterozygosity increases and genetic

differentiation, as measured by F_{ST}, decreases (Hedrick 1999). In addition, higher mutation rates tend to cause a faster approach toward an equilibrium (Epps & Keyghobadi 2015).

The results presented here largely follow these theoretical expectations. Most notably, an increase in the mutation rate causes a noticeable increase in H_o. In addition, the results appear to follow the expectation that F_{ST} decreases with increased mutation, as visible in Figure 20, although the difference between the four simulations appears to be fairly small. However, the rate of convergence does not appear to vary among the different mutation rates. One potential cause for why this trend is not visible in the results is that the simulations may not have been run long enough for an equilibrium to be reached – Figure 20 shows that the simulations with mutation rates of 0.0003 and 0.0005 are still showing a steady increase in H_o at the end of 25,000 years. Therefore, it is possible that if the model were run for a longer period of time, H_o would reach an equilibrium.

Habitat surface vs. neutral surface

As noted in the results, there were differences in the simulations run on the habitat surface and the simulations run on the neutral surface. The neutral landscape likely affects the results in two possible ways. First, because of how the movement algorithm works, a tortoise can move farther when it is in low resistance habitat. The neutral surface offers no additional resistance beyond the distance between points, which means that on average the movements on the neutral surface will be longer than

the movements on the habitat surface. Therefore, even if simulations on the two surfaces use the same movement parameter, the actual movement distances in the neutral surface simulation will be longer than the habitat simulation. Second, the neutral landscape has no landscape structure to impede movement like the habitat surface does, likely resulting in increased connectivity.

The differences between simulations run on the neutral and habitat surfaces appear similar to the effects of the distance parameter described previously, likely because of the increased movement distances occurring on the neutral surface. However, the lack of landscape structure in the neutral landscape could also be having an impact on the results, which makes it difficult to parse out the exact cause for the differences between the simulations run on the two surfaces.

Overall sensitivity and performance of the model

Overall, it is clear that movement distance and mutation rates have an impact on the results, although in different ways. It is difficult to determine if the landscape configuration significantly changes the results, since the neutral landscape results in increased movement distances in addition to landscape structure.

However, despite the impacts that these parameters have on the output, the correlation remains largely the same across all simulations. This suggests that although the distance and mutation parameters change certain characteristics of the output, the overall genetic pattern resulting from the simulations remains largely the same. Ultimately, all simulations show good correspondence with the empirical data, as

demonstrated by the high correlation values and significant *p*-values from the Mantel tests, and therefore provide evidence that the simulation is able to reproduce real-world patterns.

Selection of parameters for use in future simulations

In the absence of a clear best parameter set using the correlation, other metrics can be used to distinguish between parameter sets – indeed, using multiple patterns to calibrate a model can help make a model more "structurally realistic" (Grimm et al. 2005). Even when using multiple metrics, however, selecting a single parameter set is challenging because of the ways the parameters can interact with each other. For example, both mutation and movement distance influence the heterozygosity (Figure 19, Figure 20). Therefore, it is likely that there are multiple combinations of these two parameters that would result in a rate of heterozygosity that matches the empirical value, complicating the selection of a "best" value for each parameter. In addition, it's possible that the different metrics do not uniformly suggest that one value is best. All of this makes selecting a value for each parameter a challenging process.

Movement distance

At the end of 25,000 years, "base $\times$ 5" produces the overall F_{ST} value that is most similar to the observed value as well as the lowest RMSE, meaning that it may be justified to consider this distance parameter to be the "best". However, there are several caveats that complicate this logic. As previously mentioned, the results presented here show that multiple parameters can influence the same genetic metric.

Indeed, the results shown in Figure 19 were all run with no mutation, and an increase in mutation rate appears to cause a slight but noticeable decrease in overall F_{ST} (Figure 20). Therefore, it's reasonable to predict that including mutation in the simulations presented in Figure 19 may change which distance parameter is deemed the "best".

The problem is made more challenging because only a few parameters were adjusted in this analysis due to computational limitations, and other parameters not adjusted here will almost certainly have some influence on the results as well. For example, a parameter like the mating radius will likely impact genetic connectivity in some way, which, if tested, would introduce yet another factor that could influence the model output. Therefore, it is entirely possible that a certain parameter value (like "base $\times$ 5") could appear to be the "best" value not because it most accurately reflects the real-life movement distribution but because it interacts with the other parameters in a way that produces a value similar to that observed in empirical data.

Another complication is that the values for the F_{ST} and H_o are still increasing after 25,000 years. For example, while "base $\times$ 5" produces the overall F_{ST} value closest to the empirical value and the lowest RMSE at year 25,000, it is not clear from the plots in Figure 19 if an equilibrium has yet been reached. If run for longer, it is possible that the overall F_{ST} would continue to increase and therefore exceed the observed value. In fact, the "best" parameter will vary depending on how long the simulation is run – given the fairly arbitrary selection of 25,000 years, there is no reason to assume that whichever parameter produces the best results at year 25,000 is therefore the best.

Unfortunately, when these issues are all taken into consideration, it is difficult to select a single "best" movement parameter since the issues discussed above introduce a large degree of uncertainty into the interpretation of the results. However, it is necessary to choose a parameter value for future simulations. I suggest that the purpose of the model must play an important role in determining how much importance to attach to each of the metrics presented here. In Chapter 3, this model is used to predict range-wide genetic connectivity of the desert tortoise, and multiple scenarios are compared to assess the impact of human development on genetic connectivity. Given that the emphasis is on the spatial genetic pattern, I suggest that the correlation with the real data is the metric that should be given the most weight because this assesses how well the simulated spatial genetic pattern matches the observed spatial genetic pattern. Since four of the five movement distances tested here converge on the same correlation, this suggests that any of these parameter values would be suitable.

In the absence of definitive support for one parameter over another, I have chosen to use the "base $\times$ 3" distance in Chapter 3. The "base $\times$ 1" and "base $\times$ 2" distances show very dramatic overestimates of the F_{ST} and pairwise F_{ST}, but the three other distance parameters show more similar results. In addition, the actual movement distances that occur in the simulation will always be less than those used as the parameter value, as described in the "Methods" section. Therefore, it is possible that "base $\times$ 3" may actually result in movement distances that reflect the original values calculated from the home-range displacement data, and if there is little distinction in

the simulation output it makes sense to choose the result most similar to empirical data. That being said, because I did not save the movement distances, I cannot directly assess the similarity between the simulated movement distances and the empirical movement distribution.

An argument against the "base × 3" distance parameter could be made based on the fact that it overestimates both the overall and pairwise F_{ST}, as demonstrated in Figure 19. However, I suggest that given the intended use of the model in Chapter 3 this is unlikely to negatively impact the results. The primary goal of Chapter 3 is assessing the difference between three different scenarios using the pairwise F_{ST}, not making a prediction of the raw value of the pairwise F_{ST}. In fact, I predict that the results of Chapter 3 would be largely the same regardless of which distance parameter is used. Because the different scenarios in Chapter 3 all use the same distance parameter, they will likely all overestimate or underestimate the pairwise F_{ST} values. If this is indeed the case, then the relative differences between the simulations will likely show a very similar pattern regardless of which distance is used.

Overall, the results for the movement distance parameter are ambiguous, but the need to choose a parameter for the simulations in Chapter 3 forces a selection of one of the five distance distributions used. Arguments could undoubtedly be made for the other distributions, but I argue that in the absence of more definitive results, the "base × 3" parameter is a reasonable choice for use in future simulations.

Mutation

Fortunately, the values for the other two parameters are easier to select. For mutation rate, I suggest that a rate of 0.0001 is clearly the most supported, given that it results in a rate of heterozygosity that very closely resembles the value derived from empirical data (Figure 20). It is important to note that the different mutation rates were tested on the "base $\times$ 4" distance distribution, and given that the movement distance influences the heterozygosity (Figure 19) the results would likely change when a mutation rate of 0.0001 is used with "base $\times$ 3", making it possible that a different mutation rate would actually be superior. However, since the difference in the heterozygosity between "base $\times$ 3" and "base $\times$ 4" is relatively small (Figure 19), I predict that the results of using a mutation rate of 0.0001 with "base $\times$ 3" will be very similar to the results observed using "base $\times$ 4", making 0.0001 a reasonable choice for the mutation rate.

Landscape

For the landscape representation, I have chosen to use the habitat surface. It appears that the habitat surface may produce slightly higher correlations with the real data than the neutral surface, although as mentioned before, these differences are not statistically significant. In the absence of a definitive distinction between the two surfaces, I suggest that the habitat surface should be used. The neutral landscape is simply unrealistic – tortoises are known to be influenced by the landscape and cannot move anywhere, which is an assumption made when using the neutral surface.

Therefore, I have chosen to use the habitat surface for the simulations performed in Chapter 3.

Conclusion

The results presented here provide evidence that the model output approximates the observed genetic pattern on the landscape, lending support to the utility of the model for generating insights into desert tortoise genetic connectivity. In addition, modifying the parameter values provides a better understanding of the behavior of the model and helps identify parameter settings that are likely to produce realistic outputs. The results of this study suggest that the simulation can be a valuable tool for generating insights into the spatial genetic pattern of desert tortoises.

The simulation also offers opportunities for different avenues of research – agent-based models are quite flexible, so while the simulation is used here to predict spatial genetic patterns, the model could easily be used for other purposes. For example, it could be run with an emphasis on demography and movement rather than genetics – this could involve testing how different landscape scenarios influence the demography and yearly movement of tortoises. Another potential use could be the assessment of the tortoises' ability to track range shifts due to climate change – combining the model with future climate scenarios could be useful for assessing the speed at which tortoises can expand their range, which is an important characteristic for species facing climate change. In addition to its application to real-world scenarios, the simulation could also be used for more theoretical studies, allowing researchers to

investigate theoretical questions in an idealized environment. Overall, the simulation has the potential to support further investigations into desert tortoises in order to gain insights into the impact of individual tortoise behavior on large-scale spatial patterns.

References

Aben, J., Bocedi, G., Palmer, S. C. F., Pellikka, P., Strubbe, D., Hallmann, C., … Matthysen, E. (2016). The importance of realistic dispersal models in conservation planning: application of a novel modelling platform to evaluate management scenarios in an Afrotropical biodiversity hotspot. *Journal of Applied Ecology*, *53*, 1055–1065.

Allendorf, F. W., Luikart, G. H., Aitken, S. N., & Aitken, S. N. (2012). *Conservation and the Genetics of Populations*. Somerset, UNITED KINGDOM: John Wiley & Sons, Incorporated.

Andersen, M. C., Watts, J. M., Freilich, J. E., Yool, S. R., Wakefield, G. I., McCauley, J. F., & Fahnestock, P. B. (2000). Regression-tree modeling of desert tortoise habitat in the central Mojave Desert. *Ecological Applications*, *10*(3), 890–900.

Augusiak, J., Van den Brink, P. J., & Grimm, V. (2014). Merging validation and evaluation of ecological models to "evaludation": A review of terminology and a practical approach. *Ecological Modelling*, *280*, 117–128.

Averill-Murray, R. C., Esque, T. C., Allison, L. J., Bassett, S., Carter, S. K., Dutcher, K. E., … Nussear, K. E. (2021). *Connectivity of Mojave Desert tortoise populations—Management implications for maintaining a viable recovery network*. *Open-File Report*.

Averill-Murray, R. C., Darst, C. R., Strout, N., & Wong, M. (2013). Conserving population linkages for the Mojave desert tortoise (Gopherus agassizii). *Herpetological Conservation and Biology*, *8*(April), 1–15.

