Title
Exposure in motion: assessing disease risk through movement models and metrics

Author
Dougherty, Eric Robert

Publication Date
2018

Peer reviewed|Thesis/dissertation

Exposure in motion:
assessing disease risk through movement models and metrics

by

Eric R. Dougherty

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Environmental Science, Policy, and Management

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Wayne Marcus Getz, Chair
Professor Justin Brashares
Professor Perry de Valpine
Professor Michael Boots

Summer 2018

Exposure in motion:
assessing disease risk through movement models and metrics

Eric R. Dougherty

Abstract

Exposure in motion:
assessing disease risk through movement models and metrics

by

Eric R. Dougherty

Doctor of Philosophy in Environmental Science, Policy, and Management

University of California, Berkeley

Professor Wayne Marcus Getz, Chair

Exposure represents but one of several processes that underlie disease transmission dynamics in animal and human populations. Infection frequently depends on a number of complex interactions among factors related to the clinical properties of the pathogen (or the magnitude of the dose acquired upon contact) and the immune status of the host. When considering exposure, however, many of these aspects become trivial; the primary consideration is contact between a host and the infectious agent, whether it is harbored by another animal or an environmental reservoir. Contact, in turn, emerges from the space-use decisions of animals over time, potentially resulting in patterns that amplify or dilute the probability of encountering a pathogen. In this sense, the movement behavior of host animals is a fundamental determinant of disease dynamics. Using anthrax as its focal system, this dissertation aims to delve into the exposure process as it relates to the movement behaviors of host animals.

A set of movement trajectories were collected via GPS collars fastened to zebra (*Equus quagga* and springbok (*Antidorcas marsupialis*) in Etosha National Park in Namibia between 2009 and 2010. These data offer insight into the home ranging and habitat selection behaviors that characterize two ungulate species exhibiting susceptibility to anthrax infection and thus, form the basis of the analyses and models developed in this dissertation.

Spatial overlap analysis represents one of the most common methods for evaluating the potential for disease exposure when movement data is available. In the case of an indirectly transmitted pathogen, such as *Bacillus anthracis*, the overlap between individuals may be less important than other characteristics of individual home range usage. Metrics such as

revisitation (the rate at which an animal returns to a specific location) and duration rate (the length of time spent in a specific location) may be more informative, particularly if the locations of locally infectious zones (LIZs) are known. To assess the relative risk faced by zebra and springbok during the anthrax season, I developed a method that reduces the subjectivity in parameter selection when delineating home ranges using the Time Local Convex Hull (T-LoCoH) method. Using a cross-validation-based approach, the resulting site fidelity metrics are more directly comparable. The high values of the two site fidelity metrics imply that similar home ranging behavior among individuals can result in heterogeneous outcomes, contingent entirely upon the presence of a LIZ within an individual's home range.

Much like spatial overlap analyses, habitat selection approaches can offer insight into patterns of potential risk with respect to exposure to disease, particularly in the case of environmentally-borne pathogens. When certain environmental characteristics can be associated with pathogen persistence, niche models can be developed and directly incorporated into the resource selection function framework. I used remotely sensed data on soil, bioclimatic, and vegetation covariates to build such a niche model for anthrax based on soil samples from 40 carcass sites in Etosha National Park harboring viable anthrax spores two or more years after deposition. When this risk layer was applied as a predictor in a step-selection function of zebra, a behaviorally-dependent pattern was evident. When animals were in the foraging state exhibited an avoidance of high risk areas, whereas the same animals were apparently attracted to those higher risk areas when moving in a directed manner. One possible explanation for this pattern is that zebra recognize not only *where* but also *when* they are most susceptible to anthrax, and adjust their behavior to reduce their risk.

Another means of exploring the exposure process is through the use of simulation models. Due to the difficulty associated with comprehensive monitoring of susceptible host populations and infectious reservoirs, simulation models represent an ideal approach for extending general rules emerging from limited movement data to landscapes with known qualities. Using the behaviorally-contingent habitat selection framework created in Chapter 3, I explored the relationship between a set of environmental covariate layers and the exposure process whereby individuals encounter LIZs on the landscape. The method reveals that Wetness may represent a reasonable predictor of epidemic dynamics, with movement serving as the mediating process.

The general analytical methods and models applied here serve to elucidate the role of individual movement behavior in the disease exposure process. Rather than analyzing data on case incidence or prevalence, these methods offer insight into the potential contact patterns that might give rise to endemic or epidemic infections. Thus, they reveal the manner by which analysis of host movements, particularly in conjunction with comprehensive (or simulated) data on the spatial distribution of infectious agents on a heterogeneous landscape, might aid in the management of transmission *risk* before any actual infections occur.

To Mitzi, Isabel, and Clark

Contents

List of Figures

List of Tables

Acknowledgments

First and foremost, I must thank my advisor, Wayne Getz, who offered a great deal more than guidance in developing and assembling this dissertation. I am proud to call myself your advisee and hope to carry the distinction well as my career progresses. I am also exceedingly grateful to Justin Brashares and his lab for providing a second home to me during my time at Berkeley. I would also like to thank Perry deValpine for his involvement on both my qualifying exam and dissertation committees. This dissertation would have looked very different without his guiding hand and thoughtful comments, and the process was undeniably a learning experience. Finally, I would like to acknowledge the roles of both Mike Boots and John Marshall in helping to shape my ideas and interests in disease modeling.

The pursuit of a doctoral degree is an often lonely endeavor, but my friends and colleagues have made the process possible and enjoyable. Thank you Colin and Dana for everything you contributed to this piece of work and to so many other aspects of my time at Berkeley. It wasn't always easy, but we managed to get a surprising amount done during our respective tenures in the Getz Lab and I am grateful to both of you in nearly equal measure; you can decide amongst yourselves who deserves more of my gratitude. I am very grateful to Alex, Kaitlyn, and Juliet for so many wonderful memories, from Mt. Hammer concerts to hikes with your old pal, Potsy. I am also so thankful for all of my friends from past lives who have taken it upon themselves to travel all the way to miserable Berkeley to visit. Aaron, Billy, David, Kyle, and Jill, you have all played important roles in my life and I am lucky to call all of you my friends when you can't hear me. Each of you has made this achievement possible in different ways.

I am also eternally grateful to the institution of U.C. Berkeley and the Environmental Science, Policy, and Management department, in particular, for providing the opportunity to pursue my research interests and also for introducing me to my future wife. Jane, it would take another dissertation to describe just how important you are to me. I could not have asked for a better partner, and it means so much to me that we shared the ups and downs of this journey together. Thank you for supporting me in every conceivable way over the past five years. I cannot wait to see what the future holds for us!

Finally, I would like to thank my family, both new and old. To Shelly, Nick, Howie, Sheelu, Calvin, and Mia, thank you all for welcoming me into your lives and supporting me in so many ways. Especially, though, I would like to thank my parents, Adria and Craig. You have both instilled in me an inquisitiveness and persistence that ultimately produced this dissertation. Your support has been unwavering throughout, and I am so grateful to have had the opportunity to pursue my interests so freely.

Chapter 1

Introduction: The Nexus of Movement and Disease

Background

Disease ecology is generally focused on understanding the ecological drivers of epidemiological dynamics, which refer broadly to the study of the occurrence, distribution, and control of disease. Whereas epidemiology conventionally concerns human disease, wildlife epidemiology, and disease ecology in turn, take a systems-based perspective to consider the drivers of infectious disease transmission. At the core of both human and wildlife epidemiology lies individual behavior, yet investigations of dynamics in animal population tend to be more difficult because of the inability to directly interrogate the behavioral aspects underlying disease spread. Epidemiologists frequently use interviews and observational work to study how human behaviors, such as sexual activity, international travel, or outdoor labor, become risk factors for infectious disease. Animal behavior, though just as important to the disease transmission process, is harder to observe and predict in nature.

Movement ecology uses high-resolution spatiotemporal data to make sense of animal behavior. The "movement ecology paradigm" treats movement as the outcome of behavioral decisions influenced by the interplay of animals' internal states (e.g., physiological needs), external biological factors (e.g., predation or competition), and the physical environment (e.g., mountain ranges or water sources) [115]. Researchers tracking and modeling animal movement can extract behavioral states from telemetry and associated datasets, test hypotheses about what best predicts animal behavior, and explain how individual behavior scales up to landscape-level patterns of animal distributions. Recent advances in telemetry technology [83], the development of corresponding analytical methods [93], and the integration of complimentary datasets (e.g., acceleration data; [165, 143]) have all dramatically increased movement ecologists' inferential power. Especially in light of these developments, ecologists can decompose the impact of individual behavioral heterogeneity on pathogen

spread with much greater ease, making movement ecology a promising avenue for exploring the behavioral underpinnings of how and why diseases spread in wildlife.

Both movement and disease originate in animal behavior at the individual level, and a feedback loop between the two emerges over time at broader scales. The complex space-use patterns produced by movement decisions of individual animals, and how they result in shared space use, are likely to play a substantial role in influencing the subsequent spatiotemporal dynamics of disease transmission [41]. While some movement decisions place a host at relatively high risk of contracting an infection, others may effectively shield them from exposure. Similarly, a shift in the disease state of an animal host can directly impact its subsequent movement decisions, further amplifying or diluting the transmission of a pathogen through the population. The application of movement analyses to disease systems may help illuminate this reciprocal process (e.g., how disease state affects animal movement and, alternatively, how animal movement affects disease transmission; [41])

Understanding ecological links between movement and disease has direct implications for the way researchers model, forecast, and simulate wildlife disease outbreaks. The most basic models in epidemiology treat disease transmission as a function of the number of healthy and infected individuals in a population, linked by a transmission parameter (β). Doing so implicitly combines contact rates and transmission efficiency into one rate [102], but individual heterogeneity in both is universally recognized as an important contributor to disease dynamics in humans [92] and animals [125], and heterogeneity in movement can be an important predictor of this variation [144]. Where tools in movement ecology can help measure, describe, and predict heterogeneity in transmission between hosts, there are opportunities to pose novel questions relating to the effects of movement on contact, the role of contact on transmission, and the impact of infection on movement. This dissertation aims to explore the first of these sets of questions in detail, while alluding to the others as potential paths for investigation given additional or slightly different data.

Categorizing the study system

This dissertation focuses on the bacterium *Bacillus anthracis*, the causative agent of the disease anthrax. Though it can be transmitted via multiple mechanisms, the primary route of interest here (ingestion and infection through the gastrointestinal tract [69]) means that it can quite logically be categorized as an indirectly transmitted pathogen. In other words, the pathogen occupies an intermediate reservoir between hosts. This form of transmission means that spatiotemporal overlap between an infected and susceptible host is not necessary, rather spatial overlap between a host and reservoir is the only requirement for transmission. As such, the work presented in the following three chapters aims to explore the aspect of exposure of host animals to environmental reservoirs of an infectious agent. The specific nature of anthrax and the features of its transmission cycle make it simultaneously unique and generalizable: though not all indirectly transmitted pathogens have the ability to persist

in harsh environments in the same way, there are numerous infectious agents that exhibit similarly predictable distribution patterns on landscapes [24]. The extensive set of data collected from Etosha National Park in Namibia, ranging from ungulate movement data to immunological assays, enables a fairly comprehensive analysis of the potential role of movement in an endemic disease system.

Considering indirect transmission

In the case of pathogens transmitted indirectly, such as anthrax, movement data can shed light on the process by which hosts encounter infectious reservoirs on the landscape. Though infection often depends on a set of complex interactions among host and pathogen traits (e.g., minimum infectious dose, host immune capacity, etc.) [102], exposure can be more readily inferred from movement data. Various metrics may offer insight into the exposure process, particularly those that reveal spatial overlap and habitat selection patterns.

A systematic search of the literature concerning the application of movement analyses in disease research revealed that spatial overlap methods were used in 41 instances of the 91 total analyses found (spread across 70 studies; [41]). These ranged from examinations of home range dynamics (e.g., [170]) to studies that attempted to measure the number of contacts between animals (e.g., [167]), often using proximity sensors to do so (e.g., [100]). A plot of the relative frequency of each of the four classes of analysis can be seen in Figure 1.1 (adapted from the data originally presented in [41]). The importance of spatial overlap analyses derives from the fact that exposure to a pathogen will depend on a susceptible animal traversing the same area over which an infected animal previously ranged and shed the pathogen. The mechanism by which a pathogen is shed varies greatly across species, but in the case of anthrax, death and the subsequent destruction of the carcass represents the primary means by which spores exit the infected host.

Exposure, of course, is contingent upon this shedding process. This means that animals traversing the same area before carcass deposition and after face very different levels of risk, and this temporal component is difficult to incorporate into analyses of movement data unless comprehensive data of the locally infectious zones is available. Even so, evaluations of home ranging behavior can provide insight into which individuals are at relatively higher or lower risk than others. Of particular interest in this respect are metrics that quantify the fidelity of an animal to their home range. One method this is especially well suited to deriving such metrics is the recently developed Local Convex Hull (LoCoH; [60]) method. This algorithm functions like many other home range delineation methods, building minimum convex polygons (referred to as hulls) around sets of movement points that compose an animal's full trajectory. Unlike many of the more traditional home range methods, however, LoCoH has been extended to directly account for the autocorrelated nature of movement data, judging proximity as a function of both space and time when constructing hulls (T-LoCoH; [96]). The resulting home ranges are created by unioning these hulls. This procedure allows for

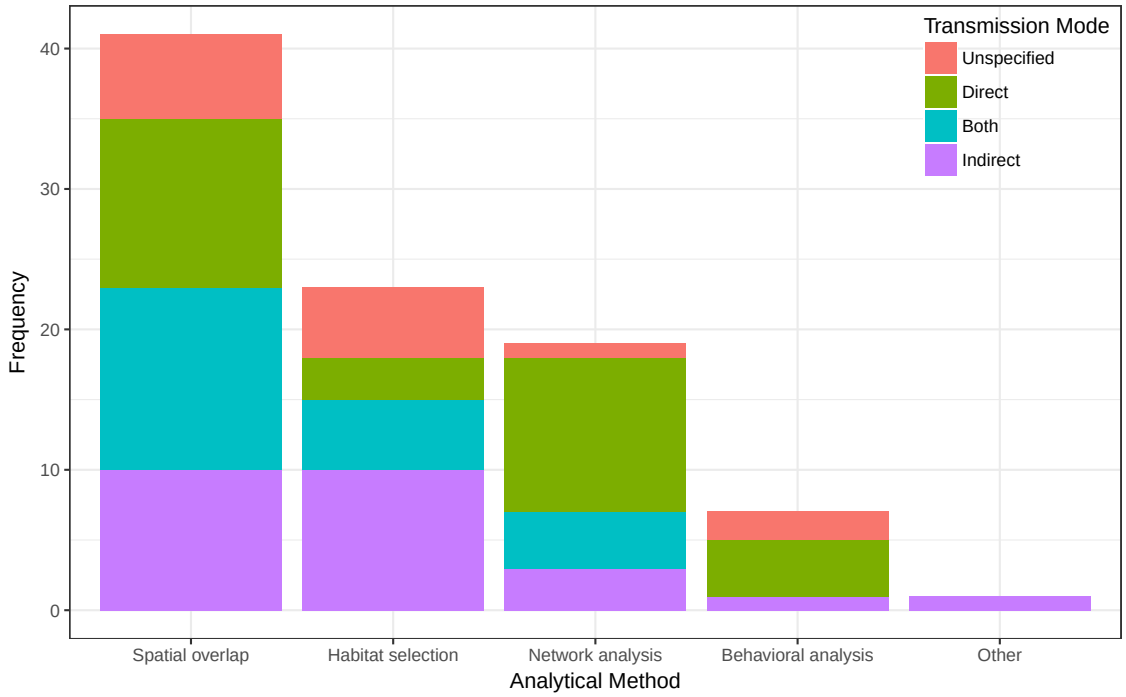


Figure 1.1: **Movement and Disease Literature Search** A frequency plot displaying the distribution of 91 total analyses (conducted in 71 separate studies) among the four predominant classes of methods. The bars are colored according to the primary transmission mode emphasized by the authors. The data that form the basis of the plot were collected and sorted with the help of Dana P. Seidel and Colin J. Carlson. A more detailed explanation of the search criteria and procedure can be found in [41].

a more nuanced interpretation of space use patterns, including the ability to differentiate between rare long visits and frequent short visits to a particular location, something that would be difficult to ascertain using conventional utilization distribution approaches [169]. Using T-LoCoH-based home ranges, several site fidelity metrics can be calculated, including the duration of a visit to a particular point or area of interest (sometimes called residence time) and the rate at which individuals return to them (known as the revisitation or return rate). Used together, these metrics can offer a means of evaluating the relative risk of contact or exposure among individuals in a population.

In Chapter 2, a site-fidelity analysis in this vein is undertaken to determine the nature of exposure risk faced by two different ungulate populations in Etosha National Park. Using the T-LoCoH approach, revisitation and duration rates are calculated for several zebra and springbok during the anthrax season. The work in this chapter also outlines an algorithm for making more consistent comparisons among individuals (both within and between species). The conclusion, with respect to the potential for exposure to an indirectly-

transmitted pathogen dispersed across the landscape, is that higher mean visitation and duration rates should indicate greater heterogeneity of infection risk across individuals in a spatially-structured population. In other words, high levels of site fidelity will effectively increase exposure risk if an infectious reservoir is present in the range, but will actually buffer an individual from exposure if the range is free of relevant pathogens or parasites. Given the fact that the carcass surveillance effort during the period of interest was likely to reveal only about 25% of the actual carcasses deposited during that time [9], we cannot draw any definitive conclusions about the relative risk of each individual. Even so, the overarching pattern suggests that even animals with very similar home ranging behaviors can face very different levels of risk.

Another set of important analytical techniques for evaluating exposure risk in systems harboring an indirectly-transmitted pathogen are habitat selection analyses. In the same literature search as mentioned previously, these methods were also applied quite frequently, with 24 cases using selection functions (e.g., [111]) or performing basic comparisons between habitat types (e.g., [121]). Selection-based approaches were developed to explore landscape-level patterns and extrapolate probabilities of use of different environmental components. These approaches consist of comparing the habitat qualities at points used by animals to the qualities of points that were available to the individual [17, 98]. Ultimately, selection functions of this variety can illuminate landscape features and types preferred by individual hosts or the population as a whole [89]. These patterns of host preference might enable predictions regarding where susceptible individuals might be exposed to disease, particularly when the infectious agent follows predictable patterns of occurrence and persistence based on abiotic environmental variables [24].

In Chapter 3, a step-selection function is developed in conjunction with a niche model of anthrax persistence in the Etosha National Park to evaluate the nature of exposure risk in the zebra population. Rather than simply assessing the overlap of theoretical selection surfaces with projections of anthrax risk on the landscape, the niche model was directly incorporated into the selection framework. The result of such an analysis could offer insight into the selection for or against areas that represent the highest suitability for anthrax persistence. Based on our knowledge of the primary mechanism of infection in ungulate hosts, it follows that susceptible individuals are most vulnerable to exposure when they are foraging [69, 152]. This expectation was evaluated by parsing the individual zebra movement trajectories into component behavioral states and conducting the step-selection function procedure on each state separately. As predicted, the analysis of points collected while the animal was in the foraging state revealed a distinct pattern from that emerging from an analysis of the directed movement state. In addition, this distinction was largely overwhelmed when all of the relocation points were analyzed together. Ultimately, the work of this chapter reveals the potential importance of explicitly accounting for behavior when evaluating habitat selection patterns. The results indicate that zebra might actively avoid high risk areas when they were foraging, but were apparently attracted to such areas when moving in a directed manner.