Beier, P., & Loe, S. (1992). A checklist for evaluating impacts to wildlife movement corridors. *Wildlife Society Bulletin*, 434–440.

Beier, P., Majka, D. R., & Spencer, W. D. (2008). Forks in the road: Choices in procedures for designing wildland linkages. *Conservation Biology*, *22*(4), 836–851.

Berry, K. H. (1986). Desert tortoise (Gopherus agassizii) relocation: implications of social behavior and movements. *Herpetologica*, *42*(1), 113–125.

Berry, K. H., Yee, J. L., Coble, A. A., Perry, W. M., & Shields, T. A. (2013). Multiple factors affect a population of agassiz's desert tortoise (Gopherus agassizii) in the Northwestern Mojave desert. *Herpetological Monographs*, *27*, 87–109.

Blair, C., Weigel, D. E., Balazik, M., Keeley, A. T. H., Walker, F. M., Landguth, E., … Balkenhol, N. (2012). A simulation-based evaluation of methods for inferring linear barriers to gene flow. *Molecular Ecology Resources*, *12*(5), 822–833.

Boulanger, E., Dalongeville, A., Andrello, M., Mouillot, D., & Manel, S. (2020). Spatial graphs highlight how multi-generational dispersal shapes landscape genetic patterns. *Ecography*, *43*, 1167–1179.

Bulova, S. J. (1994). Patterns of Burrow Use by Desert Tortoises: Gender Differences and Seasonal Trends. *Herpetological Monographs*, *8*, 133–143.

Carter, S. K., Nussear, K. E., Esque, T. C., Leinwand, I. I. F., Masters, E., Inman, R. D., … Allison, L. J. (2020). Quantifying development to inform management of mojave and sonoran desert tortoise habitat in the American southwest. *Endangered Species Research*, *42*, 167–184.

Chetkiewicz, C.-L. B., St. Clair, C. C., & Boyce, M. S. (2006). Corridors for conservation: Integrating pattern and process. *Annual Review of Ecology, Evolution, and Systematics*, *37*, 317–342.

Curtin, A. J., Zug, G. R., & Spotila, J. R. (2009). Longevity and growth strategies of the desert tortoise (Gopherus agassizii) in two American deserts. *Journal of Arid Environments*, *73*(4–5), 463–471.

Davy, C. M., Edwards, T., Lathrop, A., Bratton, M., Hagan, M., Henen, B., … Murphy, R. W. (2011). Polyandry and multiple paternities in the threatened Agassiz's desert tortoise, Gopherus agassizii. *Conservation Genetics*, *12*(5), 1313–1322.

Drake, K. K., Nussear, K. E., Esque, T. C., Barber, A. M., Vittum, K. M., Medica, P. A., … Hunter, K. W. (2012). Does translocation influence physiological stress in the desert tortoise? *Animal Conservation*, *15*(6), 560–570.

Drake, K. K., Esque, T. C., Nussear, K. E., Defalco, L. A., Scoles-Sciulla, S. J., Modlin, A. T., & Medica, P. A. (2015). Desert tortoise use of burned habitat in the Eastern Mojave desert. *Journal of Wildlife Management*, *79*(4).

Dutcher, K. E., Vandergast, A. G., Esque, T. C., Mitelberg, A., Matocq, M. D., Heaton, J. S., & Nussear, K. E. (2020). Genes in space: what Mojave desert tortoise genetics can tell us about landscape connectivity. *Conservation Genetics*.

Dutcher, K., Heaton, J., & Nussear, K. (2019). Desert Tortoise Connectivity Modeling, Project No.: 2015-UNR-1.

Edwards, T., Berry, K. H., Inman, R. D., Esque, T. C., Nussear, K. E., Jones, C. A., & Culver, M. (2015). Testing Taxon Tenacity of Tortoises: Evidence for a geographical selection gradient at a secondary contact zone. *Ecology and Evolution*, *5*(10), 2095–2114.

Edwards, T., Goldberg, C. S., Kaplan, M. E., Schwalbe, C. R., & Swann, D. E. (2003). PCR primers for microsatellite loci in the desert tortoise (Gopherus agassizii, Testudinidae). *Molecular Ecology Notes*, *3*(4), 589–591.

Epps, C. W., & Keyghobadi, N. (2015). Landscape genetics in a changing world: Disentangling historical and contemporary influences and inferring change. *Molecular Ecology*, *24*(24), 6021–6040.

Esque, T. C., Nussear, K. E., Drake, K. K., Walde, A. D., Berry, K. H., Averill-Murray, R. C., … Heaton, J. S. (2010). Effects of subsidized predators, resource variability, and human population density on desert tortoise populations in the Mojave Desert, USA. *Endangered Species Research*, *12*, 167–177.

Farnsworth, M. L., Dickson, B. G., Zachmann, L. J., Hegeman, E. E., Cangelosi, A. R., Jackson, T. G., & Scheib, A. F. (2015). Short-term space-use patterns of translocated Mojave desert tortoise in southern California. *PLoS ONE*, *10*(9), 1–18.

Freilich, J. E., Burnham, K. P., Collins, C. M., & Garry, C. A. (2000). Factors affecting population assessments of desert tortoises. *Conservation Biology*, *14*(5), 1479–1489.

Furstenau, T. N., & Cartwright, R. A. (2016). The effect of the dispersal kernel on isolation-by-distance in a continuous population. *PeerJ*, (3).

Gauffre, B., Estoup, A., Bretagnolle, V., & Cosson, J. F. (2008). Spatial genetic structure of a small rodent in a heterogeneous landscape. *Molecular Ecology*, *17*(21), 4619–4629.

Gregory, A. J., & Beier, P. (2014). Response variables for evaluation of the effectiveness of conservation corridors. *Conservation Biology*, *28*(3), 689–695.

Grimm, V., & Railsback, S. F. (2012). Pattern-oriented modelling: A "multi-scope" for predictive systems ecology. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1586), 298–310.

Grimm, V., Revilla, E., Berger, U., Jeltsch, F., Mooij, W. M., Railsback, S. F., … DeAngelis, D. L. (2005). Pattern-oriented modeling of agent-based complex systems: Lessons from ecology. *Science*, *310*(5750), 987–991.

Hagerty, B. E., Peacock, M. M., Kirchoff, V. S., & Tracy, C. R. (2008). Polymorphic microsatellite markers for the Mojave desert tortoise, Gopherus agassizii. *Molecular Ecology Resources*, *8*(5), 1149–1151.

Hagerty, B. E., Nussear, K. E., Esque, T. C., & Tracy, C. R. (2011). Making molehills out of mountains: Landscape genetics of the Mojave desert tortoise. *Landscape Ecology*, *26*(2), 267–280.

Harless, M. L., Walde, A. D., Delaney, D. K., Pater, L. L., & Hayes, W. K. (2009). Home Range, Spatial Overlap, and Burrow Use of the Desert Tortoise in the West Mojave Desert. *Copeia*, (2), 378–389.

Hedrick, P. W. (1999). Highly variable loci and their interpretation in evolution and conservation. *Evolution*, *53*(2), 313–318.

Hromada, S. J., Esque, T. C., Vandergast, A. G., Dutcher, K. E., Mitchell, C. I., Gray, M. E., … Nussear, K. E. (2020). Using movement to inform conservation corridor design for Mojave desert tortoise. *Movement Ecology*, *8*(38), 1–18.

Jaquiéry, J., Broquet, T., Hirzel, A. H., Yearsley, J., & Perrin, N. (2011). Inferring landscape effects on dispersal from genetic distances: How far can we go? *Molecular Ecology*, *20*(4), 692–705.

Keeley, A. T. H., Beier, P., Keeley, B. W., & Fagan, M. E. (2017). Habitat suitability is a poor proxy for landscape connectivity during dispersal and mating movements. *Landscape and Urban Planning*, *161*, 90–102.

Landguth, E. L., Cushman, S. A., Schwartz, M. K., McKelvey, K. S., Murphy, M., & Luikart, G. (2010). Quantifying the lag time to detect barriers in landscape genetics. *Molecular Ecology*, *19*(19), 4179–4191.

Latch, E. K., Boarman, W. I., Walde, A., & Fleischer, R. C. (2011). Fine-scale analysis reveals cryptic landscape genetic structure in desert tortoises. *PLoS ONE*, *6*(11).

Longshore, K. M., Jaeger, J. R., & Sappington, J. M. (2003). Desert tortoise (Gopherus agassizii) survival at two eastern Mojave Desert sites: death by short-term drought? *Journal of Herpetology*, *37*(1), 169–177.

Lovich, J. E., Agha, M., Ennen, J. R., Arundel, T. R., & Austin, M. (2018). Agassiz's desert tortoise (Gopherus agassizii) activity areas are little changed after wind turbine-induced fires in California. *International Journal of Wildland Fire*, *27*(12), 851–856.

Mateo-Sánchez, M. C., Balkenhol, N., Cushman, S., Pérez, T., Domínguez, A., & Saura, S. (2015). A comparative framework to infer landscape effects on population genetic structure: are habitat suitability models effective in explaining gene flow? *Landscape Ecology*, *30*, 1405–1420.

Medica, P. A., Nussear, K. E., Esque, T. C., & Saethre, M. B. (2013). Long-Term Growth of Desert Tortoises (Gopherus agassizii) in a Southern Nevada Population. *Journal of Herpetology*, *46*, 213–220.

Mitchell, C. I., Friend, D. A., Phillips, L. T., Hunter, E. A., Lovich, J. E., Agha, M., … Shoemaker, K. T. (2021). 'Unscrambling' the drivers of egg production in Agassiz's desert tortoise: climate and individual attributes predict reproductive output. *Endangered Species Research*, *44*, 217–230.

Murphy, R. W., Berry, K. H., Edwards, T., & McLuckie, A. M. (2007). A genetic assessment of the recovery units for the mojave population of the desert tortoise, Gopherus agassizii. *Chelonian Conservation and Biology*, *6*(2), 229–251.

Nafus, M. G., Esque, T. C., Averill-Murray, R. C., Nussear, K. E., & Swaisgood, R. R. (2017). Habitat drives dispersal and survival of translocated juvenile desert tortoises. *Journal of Applied Ecology*, *54*, 430–438.

Nagy, K. A., Scott Hillard, L., Tuma, M. W., & Morafka, D. J. (2015). Head-started desert tortoises (Gopherus agassizii): Movements, survivorship and mortality causes following their release. *Herpetological Conservation and Biology*, *10*(1), 203–215.

Nussear, K. E., Esque, T. C., Inman, R. D., Gass, L., Thomas, K. A., Wallace, C. S. A., … Webb, R. H. (2009). *Modeling habitat of the desert tortoise (Gopherus agassizii) in the Mojave and parts of the Sonoran Deserts of California, Nevada, Utah, and Arizona* (Vol. 1102).

Nussear, K. E., Tracy, C. R., Medica, P. A., Wilson, D. S., Marlow, R. W., & Corn, P. S. (2012). Translocation as a conservation tool for Agassiz's desert tortoises: Survivorship, reproduction, and movements. *Journal of Wildlife Management*, *76*(7), 1341–1353.

Parker, S. S., Cohen, B. S., & Moore, J. (2018). Impact of solar and wind development on conservation values in the Mojave Desert. *PLoS ONE*, *13*(12), 1–16.

R Core Team. (2022). R: A Language and Environment for Statistical Computing. Vienna, Austria.