This pattern could imply that zebra recognize both where and when they are at highest risk and adjust their behavior to reduce their risk of exposure to anthrax.

Modeling exposure

Compartmental models are a nearly universal tool for studying human and wildlife diseases [3, 84], and have been applied to a broad range of host-pathogen systems, with numerous extensions for host-age effects, pathogen-strain effects, or even the influence of pathogens on host behavior. Such models, however, are not easily adapted to account for the effects of landscape and population spatial structures on *risk of infection*. Accounting for this level of variation requires a representative sample of individuals within the population to be tracked and their contact rates with other individuals (direct transmission) or infectious environmental locations (indirect transmission) recorded. Mechanistic models allow researchers to upscale individual patterns (such as behavioral rules or contact patterns) to a broader population, and are frequently used to validate or test experimental results. However, directly upscaling animal behavioral rules into spatiotemporal patterns of disease may require researchers to build individual- or agent-based models (ABM; [65]).

More specifically, ABMs can use step length, turning angle, canonical activity mode distributions, habitat or resource preferences, or even various network-based metrics to generate likely movement paths for all individuals in the population. With basic assumptions about transmission rates as a function of contact duration, these trajectories can be used to simulate disease outbreaks on real landscapes with "real" animal movement principles. A number of ABMs that incorporate mechanistic movement rules to explore disease dynamics have been constructed [13, 38, 149, 11], though the practice of incorporating movement data into such models is by no means universal.

In Chapter 4, an simulation model of the anthrax system is developed to extend the habitat selection framework derived in the previous chapter. Such a model enables the investigation of the role of environmental heterogeneity in the exposure process whereby a susceptible host encounters a locally-infectious zone on the landscape. Using empirical movement tracks to parameterize general rules, alternative possible movement paths can be simulated in a mechanistic fashion. Each movement decision is based on the selection function associated with the behavioral state of the individual. By altering the input layers to these selection functions (e.g., increasing the mean Wetness of the Wetness layer or decreasing the mean Greenness of the Greenness layer), we can explore the manner by which movement patterns shift under differing environmental scenarios. Further, we can distribute a set of simulated LIZs across the landscape and calculate the rate at which a host individual is likely to encounter infectious agents during its normal activities. Ultimately, the simulation model offered insight into the mediating role that movement plays between environmental factors and epidemiological processes; though Wetness generally decreases linearly throughout the anthrax season, there exists a peak in anthrax mortalities in the middle of the season, and a

similar peak emerged in the calculated contact rates as a gradient of Wetness was explored. Thus, the relationship between environmental covariates and exposure was obscured without the incorporation of a host movement component. This has implications for other disease systems harboring environmentally-transmitted pathogens.

In concert, the following three chapters serve to elucidate the role of individual movement behavior in the disease exposure process. Though the metrics and models applied throughout this dissertation are by no means exhaustive, they represent a set of analytical approaches that can aid in decomposing complex disease dynamics. Rather than analyzing data on case incidence or prevalence, these methods offer insight into the potential contact patterns that might ultimately give rise to endemic or epidemic infections. Thus, considerations of host movements, particularly with comprehensive (or simulated) data on the spatial distribution of the infectious agent on a heterogeneous landscape, might serve to manage infection *risk* before any actual transmission events occur. If the movements underlying exposure can be explained or predicted, risk can be managed more effectively in the anthrax system that forms the basis of this work and beyond.

Chapter 2

A cross-validation-based approach for delimiting reliable home range estimates

Eric R. Dougherty Colin J. Carlson Jason K. Blackburn Wayne M. Getz

Originally published in *Movement Ecology* (2017; **5**:19) and reproduced here with the permission of Colin J. Carlson, Jason K. Blackburn, and Wayne M. Getz.

2.1 Abstract

With decreasing costs of GPS telemetry devices, data repositories of animal movement paths are increasing almost exponentially in size. A series of complex statistical tools have been developed in conjunction with this increase in data. Each of these methods offers certain improvements over previously proposed methods, but each has certain assumptions or shortcomings that make its general application difficult. In the case of the recently developed Time Local Convex Hull (T-LoCoH) method, the subjectivity in parameter selection serves as one of the primary impediments to its more widespread use. While there are certain advantages to the flexibility it offers for question-driven research, the lack of an objective approach for parameter selection may prevent some users from exploring the benefits of the method. Here I present a cross-validation-based approach for selecting parameter values to optimize the T-LoCoH algorithm. I demonstrate the utility of the approach using a case study from the Etosha National Park anthrax system. Utilizing the proposed algorithm, rather than the guidelines in the T-LoCoH documentation, results in significantly different values for derived site fidelity metrics. Due to its basis in principles of cross-validation, the application of this method offers a more objective approach than the relatively subjective guidelines set forth in the T-LoCoH documentation and enables a more accurate basis for the comparison of home ranges among individuals and species, as well as among studies.

2.2 Introduction

Dramatic advancements in GPS telemetry devices have enabled researchers to gain a more comprehensive understanding of animal movement behaviors [148]. The decreasing costs of such devices have resulted in their widespread deployment and a capacity for data collection at unprecedented spatial and temporal resolutions [39]. Movement ecology has emerged as a discipline in its own right [115], with numerous methods and tools being developed and disseminated to analyze the wealth of available data. Ecologists can now quantitatively characterize home ranges and space use patterns over time. Often, the purpose of applying such quantification methods to movement paths is comparison of space use among individuals or species in order to examine such processes as niche partitioning [97, 85], optimal foraging [95, 120], social aggregation [142], or even decision-making [147]. However, many methods require user-defined input parameters, and results are often highly sensitive to the selection of such values. For meaningful comparisons, standardization is required [52], yet protocols to achieve consistency across applications are often non-existent.

One of the most fundamental concepts in movement ecology is the home range, conventionally defined as "the area traversed by the individual in its normal activities of food gathering, mating, and caring for young" [22]. Despite the apparent simplicity of this definition, the statistical formalization of the home range remains challenging, with alternative approaches emphasizing different aspects of animal movement and space use. The lack of a shared underlying theoretical framework makes comparison and standardization among methods all the more difficult, and the practical implications of selecting a particular conception of the home range make such considerations important.

Methods for home range delineation have evolved substantially since the concept of the home range first emerged in the literature [22]. The Minimum Convex Polygon (MCP) method was the most commonly used in the early years of home range description [68], despite its sensitivity to outliers [42] and its inability to further partition internal space [117]. The MCP-based conception of the home range lends itself naturally to some principles of space use in behavioral ecology, such as the general rule that individuals of territorial species often exhibit larger home ranges in relatively lower quality habitat. Kernel Density Estimation (KDE; [169]) emerged as a popular alternative that overcomes some of the limitations of the MCP method, but numerous parameter choices make comparisons among studies tenuous and replication of results difficult [87]. The KDE-based conception of the home range offers a probabilistic framing of animal space use, but may obscure some of the uncertainty inherent in movement data extracted at discrete time points. Both of these methods and their descendants also treat input points as independent, an assumption that is frequently violated with regularly sampled positions from movement paths. Efforts to overcome this inherent autocorrelation have included resampling or weighting algorithms [136, 82], but more recently, methods like the Brownian Bridge Movement Model (BBMM; [75]) and autocorrelated KDE (AKDE; [53]) have been developed to explicitly incorporate the serial nature of movement data. These more nuanced conceptions of the home range and movement behaviors account statistically for uncertainty and autocorrelation, but reliance on random walk dynamics and related assumptions may not account for the behavioral dependency of animal movements [86]. While some of the earlier home range delineation methods could be built for multiple individuals simultaneously, many of these more rigorous methods are parameterized for each individual separately.

The recently developed Time Local Convex Hull method (T-LoCoH; [96]) builds upon the non-parametric LoCoH method [59] by explicitly integrating the temporal component of movement data, effectively scaling time with distance in the construction of local point sets, or hulls. Essentially, this method is governed by a simpler, MCP-based conception of the home range, but works at a finer spatiotemporal scale and enables extension to a more probabilistic description of space use. The T-LoCoH algorithm constructs a utilization distribution (UD) by aggregating local convex polygons, or hulls, built around each point. The hulls are created by selecting the k nearest neighbors of a given point and then sorted by density and merged together to form the UD. The selection of nearest neighbors can be modified by the inclusion of a dimensionless scaling parameter s, which transforms the time interval between points into a third axis in Euclidean space. The distance between points in this three-dimensional volume is called time-scaled distance (TSD), and it serves to separate points that are far apart in time despite their close proximity in two-dimensional space. Thus, an s value of zero will produce the same home range as the original LoCoH method. Guidelines exist for choosing appropriate values to construct a suitable home range, but much discretion is left to the researcher based on the particular subject of their inquiry [96].

A similar approach relies upon the parameter a, which selects nearest neighbors whose distance from the focal point sums to the value a. This method also requires the s parameter for weighting the TSD, but the alternative parameterization may be especially useful for more adaptive hull creation, such that more densely clustered areas of the movement path result in hulls with more points than areas of sparse usage [96]. A rough sensitivity analysis reveals that small differences in either of these parameters has dramatic impacts on the qualities of the resulting home range. The values of these parameters are also contingent upon the movement path itself, meaning that the paths of individuals of the same (or different) species may not result in comparable home ranges. To make such comparisons ecologically and statistically sound, the procedure must be standardized, but to date no such method exists.

Here I demonstrate the use of a novel cross-validation-based method to optimize parameter value selection for implementing the T-LoCoH algorithm based on the unique qualities of each individual movement path. This approach overcomes much of the subjectivity inherent in the recommended parameter selection protocol [96], circumventing the primary challenge to building and interpreting T-LoCoH home ranges. In addition, this method has the added benefit of enabling comparisons of home range features and derived metrics across individuals, species, and spatiotemporal scales, as the same underlying characteristics are used to select the optimal parameter values. I demonstrate the utility of this method with a case study on herbivore movement in the anthrax-dominated landscape of Namibias Etosha National Park.

2.3 Methods

Case Study

Pathogens indirectly transmitted via environmental reservoirs (e.g., water, soil, or animal excretions) represent a unique challenge for ecologists and epidemiologists. Risk of infection in such cases

will depend upon the particular conditions at reservoirs [76, 151], the feeding behavior of the host [48, 67, 69], and the spatial arrangement of reservoir sites relative to susceptible animals [111], all of which may serve to facilitate or dilute pathogen transmission. Certain characteristics of movement behavior may aid in identifying the variation in risk of infection among individuals of the same and different species, including home range size [14], site fidelity [119, 13], and contact network structure [127, 36]. Comparisons of movement-associated transmission risk across individuals may serve to guide management efforts in areas affected by environmentally borne pathogens by identifying high-risk individuals and areas [92, 153], but a failure to explicitly account for individual differences may preclude robust evaluations of epidemiologically-relevant space use patterns [15]. I applied this novel method to the movement trajectories of individuals from two herbivore species in relation to anthrax (the acute disease caused by the bacterium *Bacillus anthracis*) in Etosha National Park, Namibia. As a disease transmitted via environmental reservoirs, anthrax represents an ideal case study for exploring the connections between individual movement on the landscape and resulting disease risk.

GPS point locations were obtained for individuals of two different susceptible ungulate species during the anthrax season in Etosha National Park, Namibia. For both the plains zebra (*Equus quagga*) and springbok (*Antidorcas marsupialis*), the anthrax season was defined as the five-month period between February 1 and June 31 [153]. Due to differences in the temporal resolutions at which the data were initially collected, subsets of the data were created so that each individual had one point location per hour throughout the sampling period. The total number of points for each individual during this period ranged from 2111 to 3601 (Table 2.1). Any missing data values during the sampling period were estimated using a Kalman smoothing approach [124]. Plains zebra and springbok show no sex-related disparity in infection rate [78]. All five zebra individuals chosen for analysis were female, while four of the six springbok were female and two were male.

Existing Parameter Selection Protocol

The k (number of nearest neighbors) and s (time-scaled to distance) parameter values obtained using the proposed algorithm (below) were compared to those one might select based on the guidelines set forth in the T-LoCoH documentation [96]. In addition, the derived metrics, including visitation rate (the number of visits to a given hull, separated by a pre-defined amount of time) and mean duration (the average number of relocations within a hull during each of those visits) were compared to determine the impact of selecting these alternative parameter sets on epidemiologically and ecologically meaningful measures. Because these values are calculated at the scale of the hull, they are likely to strongly depend upon the size of the hulls themselves, with larger hulls leading to relatively higher duration and lower visitation rates as it becomes more difficult to "leave" a hull. The selection of values for the k and s parameters will therefore have implications on the mean values calculated for each individual.

To select appropriate k and s values using the guidelines, the proportion of time-selected hulls (PTSH) method was used. The PTSH approach calculates the distances between pairs of points under a set of alternative s values, and notes the proportion of pairs that are selected due to their temporal proximity rather than their spatial proximity. Ten repetitions of the method were

implemented for each trajectory and all s values associated with a PTSH between 0.4 and 0.8 were obtained from each run. The median value was then chosen from this set and assigned as the s value for that individual. Using these s values, six potential isopleth sets were created, ranging from k=5 to k=30 in increments of 5. Isopleths are created after the hulls are merged together by taking their union, whereby the i^{th} isopleth contains i-percent of points. The k values used in subsequent analyses were chosen using two independent researchers who were asked to select an isopleth set (or range of sets) that satisfied the minimum spurious hole covering criteria, which calls for the selection of the smallest k value that minimizes the holes present in the core area of the individuals home range. To convey the subjectivity associated with the k selection procedure, both the lower and upper bounds of the ranges of k values selected by the independent researchers were mapped and derived metrics extracted.

Cross-Validation-based Parameter Selection

In developing a cross-validation-based approach to parameter selection, I aim to remove much of the subjectivity in the process and enable the data to inform appropriate values. The cross-validation method depends upon the creation of a series of training and testing data sets. For each set, test points were chosen randomly from the full movement path such that approximately one out of every 450 sampled points was selected as a test point (thus, each point had a probability of 0.002222 of being a test point). To ensure independence of the testing points, the 50 points preceding and following each selected test point were removed from the full dataset, and the remainder was considered the training data. For a path with 3600 points, this results in the selection of 8 test points, on average, for each testing set, leaving 2792 points in the training set. The resulting training datasets therefore consisted of approximately 80% of the original data points (Figure 2.1). To minimize variation in the procedure, this stochastic splitting process is repeated n times (in this case, 100) for each movement path.

A grid-based exploration of parameter space was then conducted (Figure 2.2), whereby each of the training/testing datasets ($i = \{1, ..., n\}$) was analyzed at every combination of k and s values on the grid. This analysis entailed the creation of local convex hulls with k nearest neighbors and a scaling factor of s. In all subsequent analyses, we assume that the scaling of time follows a linear formulation; however, when movement patterns more closely exemplify diffusion dynamics, an alternative equation for the TSD may be more appropriate [96]. The test points ($j = \{1, ..., m\}$) were then laid upon the resulting hulls.

We formulate the probabilities for out-of-sample points by normalizing the LoCoH surface so that the probability of an observation occurring at a particular location can be calculated. This value is obtained by dividing the number of training hulls that contain the test point location ($g_{i,j}$) by the summed area of all training hulls (A_i). Then, the log probability was calculated for each point per training hullset. To avoid log probability values of -∞, test points that were not contained within any hulls were assigned a probability value equal to the inverse of A_i^2, resulting in substantially lower log probabilities. Finally, a single value ($P_{k,s}$) was assigned to each combination of k and s value by summing across all of the test points in all of the training/testing datasets:

$$P_{k,s} = \sum_{i=1}^{n} \sum_{j=1}^{m} \log\left(\frac{g_{i,j}}{A_i}\right)$$

Because the probability of each test point is normalized based on the total area contained within all of the training hulls, there exists a natural penalty for high k values. For example, a k value equal to the number of training points (k_{max}; regardless of the s value) will result in all hulls being identical and each test point overlapping all of the hulls. However, the large total area of the hullset when $k = k_{max}$ will result in relatively small probability values for each test point (i.e., independent probability values equal to the inverse of the area of one of the hulls), effectively penalizing the parameter set containing k_{max}. The underlying cross-validation procedure could very easily be extended for the optimization of the the adaptive parameter in the a-method (as opposed to the k-method) because of its scaling with the total area of the hullset.

Despite the use of a testing and training dataset in the creation of the hullsets, we deemed that the use of a measure of sensitivity versus specificity, such as the receiver operator characteristic (ROC) curve, would not serve as an effective means of comparing alternative parameter sets. While false negatives (i.e., test points that are not contained within any hulls) are certainly easy to measure, but without some form of pseudo-absence point, one cannot easily obtain a false positive rate (i.e., points that fall within the home range defined by the hulls, but not actually a point occupied by the animal). Rather, the log probability measure was chosen, as test points can be penalized for being false negatives by assigning a consistent small value as its probability, but there is no need to create pseudo-absence points or account for false positives in any way.

The grid-based search of parameter space allows for the identification of the combination of s and k values that maximizes probability (Figure 2.2). In the case that multiple k or s values result in the same probability, a ridge will appear in the surface. Along these ridges, any of the values can be used and treated as optimal, but for our purposes, we will select the minimum value of k or s associated with the maximum probability value.

2.4 Results

An efficient grid-based search algorithm covered s values from 0 to 0.05 and k values between 4 and 800. The algorithm searches across the broadest set of k values in intervals of 20 and s values in intervals of 0.01. Upon identifying a peak in the probability surface, the algorithm selects a range of 40 k values around the peak and refines the search there in k value increments of 5. Finally, another range of 10 possible k values is selected and the finest scale grid-search is conducted in intervals of 1 and s value intervals of 0.001 before selecting the optimal parameter set.

In the subsequent statistical analyses, the results of paired t-tests are presented to demonstrate the significance of differences when the proposed method was used relative to the guide-based parameter selection criteria, beginning with the k and s parameters themselves (Table 2.1). The mean k value selected using the algorithm for springbok ($N = 6$) was 225.5 (SE = 66.83) and

for zebra ($N = 5$) was 347.2 (SE = 54.36). Because a range of k values were selected using the T-LoCoH guidelines, the mean k values from each individual are used for comparison with the algorithm parameters. The mean of these mean values was 22.5 (SE = 1.71) for the springbok and 20 (SE = 1.58) for the zebra. In both springbok and zebra, these k values were significantly different from those selected by the algorithm ($p = 0.03$ for springbok and $p = 0.004$ for zebra). The mean s value selected using the algorithm for springbok was 0.02 (SE = 0.008) and for zebra was 0.0012 (SE = 0.0005). The mean s value selected using the guidelines was 0.005 (SE = 0.002) for springbok and 0.017 (SE = 0.002) for zebra. In the case of the zebra, the optimal s values according to the two approaches were significantly different ($p < 0.001$), but in the springbok, the s values were not significantly different ($p = 0.10$).

In terms of the area of the home ranges resulting from each parameter set (Table 2.2; Figure 2.3), comparisons were conducted using both the low and high values from the range of the guideline-based parameters relative to the algorithm-based parameter set. The mean home range area for springbok using the algorithm was 401.64 km^2 (SE = 127.56 km^2) and 1081.29 km^2 (SE = 162.17 km^2) for zebra. The mean home range area for springbok using the low value of the range based on the guidelines was 251.22 km^2 (SE = 72.51 km^2) and 660.84 km^2 (SE = 74.30 km^2) for zebra. The mean home range area for springbok using the high value of the guideline-based range was 265.41 km^2 (SE = 76.23 km^2) and 694.43 km^2 (SE = 80.81 km^2) for zebra. Whether considering the lower or upper value from the range of k values based on the guidelines, the difference between home range sizes derived using the algorithm were significant for both springbok ($p = 0.04$ and $p = 0.05$, respectively) and zebra ($p = 0.01$ and $p = 0.01$, respectively). The algorithm-based home ranges were larger for both species, likely because of the significantly higher number of nearest neighbors used in constructing hulls.

For the derived fidelity metrics, duration (Table 2.3) and visitation (Table 2.4), comparisons only concern the mean values of each metric for each individual, though other descriptive statistics of the distribution of all duration and visitation values may be of interest in some cases. The mean duration (MNLV) for springbok using the algorithm values was 57.91 (SE = 14.65) and for zebra was 49.20 (SE = 13.57). Mean duration derived using the low values in the range of s and k values obtained based on the guidelines were between 21.47 (SE = 3.84) for springbok and 9.72 (SE = 0.47) for zebra. The mean duration derived using the high values in the guideline-based range were 24.35 (SE = 4.20) for springbok and 11.11 (SE = 0.49) for zebra. Whether considering the lower or upper end of the ranges for springbok, the mean duration values were not significantly different from the values derived using the algorithm ($p = 0.10$ and $p = 0.08$, respectively). In the case of the zebra, on the other hand, the guideline-based duration values were significantly different from the algorithm-based values, no matter the selection of the lower or upper k value from the ranges ($p = 0.04$ and $p = 0.05$, respectively).

The mean visitation rate (NSV) for springbok using the algorithm values was 46.18 (SE = 34.47) and 23.61 (SE = 9.62) for zebra. Mean visitation rates derived using the low value from the range of k values obtained using the guidelines were 8.38 (SE = 2.06) for springbok and 8.39 (SE = 1.71) for zebra. Using the high value from the guideline-based range, the mean visitation rate is 9.00 (SE = 2.27) for springbok and 9.40 (SE = 1.77) for zebra. In springbok, the visitation rate derived from the algorithm is not significantly different from the values derived using either the

lower or upper k values from the guideline-based ranges ($p = 0.33$ and $p = 0.33$, respectively). The same is true of the visitation rates in zebra, where the algorithm-based value was not significantly different from those derived from the lower or upper values from the ranges of k based on the guidelines ($p = 0.13$ and $p = 0.15$, respectively).

2.5 Discussion

The concept of the home range remains a contentious one, with some researchers suggesting that the choice of delineation method should be defined by the question at hand [52]. When comparison is an element of an analysis, however, standardization of sampling protocols and estimation techniques is required [15, 117]. Considering the multitude of statistical issues overcome by the T-LoCoH method, it should become an increasingly prevalent tool for such analyses. Therefore, eliminating subjectivity from the procedure represents an important step for enabling comparisons both within and among species and studies.

One important consideration is that the "true" k and s values are inherently unknowable. Even the use of simulation methods, which would offer perfect knowledge of the position of an agent at any given time, would not enable the construction of a "true" home range because that would entail the selection of one particular conception of the home range. The approach laid out here offers one such conception, where consistency, as measured by the ability to capture testing points in home ranges created using a subsample of the full movement trajectory, is valued above other measures, such as contiguity or inclusion. By applying this conception of the home range to movement data from different individuals or species, the proposed method effectively unifies the resulting home ranges, enabling further comparison.

Recent empirical studies utilizing the T-LoCoH algorithm for delineating home ranges illustrate the subjectivity involved in parameter value selection [45, 80, 162]. While many studies rely upon the guidelines set forth in the Tutorial and Users Manual provided by the creators of the `tlocoh` package in R [96], there was some variation among studies regarding the selection of s values (i.e., choosing different proportions of hulls that are considered time-selected) and whether the k or a approach was used for selecting nearest neighbors. Most of the home range studies applying the T-LoCoH method do so across multiple individuals, and researchers must decide whether to select separate parameter values for each individual or to have a single overarching parameter set. This decision is particularly important in cases where multiple species are being compared [138], as attribution of differences in home ranges to actual ecology rather than parameter choice may be muddled. Most troubling, however, is the fact that several studies implementing T-LoCoH neglect to specify the parameter values they ultimately used for their analyses, making replication of results nearly impossible.

With regard to the decision about a single parameter set used across individuals or separate sets for each movement path, we argue that consistency and comparability does not emerge from the parameter sets themselves. Rather, the resulting home ranges can be unified by the home range conception that guided their creation. As previously mentioned, the method proposed here serves as that unifying conception, prioritizing consistency in the home range through the use of a

cross-validation approach. In order to construct such a home range for a particular individual, a very different parameter set from another individual may be necessary. Thus, we recommend the use of the proposed algorithm (and the underlying conception of the home range upon which it is built) to make home range analyses more readily comparable between movement tracks.

The results from this case study indicate several important trends. The first is that the s values selected using the algorithm are similar to those selected using the proportion time-selected hulls (PTSH) method, but the optimal k values are significantly different when using the cross-validation based approach. This difference, in turn, affects the sizes of the home ranges delineated by the T-LoCoH method. The home ranges generated from the example paths used here were larger when the algorithm-based parameter sets were applied, without exception. There is a natural inclination to reject larger home ranges as they approach the size and shape of the MCP, but in certain cases (e.g., relatively circular home ranges), the MCP might represent a very reasonable delineation and should not be dismissed simply because of its size.

Using the algorithm, the derived metrics of mean visitation and duration rate lead to some notable clusters that do not follow the division between species. In particular, one of the zebra and one of the springbok were assigned an optimal s value of zero. Though the optimal k values for the two individuals were very different (156 for the zebra, on the lower end of the spectrum and 554 for the springbok, the highest k value selected), the derived metrics of these two individuals stand in stark contrast to the nine other paths examined here. These two individuals had substantially smaller duration values and much higher visitation rates. This trend might indicate that there exists a cluster of more exploratory individuals and another cluster of more residential animals, with both clusters containing individuals from both species. This pattern is not readily observed using the guideline based parameter set, meaning that a very interesting dynamic could be missed if the cross-validation approach is not used. It is unclear exactly what leads to the selection of an s value of zero; it could be some particular geometric pattern or a particular pattern of space use throughout the range. No matter the cause, the cross-validation based approach might be revealing fundamental characteristics or clusters that cut across species. Such clusters could be of great significance in ecological terms, especially when considering the potential for exposure to pathogens.

Differences in site fidelity metrics can have important ecological implications. For diseases like anthrax, which are caused by indirect pathogen transmission at environmental reservoirs, if a locally infectious zone (LIZ; [27]) is present within the home range of an individual, a greater level of site fidelity is likely to place the individual at repeated and extended risk of encountering the pathogen. However, this same high site fidelity may protect an individual against exposure if there are no LIZs in the home range. Consequently, higher mean visitation and duration values are likely to produce a greater level of heterogeneity of infection risk for individuals within a spatially structured population. More research will be required to determine whether duration or visitation is most likely to result in contact with locally-infectious zones, but the fact that similarities in these site fidelity metrics are shared across species suggests that certain movement types, rather than species, may be more vulnerable to exposure.

In the case of these particular herbivores in the Etosha system, this difference in heterogeneity

may be observed in the relative likelihood of a lethal versus non-lethal infection in the two species. The zebra population in Etosha is approximately 13,000 and the springbok population is estimated at 15,600 [27]. After accounting for imperfect detection [9], carcass surveillance data from 2000-2013 suggest that the mean annual mortality rate directly linked to anthrax is approximately 1.34% (95% CI: 0.80% - 1.88%) in zebra and 0.26% (95% CI: 0.18% - 0.35%) in springbok. Additionally, the rate of sub-lethal exposure as indicated by the existence of antibodies in blood serum samples is between 52% and 87% for zebra and between 0% and 15% for springbok [27]. Based on the high values of non-lethal infection, the annual rate of a zebra exposed to anthrax experiencing a lethal dose is approximately 1.5% whereas exposed springbok experience a lethal dose at an annual rate of approximately 1.8%. This suggests that the zebra population may experience higher overall exposure rates to the pathogen, but because of their relatively low mean duration, a large proportion of the exposed population will contract a non-lethal dose, as they will move on from LIZs relatively quickly. The greater mean duration value observed in the springbok population would lead to expectations that some individuals will experience high doses based on repeated and lengthy visits to LIZs or no exposure, with moderate, non-lethal exposure being fairly rare.

The same principles can be applied to other disease systems, where indirect pathogen transmission may be linked to the spatial overlap of a species shedding a pathogen into the environment and naive hosts of another species contacting the pathogen during commingling, as in the case of brucellosis [131]. Commingling, frequently calculated as a function of home range overlap, is a common measure of inter-specific transmission risk, particularly between livestock and wildlife (e.g., bovine tuberculosis [20, 132]). The use of the algorithm enables the construction of comparable home ranges among different species with greater confidence, thereby overcoming one of the most important challenges of using and interpreting T-LoCoH and allowing for a broader application in multi-species disease systems.

Finally, the concept of the probabilistic home range was an important advancement in the home range literature [87], but in the case of T-LoCoH, where isopleths are built atop a series of hulls, the resulting home range may represent an overfitting to the data (Figure 2.4; panel b). As such, this process may be useful for identifying core areas, but may overlook corridors or treat such outlying landscape features as part of the core area by altering the parameter set to fill in 'holes' in the home range. The guidelines aim to minimize holes in the core area of the home range, but because they are based on the probabilistic isopleths, the hulls may need to grow considerably (i.e., the k value must increase) before the underlying hulls predict presence in those areas. Using the hulls underlying those isopleths themselves may represent an underfitting to the data (Figure 2.4; panel e), in essence, a return to the MCP concept whereby too much unused space would be considered suitable. The algorithm circumvents the intermediate step of using isopleths by minimizing holes in the hullset itself. The home range that one builds from these hulls may therefore represent an ideal trade-off between the overfitting of the isopleths and the underfitting of the hulls at an inflated k value.

2.6 Conclusion

Here I present a unifying protocol for parameter selection based on a cross-validation approach. Using the hulls created by the T-LoCoH method as the guiding element for choosing appropriate s and k values, one can maximize the information content of the home range, penalizing parameter sets that resemble the uninformative MCP while maintaining a level of generality that allows for inference beyond the telemetry points themselves. This approach enables consistent comparisons among the derived metrics of different individuals and species, as well as among different time periods, removing subjectivity from the T-LoCoH parameter selection process. The lack of a unifying conception of the home range contributes to the broad and inconsistent application of the term throughout the movement ecology literature and beyond. While the method proposed here has its own assumptions, it offers an objective alternative that can be applied across taxa and study sites to unify results. Ultimately, standardization will facilitate a more explicit connection between animal movement and a particular conception of space use patterns with major implications for the conservation and management of wildlife.

2.7 Tables

Table 2.1: Parameter values for analysis. The s and k values selected using the algorithm and the guidelines in the T-LoCoH documentation. A range of k values were used for the Guide due to the subjective nature of parameter selection.

ID	Species	Sample Points	s (Algo)	k (Algo)	s (Guide)	k Range (Guide)
AG063	Zebra	2111	0.003	355	0.023125	20-25
AG252	Zebra	3601	0.001	485	0.0140625	20-25
AG253	Zebra	3601	0	156	0.0140625	25-30
AG255	Zebra	3601	0.001	405	0.0184375	20-25
AG256	Zebra	3601	0.001	335	0.0171875	15-20
AG205	Springbok	2887	0.05	182	0.003125	25-30
AG206	Springbok	3601	0.023	187	0.00875	25-30
AG207	Springbok	3601	0.036	155	0.01140625	20-25
AG209	Springbok	2887	0.013	171	0.002421875	25-30
AG214	Springbok	2887	0.001	104	0.00265625	15-20
AG215	Springbok	2883	0	554	0.00328125	25-30

Table 2.2: Home range areas (in square kilometers). The total area of the home range obtained using the parameter sets recommended by the algorithm and by the guidelines set forth in the T-LoCoH documentation.

ID	HR Area (Algo)	HR Area (Guide Low)	HR Area (Guide High)
AG063	1092.66	570.65	602.61
AG252	1486.18	913.26	958.41
AG253	593.08	501.23	513.06
AG255	871.31	578.60	600.14
AG256	1363.21	740.47	797.94
AG205	369.61	256.03	268.42
AG206	972.84	557.74	587.54
AG207	429.81	298.64	317.98
AG209	347.19	207.30	215.62
AG214	32.22	23.01	25.46
AG215	258.16	164.57	177.42

Table 2.3: Mean duration (MNLV) values. The derived metrics obtained using the parameter sets recommended by the algorithm and by the guidelines set forth in the T-LoCoH documentation.

ID	MNLV (Algo)	MNLV (Guide Low)	MNLV (Guide High)
AG063	48.94	10.02	11.32
AG252	77.32	10.37	11.74
AG253	2.61	10.71	12.45
AG255	75.11	9.50	10.34
AG256	42.04	8.00	9.70
AG205	92.61	24.38	27.10
AG206	80.78	14.32	16.42
AG207	67.92	12.26	14.47
AG209	78.92	23.41	26.04
AG214	24.68	16.54	19.43
AG215	2.57	37.89	42.63

Table 2.4: Mean visitation (NSV) values. The derived metrics obtained using the parameter sets recommended by the algorithm and by the guidelines set forth in the T-LoCoH documentation.

ID	NSV (Algo)	NSV (Guide Low)	NSV(Guide High)
AG063	13.82	5.82	6.58
AG252	9.11	5.64	6.30
AG253	61.49	15.00	16.04
AG255	19.70	8.07	9.46
AG256	13.95	7.43	8.62
AG205	7.06	4.24	4.50
AG206	8.19	6.46	6.90
AG207	17.83	14.85	15.67
AG209	5.71	3.60	3.80
AG214	20.21	14.56	16.3
AG215	218.07	6.57	6.83

2.8 Figures

Figure 2.1: **Conceptual Figure of the Proposed Algorithm** A test case of the algorithm using a simulated movement trajectory of 1000 relocation points (a). Three of the subsets of those points, with red points indicating those locations that remain in the training sets and blue points representing the test points for the later probability calculation (b,f,j). For each subset of points, a hullset is created using T-LoCoH, with an arbitrarily chosen s value of 0.5 and k values of 5 (c,g,k), 15 (d,h,l), or 25 (e,i,m). These three subsets serve to illustrate three possible scenarios as the k values increases: either test points that are not covered by the hull set at low k values continue to be uncovered with high k values (left-hand column), test points that were not originally covered by the hull set at smaller k values becomes covered (center column), or test points are covered at low k values and continue being covered at higher k values (right-hand column).

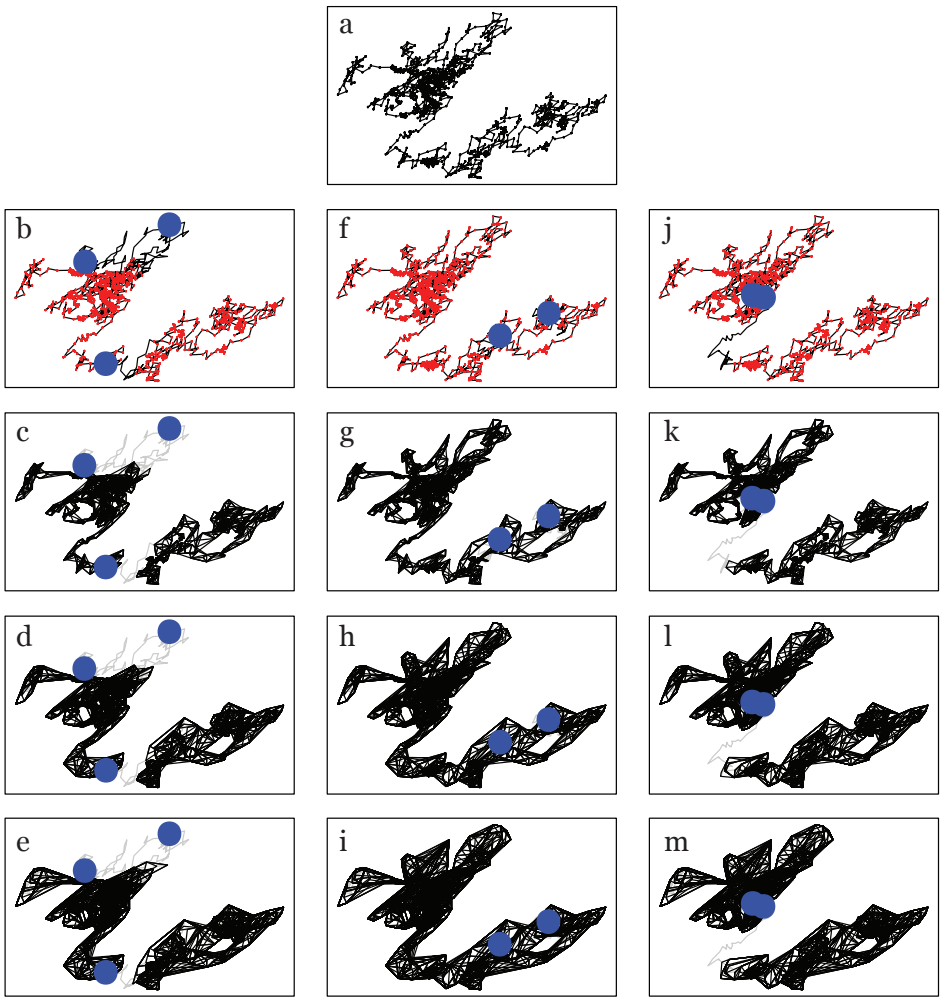


Figure 2.2: **Conceptual Figure of Grid-based Search** A probability surface is generated as the algorithm searches over a grid of alternative s and k values for each individual movement path. The increments of the grid can be chosen by the user. The peak in the surface indicates that the home range associated with the particular parameter set offers the highest probability given the test points. Here, the white boxes denote the maximum probability value, and thereby, the optimal parameter set.

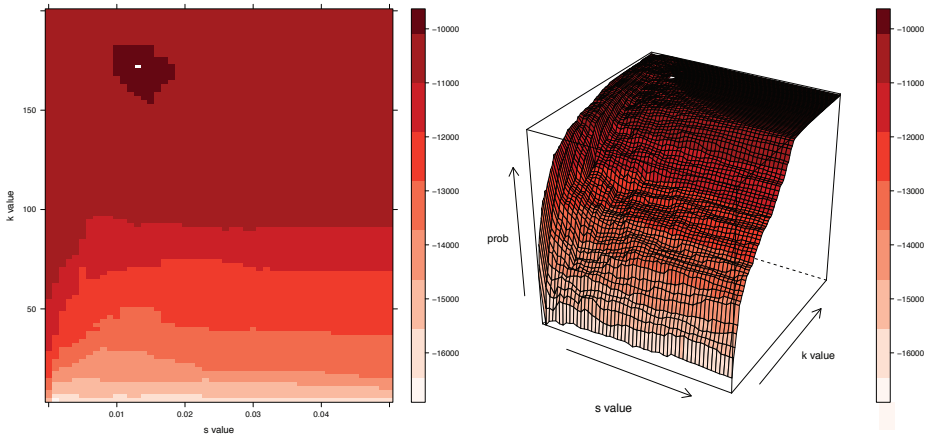


Figure 2.3: **Comparison of Resulting Home Ranges** An illustration of two sets of home ranges that result from the parameter sets chosen by the algorithm (red), the low range of the guide (blue), and the high range of the guide (black). The home range set on the left is based on the sample points from the springbok AG207, and the largest home range covers 429.81 km^2. The home range set on the right is based on the GPS fixes from zebra AG256, and the largest home range covers 1363.21 km^2.