Ruby, D. E., Zimmerman, L. C., Bulova, S. J., Salice, C. J., O'Connor, M. P., & Spotila, J. R. (1994). Behavioral responses and time allocation differences in desert tortoises

exposed to environmental stress in semi-natural enclosures. *Herpetological Monographs*, *8*, 27–44.

Safner, T., Miller, M. P., McRae, B. H., Fortin, M. J., & Manel, S. (2011). Comparison of Bayesian clustering and edge detection methods for inferring boundaries in landscape genetics. *International Journal of Molecular Sciences*, *12*(2), 865–889.

Sah, P., Nussear, K. E., Esque, T. C., Aiello, C. M., Hudson, P. J., & Bansal, S. (2016). Inferring social structure and its drivers from refuge use in the desert tortoise, a relatively solitary species. *Behavioral Ecology and Sociobiology*, *70*, 1277–1289.

Sánchez-Ramírez, S., Rico, Y., Berry, K. H., Edwards, T., Karl, A. E., Henen, B. T., & Murphy, R. W. (2018). Landscape limits gene flow and drives population structure in Agassiz's desert tortoise (Gopherus agassizii). *Scientific Reports*, (8).

Schwartz, T. S., Osentoski, M., Lamb, T., & Karl, S. A. (2003). Microsatellite loci for the North American tortoises (genus Gopherus) and their applicability to other turtle species. *Molecular Ecology Notes*, *3*(2), 283–286.

Shirk, A. J., Cushman, S. A., & Landguth, E. L. (2012). Simulating pattern-process relationships to validate landscape genetic models. *International Journal of Ecology*.

Thatte, P., Joshi, A., Vaidyanathan, S., Landguth, E., & Ramakrishnan, U. (2018). Maintaining tiger connectivity and minimizing extinction into the next century: Insights from landscape genetics and spatially-explicit simulations. *Biological Conservation*, *218*, 181–191.

Theobald, D. M., Travis, W. R., Drummond, M. A., & Gordon, E. S. (2013). The Changing Southwest. In G. Garfin, A. Jardine, R. Merideth, M. Black, & S. LeRoy (Eds.), *Assessment of Climate Change in the Southwest United States: A Report Prepared for the National Climate Assessment* (pp. 37–55). Washington, DC: Island Press.

Thiele, J. C., Kurth, W., & Grimm, V. (2014). Facilitating Parameter Estimation and Sensitivity Analysis of Agent-Based Models: A Cookbook Using NetLogo and R. *Journal of Artificial Societies and Social Simulation*, *17*(3).

U.S. Geological Survey. (2015). National Watershed Boundary Dataset (WBD).

Varvio, S., Chakraborty, R., & Nei, M. (1986). Genetic variation in subdivided populations and conservation genetics. *Heredity*, *57*, 189–198.

Wang, J. (1997). Effective size and F-statistics of subdivided populations. II. Dioecious species. *Genetics*, *146*(4), 1465–1474.

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-Statistics for the Analysis of Population Structure. *Evolution*, *38*(6), 1358–1370.

Wright, S. (1943). Isolation by distance. *Genetics*, *28*(2), 114–138. Retrieved from

Zeller, K. A., Mcgarigal, K., & Whiteley, A. R. (2012). Estimating landscape resistance to movement: a review. *Landscape Ecology*, *27*, 777–797.

CHAPTER 3

Using a range-wide agent-based model to identify areas of decreased connectivity for the Mojave desert tortoise

Introduction

Habitat fragmentation resulting from anthropogenic development is a major conservation concern to many species. In addition to outright population loss, habitat loss that causes fragmentation can hinder gene flow and make subpopulations more susceptible to disturbances by preventing the "rescue effect" from nearby subpopulations in the event of a catastrophic disturbance (Haddad et al. 2015; Keyghobadi 2007). One species facing severe habitat fragmentation is the Mojave desert tortoise (U.S. Fish and Wildlife Service 2011). Historically, desert tortoise habitat was largely connected, resulting in a high degree of genetic connectivity across their range despite their relatively limited dispersal capabilities (Dickson et al. 2017; Hagerty &Tracy 2010). In recent years, however, human development in the region has fragmented this once-contiguous area and poses a serious conservation concern for the desert tortoise. Urban growth, roads, railroads, and solar facilities all result in habitat loss and fragmentation, and human activities are likely to increase due to continued human population growth (Parker et al. 2018).

Because of this, conservation efforts for the desert tortoise include mitigating the impacts of human development on the connectivity of desert tortoise habitat (U.S. Fish and Wildlife Service 2011; Averill-Murray et al. 2021). Making informed and

strategic decisions aimed towards improving or maintaining connectivity requires an understanding of how different landscape features impact connectivity, which can lead to the identification of areas that are experiencing fragmentation as well as areas that are important to protect in order to maintain connectivity.

A major challenge in evaluating the impact of habitat fragmentation, however, is that some of its effects may take time to manifest (Tilman et al. 1994). In particular, the genetic effects of fragmentation may not be detectable until many generations after the disturbance occurs (Blair et al. 2012; Epps & Keyghobadi 2015; Landguth et al. 2010). Given that desert tortoises have long generation times, this is a particularly relevant issue for this species (Dutcher et al. 2020). Informed conservation actions require knowledge of how development will impact the desert tortoise – ideally, data and evidence about how development has affected genetic connectivity in the past could be used to inform future decisions. However, because the genetic effects of fragmentation may not be detectable for many years, little evidence of genetic differentiation may exist even if a landscape feature does in fact severely limit gene flow (Edwards et al. 2004).

Researchers often use simulations to predict the effects human development may have on future genetic connectivity (Landguth et al. 2010). In particular, agent-based models can be used to simulate the movements and life history of individual animals in order to predict emerging genetic patterns. Many studies make use of agent-based models in this way (Cushman et al. 2016; Kaszta et al. 2020; Thatte et al. 2018; van Strien et al. 2014), and software packages like CDPOP and SimAdapt have been

developed for running landscape-genetic simulations (Landguth & Cushman 2010; Rebaudo et al. 2013). By using simulations with realistic landscape scenarios, predictions can be made regarding the impact of landscape change on a species' genetic connectivity that can help guide conservation actions.

A major difficulty with using agent-based models, however, is their computational intensity. Existing software packages, like CDPOP and SimAdapt, are limited to relatively small areas and population sizes. While smaller simulations can still be very useful and informative, a computationally inefficient model severely limits the scope at which it can be used. A more computationally efficient model, on the other hand, could be used to perform simulations at a much larger scale. When applied to landscape genetics, for example, a range-wide model for a species would allow for insights into large-scale genetic patterns that smaller simulations do not provide. In the context of habitat fragmentation, it could be used to identify areas that are important for maintaining connectivity as well as areas that are likely to experience a loss in connectivity due to anthropogenic development.

In this chapter, I use the agent-based model presented in Chapter 2 at a range-wide scale to evaluate the impact of current development on genetic connectivity throughout the Mojave Desert. I run the simulation for three different scenarios representing an undisturbed landscape, a landscape degraded by development, and a landscape where all development is treated as a complete barrier. The overall goal of this chapter is to determine the impact of barriers on tortoise connectivity and to

identify areas in the Mojave exhibiting increased risk of genetic isolation due to anthropogenic development.

Methods

Study area and landscape representation

Landscape extent

The same habitat model used in Chapter 2 (Nussear et al. 2009) was used to represent the landscape and the entire extent of the model was used (Figure 23), with one exception. The area east of the Colorado river was not included as the river is generally considered the eastern edge of Mojave desert tortoise range, although there are some Mojave desert tortoises in and around the Black Mountains of Arizona (Edwards et al. 2015; McLuckie et al. 1999).

Therefore, the modelled area extended from Bakersfield in the west to about 100km east of St. George, Utah, and from the Mexican border in the south to about 200km north of Las Vegas, covering parts of the Sonoran Desert and most of the Mojave Desert. Desert tortoises do not occur across the entirety of this region – see Figure 24 for a representation of the modelled habitat quality across the study area.

Within the study area are some major anthropogenic features likely impacting desert tortoises. Las Vegas exists within the region of suitable habitat and occupies most of the Las Vegas Valley, and other urban areas are present in suitable tortoise habitat as

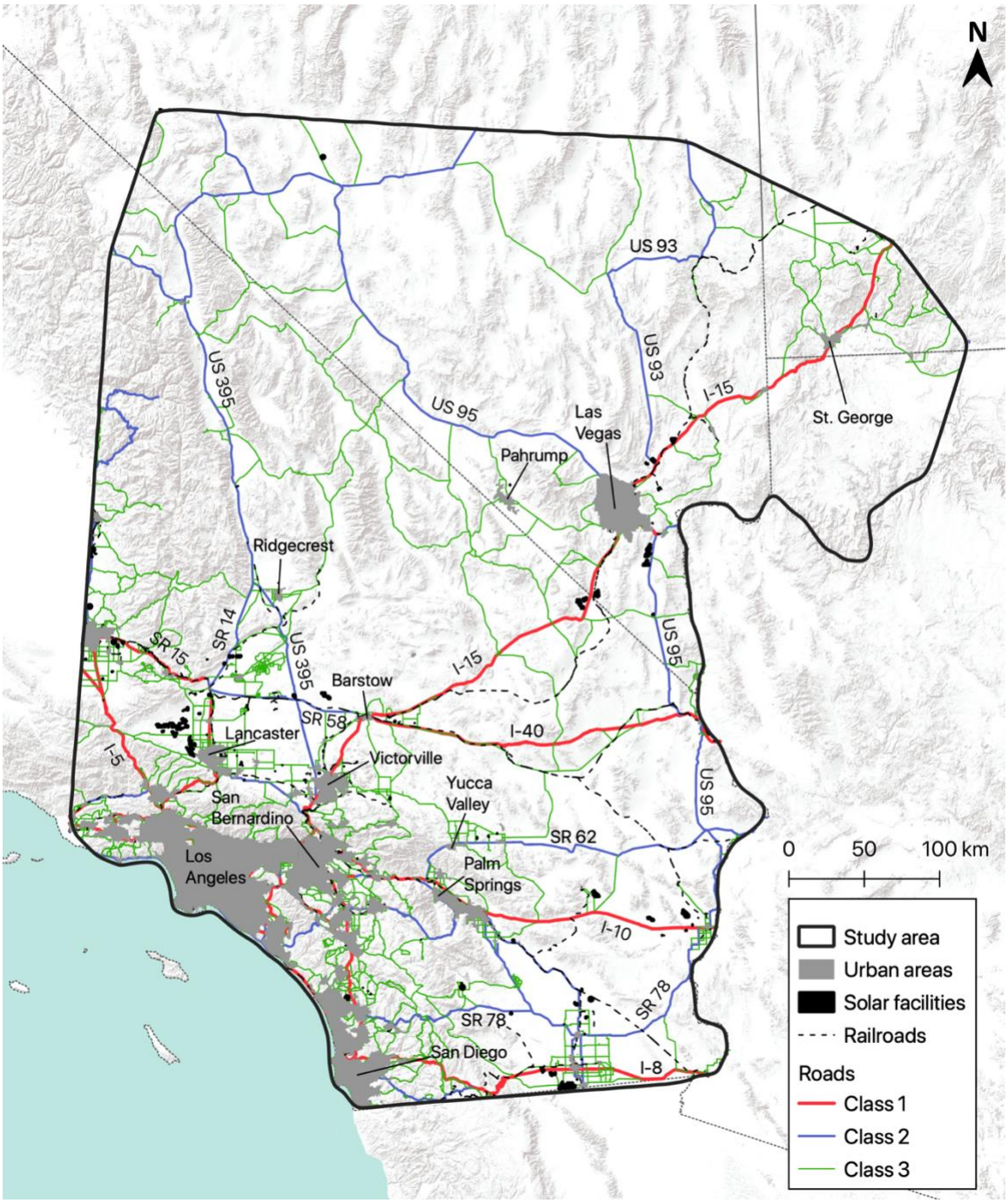

Figure 23 – The extent of the study area and the anthropogenic landscape features used in the simulations.

well. In addition, several major roads cut across the region, like I-15, I-40, and I-10, and solar facilities are present across the area.

Habitat representation

As previously mentioned, the landscape was represented by a raster representing habitat suitability, which had an approximate pixel resolution of 800m. Smaller cell sizes result in longer run times (see Chapter 1) – to decrease computation time, the raster was aggregated by a factor of two, resulting in a raster with a pixel resolution of 1589m $\times$ 1589m. After the landscape scenarios were constructed using this raster, they were converted to quadtrees (see below).

Anthropogenic development

In order to represent human development on the landscape, I obtained GIS data of anthropogenic features known to impact habitat connectivity (Figure 23).

I used a dataset of major and minor highways that were categorized into multiple classes (UCLA Institute for Digital Research and Education 2015). Class one represents limited access roads, class two represents highways, and class three represents minor highways. Data on railroads and urban areas were obtained from the U.S. Census Bureau (U.S. Census Bureau 2017, 2019).

An existing dataset was used for data on solar facilities. To construct the dataset, point data for all solar facilities with a combined nameplate capacity of one megawatt or more had been downloaded from the U.S. Energy Information Administration. The solar sites located at these points had then been digitized onscreen to create polygons representing each site (Wright 2021).

Scenarios

I created three scenarios called “no barrier”, “degraded”, and “barrier”, which are defined by the landscape raster used. Each raster is based on the habitat raster, but they differ in their representation of anthropogenic disturbance.

The “no barrier” scenario consists of only the habitat model with no anthropogenic features included. Conversely, the “barrier” scenario used all the features described above and treated them as impassable barriers to tortoise movement. These two scenarios represent the extremes and are useful for evaluating the connectivity in the best- and worst-case scenarios.

The third scenario, “degraded”, factors in the spatial features by altering the cost of the cells that overlap with anthropogenic landscape features. Urban areas are impassable by desert tortoises, as are solar facilities, which are typically fenced off and inaccessible to desert tortoises (Hromada et al. 2020); therefore, these areas were impassable to the simulated tortoises. Because roads and railroads limit movement but do not necessarily represent complete barriers, they were assigned “conversion factors” which were used to decrease the habitat values of overlapping cells according to the following formula (Dutcher et al. 2019; Inman et al. 2013):

$$\textit{degraded value} = \textit{habitat value} \times (1 - \textit{conversion factor})$$

After applying the formula for all anthropogenic features, the values were subtracted from one (as in Chapter 2) to obtain a resistance surface where zero indicates no resistance to movement and one indicates high resistance to movement. Once converted to a resistance value, a value that has been modified by a conversion factor

will always be equal to or greater than the conversion factor. For example, any cells containing a class one road will have a minimum resistance value of 0.9 (the conversion factor for class one roads), which would occur when the habitat quality is one, the highest value possible ($1 - (1 \times (1 - 0.9)) = 0.9$). Thus, the conversion factor essentially adjusts the value such that the resulting resistance falls between the conversion factor and one.

Conversion factors were chosen to represent the hypothesized influence of the given landscape feature on tortoise movement. Roads and railroads are known to impact tortoise distribution and connectivity – previous studies have found reduced tortoise densities near roads (Boarman & Sazaki 2006; Nafus et al. 2013; Peaden et al. 2015), and tortoises have been shown to avoid movement near roads, including low traffic roads (Peaden et al. 2017; Hromada et al. 2020; Gray et al. 2019; Sadoti et al. 2017). Genetic studies have also found that roads and railroads are associated with increased genetic differentiation, indicating decreased tortoise movement (Latch et al. 2011; Dutcher et al. 2020). In addition, studies have shown that different classes of roads have variable impacts on tortoise distribution (Nafus et al. 2013; Peaden et al. 2015).

I therefore assigned higher resistance values to cells containing roads and used different conversion factors based on the class of the road (see Table 3). Higher resistance values

Feature	Conversion factor
Road, class 1	0.9
Road, class 2	0.7
Road, class 3	0.4
Railroad	0.4

Table 3 – The conversion factors used to represent anthropogenic features in the "degraded" landscape.

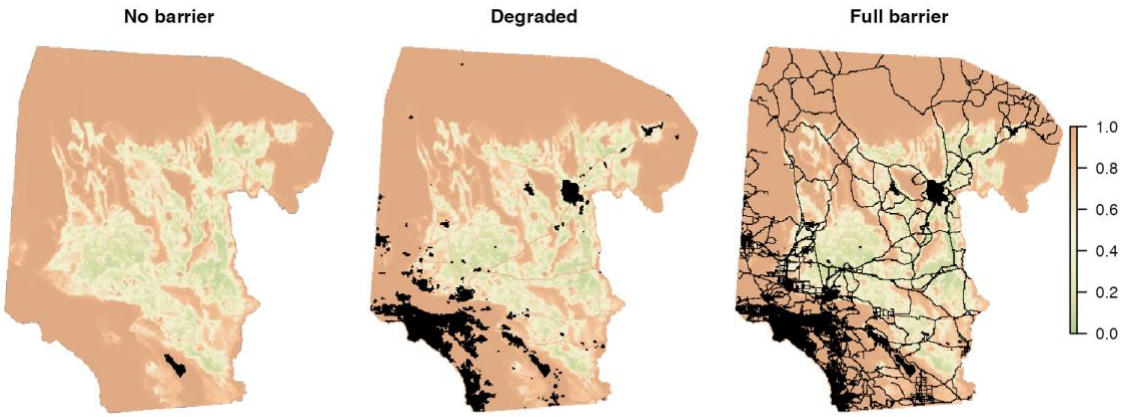

Figure 24 – The resistance surfaces for each of the scenarios. Black areas represent impassable barriers.

decrease the movement potential and increase the mortality risk, making it less likely that a tortoise will pass through the cell. Class one roads, which primarily consist of interstates, received the highest conversion factor, with class two and three roads and railroads receiving lower conversion factors.

Quadtree representation

After the three landscapes were created, they were converted to quadtrees using the "range" method described in Chapter 1 – in the process of creating the quadtree, any quadrant whose maximum and minimum values were greater than 0.1 apart was split. The resulting quadtrees can be seen in Figure 24.

Table 4 – Parameter values used in the simulations.

Parameter	**Value**
Number of years	10,000
Number of replications	10
Mating age	17
Number of offspring (lambda of Poisson distribution)	6
Mating distance	1000m
Age risk score functions	juveniles: $14 / (1 + e^{2 * (x - 10)})$ adults: $6 / (1 + e^{0.2 * (x - 60)})$

Habitat risk score function	$10 / (1 + e^{15 * ((1 - x) - 0.3)})$
Density risk score function	$10 / (1 + e^{0.2 * (x - 60)})$

Simulation

Simulation design and parameterization

The same simulation described in Chapter 2 was used. Based on the results from Chapter 2, I used the "base $\times$ 3" movement parameter and a mutation rate of 0.0001. All other parameters were the same as those used in Chapter 2 and can be seen in Table 4.

Initialization

To create a set of demographically stable tortoises to use as the initial state, a set of tortoises was distributed uniformly across the "no barrier" landscape, and then the model was run until the population size stabilized. The resulting agents were used as the starting set of agents for all simulations.

Genetics

To initialize the genetics, I used a dataset of 748 genetic samples of tortoises across the Mojave (Hagerty & Tracy 2010). As in chapter 2, the genetic data consisted of 20 diploid loci. I calculated the allele frequencies for each locus, which were then treated as the probability of being selected for each allele and were used to initialize the starting agents with random genetic data. This results in a panmictic set of agents – that is, the genetics show no spatial pattern.

Simulation runs and output

For each landscape, 10 repetitions of 10,000 years were run. The complete set of tortoises were saved at 1,000 year increments, which, for every tortoise, included the coordinates, sex, age, and genetics.

Data analysis

Sampling locations

To assess the genetic pattern, I sampled tortoises from the complete set of tortoises for year 10,000 using two different sampling schemes. In each sampling scheme, 2km × 2km squares were distributed across the landscape, and tortoises that fell within each of these sites were selected. For the purpose of the genetic statistics, the tortoises in a site were considered to be a single population.

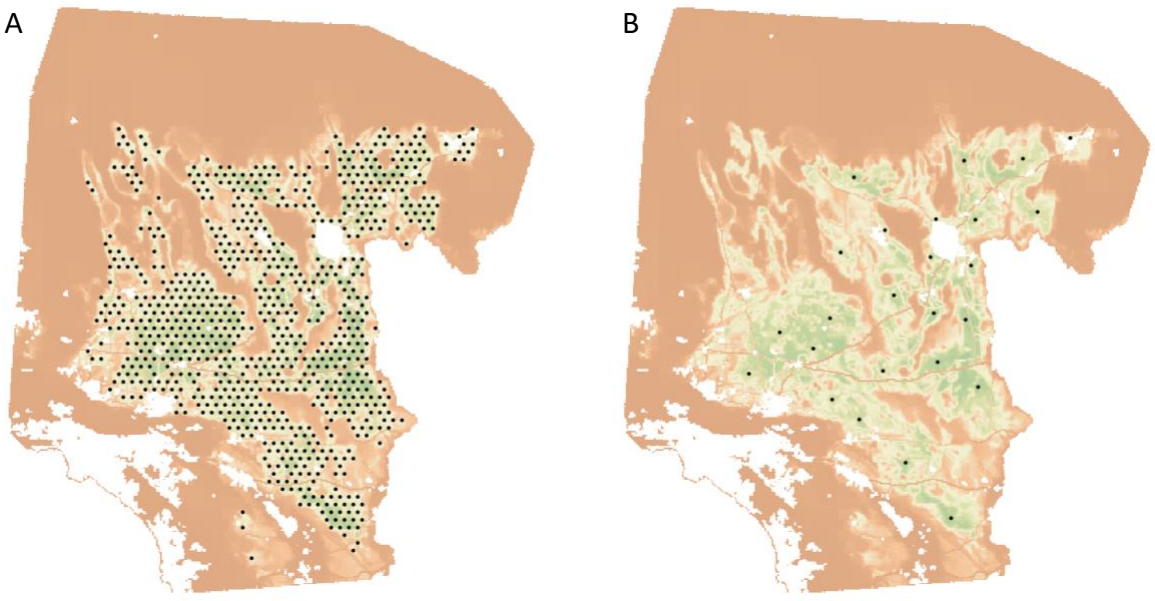

Figure 25 – The two sets of locations used for sampling tortoises. (A) The small-scale sampling scheme. (B) The large-scale sampling scheme.

In order to assess the genetic pattern at different scales, I created a "small-scale" and a "large-scale" sampling scheme (Figure 25). In the small-scale sampling scheme, the squares were arranged in a hexagonal grid, where each square was 10km from its nearest neighbors. To examine spatial relationships at a larger scale, I created a second set of squares by manually selecting 25 locations across the tortoises' range, which I refer to as the large-scale sampling scheme. The sampling locations were chosen to be in high-quality habitat to ensure enough tortoises were present in all three simulations for the metrics calculated to be reliable.