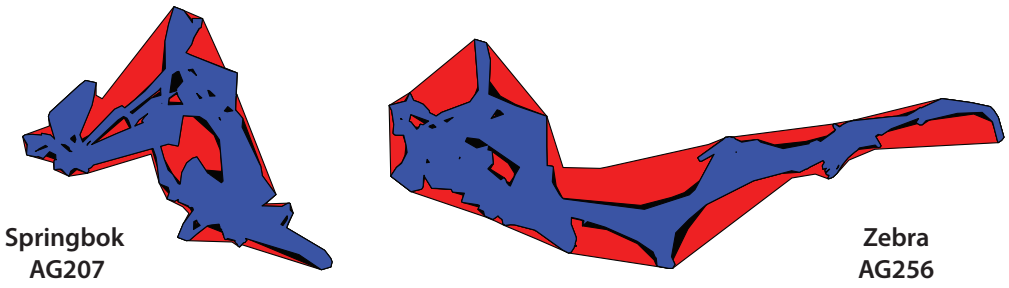
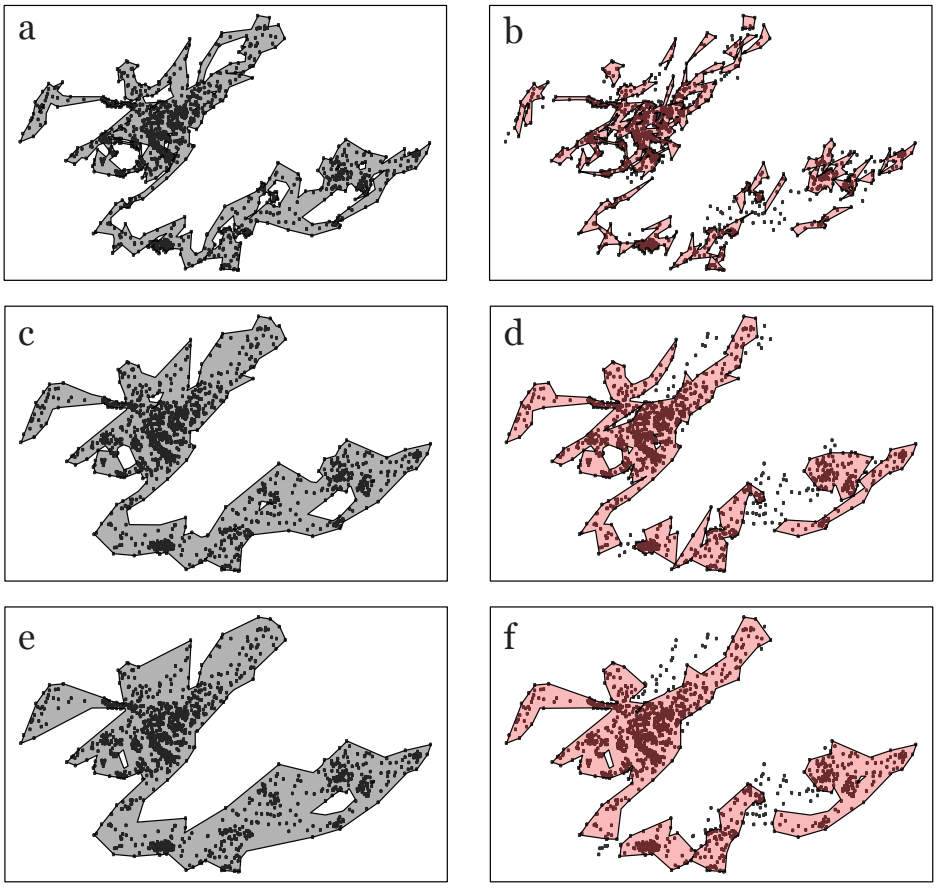


Figure 2.4: **Hulls versus Isopleths** Using the simulated movement trajectory, home ranges can be delimited using the hulls themselves (a,c,e) or the isopleths (b,d,f) derived from the level of overlap among hulls (in this case, the 95% isopleth is displayed). When the k value is relatively small (k=5), the hulls (a) outline the movements of the animal very closely, offering insight, not only into core areas, but also potentially important movement corridors. Using isopleths (b) at low k values may result in large holes throughout the home range while failing to capture corridors. At moderate and high k values (c,d,e,f), both the hulls and isopleths begin to fill in many of the ancillary features, delimiting similar home ranges at slightly different rates (i.e., at k=25, the isopleths (f) resemble the home range outlined by the hulls at k=15 (c)). This illustrates the issue of underfitting when using hulls at high k values and overfitting when using isopleths at low k values. The algorithm proposed here serves to balance these two scenarios as effectively as possible.



Chapter 3

Do zebra mitigate exposure to anthrax when foraging? Incorporating behavior into spatial disease risk models

Eric R. Dougherty Dana P. Seidel Wayne M. Getz

Originally formatted for submission to the *Journal of Animal Ecology* and reproduced here with the permission of Dana P. Seidel and Wayne M. Getz.

3.1 Abstract

Despite behavior being an important contributor to habitat selection, the incorporation of behavioral analyses in spatial disease risk models has lagged behind other disciplines (e.g., conservation biology). In disease systems where spatial distribution models can be developed for a pathogen, parasite, or vector, the particular habitat preferences of host individuals will affect their probabilities of exposure. If habitat selection at behavioral time scales differs markedly among behaviors, then exposure risk may change dramatically when behavior is taken into account. Here I parse animal movement trajectories into behavioral states and construct step-selection functions (SSFs) conditioned on these behavioral states, incorporating pathogen exposure risk as one of the variables in these SSFs. For the model of anthrax persistence, I constructed a Maximum Entropy (MaxEnt) model to associate soil, bioclimatic, and vegetation variables with spore prevalence data from 40 zebra carcass sites detected between 2010 and 2012. I then used a hidden Markov model (HMM) to identify foraging and non-foraging behavioral states along the movement pathways of nine zebra during the 2009 and 2010 anthrax seasons in Etosha National Park, Namibia. Using the MaxEnt

output as a potential predictor variable, I construct a set of SSFs to describe zebra movements and compare the patterns of selection arising during the foraging and directed movement states. Finally, these were compared to the results of an analysis of the entire movement trajectory to demonstrate the value of behaviorally-conditioned analyses. During foraging, individuals actively avoid areas where the risk of exposure to anthrax is relatively high. Conversely, during the directed movement state, they appear to actively select for these high risk areas, perhaps to move through them when in their less vulnerable state. Pathogen (or vector) distribution models may be misleading with regard to the actual risk faced by host animal populations when specific behavioral states are not explicitly accounted for in selection analyses. To more accurately evaluate exposure risk, especially in the case of environmentally transmitted pathogens, selection functions should be built for each identified behavioral state and then used to assess the comparative exposure risk across relevant states. Caution, however, is required when interpreting results, particularly as they relate to scale.

3.2 Introduction

The manner in which animals use space is dependent on the dynamic interplay between the internal state of an individual and the heterogeneous landscape over which it moves [115]. Heterogeneity, however, has a vast number of contributing factors, ranging from readily measurable features such as vegetation type or canopy cover to more elusive features like infection risk. Ultimately, animal movement decisions must be made based on trade-offs between the benefits of satisfying physiological needs and the costs of potential encounters with competitors, predators, or pathogens [58].

With recent advancements in the technologies that track animal positions through time [165], a number of path segmentation methods have been developed to parse movement paths into behavioral states and more clearly interpret the motivations underlying the decision to move [46, 41]. Such analytical methods offer insight into the space use patterns of animals during specific activity modes, allowing researchers to understand the manner by which resource selection differs depending on the internal state of an individual.

In an impressive meta-analysis of 859 habitat selection studies, [103] identified only nine in which the researchers conducted a multi-level analysis wherein different behavioral states were treated as giving rise to differential habitat selection. These ranged across taxa and geographical regions, from the wandering albatross (*Diomedea exulans*) in the Southern Ocean [94] to the Canada lynx (*Lynx canadensis*) in the Northern Rocky Mountains [145]. In these cases and others, the incorporation of behavioral state resulted in distinctly different conclusions regarding the space use patterns of the animals (e.g., [50, 31, 5, 25, 8, 7, 171]). Noting this important trend, others have emphasized the potential implications of ignoring behavioral state when considering habitat selection and animal space use patterns [166], particularly in the context of conservation [135, 1]. Despite this recognition, applications of behavioral analysis methods in habitat selection studies are still rare.

Notably, considerations of behavioral state have not yet permeated the literature regarding the transmission of disease, where host behavior is a fundamental element of pathogen spread. This is

particularly the case for pathogens that are transmitted via environmental reservoirs. The spatial distribution of such pathogens can be readily modeled using remotely-sensed proxies of various environmental factors [24], making studies of their overlap with host animals especially fruitful (e.g., [111]). The importance of particular behavioral states has not been explicitly acknowledged in studies of disease risk thus far, but the implications of excluding this information could be significant. For example, habitat selection studies may offer insight into the evolutionary struggle between host and pathogen. If a pathogen is able to persist in areas that are favored by a host species, they are likely to induce more infections in the host population (i.e., they represent ecological traps; [88]). On the other hand, host animals may be actively avoiding the areas that present the highest risk of exposure (e.g., [174]), especially during the behavioral states in which they are most susceptible. An analysis performed on a full movement trajectory might result in dramatically different estimates of selection coefficients than an analysis of only periods during which the host is in the vulnerable behavioral state. When ignoring behavior, relatively high selection coefficients for the habitat types that are suitable for the pathogen would result in overestimates of the risk faced by the animal. On the other hand, an analysis that disregards behavior might result in relatively low selection coefficients for those high risk areas, leading to the incorrect conclusion that the animal is safe from exposure. No matter the direction of the difference, the exclusion of behavioral information is likely to lead to an inaccurate interpretation of the relationship between the host and pathogen. Similarly, the direct consideration of alternative behavioral states could illuminate important differences in the ways individuals mitigate risk at particular times. A more accurate assessment of the ways animals utilize available resources for different purposes can inform wildlife managers, particularly in systems that harbor environmentally-transmitted pathogens.

The unique nature of environmentally-transmitted pathogens makes them ideal case studies for demonstrating the impact of the incorporation of behavioral analysis into habitat selection studies. Here I use a set of movement trajectories collected from a system harboring one such pathogen, *Bacillus anthracis*, the causative agent of the disease anthrax. Based on extensive research suggesting that the primary route of anthrax infections in ungulate species is through the gastrointestinal tract [153], one can deduce that the most important behavioral state for judging the risk of infection is foraging [69]. Thus, I compare habitat selection models constructed from points assigned to the 'foraging' versus the 'directed' movement states; where the risk of exposure is likely significantly lower in the latter. In addition, I compare these interpretations of space use to the conclusions one would draw by analyzing the entire movement path as opposed to the behaviorally-conditioned selection analysis. By incorporating maps of predicted pathogen persistence in all of these models, direct comparisons can be made among the differential use of resources across behavioral states. Ultimately, I demonstrate that zebra appear to mitigate their risk of exposure when they are in the behavioral state during which they are most vulnerable. I also illustrate how these patterns may not be as evident when the analysis does not explicitly account for behavior.

3.3 Methods

Movement data preparation

Analyses were conducted on GPS trajectories collected from zebra in Etosha National Park in Namibia between 2009 and 2010. Step-selection functions were developed for the nine zebra (*Equus quagga*) for which GPS points were recorded during the anthrax seasons of those years. The anthrax season was defined as the five-month period between February 1 and June 30, following [153, 40]. This temporal criterion resulted in a dataset consisting of five paths recorded during the 2009 season and six during the 2010 season. By splitting up tracks by season, the nine zebra produced eleven separate paths, with two individuals having long enough trajectories to be represented during both seasons (Table 3.1). The original sampling procedure resulted in two parallel sets of data offset by one minute. The short interval could offer interesting insight into the movement patterns of the zebra (e.g., measuring the degree to which zebra movements are diffusive versus dispersive). However, for the purposes of the analyses here, I focused on the first of the 20-minute data sets by using only the first, third, and subsequent fixes that were all 20 minutes apart.

Environmental covariate derivation

Several covariate layers were used in constructing the step-selection functions. Three continuous variables were derived directly from the available bands of the Landsat 4-5 Thematic Mapper remote sensing data. The Normalized Difference Vegetation Index (NDVI) was calculated using the standard formula:

$$\text{NDVI} = \frac{\text{NIR} - \text{Red}}{\text{NIR} + \text{Red}}$$

where Red is Band 4 (B_4) and NIR is Band 5 (B_5). Greenness (G) and Wetness (W) were calculated based on the tasseled-cap transformation equation presented by [32], which utilizes 6 of the 7 bands in a regression framework to calculate several measures (weights rounded to 2 decimal places):

$$G = -0.29B_1 - 0.24B_2 - 0.54B_3 + 0.72B_4 + 0.08B_5 - 0.18B_7$$

$$W = 0.15B_1 + 0.18B_2 + 0.33B_3 + 0.34B_4 - 0.71B_5 - 0.46B_7$$

These measures were calculated for the four cloudless images available during the 2009 anthrax season (March 22, April 23, May 9, and May 25) and the four cloudless images available during the 2010 anthrax season (February 5, April 10, May 12, and May 28). A single mean layer was then calculated for the Greenness and Wetness metrics in each year. All of these layers were obtained at a resolution of 30 meters.

The mean Wetness and mean NDVI measures were highly negatively correlated at the landscape scale in the region through which the zebra moved ($R = -0.95$ for the 2009 season and $R = -0.94$ for the 2010 season). Although the correlation was not observed in the values associated with the 'used' and 'available' points of every individual, the mean NDVI covariate was eliminated from subsequent analyses. The correlations between mean Greenness and mean Wetness at the landscape scale were relatively low ($R = -0.59$ in 2009 and $R = -0.47$ in 2010), so both were maintained as potential predictors.

To account for the potential impact of human development in the area, primary roads (i.e., those made of tar or gravel) were mapped and recast in the form of a road density layer. The road density layer was created in ArcMap 10.3.1 by calculating the length of road (in meters) per unit area (square meters) in each 30 meter raster cell. This layer exhibited low covariance with the other two continuous variables that were maintained in the predictor variable set, resulting in a total of three potential predictor layers. Other frequently used continuous variables, such as elevation, slope, and aspect, were eliminated *a priori* due to the natural homogeneity of the study site with regard to those variables. Though potentially important, particularly in terms of aggregating anthrax spores, the minute differences in elevation across Etosha National Park were not detectable at the resolution at which data were available.

In order to facilitate direct comparisons among the effects of the predictor variables, the layers were standardized such that the resulting layers had a mean value of zero and unit variance (i.e., the overall mean was subtracted from each value and the result was divided by the standard deviation).

Anthrax risk map

A predictive layer of anthrax risk was created using an ensemble ecological niche modeling approach. Separate maps were created for the 2009 and 2010 anthrax seasons based on the presence-only data gathered from sites in Etosha National Park that contained anthrax spores at least one year after the deposition of a carcass [152]. The carcass data consisted of 40 points at sites that contained non-zero concentrations of anthrax spores (in colony-forming units per gram) during sampling one and two years following initial deposition. Of these 40 sites, 26 were associated with carcasses deposited in 2010, 4 with carcasses deposited in 2011, and 11 with carcasses deposited in 2012. Studies show that individual zebra avoid carcass sites for several months after they are created but are attracted to them during subsequent years, when these sites are still highly infectious [152]. Thus, the risk of infection with anthrax in 2009 and 2010 will depend upon carcass sites from 2007-2008 and 2008-2009, respectively. It should be noted that the carcasses used to derive the anthrax risk layers do not represent an exhaustive record of anthrax-positive carcasses in the Etosha region. In fact, it has been estimated that less than 25% of carcasses from zebra that have died of anthrax are actually observed [9], and this estimate does not account for individuals of other species, such as springbok, elephant and wildebeest, all of which are also susceptible to anthrax. In addition, the set of carcasses that serve as the basis for this model exhibit spatial bias, as they were selected largely because of their proximity and accessibility from the research base in Etosha. This accessibility is directly related to the density of the road network in the vicinity of the base, possibly leading to a confounding relationship. The potential effect of this and possible solutions

in larger datasets are discussed below.

Others have created niche models for the bacterium *Bacillus anthracis*, but due to the site-specific nature of the data used, they tend to be applicable only in the region for which they are built (e.g., [111], [6], [146]). Despite their specificity, these models do offer insight into potential predictor variables for anthrax persistence and can inform the niche model constructed here (Table 3.3). Because the carcass data used in developing this particular niche model represent sites at which anthrax spores were able to persist for multiple years, the risk map does not simply serve as a predictive map of carcasses. Rather, it relates anthrax *persistence* to the soil, bioclimatic, and vegetation covariates at sites previously occupied by a carcass. The presence-only data recorded here lends itself to the application of Maximum Entropy methods [129].

The initial predictor variable set consisted of three general categories: soil characteristics, bioclimatic variables, and vegetation indices. The soil quality was summarized using soil pH in H_2O (pH), organic carbon content (OC), and cation exchange capacity (CEC). All three of these layers were obtained from the SoilGrids database at a resolution of 250m. The five bioclimatic variables consisted of mean annual temperature (bio1), mean temperature range (bio7), annual precipitation (bio12), precipitation of the wettest month (bio13) and precipitations of the driest month (bio14). The latter variable was removed because it lacked variation in the region of interest (i.e., every cell had the same value). The bioclimatic variable layers were obtained from the WorldClim database at a resolution of 30 arc-seconds ($\approx$1 km). The potential vegetation indices were calculated based on the Landsat 7 NDVI 8-Day Composite layers, which were obtained at a resolution of 30m. Due to the disparity in resolution, the soil characteristic and bioclimatic variables were resampled to the finer resolution of the vegetation data. Because anthrax persistence was likely dependent on vegetation trends during the entire period of sampling (i.e., the year of deposition and the two subsequent years), the vegetation metrics were calculated for the three year periods between initial discovery and final sampling. The vegetation indices that served as potential predictors were the mean normalized difference vegetation index (mean_ndvi), maximum NDVI (max_ndvi), minimum NDVI (min_ndvi), and the range of NDVI over the period of observation (range_ndvi). To minimize the covariance between predictor variables used in the model for each year, a Pearson correlation matrix was calculated, and any pairs of variables whose coefficient of correlation was > 0.8 were reduced to a single variable as explained below.