For each of the three scenarios, I sampled tortoises at the sampling sites and then calculated the average number of tortoises in the site across the 10 repetitions for each scenario. When analyzing a single scenario, I only used sites that had more than 50 tortoises on average, and when comparing two scenarios I only used sites where both scenarios had more than 50 tortoises on average. Given that adult tortoises made up about 10% of the population (see "Results"), 50 tortoises in a 2km $\times$ 2km square results in a density of approximately 1.25 adults per km^2.

The analyses performed involve calculating pairwise values between sampling sites, and the large number of pairwise values makes visualization difficult, especially for the small-scale sampling scheme. To reduce the comparisons to only nearby points and allow for visualization of the results, I calculated the Delaunay triangulation for each set of sampling locations, which essentially connects points to nearby points without any overlapping segments (Lee & Schachter 1980). This allows for a drastic reduction in the

number of pairwise values while maintaining the relationships between nearby points, making for more interpretable visualizations.

Genetic data analysis

For each scenario I calculated the overall expected heterozygosity (H_e), overall observed heterozygosity (H_o), and overall F_{ST}, and took the average of these metrics across the 10 repetitions to obtain a single value for each scenario. In addition, I calculated H_e and H_o for each site in the two sampling schemes. H_e is calculated using the allele frequencies for each locus and indicates the percent of individuals that would be expected to be heterozygous if mating was random. H_o, on the other hand, is the actual rate of heterozygosity, and decreases in H_o can indicate increasing fragmentation. F_{ST} measures the genetic divergence among subpopulations, with higher values indicating increased genetic divergence. I used the Weir and Cockerham method to calculate F_{ST} values (Weir & Cockerham 1984). I only present these per-population metrics for the small-scale sampling scheme given that it represents a more thorough sampling of the tortoises than the large-scale sampling scheme.

The primary method used for investigating the genetic connectivity across the landscape was the pairwise F_{ST} (Weir & Cockerham 1984), which indicates the degree of genetic differentiation between two populations. F_{ST} generally responds quickly to changes in connectivity (Epps & Keyghobadi 2015), making it a good metric for detecting the genetic effects of landscape change. The pairwise F_{ST} was calculated for each pair of sampling sites for each scenario and repetition. To obtain a single set of pairwise F_{ST}

values per scenario, I calculated the average for each pairwise value across the ten repetitions of each scenario.

Comparison between scenarios

To assess the effect of anthropogenic landscape features on genetic connectivity, I subtracted the mean "no barrier" pairwise F_{ST} values from the corresponding values for the "barrier" and "degraded" scenarios and then plotted these differences to identify areas where genetic connectivity is reduced in the "barrier" and "degraded" scenarios.

In addition, I identified values with significant differences between scenarios by using a *t*-test to test for a significant difference between the 10 repetitions of one scenario and the 10 repetitions of another scenario. Because I performed multiple comparisons, I used the Holm-Bonferroni method to adjust the *p*-values (Holm 1979) and then identified all values where the corrected p-value was below 0.05. This method was used to identify significant differences in the per-population heterozygosity values as well as the pairwise F_{ST} values. In the latter case, I only tested the significance of the Delaunay edges.

Impact of linear barriers

To assess the impact of roads and railroads on gene flow, I found the pairwise F_{ST} between sites in the small-scale sampling scheme divided by a road or railroad. I found all 10km-long Delaunay edges (i.e. edges that connect to one of the six "neighboring" squares) that intersected a road or railroad and then created univariate density plots of the pairwise F_{ST} by road class and scenario. If a segment intersected multiple linear

features of different classes, it was assigned to the most impermeable class of linear barrier intersected by the segment.

Simulation code and computation

The simulation was written in C++ and run on an 80-core Linux machine with 377GB of RAM. Multiple simulation runs were run in parallel, and computation time for a single run ranged from about 334-364 hours. All data analysis was performed in R (R Core Team 2022).

Results

Connectivity by scenario

No barrier

The "no barrier" scenario had the lowest overall F_{ST} (0.042) and the highest overall H_o (0.772), indicating that this scenario is the most connected of the three (Table 5). It also had the highest number of tortoises of the three scenarios (396,253 adult

	Avg. number of tortoises	**Avg. number of adult tortoises**	H_o	H_e	F_{ST}
Barrier	3,364,606 *(3404)*	340,475 *(482.4)*	0.724 *(0.0015)*	0.797 *(0.0013)*	0.101 *(0.0011)*
Degraded	3,529,530 *(4487)*	357,374 *(514.1)*	0.765 *(0.0007)*	0.797 *(0.0006)*	0.051 *(0.0005)*
No barrier	3,909,239 *(4167)*	396,253 *(359.7)*	0.772 *(0.0011)*	0.798 *(0.0012)*	0.042 *(0.0004)*

Table 5 – Summary statistics for each scenario at year 10,000. Values are averaged across the 10 repetitions. Values in parentheses provide the standard deviation among the 10 repetitions. H_o, H_e, and F_{ST} were calculated using the tortoises from the small-scale sampling scheme.

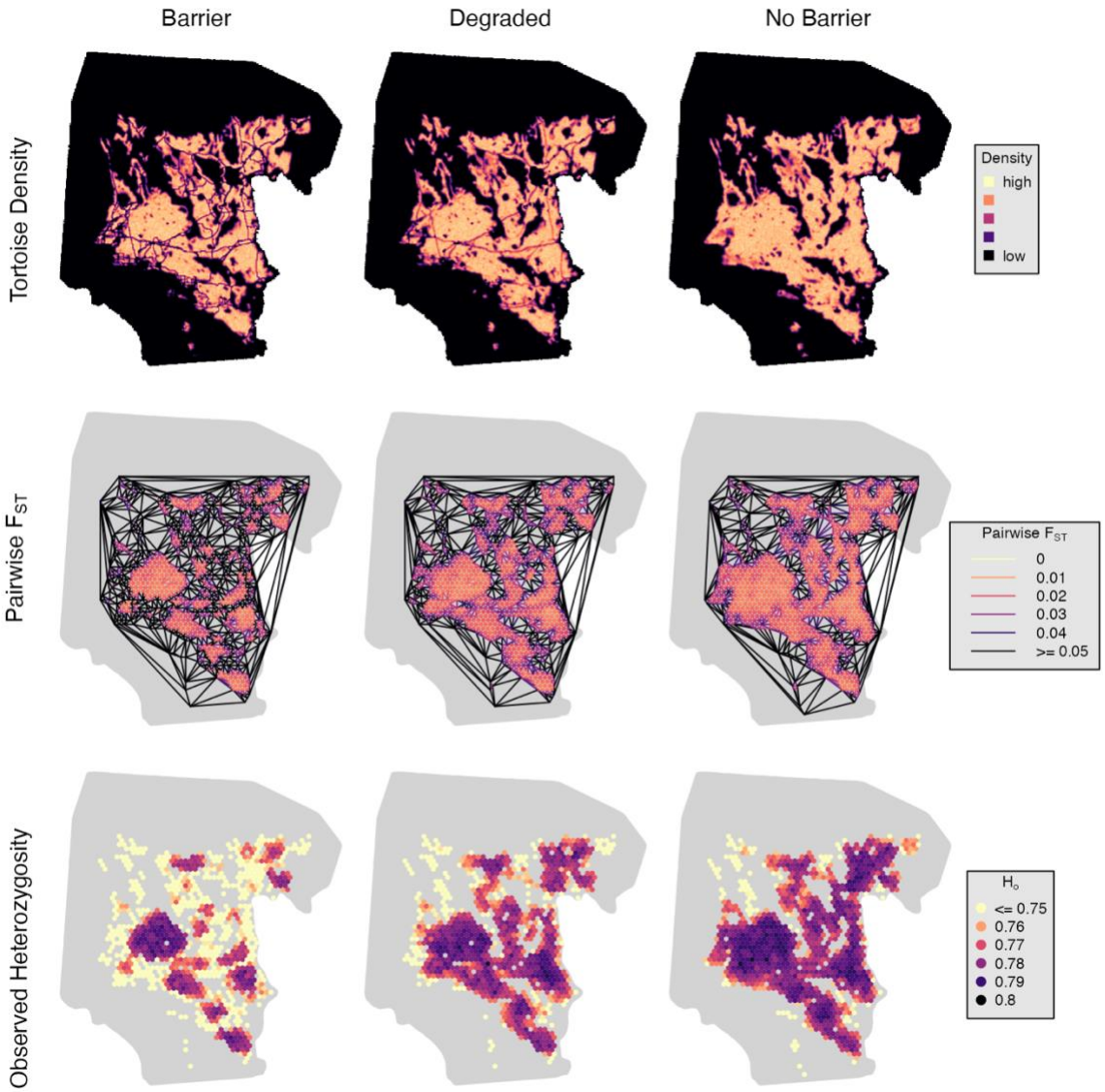

Figure 26 – The tortoise density, pairwise F_{ST}, and observed heterozygosity in year 10,000 of each scenario

tortoises; Table 5), which is as expected given that the other two scenarios include habitat loss and habitat degradation.

The results in Figure 26 depict a landscape that is largely connected. This is especially evident in the plot of the pairwise F_{ST}, which shows that there are very few areas that are cut off from the main body of the habitat. The per-population H_o provides

similar results – most populations have a high level of heterozygosity, with isolated populations and edge populations having lower H_o.

Barrier

The "barrier" scenario shows an extreme level of fragmentation. It has the lowest overall H_o (0.724) and the highest overall F_{ST} (0.101), indicating lower gene flow (Table 5). In addition, it has the lowest number of adult tortoises (340,475) of the three

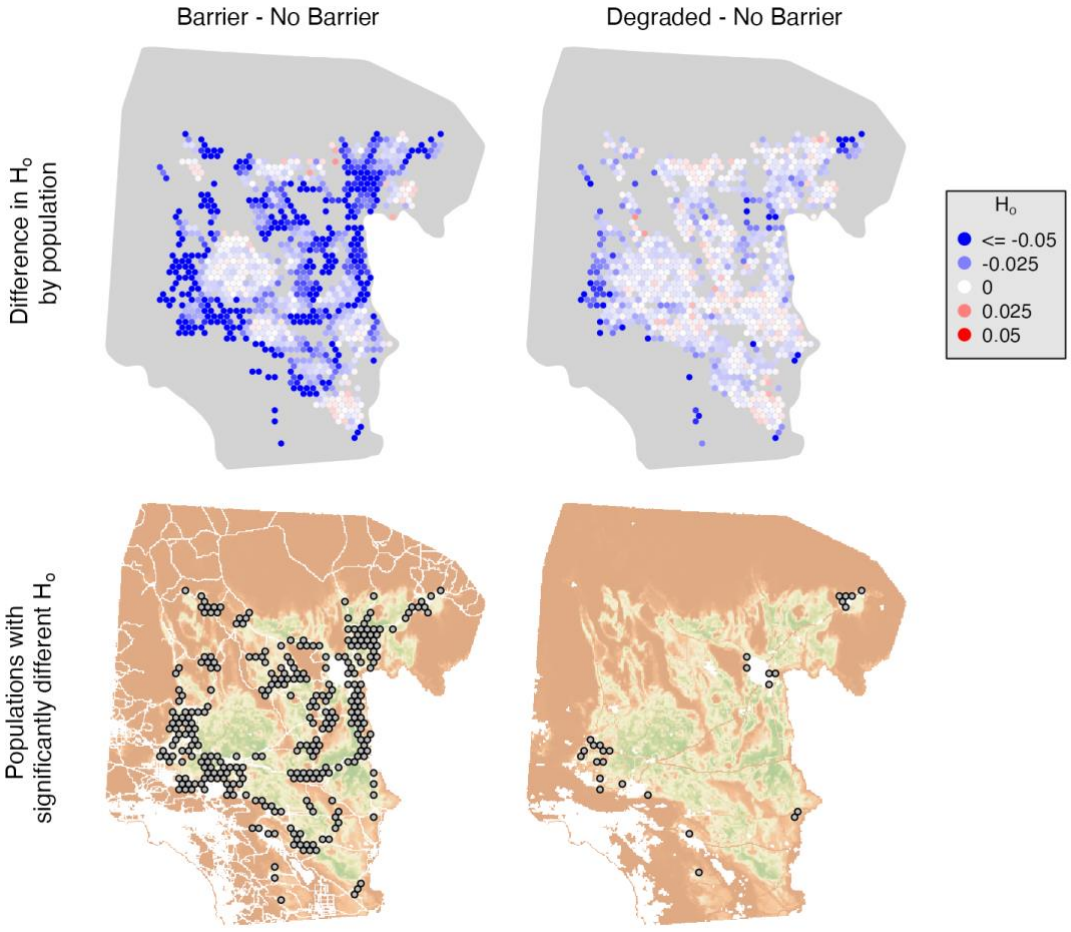

Figure 27 – The differences between the "barrier" and "no barrier" scenarios and the "degraded" and "no barrier" scenarios for per-population H_o using the small-scale sampling points. For each comparison, the raw differences are shown as well as the differences that were statistically significant at the 0.05 level (after correction using the Holm-Bonferroni method).

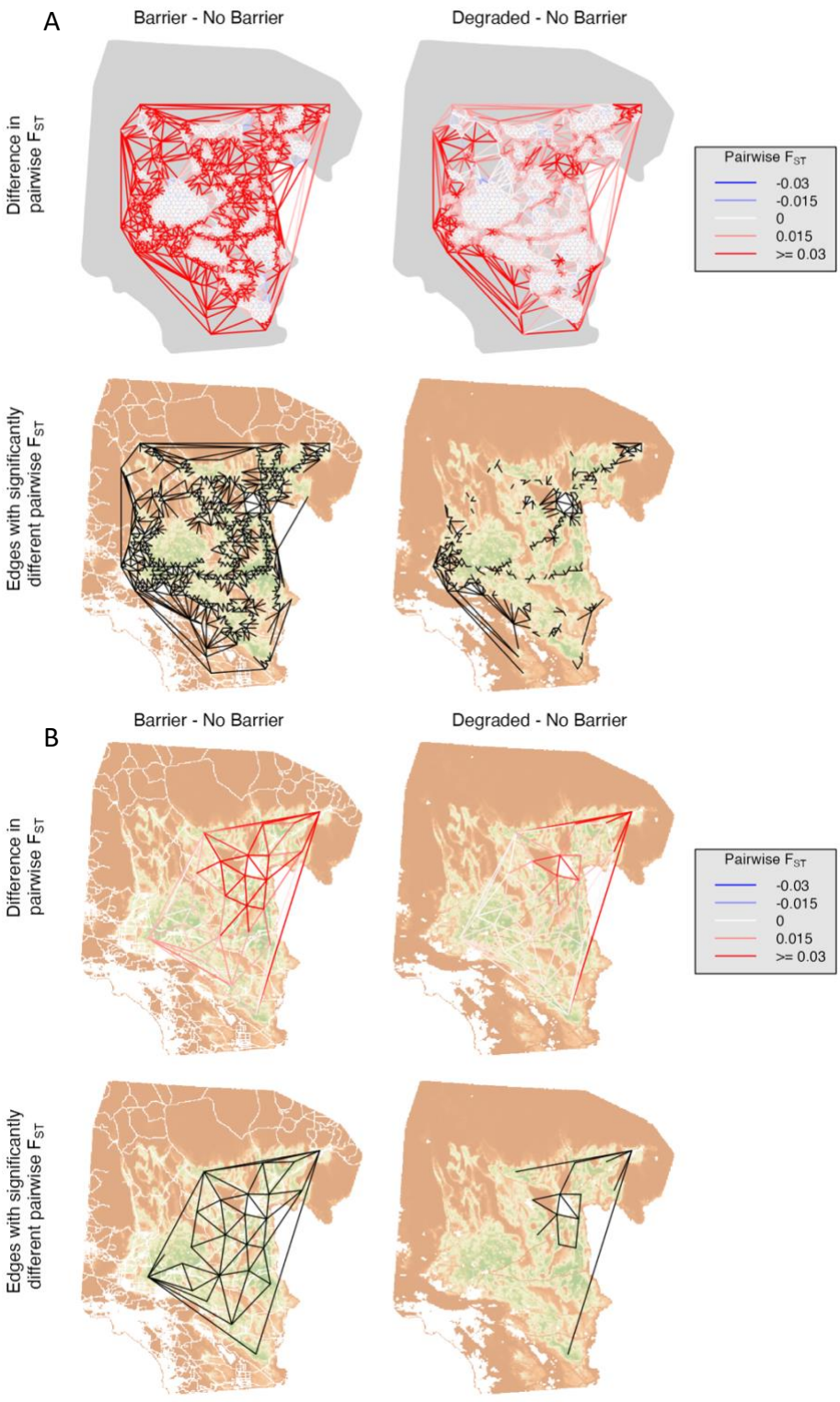

Figure 28 – The differences between the "barrier" and "no barrier" scenarios and the "degraded" and "no barrier" scenarios for pairwise F_{ST} for the (A) small-scale and (B) large-scale sampling points. For each comparison, the raw differences are shown as well as the differences that were statistically significant at the 0.05 level (after correction using the Holm-Bonferroni method).

scenarios (Table 5). These results are as expected – because all human development is treated as a complete barrier, this scenario essentially consists of many completely isolated populations, thereby resulting in a very high degree of genetic differentiation.

The pairwise F_{ST} and per-population H_o also indicate extreme fragmentation (Figure 26). The map of pairwise F_{ST} visibly outlines the fragmented populations – the lower pairwise F_{ST} within isolated areas clearly contrasts with high F_{ST} between populations in different fragments. The per-population H_o also highlights the isolation of the various fragments. Larger fragments, like the one in the western Mojave, still have populations with fairly high H_o. However, smaller fragments have noticeably reduced H_o. This aligns well with theoretical results, which suggest that small populations are likely to have less genetic diversity (Allendorf et al. 2012).

The comparisons with the "no barrier" scenario (Figure 27, Figure 28) show the same patterns just discussed – increased differences in the pairwise F_{ST} between fragmented populations and decreased H_o, with the most dramatic differences occurring in small fragments. Both the large-scale and small-scale sampling schemes show a large number of significant differences across the Mojave (Figure 28).

Degraded

The "degraded" scenario shows increased genetic differentiation, but much less than the "barrier" scenario. The overall F_{ST} (0.051) and overall H_o (0.765) indicate more fragmentation than the "no barrier" scenario but less than the "barrier" scenario (Table

5). Similarly, the total number of adult tortoises (357,374) was more than "barrier" but less than "no barrier" (Table 5).

The maps of pairwise F_{ST} and H_o reflect the addition of anthropogenic development (Figure 26). Noticeably higher pairwise F_{ST} values occur between sites divided by class one roads such as I-15, I-10, and I-40. Overall, the maps show a landscape that is not completely fragmented like the "barrier" scenario but is not as continuous and connected as the "no barrier" scenario.

It is when the results are directly compared with "no barrier" that the effects of the added movement barriers become most evident (Figure 27). The increased pairwise F_{ST} values due to class one roads are very evident, and many of the edges intersecting a road show significant differences with the "no barrier" scenario. In addition, a few larger areas of increased genetic differentiation can be identified, as evidenced by increased pairwise F_{ST} values and decreased H_o. There is a significant loss of connectivity and heterozygosity in the western Mojave, in particular around the cities of Victorville and Ridgecrest, California, where there is a higher concentration of roads, urban areas, and

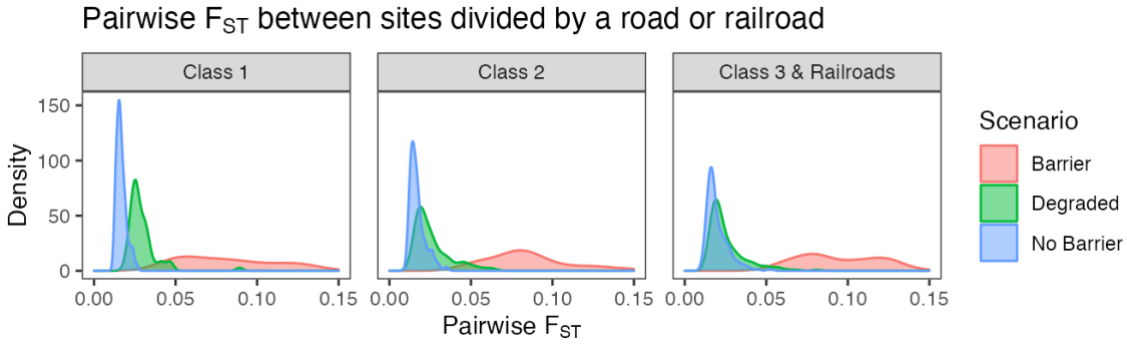

Figure 29 – The density of pairwise F_{ST} values for 10km-long Delaunay edges intersected by a road or railroad. Class 3 roads and railroads are grouped together because they use the same conversion factor (see Table 3).

solar facilities. Las Vegas, Nevada, also results in fragmentation – it blocks an area connecting a few different patches of habitat, and as a result causes increased differentiation between populations on opposite sides of the city. Finally, the populations in Utah, which are already on the edge of the suitable habitat, appear to be mostly cut off from the rest of the range, likely because of I-15 and St. George as well as some smaller roads in the area.

It is also interesting to note the differences in the spatial location of significant edges identified between the small-scale and large-scale sampling schemes. Unlike the small-scale sampling scheme, the large-scale sampling scheme only shows significant differences in pairwise F_{ST} in the areas around Las Vegas and St. George, suggesting that the genetic differentiation caused by roads may be detectable only at sites close to a road and not at more distant sites.

Impact of roads on connectivity

The road classes have differential effects on the pairwise F_{ST}, as shown in Figure 29, which compares edges intersecting roads across the three scenarios. Class one roads (primarily consisting of interstates) have the biggest impact, while class two and three roads as well as railroads have noticeable but less distinct impacts. However, while class one roads clearly have a major impact on connectivity, some gene flow appears to be happening in the “degraded” scenario – the pairwise F_{ST} is still significantly lower than the “barrier” scenario, where all roads and railroads are impermeable, suggesting that some gene flow is still occurring.

Discussion

Simulated population size compared with population estimates

The overall number of tortoises ranged between 3.3 and 4 million tortoises; however, for each simulation, only about 10% were adults (i.e. tortoises 17 years or older), with the average number of tortoises ranging from 340,475 ("barrier") to 396,253 ("no barrier") (Table 5). This corresponds with estimates of real-world population size – a recent study estimated the range-wide number of adult tortoises in 2004 and 2012 to be 336,393 (standard error = 51,596) and 212,343 (standard error = 31,391), respectively (Allison & McLuckie 2018). While the number of simulated tortoises is significantly higher than the estimate for 2012, the estimated total population size for all three scenarios is similar to the 2004 estimate. The difference between the 2012 estimate and the number of tortoises in the simulation could be caused by human impacts not included in the simulation. For example, the simulation does not take into account activities like livestock grazing or off-road vehicle use, which both serve to reduce the quality of tortoise habitat (Fleischner 1994; Lovich & Bainbridge 1999).