To parameterize a MaxEnt model, the background distribution of covariates must be considered. This requires a sampling protocol in which points are randomly dispersed throughout the region of interest and the values of the environmental covariates at those points (i.e., pseudo-absence points) are recorded for comparison with the presence points. In this case, 500 background points were selected. Due to the association of carcasses with a particular deposition year, these background points were divided into three groups in proportion to the number of observed carcasses in each of the three years (2010, 2011, 2012). The result was a set of pseudo-absence points consisting of 312, 50, and 138 points associated with 2010, 2011, and 2012 seasons, respectively. The covariate values were extracted for each presence and pseudo-absence point according to its deposition year. An initial MaxEnt model was run on these data and the full candidate predictor set using the `dismo` package [71] in R (version 3.4.3; [133]). Following an investigation of the variable contributions to this full model (generated as a standard output of the `maxent` function), variables exhibiting

covariance with another predictor were culled such that the variable in the pair with the higher contribution to the MaxEnt model was maintained and its counterpart eliminated. Finally, another MaxEnt model was run on the reduced predictor variable set.

In order to obtain an anthrax risk map for both 2009 and 2010, the MaxEnt model was projected onto the environmental covariate sets associated with those years (Supplementary Figures 1 and 2). In 2009, this meant that the vegetation indices were calculated over the period from 2007 to 2009, and for the 2010 risk map, the vegetation indices were calculated over the period between 2008 and 2010. These risk layers were then directly incorporated into the selection functions described below.

Behavioral analysis

Each animal's movement path was analyzed using a hidden Markov model (HMM; [123, 122]) in the `moveHMM` package [109]. to probabilistically assign each relocation point to one of three different behavioral states that generally corresponded to resting (state 1), foraging (state 2) and directional movement (state 3). The foraging state is defined by medium length steps (on the order of 10^1 to 10^2) and turning angles with a mean close to zero and moderate directionality (concentration). This state can be differentiated from resting (smaller step sizes on the order of 10^0 to 10^1 and low concentration values indicating a lack of directionality) and directed movement (larger step sizes on the order of 10^2 to 10^3 and higher concentration values indicating persistent direction). The points that were assigned to the 'foraging' and 'directed movement' states formed the basis of the two reduced datasets used for the behaviorally-conditioned step-selection function described below. Each path was analyzed separately to more accurately reflect the variability among individuals and properly parameterize the animal-specific step length distributions used in subsequent analyses. The parameters governing the step length and turning angle distributions of the three behavioral states can be seen in Supplementary Tables 1-11. In subsequent analyses, the means (μ) and standard deviations (σ) in these tables are transformed into the more traditional shape and rate (i.e., the inverse of scale) parameters of the gamma distribution according to:

$$\text{shape} = \frac{\mu^2}{\sigma^2}, \qquad \text{rate} = \frac{\mu}{\sigma^2}$$

Though the three-state model described above was ultimately used for the subsequent analyses, a two-state model was also tested. A simple AIC-based model selection procedure revealed that the three-state model was a more accurate reflection of the animal behavior in question, so it was deemed the most appropriate.

Despite uncertainty in the model estimates of parameters and state assignments, this uncertainty was not propagated between steps in the analysis. In other words, the state assignments were treated as certain (binary) despite their probabilistic nature.

Step-selection function

The step-selection function (SSF) procedure implemented here follows that of [171] (later used in [173] and [172]) with some minor adjustments. Conventional SSF approaches often select a certain number of 'available' points for each 'used' point based on empirical step length and turning angle distributions. This represents an extension of the traditional resource selection framework, wherein 'available' points are selected from within the home range of the animal [81, 17, 99]. Thus, the SSF approach eliminates much of the subjectivity associated with home range delineation methods [87] and directly incorporates the temporally-autocorrelated nature of movement data. The 'available' point selection process of the conventional SSF is modified here to more accurately sample the covariates of the area *within reach* of the animal from its actual location at a specific time. The method proposed by [171] has an additional advantage in that it overcomes the potential bias in inference associated with inappropriate sample sizes of 'available' points [118]. By censusing the entire available area, one can estimate the correct proportions (in the case of categorical variables) and accurately reflect the distribution (in the case of continuous variables) of covariates associated with a given 'used' point. In this sense, the proposed method abides by the context-dependent modeling approach [35], as the value of a cell reflects the attributes of that cell as well as the attributes of the surrounding cells (e.g., the probability of a grassland cell will likely differ if the cell is situated near an urban area rather than other grassland cells).

The SSF method incorporated the following steps:

1. The empirical step length distribution is fitted using an appropriate density kernel. In [171], the authors utilize a generalized Pareto distribution, which has a steep curve and long right tail, to reflect the step lengths of their study species. Here I approximate the empirical steps lengths using a gamma distribution, which has a similar shape to the Pareto distribution but with a shorter right tail. This approach was chosen in light of the recommendation by Zeller *et al.* to threshold the kernel at the 97.5 percentile or use the maximum observed displacement distance. In every instance, the 97.5 percentile value of the fitted Pareto distribution was substantially larger than the maximum displacement distance. The tail of the gamma distribution reflected the observed distribution of step lengths more accurately.

2. After fitting an appropriate kernel, a radius value was calculated as the 97.5 percentile of the distribution, effectively representing the perceptual range at the scale of each 20 minute time step (Table 3.2). This radius was then used to construct a buffer around each 'used' point. The area within this buffer was treated as the 'available' area. Unlike in the [171] method, I chose to forgo the construction of a separate 30-meter buffer around each 'used' point, which is intended to account for the potential error associated with the GPS fixes. Due to the resolution of the underlying predictor layers relative to the locational error of the GPS units, it was deemed unnecessary to incorporate this additional component of uncertainty.

3. A weighted mean was calculated for each continuous variable, where the weight of each cell (represented by the point at the center of the cell) was calculated based on the distance of that cell from the 'used' point and multiplied by the value of the kernel density function at that distance. This procedure places higher weights on areas closer to the 'used' point

and thus, the perceptual range should not be viewed as a uniform buffer. These weighted mean values are paired with the values extracted based on the position of the 'used' points, resulting in two sets of covariate values associated with each locational fix.

4. A selection model is fitted in a conditional logistic regression framework (also known as a case-controlled or paired logistic regression; [19]). This method pairs the set of values arising from each 'used' area with the corresponding set of values calculated in the associated 'available' area [30, 18, 140]. An SSF involving n covariates, when projected onto a landscape, follows the structure:

$$w(x_1, ..., x_n) = e^{\beta_1 x_1 + \cdots + \beta_n x_n}$$

where the β coefficients are estimated by a conditional logistic regression [55, 99]. The β_i, $i = 1, ...n$, can be interpreted as selection ratios (i.e., the relative magnitude of selection) and e^{β_i} can be thought of as an odds ratio. Thus, $\beta_i > 0$ indicates selection in favor of a particular covariate (x_i) compared to the expectation based on the availability of the covariate, whereas $\beta_i < 0$ indicates selection against covariate x_i. The w of a given cell is its SSF score, which can be viewed as the relative desirability based on the combination of environmental predictor values at that location and weighted by the animal's preferences for those predictors.

In generating the selection functions, I note that individual ID was treated as a random effect [63, 44], alongside the fixed effects (Greenness, Wetness, Road Density, and Anthrax Risk). However, I repeated this analysis separately for each of the two seasons, thereby by deriving separate $w_{2009}(x_1, ..., x_n)$ and $w_{2010}(x_1, ..., x_n)$ SSFs. This allowed us to establish whether or not the results were consistent across consecutive years. I also note that in this conditional logistic framework, the 'available' area around the point at time t would be associated with the 'used' point at time $t+1$, thereby linking the foregone options available from the previous point with the point ultimately selected by the animal. For inclusion in the behaviorally-conditioned models, the animal must be assigned to the pertinent behavioral state at time $t + 1$, indicating that the choices available at time t (no matter their state) resulted in that particular behavior.

To allow us to evaluate the sensitivity of the results to the scale of analysis, I increased the scale of the SSF by a factor of three. Thus, I generated a second set of step-selection functions using hourly rather than 20 minute fixes. To determine the behavioral states associated with each hourly fix, I simply used a majority rule, such that the point would be assigned to whichever state the animal was in for two or more of the twenty minute periods. If there was one of each, the animal was assigned to the foraging state, as this represented the most conservative approach. For the hourly step selection function, a new set of buffer lengths were also calculated by fitting a gamma distribution to the hourly step length distributions using all of the points, only the foraging points, or only the directed movement points. These buffer radii are presented in Supplementary Table 12.

The resting state was omitted from the step-selection function analyses because selection is contingent on animals making explicit movement decisions. This determination was made based on the mean step lengths associated with points recorded while an animal was in each state. Because the mean step length during the resting phase was less than 42.4 meters (representing the diagonal distance required to cross a 30 meter resolution cell of the environmental covariate layers used for

this analysis), the individual was deemed unlikely to be making a concerted decision to move (or not move). Alternative thresholds for inclusion could be developed, but given the emphasis on behavioral state here, this seemed most appropriate.

3.4 Results

Anthrax risk map

The anthrax risk maps produced for the 2009 and 2010 seasons reflect the output of the MaxEnt model, built using 40 presence points and 422 background points (after 78 were removed for falling within a region in which data was not available for at least one of the environmental covariates; Figure 3.1, Panel a). These removed points were distributed haphazardly among the different years, so additional adjustments were foregone. The final model was built on a set of nine continuous predictors after the elimination of annual precipitation (bio12) and mean NDVI, which were highly correlated with precipitation of the wettest month (bio13) and maximum NDVI, respectively (see Supplementary Tables 13-15 for the covariance matrices in 2010, 2011, and 2012, and Supplementary Table 16 for the variable contributions associated with the full model). The final model had an AUC value of 0.937. The variable importance table indicates that the bioclimatic and soil characteristics were larger contributors than the vegetation indices (Table 3.4). Mean temperature range dominated the model, contributing 73% to the final model. Soil organic carbon content was the next highest contributor at 11.2%. The vegetation measures contributed a total of only 7% to the model, despite being more temporally specific than the other measures. For the sake of reproducibility, the model coefficients (i.e., lambdas) and associated feature classes emerging from the MaxEnt algorithm are available in Supplementary Table 17.

The final predictive maps for both 2009 and 2010 show that the greatest level of risk occurs at the southwestern edge of the Etosha pan (Figure 3.1, Panels b and c). The geographical range of risk appears to be considerably larger during the 2010 season than during the 2009 season. These differences are likely driven by differences in the vegetation layers because the soil characteristic and bioclimatic variables are static between the years. To quantify the differences in the geographic range of risk, or pertinent transmission zone (PTZ), I set three thresholds representing liberal, moderate, and conservative cutoffs above which the pixel is treated as risky (Figure 3.2). In 2009, the area defined as the PTZ based on the most liberal definition of risk (associated with a suitability value of > 0.1) was approximately 730 km^2. The same liberal cutoff in 2010 results in a PTZ of over 943 km^2. The moderate threshold (with a suitability value of > 0.25) offers a similar impression of the disparity in PTZ size across seasons, with 2009 having a PTZ of about 344 km^2 and 2010 having one of over 463 km^2. Finally, the difference between anthrax seasons is even more pronounced when the most conservative threshold (a suitability value > 0.5) is applied, with 2009 having a PTZ of about 76.6 km^2 and 2010 having a PTZ that is nearly twice as large (133 km^2). The PTZs in 2009 represent 10.4%, 4.9%, and 1.1% of the total area in the $\approx$ 7000 km^2 region of interest in Etosha National Park for the liberal, moderate, and conservative thresholds, respectively. The PTZs in 2010, however, represent 13.5%, 6.6%, and 1.9% of the total area for the same thresholds.

It is worth noting that both risk layers exhibit a minor diagonal striping pattern that is an

artifact of a malfunction in the scan line detector during the Landsat 7 mission. The focus in this case is to demonstrate a methodology for approaching a model of this sort, so I will use the Landsat images without applying any form of correction. Despite the limitations described, the environmental layers upon which the risk layer is built represent the best available data.

Step-selection function

To verify the efficacy of the step-selection function method applied here, two rasters were generated with random values assigned to each cell (based on draws from a normal distribution with a mean of 0 and a standard deviation of 1). These layers were added as potential predictors in the analysis of the largest dataset (consisting of all movement points in 2010). This test allowed us to verify that methodological artifacts, including sample size, were not artificially inflating the significance of various predictors. Though this basic test cannot rule out the possibility of inflated Type I error rates, the results indicate that there are no overt issues that need to be addressed upfront. These results are presented in Supplementary Table 18.

For the sake of comparison, the results of the analysis conducted using all of the movement points, irrespective of the behavioral state, will be presented first, followed by the results emerging from the analysis that incorporates behavior. In the latter case, I will distinguish between the foraging and the directed movement state to determine how selection patterns compare across behavioral modes as well as across years (Table 3.5).

Applying the step-selection framework to all relocation points revealed some consistent trends across the 2009 and 2010 anthrax seasons. Wetness represented the covariate with the largest effect on selection, and in both years, animals actively avoided areas with higher Wetness. In 2009, the avoidance pattern was slightly weaker ($\beta_{\mathrm{W}} = -0.52 \pm 0.039$; $p < 0.001$) than in 2010 ($\beta_{\mathrm{W}} = -0.85 \pm 0.023$; $p < 0.001$), but in both cases, the effects were highly significant. In both years, animals appeared to be slightly, though significantly, attracted to areas with greater Road Density. In 2009, this attraction was also slightly smaller in magnitude ($\beta_{\mathrm{RD}} = 0.01 \pm 0.006$; $p = 0.04$) than in 2010 ($\beta_{\mathrm{RD}} = 0.03 \pm 0.004$; $p < 0.001$). Avoidance of areas with relatively high anthrax risk was consistent across years, although only significant in 2010. In 2009, the negative trend was nearly significant ($\beta_{\mathrm{AR}} = -0.02 \pm 0.012$; $p = 0.06$), but in 2010, it was highly significant ($\beta_{\mathrm{AR}} = -0.06 \pm 0.0108$; $p < 0.001$). The only pattern that was not maintained across years when all of the points were considered was the role of Greenness. In 2009, Greenness was not a significant contributor ($p = 0.86$), meaning that animals were just as likely to select a point with relatively low Greenness values as one with relatively high values given the option. In 2010, however, Greenness was a highly significant predictor ($\beta_{\mathrm{G}} = 0.36 \pm 0.031$, $p < 0.001$), second only in magnitude to Wetness.

When the dataset is parsed into different behavioral states, the results offer a slightly different picture, and also offer insight into the various factors that animals consider when in the foraging versus directed movement state. When considering only the foraging state, Wetness was no longer as consistent a predictor of habitat selection. In 2010, the role of Wetness was negative and highly significant ($\beta_{\mathrm{W}} = -0.30 \pm 0.030$; $p < 0.001$), but in 2009, the effect was actually positive and significant ($\beta_{\mathrm{W}} = 0.23 \pm 0.052$; $p < 0.001$). Comparatively, Greenness was the factor with the

greatest impact on movement decisions during the foraging phase in 2010 ($\beta_{\mathrm{G}} = 0.43 \pm 0.044$; $p < 0.001$). In 2009, the effect of Greenness was not significant ($p = 0.22$). This pattern is consistent with results when all of the points are analyzed. Unlike when behavioral states were ignored, though, the role of Road Density is negligible in both 2009 ($p = 0.61$) and 2010 ($p = 0.75$). Importantly, in both years, foraging zebra appear to consistently avoid the areas of highest risk of exposure to anthrax. Though the effect is slightly stronger in 2009 ($\beta_{\mathrm{AR}} = -0.11 \pm 0.018$; $p < 0.001$) than in 2010 ($\beta_{\mathrm{AR}} = -0.06 \pm 0.011$; $p < 0.001$), the avoidance behavior when animals are foraging is significant across both seasons.

Distinctions between the different behavioral states are clarified by considering the selection patterns that emerge from an analysis of the directed movement points in addition to the foraging points. When animals exhibit directed movements, with longer steps lengths and relatively little variance in their heading, they seem to actively avoid areas with high Wetness. The effect was highly significant in both 2009 ($\beta_{\mathrm{W}} = -2.17 \pm 0.145$; $p < 0.001$) and 2010 ($\beta_{\mathrm{W}} = -2.19 \pm 0.075$; $p < 0.001$). It was this large effect that likely drove the relatively high avoidance patterns when all of the points were analyzed. There were several factors that exhibited notably different effects during the directed movement state than during the foraging state. In 2009, the effect of Greenness was negligible during the foraging state, but animals appeared to actively avoid areas of high Greenness during the directed movement state ($\beta_{\mathrm{G}} = -0.54 \pm 0.130$; $p < 0.001$). The oppositional trend was repeated in 2010, where foraging animals demonstrated a significance preference for higher Greenness, but animals in the directed movement state seemed ambivalent to the level of Greenness ($p = 0.80$). Similarly, Road Density was a significant predictor of selection during directed movement in both 2009 ($\beta_{\mathrm{RD}} = 0.04 \pm 0.012$; $p = 0.001$) and 2010 ($\beta_{\mathrm{RD}} = 0.07 \pm 0.008$; $p < 0.001$), where animals actively selected to be in areas with higher Road Density, but the factor did not appear to significantly affect movement decisions during the foraging state. A possible explanation for this difference is that roads may facilitate directed movement by eliminating potential barriers, making them more attractive for longer distance 'steps'. However, the disparity between the behavioral states was perhaps most notable with regard to role of Risk in movement decisions. When animals were in the directed movement state, they consistently selected areas that correlated with greater risk of exposure, whereas animals tended to actively avoid such high risk areas when they were foraging. This was the case in 2009: β_{AR}: direct $= 0.13 \pm 0.043$ ($p = 0.003$) vs. foraging $= -0.11 \pm 0.018$ ($p < 0.001$). It was also the case in 2010: β_{AR}: direct $= 0.09 \pm 0.024$ ($p < 0.001$) vs. foraging $= -0.06 \pm 0.011$ ($p < 0.001$)

To test the robustness of the results from the above analysis, a parallel analysis was conducted at a coarser scale. Using an hourly fix rate, rather than a 20 minute fix rate, the same general patterns emerge (Table 3.6). Of particular importance are those trends that were consistent across seasons in the fine-scale analysis: the selection coefficients for regions with high road density and areas of high suitability to anthrax. As in the analysis of the fine-scale data, foraging animals appeared to be affected negligibly by Road Density ($p = 0.95$ in 2009; $p = 0.11$ in 2010), but significantly avoided areas with high anthrax risk in both 2009 ($\beta_{\mathrm{AR}} = -0.10 \pm 0.032$; $p = 0.001$) and 2010 ($\beta_{\mathrm{AR}} = -0.10 \pm 0.018$; $p < 0.001$). Similarly, the attraction to higher Road Density during the directed movement state was apparent in both 2009 ($\beta_{\mathrm{RD}} = 0.06 \pm 0.025$; $p = 0.02$) and 2010 ($\beta_{\mathrm{RD}} = 0.08 \pm 0.016$; $p < 0.001$), as was the attraction to areas of higher suitability for anthrax

persistence in 2009 ($\beta_{\text{AR}} = 0.28 \pm 0.057$; $p < 0.001$) and 2010 ($\beta_{\text{AR}} = 0.07 \pm 0.028$; $p = 0.01$).

3.5 Discussion

The selection functions reveal interesting dynamics with regard to the behaviorally-contingent space use patterns of zebra. In both 2009 and 2010, the zebra exhibit distinct avoidance of areas that are most suitable for anthrax persistence when they are in the foraging state (indicated by a negative selection coefficient). Notably, they demonstrate a pattern of attraction (indicated by a position selection coefficient) to these areas of high risk when they are in the directed movement state. This pattern was also reflected when the coarser dataset (hourly fixes rather than 20 minute fixes) were analyzed. Overall, these results suggest that zebra not only recognize where they face the greatest risk of exposure to anthrax, but that they also recognize their own increased vulnerability in the foraging state. Further, the results of the behaviorally-conditioned step-selection functions indicate that foraging animals that approach areas of high risk might actually intentionally alter their behavior, shifting to a more directed mode of movement with longer step lengths and less variance in their heading in order to move quickly through the area and avoid contact with a pathogen. This would explain the apparent selection in favor of these riskier areas during that state. It is important to note that these general trends do not indicate that animals never forage in areas with some risk of exposure; it merely implies that animals exhibit a statistically meaningful preference for areas with lower risk over areas with higher risk while in the foraging state. Similarly, zebra will, on occasion, exhibit directed movements outside of high risk areas, but there exists a meaningful preference such that they are more likely to select for areas of high risk than areas with low risk when they are in the directed movement state. In effect, the zebra analyzed here exhibited an ability to reduce their risk by altering their behavior depending on their proximity to areas with higher potential for exposure to anthrax.

It is unclear exactly what the mechanism underlying this avoidance trend during foraging would be. After a year — and especially after two — carcasses are unlikely to leave visible cues in the vicinity of the deposition site. The hemorrhagic fluid released during the period of decomposition following death is unlikely to leave notable signs after that amount of time, and carcasses themselves tend to be dragged and destroyed by scavenging animals within a few days or weeks (except in the case of larger elephant carcasses, which were not included in the MaxEnt model). In fact, a camera trap study conducted in Etosha National Park actually indicated the opposite pattern, that animals tended to be attracted to the locally infectious zones (LIZs; [56]) due to a vegetation green-up that occurred in seasons following the release of the nutrient-rich fluids from the carcass [151]. One possible explanation for the avoidance behavior is the existence of a confounding variable that was not considered in this analysis. For example, if a particular soil or vegetation type is frequently associated with carcasses, an association might develop in the minds of the vulnerable host population. Avoiding these more obvious and more permanent cues may give rise to the behavior observed.

The alternative interpretation that arises based on the results of the selection analysis conducted here might concern the specific scale at which the analysis was conducted. Given the average size of the locally infectious zone caused by a zebra carcass, it is not especially surprising that an analysis

at a 30 meter resolution could miss some of the finer-scale dynamics. At this scale, the signature of a LIZ site is likely overwhelmed by the averaging of the characteristics of a cell, as the LIZ itself represents only about one-tenth of the area of a cell (though this depends on the species of the animal that succumbed at that location). Thus, at the scale of this analysis, animals may, in fact, select for areas that present lower risk of exposure to a pathogen. This pattern of avoidance has been noted in other species [34, 130], and has given rise to the concept of the 'landscape of disgust' [163]. If, however, an animal ends up in a high risk cell, they might be attracted to the LIZ site within that cell. Thus, an overall avoidance pattern may be observed across the landscape, but attraction at the sub-cell scale. This implies a very important point about habitat selection analyses that are frequently conducted at the finest scale allowed by the environmental data as opposed to the most meaningful scale from a biological perspective. Interpretations regarding pathogen exposure risk might depend not only on the behavioral state of the animal, as investigated here, but could also be influenced by the scale of analysis. It should also be noted, however, that to make use of such fine-scale environmental data, the temporal resolution of the movement tracks would also likely need to be finer, perhaps on the order of 1 minute per fix. At the time that these data were collected in 2009 and 2010, this technology was not widely available, but recent advancements in GPS devices makes such fine-scale movement data eminently collectible.

In this case, I was only able to account for the potential effect of scale by decreasing the resolution of the environmental or movement data. The results presented here are fairly robust when the scale of the movement data is coarsened from a 20 minute fix rate to hourly fixes. Similar patterns were borne out, including the general attraction to areas of high anthrax exposure risk during the directed movement state and avoidance of those areas whilst foraging. Insights regarding the attraction towards areas of high road density and a more pronounced avoidance of wet areas during directed movement were also replicated at this scale.

These patterns, however, are not as stark when the entire movement path is analyzed without accounting for different behavioral states. The distinct divergence in the selection coefficients for the anthrax risk layer that emerges from the behaviorally-conditioned step-selection function is lost in an analysis of the full trajectories. Though the avoidance pattern is still significant in 2010 and nearly significant in 2009, the dynamic in which animals actively select for the higher risk areas during directed movements is no longer apparent. This is likely because of the relatively large proportion of points that were defined as foraging compared to the proportion assigned to the directed movement state. The attraction is effectively overshadowed by the avoidance associated with the state that had the larger sample size, further demonstrating the importance of explicitly considering behavioral states when possible.

The explicit consideration of particular behavioral states in habitat selection studies can offer important insights, especially in systems with environmentally-transmitted pathogens. The unique biology of these pathogens enables them to persist in reservoirs outside of hosts for relatively long periods of time. Anthrax spores, for example, may remain viable in the soil in Etosha for up to seven years [152]; it is possible that they can persist even longer in systems with more vegetation cover, potentially giving rise to episodic infection dynamics [24]. Where environmental persistence is possible, the ability to predict the presence of the pathogen is directly related to the dependence of the pathogen on particular environmental factors. Anthrax exhibits a dependence on soil with a

slightly alkaline pH, high organic matter, and high calcium content [76]. The availability of these remotely sensed data makes it feasible to map the likely niche of anthrax. Ideally, soil samples at carcass sites in the years following deposition are used to map anthrax suitability, as in the MaxEnt model built here. However, this strict criterion substantially limited the number of data points available to inform the anthrax risk layer, from approximately 300 carcasses observed between 2010 and 2012 to only 40 that were subsequently samples for spore persistence. In this way, however, I was able to directly map suitability for anthrax persistence, and thus, exposure risk, as opposed to mapping some proxy, such as habitat preferences of animals weakened by anthrax infection. Importantly, the 40 sites that were sampled in subsequent years were not randomly or haphazardly chosen from the many carcasses in the region of interest. All of these sites were easily accessible from the research base at Okaukuejo, and thus, were spatially biased. Because the region closest to the research base has the highest density of roads in the park, the relationships observed in the SSF models between Road Density and Anthrax Risk (particularly during the directed movement state) may be confounded. Due to the relatively small sample size of LIZ sites, I chose to maintain all of the 40 points rather than applying a subsampling procedure. This is a potentially serious flaw in the derivation of the risk map, and when sample size allows, appropriate methods for reducing spatial bias should be applied. Another alternative approach would involve modeling the probability of spore persistence at a carcass site and projecting that across all observed carcasses (rather than only those that were sampled in subsequent years). This might increase the sample size and enable subsampling approaches to reduce bias, but the propagation of uncertainty between stages of analysis will be important when interpreting the results.

The interpretation of a habitat selection analysis will depend heavily on the environmental factors that serve as inputs to the model. In this case, remotely sensed data from 2009 and 2010 were limited at the resolution of the analysis. Though Landsat data enabled the creation of mean Wetness and Greenness layers to help characterize the region, there are likely a number of other potential abiotic factors that influence animal habitat selection that could not be summarized meaningfully at the the spatial and temporal scale of this analysis. On a similar note, the unique nature of the study site in question limited the number of potentially meaningful environmental covariates. For example, elevation, slope, or aspect are very commonly incorporated into models of habitat preference, but the lack of heterogeneity within the region of interest investigated here precluded definitive conclusions regarding such predictors. There are numerous biotic factors that are very difficult to incorporate meaningfully into such models as well: competition and predation, for example, are likely to influence movement patterns, particularly at the scale of 20 minute steps, but these could not be derived from the available data for this period.

It is important to note some additional drawbacks and potential shortcomings of the methods applied here beyond the issues of scale and data availability. For instance, the selection of a particular analytical method to parse a movement trajectory into the canonical activity modes (CAMs; [58]) or behavioral states of which it is composed might introduce an additional layer of uncertainty. For this study, I have chosen a Hidden Markov modeling (HMM) approach, but alternative approaches might have resulted in different state assignments. In addition to the uncertainty associated with assigning points to particular behavioral states, there is uncertainty in the parameter estimates derived from the model (e.g., step length and turning angle distribution

parameters). In the procedure above, I treated these parameter estimates and state assignments as the basis for subsequent analyses without propagating the associated uncertainty between stages. Though it is difficult to determine the effect of this simplification, it is likely to bias the results to some degree. Evaluating the effect of disregarding uncertainty and devising methods to explicitly account for it will be important considerations when applying behavioral analyses to multi-level selection functions in the future.

In the case of HMMs, the models that emerge are dependent on user inputs, including an *a priori* decision regarding the number of states to which points can be assigned. Here I tested a two- and three- state model and chose the latter based on AIC (and interpretability), but alternative models may have fit the data better than the one that I ultimately applied across all of the individuals. This is another important consideration when researchers intend to incorporate behavior into models of habitat selection. Recently, telemetry devices have been fitted with auxiliary sensors, such as accelerometers [116, 165] that might offer additional clarity to researchers wishing to parse movement tracks. Similarly, several new tracking devices directly account for the movement mode of an animal by altering the positional fix rate based on the current speed of movement, offering classification of steps without additional analyses [1]. However, it is unclear exactly what the ramifications of misclassification would be, and the identification of these effects will be difficult without definitive knowledge of the "true" behavioral states of an animal through time.

Mapping the host selection of potential locally infectious zones on the landscape may help guide managers in identifying areas of high anthrax risk and individuals with high exposure potential given their selection patterns. If animal movements can be isolated to a single behavioral state during which individuals are vulnerable, such analyses can be made even more accurate. Ultimately, integration LIZ and behavioral information might aid in preventing outbreaks of the endemic pathogen. Though other disease systems, including those characterized by transmission via environmental reservoirs, might not involve a particular behavioral state that exhibits a definitively higher level of vulnerability, the consideration of behavior could be important for judging other epidemiological processes, such as contact or succumbing to infection [41]. In the case of the former, particular behavioral modes might result in shifting selection patterns that lead to large aggregations of individuals, thereby placing animals at a higher risk of contacting an infectious conspecific [54]. When investigating the infection process itself, novel selection patterns may be induced by infection with a parasite or pathogen, and these shifts might be apparent in a movement trajectory [114, 33]. The growing availability of fine-scale GPS data and the growing set of analytical methods to infer behavior from such data makes the direct incorporation of behavior an important and exciting avenue for future exploration.

3.6 Tables

Table 3.1: Summary of the eleven regularized zebra tracks for which step-selection functions were developed. Note that individuals AG063 and AG068 had paths that spanned two anthrax seasons, resulting in two separate entries here.

Animal ID	Number of Points	Missing Points	Start Date	End Date
AG059	4,824	5	2009-04-25	2009-06-30
AG061	4,824	152	2009-04-25	2009-06-30
AG062	4,824	646	2009-04-25	2009-06-30
AG063	4,824	7	2009-04-25	2009-06-30
AG068	4,824	11	2009-04-25	2009-06-30
AG063	6,331	86	2010-02-01	2010-04-30
AG068	10,800	2,072	2010-02-01	2010-08-29
AG252	10,800	39	2010-02-01	2010-08-29
AG253	10,800	739	2010-02-01	2010-12-17
AG255	10,800	28	2010-02-01	2010-08-29
AG256	10,800	2	2010-02-01	2010-08-29

Table 3.2: Radii of the kernels (in meters) used in producing the step-selection functions for each individual. Separate radii were used for the full datasets, the foraging only dataset, and the directed movement only datasets.

Animal ID	Kernel Radius (All)	Kernel Radius (Foraging)	Kernel Radius (Directed)
AG059_2009	1131	667	1532
AG061_2009	739	273	1190
AG062_2009	837	240	1148
AG063_2009	985	581	1534
AG068_2009	1183	607	1595
AG063_2010	1256	626	1686
AG068_2010	1236	590	1636
AG252_2010	1012	341	1450
AG253_2010	1101	499	1702
AG255_2010	1056	324	1502
AG256_2010	1014	376	1479

Table 3.3: Set of potential predictor variable layers used in creating the anthrax risk map. These covariates were compiled based on their use in similar ecological niche modeling efforts of *Bacillus anthracis* (see [6] and [111] for more details). Several of these variables were eliminated, however, due to collinearity with other, more important, variables in the set. An 'X' in the 'Final Model' column indicates the inclusion of that variable in the final MaxEnt model.

Environmental variable (units)	**Covariate name**	**Data source**	**Final Model**
Soil pH x 10 in H_2O	pH	SoilGrids*	X
Soil Organic Carbon Content (g/kg)	OC	SoilGrids*	X
Soil Cation Exchange Capacity (cmolc/kg)	CEC	SoilGrids*	X
Mean annual temperature (C°)	bio1	WorldClim†	X
Annual temperature range (C°)	bio7	WorldClim†	X
Annual precipitation (mm)	bio12	WorldClim†	
Precipitation of the wettest month (mm)	bio13	WorldClim†	X
Precipitation of the driest month (mm)	bio14	WorldClim†	
Mean NDVI	NDVI	Landsat 7‡	
Maximum NDVI	max_ndvi	Landsat 7‡	X
Minimum NDVI	min_ndvi	Landsat 7‡	X
Range NDVI	range_ndvi	Landsat 7‡	X

[70] † [72] ‡ data courtesy of the U.S. Geological Survey

Table 3.4: Variable contribution and importance results from the final MaxEnt model, built on the reduced environmental covariate set following the elimination of annual precipitation (bio12) and mean_ndvi due to covariance.

Variable	**Name**	**Percent contribution**	**Permutation importance**
Mean temperature range	bio7	73	80
Soil Organic Carbon Content	OC	11.2	2.6
Precipitation of the wettest month	bio13	6.5	7.1
Range of NDVI	range_ndvi	4.7	2.3
Maximum NDVI	max_ndvi	2	1.6
Mean annual temperature	bio1	1.2	1.5
Soil Cation Exchange Efficiency	CEC	0.6	2.1
Soil pH	pH	0.5	0.5
Minimum NDVI	min_ndvi	0.3	2.3

Table 3.5: Results of the conditional logistic mixed effects models as applied to all of the movement points (n = 22,949 in 2009 and n = 56,495 in 2010), only the foraging points (n = 11,733 in 2009 and n = 27,898 in 2010), and only the directed movement points (n = 4,381 in 2009 and n = 11,486 in 2010).

2009 All Points	coef	exp(coef)	se(coef)	z	p	sig
Wetness (β_{W})	-0.52	0.59	0.039	-13.26	<2e-16	***
Greenness (β_{G})	-0.01	0.99	0.031	-0.18	0.86	
Road Density (β_{RD})	0.01	1.01	0.006	2.03	0.04	*
Anthrax Risk (β_{AR})	-0.02	0.98	0.012	-1.91	0.06	.

2010 All Points	coef	exp(coef)	se(coef)	z	p	sig
Wetness (β_{W})	-0.85	0.43	0.023	-36.58	<2e-16	***
Greenness (β_{G})	0.36	1.43	0.031	11.38	<2e-16	***
Road Density (β_{RD})	0.03	1.03	0.004	6.16	7.1e-10	***
Anthrax Risk (β_{AR})	-0.06	0.95	0.008	-7.57	3.8e-14	***

2009 Foraging Points	coef	exp(coef)	se(coef)	z	p	sig
Wetness (β_{W})	0.23	1.25	0.052	4.34	1.4e-05	***
Greenness (β_{G})	0.05	1.05	0.042	1.22	0.22	
Road Density (β_{RD})	-0.00	1.00	0.008	-0.50	0.61	
Anthrax Risk (β_{AR})	-0.11	0.90	0.018	-6.05	1.4e-09	***

2010 Foraging Points	coef	exp(coef)	se(coef)	z	p	sig
Wetness (β_{W})	-0.30	0.74	0.030	-9.98	<2e-16	***
Greenness (β_{G})	0.43	1.54	0.044	9.88	<2e-16	***
Road Density (β_{RD})	-0.00	1.00	0.006	-0.31	0.75	
Anthrax Risk (β_{AR})	-0.06	0.94	0.011	-5.85	4.8e-09	***

2009 Directed Points	coef	exp(coef)	se(coef)	z	p	sig
Wetness (β_{W})	-2.17	0.11	0.145	-14.97	<2e-16	***
Greenness (β_{G})	-0.54	0.58	0.130	-4.15	3.3e-05	***
Road Density (β_{RD})	0.04	1.04	0.012	3.26	0.001	**
Anthrax Risk (β_{AR})	0.13	1.14	0.043	3.01	0.003	**

2010 Directed Points	coef	exp(coef)	se(coef)	z	p	sig
Wetness (β_{W})	-2.19	0.11	0.075	-29.37	<2e-16	***
Greenness (β_{G})	-0.02	0.98	0.097	-0.25	0.80	
Road Density (β_{RD})	0.07	1.07	0.008	8.36	<2e-16	***
Anthrax Risk (β_{AR})	0.09	1.09	0.024	3.81	1.4e-04	***

Table 3.6: Results of the conditional logistic mixed effects models as applied to all of the movement points (n = 7,650 in 2009 and n = 18,908 in 2010), only the foraging points (n = 3,955 in 2009 and n = 9,546 in 2010), and only the directed movement points (n = 1,420 in 2009 and n = 3,735 in 2010), using hourly fixes rather than the finer-resolution 20 minute intervals.