Impact of population size on connectivity

In the "no barrier" scenario, the large-scale sampling scheme shows larger differences in pairwise F_{ST} in the northeastern Mojave (Figure 28). In some ways, this may seem unexpected given the fragments are completely isolated from each other, so no two fragments are more connected than any other. In fact, all else being equal, there

should be no overall spatial pattern whatsoever between fragments since there was no spatial genetic pattern present when the simulations began and no interaction between populations in separate fragments. However, the observed differences have nothing to do with spatial arrangement and instead are a result of varying fragment sizes – the northeastern Mojave consists mostly of small fragments, while the western and southern portions have some relatively large fragments, and the sampling areas in this region all fall in these larger fragments (see Figure 25).

Larger populations are more resilient to genetic drift, and so the populations in large fragments are more likely to retain the breadth of genetic diversity present in the initialized genetics (Allendorf 1986) – this is evidenced by the higher values of H_o in large fragments. Genetic differentiation still occurs between these areas – Figure 28 shows that nearly all the pairwise F_{ST} values in the large-scale sampling scheme are significantly different than the "no barrier" scenario – but occurs more slowly in large populations (Varvio et al. 1986; Epps & Keyghobadi 2015). The small fragments lose much of the original genetic variation, and therefore will be left with a smaller subset of the original set of alleles (Allendorf et al. 2012). Given that genetic drift is an inherently stochastic process, the set of alleles kept or lost will vary between the small fragments. Therefore, it is likely that small fragments will retain different sets of alleles, which will then increase the pairwise F_{ST} between populations in these fragments. In this way, the apparent spatial pattern of increased pairwise F_{ST} in the northeast (Figure 28B) is simply a product of the smaller fragments being grouped together.

Isolation vs. separation

The previously discussed phenomenon is important to keep in mind when interpreting the results of the simulations, especially when interpreting the differences in pairwise F_{ST} between the two disturbed landscapes and the "no barrier" landscape. One possible interpretation of a significant difference in pairwise F_{ST} with the "no barrier" simulation could be that less gene flow is occurring between the two populations in the disturbed landscape than in the undisturbed landscape, and this may be the case. However, this may not always be a suitable explanation. A good example is evident in Figure 28 in the comparison between the "degraded" and "no barrier" scenarios – in the large-scale sampling scheme there is a significant difference detected between a location in Utah and a location at the very southern tip of the tortoises' range, which are approximately 450km apart. In the "no barrier" simulation, no direct gene flow can occur between these two sites, so reduced gene flow is not a plausible explanation given that no gene flow occurred in the first place. Instead, the explanation likely lies in the pattern discussed above. The sampled population in Utah is mostly cut off by St. George, forming a small and isolated population which likely retains only a small subset of the original alleles. Thus, the significant difference is likely being caused by the extreme genetic drift occurring in that area. This results in the population having significant differences in the pairwise F_{ST} between virtually every other population, regardless of spatial location – indeed, every edge connecting to the Utah population shows a significant difference in pairwise F_{ST} with the "no barrier" scenario. In this case these differences likely reflect the Utah population's isolation rather than a decreased

amount of gene flow between the two regions given that direct gene flow was not occurring in the “no barrier” scenario.

In fact, for the purpose of interpretation, it may be useful to distinguish between two different reasons for significant differences in pairwise F_{ST} detected between scenarios. The previous discussion illustrates that some of these differences may occur because of the isolation of a single area. In the results presented here isolated populations can be detected by significant decreases in H_o (Figure 27). In addition, these populations typically show that every pairwise F_{ST} value is significantly different, as is evident in the previously discussed example. Isolation can be detected to some degree by both H_o and the pairwise F_{ST}, which is illustrated by the large degree of spatial overlap between regions with significantly different H_o and regions with significantly different pairwise F_{ST} (Figure 27, Figure 28).

However, other significant differences in pairwise F_{ST} between scenarios may be caused by separation between the two populations rather than the extreme isolation of one of the populations. This is evident in sites divided by an interstate. Figure 28 shows that many of the edges in the degraded scenario that cross an interstate are significantly different than those same edges in the no barrier scenario. Comparison with the differences in H_o displayed in Figure 27 shows that these sites do not show dramatic decreases in H_o. Taken together, this indicates that the differences are not simply a result of the extreme isolation of one population.

The difference between these two explanations (what I have termed “isolation” and “separation”) should not be carried too far, as isolation obviously implies

separation. However, in this case it is a useful distinction to make when attempting to identify the reasons for differences in pairwise F_{ST} detected between scenarios.

Implications of results

Long-distance vs. short-distance effect of roads

In the "degraded" simulation, significant differences were found with the "barrier" simulation in nearby sites (e.g. sites from the small-scale sampling scheme) divided by a road, but sites at a greater distance divided by the same road (e.g. sites from the large-scale sampling scheme) did not have significantly different pairwise F_{ST} values with the "no barrier" simulation. For example, the small-scale sampling scheme had several edges intersecting I-40 that were significantly different than the "no barrier" scenario, while none of the edges crossing I-40 in the large-scale sampling scheme had significant differences with the "no barrier" scenario (Figure 28). This is in line with previous studies that have shown that the genetic effects of barriers manifest more slowly between sites at greater distances (or when dispersal distance decreases) (Epps & Keyghobadi 2015; Landguth et al. 2010).

The practical implication of this finding is that research efforts aiming to assess the impact of roads on genetic connectivity should ensure that the genetic data are collected close to the roads whose impacts are being investigated. This is especially relevant given that the relatively recent addition of roads to the landscape may mean that if there is any detectable genetic signal caused by a road, it is likely to be weak.

Reduced movement leads to increased genetic differentiation

Comparing the impact of roads on connectivity shows that road class has a noticeable impact on genetic connectivity, as seen in Figure 29. Class one roads clearly cause genetic differentiation, class two roads lead to genetic differentiation but not as much as class one roads, and class three roads and railroads appear to cause very slight but noticeable differentiation, illustrating the impact that movement reduction has on genetic connectivity. In the simulation, the increased resistance values both reduce movement distances and increase mortality, thus decreasing the number of tortoises that can cross from one side to another. This clearly shows that barriers that reduce tortoise movement increase genetic differentiation.

Another interesting result is the vast difference between the pairwise F_{ST} values of the "barrier" and "degraded" scenarios evident in Figure 29. While the difference between the "degraded" and "no barrier" values show the impact that reduced movement has on connectivity, the even more dramatic difference between the "degraded" and "barrier" scenarios shows that even limited movement is vastly superior to no movement. This provides support for conservation actions that may allow even small numbers of tortoises to cross an otherwise-impermeable barrier (for example, ensuring that culverts underneath interstates are passable by tortoises).

It should be noted that this result should be treated with some caution. Given that I did not save individual movement paths, I cannot assess the number of tortoises that crossed the barrier, and therefore cannot know how many tortoises crossed over any of the roads. Because of this, I cannot determine the actual rate of crossings

required to maintain an acceptable amount of gene flow. An interesting avenue of future research would be to design a simulation that specifically assesses how different crossing rates affect genetic connectivity and what crossing rate is necessary to maintain a suitable level of gene flow.

Areas of concern

As mentioned in the results, the "degraded" scenario appears to show at least three large areas that are negatively impacted – the western edge of the Mojave, the Las Vegas area, and the St. George area. In each of these three areas, there are a number of populations that show significant decreases in H_o (Figure 27), suggesting that development in these areas is creating isolated subpopulations of tortoises that are "trapped" because of a combination of urban development and topography. Because of their isolation, these populations experience increased genetic drift and a loss of genetic diversity.

While all three areas show increased fragmentation, Las Vegas is of particular concern because of its central location – the valley in which it is situated was previously a transitional corridor (Britten et al. 1997; Hagerty et al. 2011). Given that the city now occupies the vast majority of this valley, it acts as a major barrier to gene flow, resulting in genetic differentiation between populations located around the city (Figure 28). In fact, it appears to largely isolate the area northeast of Las Vegas, which makes up the Northeastern Mojave Recovery Unit (U.S. Fish and Wildlife Service 2011). A narrow corridor north of Las Vegas could potentially provide connectivity to the habitat

northwest of Las Vegas, but this area already has some human development, and access to this corridor is made more challenging because of major roads as well as solar facilities. Therefore, it appears that this recovery unit is at risk of being cut off from the rest of the desert tortoises' range.

In addition to these urban areas, major roads like interstates cause increased genetic differentiation, highlighting the importance of exploring conservation actions that can increase the permeability of roads. The number of major roads in the area as well as their location in core tortoise habitat raises concerns about their long-term impacts on the tortoises' genetic connectivity.

Sources of uncertainty in the results

Current and future connectivity

It should be noted that while the simulations ran for 10,000 years into the future, they did so on a static landscape and did not take into account future development or climate change, both of which are occurring rapidly. Therefore, the results should not be interpreted as a prediction of what will happen, but rather as a way of inferring the not-yet-visible impacts of current habitat degradation and fragmentation. Indeed, given the rate of landscape change in this area, the "degraded" landscape may actually be the best-case scenario since future landscapes will be impacted by new development as well as climate change.

Parameterization

The parameterization of a simulation model is notoriously difficult, and often requires subjective definitions of certain parameters. There are many parameters in this model whose value was chosen largely subjectively. In particular, the choice of conversion values to apply for different landscape features is inherently subjective. As mentioned in the "Methods" section, different road classes are known to affect tortoises to varying degrees, and so varying the resistance value based on road class has ample justification. However, the actual values selected is much more subjective. The problem is made particularly challenging in that there is no clear real-world correlate to the "resistance" of a road as defined in the model, which means that even with a plethora of data about tortoise movement across roads it would be difficult to determine how to translate it to the representation used in the simulation.

Unrealistic starting conditions

The simulations began with tortoise genetics randomly assigned across the entire Mojave, which is clearly an unrealistic starting condition – current tortoise genetic patterns are a result of complex historical factors like past climates and geological history, and at no point were the genetics completely panmictic across the entire range. Also, in the real world the addition of barriers on the landscape occurred after a spatial genetic pattern already existed, unlike the initial panmictic state of the "barrier" and "degraded" scenarios. Thus, caution is required when comparing the simulated genetic pattern to the empirical genetic pattern.

That being said, even if the starting condition is unrealistic, areas of increased genetic differentiation in the simulation still likely reflect actual genetic differentiation so long as the parameterization and structure of the model is correct. For example, regardless of whether the panmictic starting condition is realistic, a linear barrier (like a class one road) will increase the genetic differentiation in the simulation, as it does in reality.

Other landscape features

While the "barrier" and "degraded" scenarios include many of the most prominent barriers to tortoise movement, there are other features not included that could change the results. For example, the models do not include minor roads or unpaved roads, which are known to impact tortoise movement (Latch et al. 2011; Hromada et al. 2020). Other features that influence tortoise habitat and therefore could be included in the model are mines, livestock grazing areas, areas damaged due to military maneuvers, and OHV areas, among others.

Conclusion

Habitat fragmentation and degradation is a serious concern for the Mojave desert tortoise. The results presented here highlight the negative impact that current anthropogenic development has on the tortoises' genetic connectivity. As the first range-wide genetic simulation of desert tortoises, this work provides insight into the

way that human development influences the genetic connectivity of desert tortoises across the entire Mojave.