2009 All Points	coef	exp(coef)	se(coef)	z	p	sig
Wetness (β_{W})	-1.07	0.34	0.074	-14.51	<2e-16	***
Greenness (β_{W})	0.11	1.11	0.058	1.83	0.07	.
Road Density (β_{RD})	0.02	1.02	0.010	1.91	0.06	.
Anthrax Risk (β_{AR})	-0.04	0.96	0.022	-1.98	0.05	*

2010 All Points	coef	exp(coef)	se(coef)	z	p	sig
Wetness (β_{W})	-1.63	0.20	0.043	-37.73	<2e-16	***
Greenness (β_{G})	0.84	2.32	0.057	14.64	<2e-16	***
Road Density (β_{RD})	0.04	1.04181	0.008	5.16	2.5e-07	***
Anthrax Risk (β_{AR})	-0.08	0.92751	0.014	-5.55	2.9e-08	***

2009 Foraging Points	coef	exp(coef)	se(coef)	z	p	sig
Wetness (β_{W})	-0.18	0.84	0.094	-1.88	0.06	.
Greenness (β_{G})	0.07	1.08	0.076	0.96	0.34	
Road Density (β_{RD})	0.00	1.00	0.015	0.06	0.95	
Anthrax Risk (β_{AR})	-0.10	0.90	0.032	-3.19	1.4e-03	**

2010 Foraging Points	coef	exp(coef)	se(coef)	z	p	sig
Wetness (β_{W})	-0.94	0.39	0.056	-16.63	<2e-16	***
Greenness (β_{G})	0.72	2.05	0.080	9.00	<2e-16	***
Road Density (β_{RD})	0.02	1.02	0.011	1.58	0.11	
Anthrax Risk (β_{AR})	-0.10	0.91	0.018	-5.29	1.3e-07	***

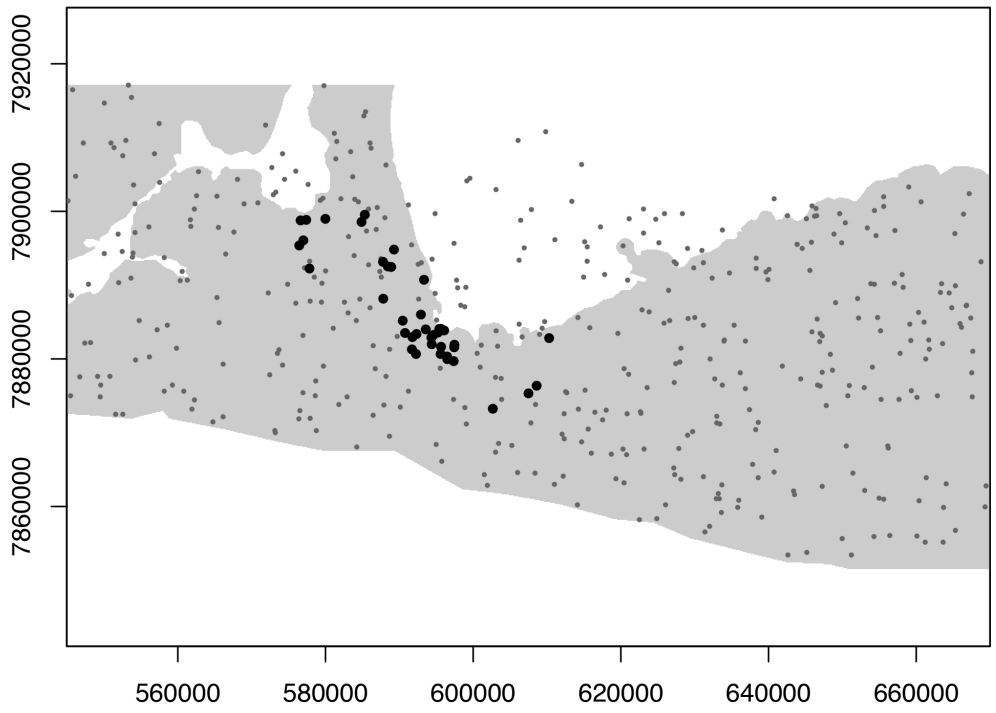
2009 Directed Points	coef	exp(coef)	se(coef)	z	p	sig
Wetness (β_{W})	-2.81	0.06	0.226	-12.47	<2e-16	***
Greenness (β_{G})	0.21	1.24	0.164	1.29	0.20	
Road Density (β_{RD})	0.06	1.06	0.025	2.37	0.02	*
Anthrax Risk (β_{AR})	0.28	1.32	0.057	4.87	1.1e-06	***

2010 Directed Points	coef	exp(coef)	se(coef)	z	p	sig
Wetness (β_{W})	-1.66	0.19	0.090	-18.39	<2e-16	***
Greenness (β_{G})	0.41	1.50	0.112	3.64	2.7e-04	***
Road Density (β_{RD})	0.08	1.08	0.016	4.98	6.2e-07	***
Anthrax Risk (β_{AR})	0.07	1.07	0.028	2.55	0.01	*

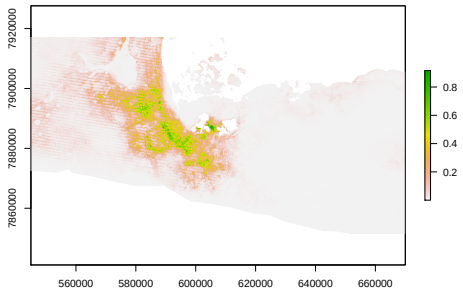
3.7 Figures

Figure 3.1: **Predicted Anthrax Suitability Maps** MaxEnt derived maps of suitability for anthrax persistence within the region of interest in Etosha National Park, Namibia. Panel **a** illustrates the spatial distribution of presence points (large black dots) and background sampling points (small gray dots) for the MaxEnt algorithm. Panels **b** and **c** are the resulting predictive maps of suitability for anthrax spores in 2009 and 2010.

(a) Anthrax Presence and Background Points



(b) Predicted Anthrax Risk Map (2009)



(c) Predicted Anthrax Risk Map (2010)

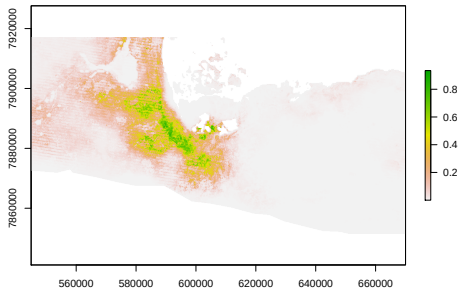
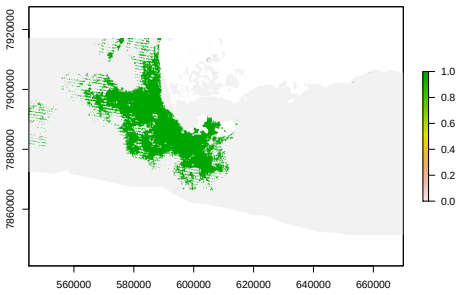
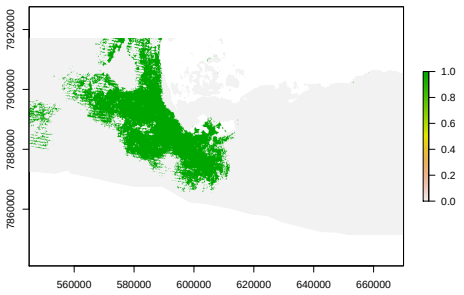


Figure 3.2: **Pertinent Transmission Zones** PTZs for anthrax as delimited using three different thresholds: >10%, >25%, and >50% probability of suitability, corresponding to a liberal, moderate, and conservative estimate of the area in which anthrax is likely to persist, respectively. The two columns represent the same three thresholds applied to the 2009 season (left column) and 2010 season (right column).

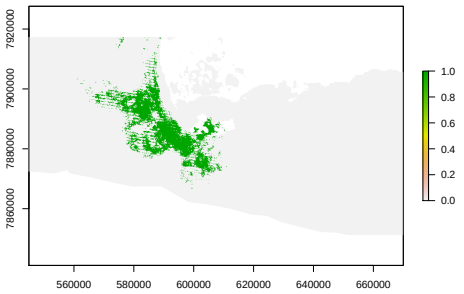
(a) 2009 Liberal PTZ for Anthrax



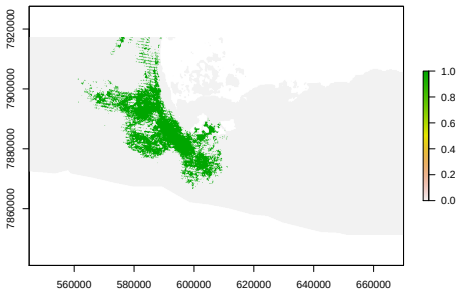
(b) 2010 Liberal PTZ for Anthrax



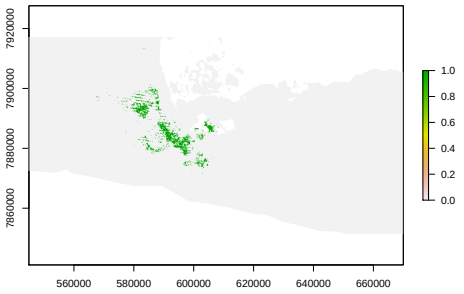
(c) 2009 Moderate PTZ for Anthrax



(d) 2010 Moderate PTZ for Anthrax



(e) 2009 Conservative PTZ for Anthrax



(f) 2010 Conservative PTZ for Anthrax

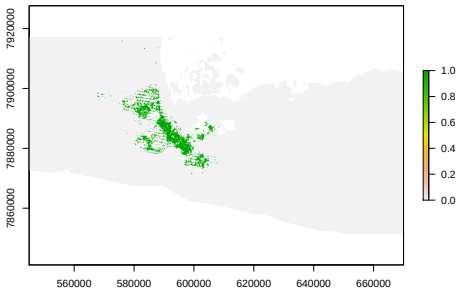
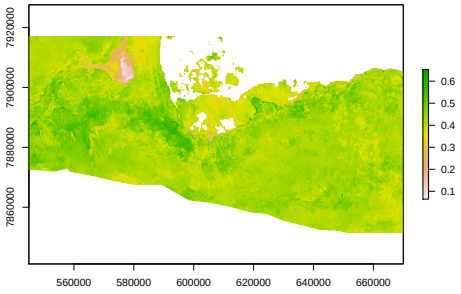
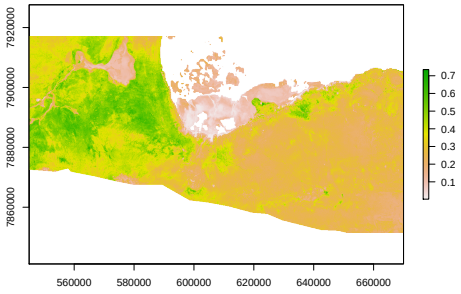


Figure 3.3: **Step Selection Function Projected Surfaces** Step selection functions projected within the region of interest in Etosha National Park, Namibia. Panels **a** and **b** illustrate the selection functions for anthrax seasons 2009 and 2010, respectively, when all of the recorded movement points are used. Panels **c** and **d** represent the selection functions during the same time periods, but using only the points during which the individual was in the foraging behavioral state. Finally, panels **e** and **f** illustrate the selection surfaces when the animals were in the directed movement state in 2009 and 2010.

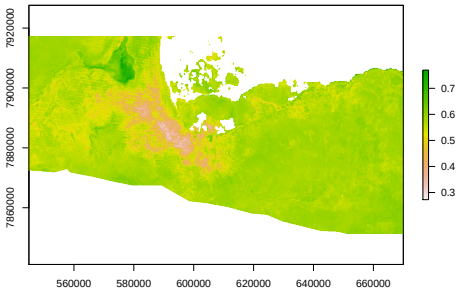
(a) Zebra SSF (2009; All Points)



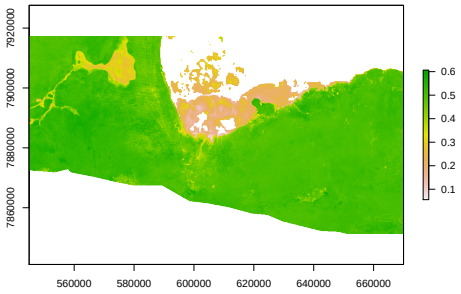
(b) Zebra SSF (2010; All Points)



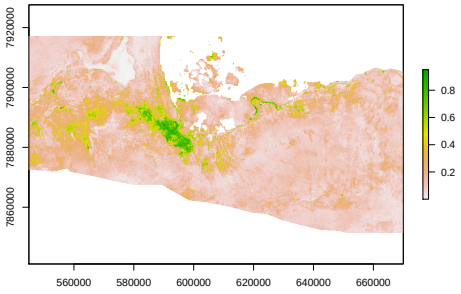
(c) Zebra SSF (2009; Foraging Points)



(d) Zebra SSF (2010; Foraging Points)



(e) Zebra SSF (2009; Directed Points)



(f) Zebra SSF (2010; Directed Points)

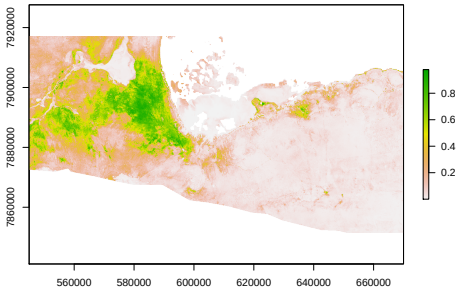
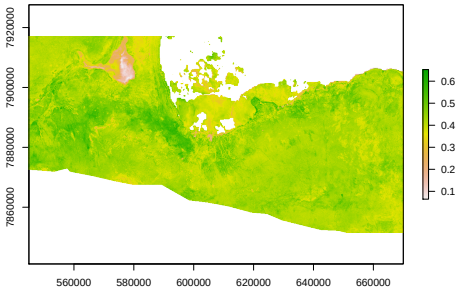
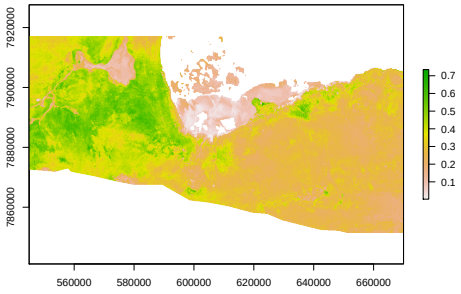


Figure 3.4: **Hourly Step Selection Functions** Alternative SSFs within the region of interest in Etosha National Park, Namibia based on hourly fixes. Panels **a** and **b** illustrate the selection functions for anthrax seasons 2009 and 2010, respectively, when all of the recorded movement points are used. Panels **c** and **d** represent the selection functions during the same time periods, but using only the points during which the individual was in the foraging behavioral state. Finally, panels **e** and **f** illustrate the selection surfaces when the animals were in the directed movement state in 2009 and 2010.

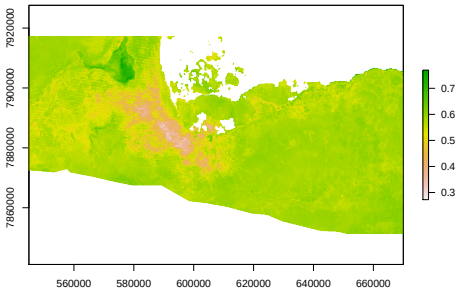
(a) Hourly SSF (2009; All Points)



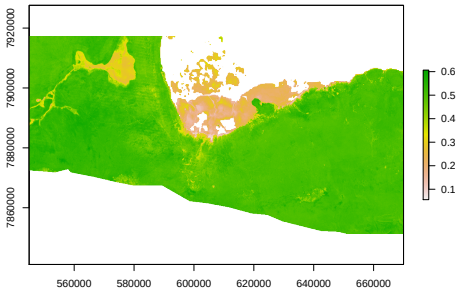
(b) Hourly SSF (2010; All Points)



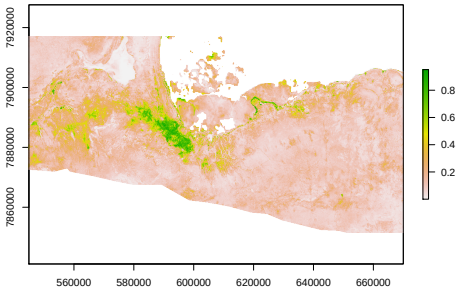
(c) Hourly SSF (2009; Foraging Points)



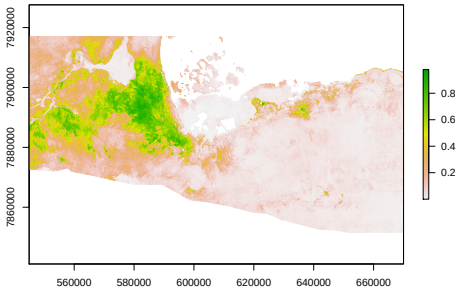
(d) Hourly SSF (2010; Foraging Points)



(e) Hourly SSF (2009; Directed Points)



(f) Hourly SSF (2010; Directed Points)



3.8 Supplementary Materials

Table 3.7: HMM results for AG059 during 2009 season

Behavioral State	**Step Lengths (Gamma)***		**Turning Angles (vonMises)**	
	mu (m)	sigma (m)	mean (radians)	concentration
Resting	23.5	27.6	-2.5	0.03
Foraging	193.6	178.6	0.01	1.8
Directed	853.8	299.0	-0.01	20.5

*Gamma distributions may be parameterized with an additional zero-mass value, but these are excluded here because their magnitude was negligible in all models

Table 3.8: HMM results for AG061 during 2009 season

Behavioral State	**Step Lengths (Gamma)***		**Turning Angles (vonMises)**	
	mu (m)	sigma (m)	mean (radians)	concentration
Resting	6.0	4.7	-3.0	0.27
Foraging	78.3	73.4	0.02	0.89
Directed	405.8	306.8	0.01	3.8

*Gamma distributions may be parameterized with an additional zero-mass value, but these are excluded here because their magnitude was negligible in all models

Table 3.9: HMM results for AG062 during 2009 season

Behavioral State	**Step Lengths (Gamma)***		**Turning Angles (vonMises)**	
	mu (m)	sigma (m)	mean (radians)	concentration
Resting	2.9	2.1	2.9	0.28
Foraging	68.0	64.7	-0.02	0.57
Directed	386.4	297.0	0.01	2.6

*Gamma distributions may be parameterized with an additional zero-mass value, but these are excluded here because their magnitude was negligible in all models

Table 3.10: HMM results for AG063 during 2009 season

Behavioral State	**Step Lengths (Gamma)***		**Turning Angles (vonMises)**	
	mu (m)	sigma (m)	mean (radians)	concentration
Resting	30.7	36.4	-0.15	0.17
Foraging	188.1	151.8	0.01	1.8
Directed	758.0	332.0	-0.03	5.6

*Gamma distributions may be parameterized with an additional zero-mass value, but these are excluded here because their magnitude was negligible in all models

Table 3.11: HMM results for AG068 during 2009 season

Behavioral State	**Step Lengths (Gamma)***		**Turning Angles (vonMises)**	
	mu (m)	sigma (m)	mean (radians)	concentration
Resting	11.5	13.1	-2.97	0.15
Foraging	173.4	163.0	0.03	1.4
Directed	907.7	304.5	-0.01	12.3

*Gamma distributions may be parameterized with an additional zero-mass value, but these are excluded here because their magnitude was negligible in all models

Table 3.12: HMM results for AG063 during 2010 season

Behavioral State	**Step Lengths (Gamma)***		**Turning Angles (vonMises)**	
	mu (m)	sigma (m)	mean (radians)	concentration
Resting	23.9	27.4	-3.01	0.1
Foraging	205.6	163.0	< 0.01	1.4
Directed	818.9	369.5	0.01	4.1

*Gamma distributions may be parameterized with an additional zero-mass value, but these are excluded here because their magnitude was negligible in all models

Table 3.13: HMM results for AG068 during 2010 season

Behavioral State	**Step Lengths (Gamma)***		**Turning Angles (vonMises)**	
	mu (m)	sigma (m)	mean (radians)	concentration
Resting	24.2	30.2	-1.26	0.02
Foraging	187.0	155.0	0.04	1.4
Directed	852.0	339.9	-0.02	4.1

*Gamma distributions may be parameterized with an additional zero-mass value, but these are excluded here because their magnitude was negligible in all models

Table 3.14: HMM results for AG252 during 2010 season

Behavioral State	**Step Lengths (Gamma)***		**Turning Angles (vonMises)**	
	mu (m)	sigma (m)	mean (radians)	concentration
Resting	17.6	16.5	0.17	0.27
Foraging	110.9	88.9	-0.01	1.6
Directed	600.1	348.1	< 0.01	3.1

*Gamma distributions may be parameterized with an additional zero-mass value, but these are excluded here because their magnitude was negligible in all models

Table 3.15: HMM results for AG253 during 2010 season

Behavioral State	**Step Lengths (Gamma)***		**Turning Angles (vonMises)**	
	mu (m)	sigma (m)	mean (radians)	concentration
Resting	24.5	24.6	0.17	0.27
Foraging	154.1	131.8	-0.01	1.6
Directed	785.5	385.5	0.01	2.4

*Gamma distributions may be parameterized with an additional zero-mass value, but these are excluded here because their magnitude was negligible in all models

Table 3.16: HMM results for AG255 during 2010 season

Behavioral State	**Step Lengths (Gamma)***		**Turning Angles (vonMises)**	
	mu (m)	sigma (m)	mean (radians)	concentration
Resting	5.49	4.65	3.07	0.41
Foraging	97.1	86.1	-0.04	0.89
Directed	560.1	376.3	-0.01	3.7

*Gamma distributions may be parameterized with an additional zero-mass value, but these are excluded here because their magnitude was negligible in all models

Table 3.17: HMM results for AG256 during 2010 season

Behavioral State	**Step Lengths (Gamma)***		**Turning Angles (vonMises)**	
	mu (m)	sigma (m)	mean (radians)	concentration
Resting	5.98	5.69	3.09	0.30
Foraging	108.4	100.9	0.03	1.1
Directed	581.8	362.9	0.01	3.5

*Gamma distributions may be parameterized with an additional zero-mass value, but these are excluded here because their magnitude was negligible in all models

Table 3.18: Radii of the kernels (in meters) used for in producing the step-selection functions for each individual at the hourly fix rate. Separate radii were used for the datasets containing all of the movement points, only the foraging points, and only the directed movement points.