The simulation results can help identify areas where gene flow may be reduced but is not currently detectable in the genetic data, and the areas identified here may be good places to implement actions that could help reduce the negative impact of human development on gene flow. For example, fencing combined with culverts could provide a way for tortoises to safely cross roads (Averill-Murray et al. 2021; Ruby et al. 1994), so placing fencing along roads identified as barriers and ensuring that culverts are passable could be one way to improve the connectivity of areas the simulation identified as limiting gene flow. Overall, models like the one presented have the potential to help identify areas of reduced gene flow, and taking actions in these areas could ameliorate these problems before more severe effects occur.

References

Allendorf, F. W., Luikart, G. H., Aitken, S. N., & Aitken, S. N. (2012). *Conservation and the Genetics of Populations*. Somerset, UNITED KINGDOM: John Wiley & Sons, Incorporated. Retrieved from

Allendorf, F. W. (1986). Genetic drift and the loss of alleles versus heterozygosity. *Zoo Biology*, *5*(2), 181–190.

Allison, L. J., & McLuckie, A. M. (2018). Population trends in mojave desert tortoises (Gopherus Agassizii). *Herpetological Conservation and Biology*, *13*(2), 433–452.

Averill-Murray, R. C., Esque, T. C., Allison, L. J., Bassett, S., Carter, S. K., Dutcher, K. E., … Nussear, K. E. (2021). *Connectivity of Mojave Desert tortoise populations—Management implications for maintaining a viable recovery network*. *Open-File Report*.

Blair, C., Weigel, D. E., Balazik, M., Keeley, A. T. H., Walker, F. M., Landguth, E., … Balkenhol, N. (2012). A simulation-based evaluation of methods for inferring linear barriers to gene flow. *Molecular Ecology Resources*, *12*, 822–833.

Boarman, W. I., & Sazaki, M. (2006). A highway's road-effect zone for desert tortoises (Gopherus agassizii). *Journal of Arid Environments*, *65*(1), 94–101.

Britten, H. B., Riddle, B. R., Brussard, P. F., Marlow, R., & Lee Jr., T. E. (1997). Genetic Delineation of Management Units for the Desert Tortoise, Gopherus agassizii, in Northeastern Mojave Desert. *Copeia*, (3), 523–530.

Cushman, S. A., Elliot, N. B., Macdonald, D. W., & Loveridge, A. J. (2016). A multi-scale assessment of population connectivity in African lions (Panthera leo) in response to landscape change. *Landscape Ecology*, *31*(6), 1337–1353.

Dickson, B. G., Albano, C. M., McRae, B. H., Anderson, J. J., Theobald, D. M., Zachmann, L. J., … Dombeck, M. P. (2017). Informing Strategic Efforts to Expand and Connect Protected Areas Using a Model of Ecological Flow, with Application to the Western United States. *Conservation Letters*, *10*(5), 564–571.

Dutcher, K. E., Vandergast, A. G., Esque, T. C., Mitelberg, A., Matocq, M. D., Heaton, J. S., & Nussear, K. E. (2020). Genes in space: what Mojave desert tortoise genetics can tell us about landscape connectivity. *Conservation Genetics*.

Dutcher, K., Heaton, J., & Nussear, K. (2019). Desert Tortoise Connectivity Modeling, Project No.: 2015-UNR-1.

Edwards, T., Berry, K. H., Inman, R. D., Esque, T. C., Nussear, K. E., Jones, C. A., & Culver, M. (2015). Testing Taxon Tenacity of Tortoises: Evidence for a geographical selection gradient at a secondary contact zone. *Ecology and Evolution*, *5*(10), 2095–2114.

Edwards, T., Schwalbe, C. R., Swann, D. E., & Goldberg, C. S. (2004). Implications of anthropogenic landscape change on inter-population movements of the desert tortoise (Gopherus agassizii). *Conservation Genetics*, (5), 485–499.

Epps, C. W., & Keyghobadi, N. (2015). Landscape genetics in a changing world: Disentangling historical and contemporary influences and inferring change. *Molecular Ecology*, *24*(24), 6021–6040.

Fleischner, T. L. (1994). Ecological Costs of Livestock Grazing in Western North America. *Conservation Biology*, *8*(3), 629–644.

Gray, M. E., Dickson, B. G., Nussear, K. E., Esque, T. C., & Chang, T. (2019). A range-wide model of contemporary, omnidirectional connectivity for the threatened Mojave desert tortoise. *Ecosphere*, *10*(9).

Haddad, N. M., Brudvig, L. A., Clobert, J., Davies, K. F., Gonzalez, A., Holt, R. D., … Townshend, J. R. (2015). Habitat fragmentation and its lasting impact on Earth's ecosystems. *Science Advances*, *1*(2), 1–10.

Hagerty, B. E., Nussear, K. E., Esque, T. C., & Tracy, C. R. (2011). Making molehills out of mountains: Landscape genetics of the Mojave desert tortoise. *Landscape Ecology*, *26*(2), 267–280.

Hagerty, B. E., & Tracy, C. R. (2010). Defining population structure for the Mojave desert tortoise. *Conservation Genetics*, *11*(5), 1795–1807.

Holm, S. (1979). A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian Journal of Statistics*, *6*(2), 65–70.

Hromada, S. J., Esque, T. C., Vandergast, A. G., Dutcher, K. E., Mitchell, C. I., Gray, M. E., … Nussear, K. E. (2020). Using movement to inform conservation corridor design for Mojave desert tortoise. *Movement Ecology*, *8*(38), 1–18.

Inman, R. D., Esque, T. C., Nussear, K. E., Leitner, P., Matocq, M. D., Weisberg, P. J., … Vandergast, A. G. (2013). Is there room for all of us? Renewable energy and Xerospermophilus mohavensis. *Endangered Species Research*, *20*(1), 1–18.

Kaszta, Z., Cushman, S. A., Htun, S., Naing, H., Burnham, D., & Macdonald, D. W. (2020). Simulating the impact of Belt and Road initiative and other major developments in Myanmar on an ambassador felid, the clouded leopard, Neofelis nebulosa. *Landscape Ecology*, *35*, 727–746.

Keyghobadi, N. (2007). The genetic implications of habitat fragmentation for animals. *Canadian Journal of Zoology*, *85*(10), 1049–1064.

Landguth, E. L., Cushman, S. A., Schwartz, M. K., McKelvey, K. S., Murphy, M., & Luikart, G. (2010). Quantifying the lag time to detect barriers in landscape genetics. *Molecular Ecology*, *19*(19), 4179–4191.

Landguth, E. L., & Cushman, S. A. (2010). CDPOP: A spatially explicit cost distance population genetics program. *Molecular Ecology Resources*, *10*, 156–161.

Latch, E. K., Boarman, W. I., Walde, A., & Fleischer, R. C. (2011). Fine-scale analysis reveals cryptic landscape genetic structure in desert tortoises. *PLoS ONE*, *6*(11).

Lee, D. T., & Schachter, B. J. (1980). Two algorithms for constructing a Delaunay triangulation. *International Journal of Computer & Information Sciences*, *9*(3), 219–242.

Lovich, J. E., & Bainbridge, D. (1999). Anthropogenic degradation of the southern California desert ecosystem and prospects for natural recovery and restoration. *Environmental Management*, *24*(3), 309–326.

McLuckie, A. M., Lamb, T., Schwalbe, C. R., & Mccord, R. D. (1999). Genetic and Morphometric Assessment of an Unusual Tortoise (Gopherus agassizii) Population in the Black Mountains of Arizona. *Journal of Herpetology*, *33*(1), 36–44.

Nafus, M. G., Tuberville, T. D., Buhlmann, K. A., & Todd, B. D. (2013). Relative abundance and demographic structure of Agassiz's desert tortoise (Gopherus agassizii) along roads of varying size and traffic volume. *Biological Conservation*, *162*, 100–106.

Nussear, K. E., Esque, T. C., Inman, R. D., Gass, L., Thomas, K. A., Wallace, C. S. A., … Webb, R. H. (2009). *Modeling habitat of the desert tortoise (Gopherus agassizii) in the Mojave and parts of the Sonoran Deserts of California, Nevada, Utah, and Arizona* (Vol. 1102).

Parker, S. S., Cohen, B. S., & Moore, J. (2018). Impact of solar and wind development on conservation values in the Mojave Desert. *PLoS ONE*, *13*(12), 1–16.

Peaden, J. M., Nowakowski, A. J., Tuberville, T. D., Buhlmann, K. A., & Todd, B. D. (2017). Effects of roads and roadside fencing on movements, space use, and carapace temperatures of a threatened tortoise. *Biological Conservation*, *214*, 13–22.

Peaden, J. M., Tuberville, T. D., Buhlmann, K. A., Nafus, M. G., & Todd, B. D. (2015). Delimiting road-effect zones for threatened species: Implications for mitigation fencing. *Wildlife Research*, *42*(8), 650–659.

R Core Team. (2022). R: A Language and Environment for Statistical Computing. Vienna, Austria.

Rebaudo, F., Le Rouzic, A., Dupas, S., Silvain, J.-F., Harry, M., & Dangles, O. (2013). SimAdapt: An individual-based genetic model for simulating landscape management impacts on populations. *Methods in Ecology and Evolution*, *4*, 595–600.

Ruby, D. E., Spotila, J. R., Martin, S. K., & Kemp, S. J. (1994). Behavioral responses to barriers by desert tortoises: implications for wildlife. *Herpetological Monographs*, *8*, 144–160.

Sadoti, G., Gray, M. E., Farnsworth, M. L., & Dickson, B. G. (2017). Discriminating patterns and drivers of multiscale movement in herpetofauna: The dynamic and changing environment of the Mojave desert tortoise. *Ecology and Evolution*, *7*(17), 7010–7022.

Thatte, P., Joshi, A., Vaidyanathan, S., Landguth, E., & Ramakrishnan, U. (2018). Maintaining tiger connectivity and minimizing extinction into the next century: Insights from landscape genetics and spatially-explicit simulations. *Biological Conservation*, *218*, 181–191.

Tilman, D., May, R. M., Lehman, C. L., & Nowak, M. A. (1994). Habitat destruction and the extinction debt. *Nature*, *371*, 65–66.

U.S. Census Bureau. (2019). TIGER/Line Shapefile, 2019, nation, U.S., Rails National Shapefile. Retrieved from https://catalog.data.gov/dataset/tiger-line-shapefile-2019-nation-u-s-rails-national-shapefile

U.S. Census Bureau. (2017). TIGER/Line Shapefile, 2017, 2010 nation, U.S., 2010 Census Urban Area National. Retrieved from https://catalog.data.gov/dataset/tiger-line-shapefile-2017-2010-nation-u-s-2010-census-urban-area-national

U.S. Fish and Wildlife Service. (2011). *Revised Recovery Plan for the Mojave Population of the Desert Tortoise*. Sacramento, CA.

UCLA Institute for Digital Research and Education. (2015). U.S. Highways. Retrieved from https://apps.gis.ucla.edu/geodata/dataset/us_highways/resource/9f1de80f-a8f1-4c75-a23a-8cac586ca2bc

van Strien, M. J., Keller, D., Holderegger, R., Ghazoul, J., Kienast, F., & Bolliger, J. (2014). Landscape genetics as a tool for conservation planning: Predicting the effects of landscape change on gene flow. *Ecological Applications*, *24*(2), 327–339.

Varvio, S., Chakraborty, R., & Nei, M. (1986). Genetic variation in subdivided populations and conservation genetics. *Heredity*, *57*, 189–198.

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-Statistics for the Analysis of Population Structure. *Evolution*, *38*(6), 1358–1370.

Wright, S. A. (2021). *Utility Scale Solar Electric (USSE) and land use implications of meeting 2030 Renewable Energy Goals for California and Nevada in the Mojave Desert*. University of Nevada, Reno.