Animal ID	Kernel Radius (All)	Kernel Radius (Foraging)	Kernel Radius (Directed)
AG059_2009	3022	1585	3990
AG061_2009	1827	738	2782
AG062_2009	1942	691	2543
AG063_2009	2424	1391	3644
AG068_2009	3363	1548	4322
AG063_2010	2895	1469	3988
AG068_2010	3165	1438	4038
AG252_2010	2453	1009	3425
AG253_2010	2584	1384	3944
AG255_2010	2643	919	3592
AG256_2010	2498	964	3596

Table 3.19: Pearson Correlation matrix among predictor variables for carcasses deposited in 2010

	pH	OCC	CEC	bio1	bio7	bio12	bio13	mean	max	min	range
pH	1	0.32	0.46	-0.12	0.23	-0.50	-0.52	-0.40	-0.22	-0.17	-0.15
OCC	0.32	1	0.45	0.04	0.25	-0.18	-0.19	-0.35	-0.33	-0.22	-0.26
CEC	0.46	0.45	1	0.21	0.43	-0.30	-0.30	-0.39	-0.24	-0.21	-0.15
bio1	-0.12	0.04	0.21	1	0.50	-0.17	-0.11	-0.07	0.00	-0.02	0.01
bio7	0.23	0.25	0.43	0.50	1	-0.24	-0.26	-0.46	-0.31	-0.24	-0.20
bio12	-0.50	-0.18	-0.30	-0.17	-0.24	1	0.98	0.02	-0.14	-0.14	-0.07
bio13	-0.52	-0.19	-0.30	-0.11	-0.26	0.98	1	0.03	-0.16	-0.13	-0.10
mean_ndvi	-0.40	-0.35	-0.39	-0.07	-0.46	0.02	0.03	1	0.89	0.80	0.49
max_ndvi	-0.22	-0.33	-0.24	0.00	-0.31	-0.14	-0.16	0.89	1	0.71	0.74
min_ndvi	-0.17	-0.22	-0.21	-0.02	-0.24	-0.14	-0.13	0.80	0.71	1	0.05
range_ndvi	-0.15	-0.26	-0.15	0.01	-0.20	-0.07	-0.10	0.49	0.74	0.05	1

Table 3.20: Pearson Correlation matrix among predictor variables for carcasses deposited in 2011

	pH	**OCC**	**CEC**	**bio1**	**bio7**	**bio12**	**bio13**	**mean**	**max**	**min**	**range**
pH	1	0.32	0.46	-0.12	0.23	-0.50	-0.52	-0.41	-0.21	-0.16	-0.15
OCC	0.32	1	0.45	0.04	0.25	-0.18	-0.19	-0.34	-0.30	-0.22	-0.22
CEC	0.46	0.45	1	0.21	0.43	-0.30	-0.30	-0.41	-0.23	-0.16	-0.18
bio1	-0.12	0.04	0.21	1	0.50	-0.17	-0.11	-0.10	-0.04	0.03	-0.07
bio7	0.23	0.25	0.43	0.50	1	-0.24	-0.26	-0.50	-0.33	-0.21	-0.26
bio12	-0.50	-0.18	-0.30	-0.17	-0.24	1	0.98	0.06	-0.15	-0.08	-0.13
bio13	-0.52	-0.19	-0.30	-0.11	-0.26	0.98	1	0.07	-0.17	-0.06	-0.17
mean_ndvi	-0.41	-0.34	-0.41	-0.10	-0.50	0.06	0.07	1	0.85	0.72	0.57
max_ndvi	-0.21	-0.30	-0.23	-0.04	-0.33	-0.15	-0.17	0.85	1	0.61	0.83
min_ndvi	-0.16	-0.22	-0.16	0.03	-0.21	-0.08	-0.06	0.72	0.61	1	0.06
range_ndvi	-0.15	-0.22	-0.18	-0.07	-0.26	-0.13	-0.17	0.57	0.83	0.06	1

Table 3.21: Pearson Correlation matrix among predictor variables for carcasses deposited in 2012

	pH	**OCC**	**CEC**	**bio1**	**bio7**	**bio12**	**bio13**	**mean**	**max**	**min**	**range**
pH	1	0.32	0.46	-0.12	0.23	-0.50	-0.52	-0.44	-0.30	-0.20	-0.26
OCC	0.32	1	0.45	0.04	0.25	-0.18	-0.19	-0.34	-0.29	-0.26	-0.22
CEC	0.46	0.45	1	0.21	0.43	-0.30	-0.30	-0.42	-0.24	-0.21	-0.18
bio1	-0.12	0.04	0.21	1	0.50	-0.17	-0.11	-0.09	0.02	0.07	-0.01
bio7	0.23	0.25	0.43	0.50	1	-0.24	-0.26	-0.50	-0.31	-0.21	-0.26
bio12	-0.50	-0.18	-0.30	-0.17	-0.24	1	0.98	0.06	-0.13	-0.04	-0.13
bio13	-0.52	-0.19	-0.30	-0.11	-0.26	0.98	1	0.07	-0.13	0.00	-0.15
mean_ndvi	-0.44	-0.34	-0.42	-0.09	-0.50	0.06	0.07	1	0.87	0.66	0.71
max_ndvi	-0.30	-0.29	-0.24	0.02	-0.31	-0.13	-0.13	0.87	1	0.59	0.90
min_ndvi	-0.20	-0.26	-0.21	0.07	-0.21	-0.04	0.00	0.66	0.59	1	0.18
range_ndvi	-0.26	-0.22	-0.18	-0.01	-0.26	-0.13	-0.15	0.71	0.90	0.18	1

Table 3.22: Variable contribution and importance results from the full MaxEnt model, built on the full environmental covariate set. Due to covariance observed in Supplementary Tables 13-15, two pairs of covariates were considered for variable set reduction: bio12 with bio13 and mean_ndvi with max_nndvi.

Variable	**Name**	**Percent contribution**	**Permutation importance**
Mean temperature range	bio7	72.6	71.5
Soil Organic Carbon Content	OC	11.3	2.6
Precipitation of the wettest month	bio13	5.3	15.5
Range of NDVI	range_ndvi	4.7	0.9
Maximum NDVI	max_ndvi	2.4	1.5
Annual precipitation	bio12	2	4.2
Mean annual temperature	bio1	1.1	0.8
Soil pH	pH	0.5	0.7
Soil Cation Exchange Efficiency	CEC	0.1	1.7
Minimum NDVI	min_ndvi	0	0.8
Mean NDVI	men_ndvi	0	0

Table 3.23: Final MaxEnt model 'lambda' values associated with the variables (and their derivatives) ultimately included.

Variable	Lambda	Min	Max
CEC	0.222	7.075	26.000
OC	-5.052	0.000	58.429
bio1	0.669	22.225	23.194
bio13	3.481	86.000	116.000
bio7	8.804	27.285	30.536
max_ndvi	0.000	0.092	0.639
min_ndvi	3.044	-0.309	0.160
pH	0.000	72.655	83.327
range_ndvi	0.000	0.087	0.589
CEC^2	2.560	50.052	676.000
bio1^2	0.522	493.929	537.960
bio7^2	2.072	744.498	932.440
'OC	-1.015	0.000	0.350
'bio13	-0.369	86.000	95.000
'range_ndvi	-1.278	0.382	0.589
'bio1	-0.421	23.093	23.194
'bio1	-0.546	23.092	23.194
'bio7	-1.240	30.280	30.536
'range_ndvi	-0.324	0.382	0.589
'pH	-0.882	72.655	78.347
'max_ndvi	-1.746	0.092	0.355
'bio1	-0.183	23.090	23.194
'OC	-0.180	0.000	1.008
'bio7	-0.525	30.209	30.536
'bio1	-0.271	23.060	23.194
'bio7	-0.173	30.205	30.536
'bio13	-0.178	86.000	95.262
'OC	-0.229	0.000	1.003

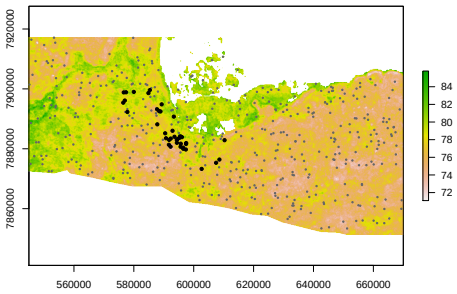
var^2 represents quadratic feature
'var represents forward hinge feature
'var represents reverse hinge feature
linearPredictorNormalizer: 15.142
densityNormalizer: 34.768
entropy: 4.561

Table 3.24: Results of the conditional logistic mixed effects models as applied to all of the movement points in 2010 (n = 56495) including two randomly generated predictor layers (Rand and Rand2) to evaluate the efficacy of the step-selection methodology employed throughout.

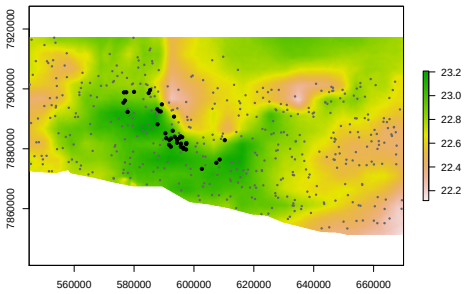
2010 All Points	coef	exp(coef)	se(coef)	z	p	sig
Rand	-0.00	1.00	0.009	-0.37	0.71	
Rand2	-0.01	0.99	0.009	-1.15	0.25	
Wet_Norm	-0.85	0.43	0.023	-36.57	<2e-16	***
Green_Norm	0.36	1.43	0.031	11.40	<2e-16	***
Road_Dens_Norm	0.03	1.03	0.004	6.16	7.1e-10	***
Risk_Norm	-0.06	0.94	0.008	-7.56	3.9e-14	***

Figure 3.5: **Supplementary Figure 1** Soil and bioclimatic variables used for the 2009 and 2010 predictive anthrax risk map based on the final MaxEnt model. Larger black points are the locations of the carcasses used as presence locations for the model, whereas smaller gray points are the 422 background locations used to parameterize the model.

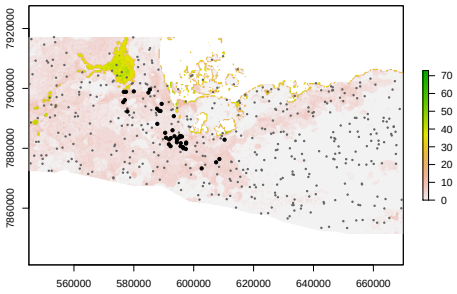
(a) Soil pH in H_2O



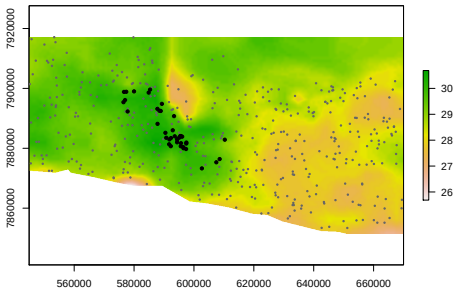
(b) Mean annual temperature (bio1)



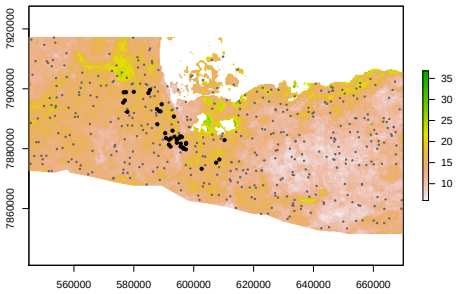
(c) Soil Organic Carbon Content



(d) Mean temperature range (bio7)



(e) Soil Cation Exchange Capacity



(f) Precipitation of the wettest month (bio13)

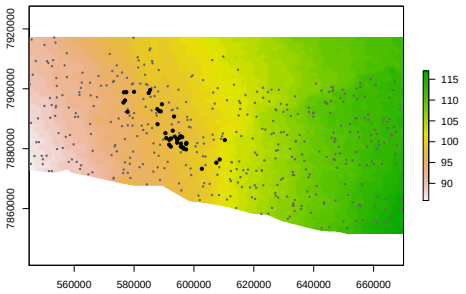
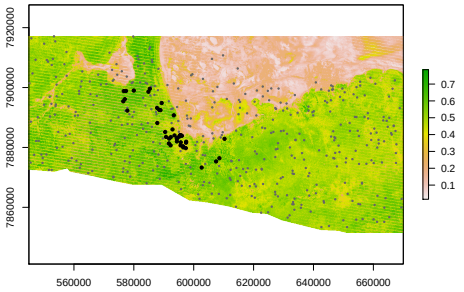
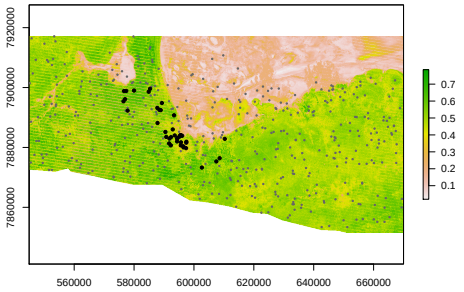


Figure 3.6: **Supplementary Figure 2** Environmental variables used for the 2009 (left column) and 2010 (right column) predictive anthrax risk map based on the final MaxEnt model. Larger black points are the presence locations and smaller gray points are the randomly generated background sampling points.

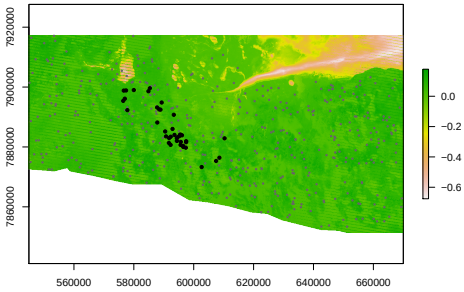
(a) Maximum NDVI 2007-2009



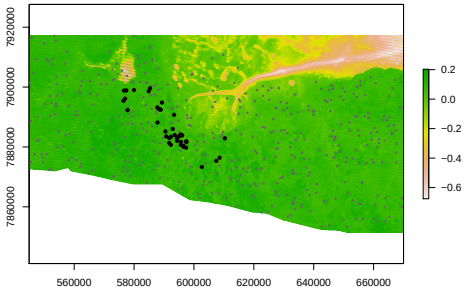
(b) Maximum NDVI 2008-2010



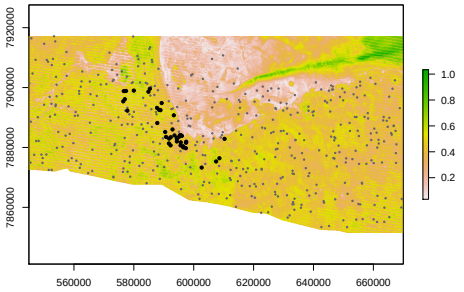
(c) Minimum NDVI 2007-2009



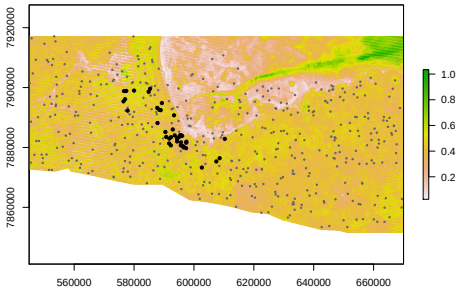
(d) Minimum NDVI 2008-2010



(e) Range NDVI 2007-2009



(f) Range NDVI 2008-2010



Chapter 4

Movement mediates environmental influences on infection dynamics of anthrax in herbivores

Eric R. Dougherty Dana P. Seidel Colin J. Carlson Wayne M. Getz

4.1 Abstract

The relationship between environmental factors and epidemic dynamics is not always straightforward. In the case of indirectly transmitted pathogens, there might exist temporal or spatial lags that obscure the link between heterogeneous environmental covariates and observed epidemiological processes. Here I apply a simulation modeling framework to explore some of these potential relationships, using a mechanistic movement model as an intermediary. Using a set of behaviorally-contingent selection functions to guide the movement decisions of a simulated host animal, I calculated the rate at which the agent contacts stationary infectious reservoirs distributed across a focal region in Etosha National Park, Namibia. A non-linear regression model reveals an apparent relationship between Wetness, one of the covariates incorporated into the selection function, and the emergent contact rates. This, in turn, can be associated with the observed pattern of anthrax-induced mortalities in Etosha. These results demonstrate the importance of movement as a mediator of the exposure process in an anthrax-endemic system and suggests that environmental heterogeneity may be directly related to temporal patterns in mortality.

4.2 Introduction

Though epidemic dynamics are often simplified using rates averaged at the population scale [57], the transmission process underlying disease spread is highly dependent on various forms of heterogeneity [102]. The influence of such heterogeneity has been recognized in both human [92] and wildlife [125] populations, but the diversity of underlying causes of individual variance often makes precise

measurement of contributing factors difficult. The impact of any one individual on the subsequent dynamics of an epidemic arises from a complex combination of host and pathogen characteristics, as well as the environment in which transmission occurs [92]. A variety of methods, often applied during or after an epidemic, have demonstrated this variation in natural systems [108, 161].

One fundamental driver of heterogeneity in disease transmission is the variance in movement behaviors among host individuals [144, 41]. Until recently, however, the tools available to observe empirical movement trajectories did not offer fine enough resolution to describe individual heterogeneity beyond differences in space-use at landscape scales. Various analytical methods offer insight into the behavioral underpinnings of movement behaviors [116, 46]. Further, the application of habitat selection frameworks [17, 173] can be used to explore associations between these specific behavioral states and various environmental covariates that might drive animal movement decisions. In this way, novel approaches from the field of movement ecology [115] can be used to reveal the role of environmental factors in driving heterogeneous movements, thereby giving rise to particular infection dynamics at the population scale.

Though other investigations of the influence of landscape features and environmental factors on disease dynamics have been undertaken (e.g., chronic wasting disease in mule deer [51, 66]; bovine tuberculosis in African buffalo [38]; and chytridiomycosis in various amphibian species [159]), the emphasis is often on broad-scale considerations of population structure. The mediating role of fine-scale animal movement behavior, however, has been relatively unexplored. This trend may be due, in part, to the difficulty associated with monitoring the movements of entire wildlife populations [164]. However, with the expansion of computational power, agent-based models (ABMs) have emerged as a useful tool for translating a set of general rules into emergent properties at broader levels of analysis. Applying such models in the pattern-oriented modeling (POM; [65, 64]) framework offers a means of validating the underlying components, lending credibility to the outputs of the model. The flexibility of this framework has led to broad applications of ABMs even within the ecological literature, including to questions in environmental resource management [16, 105], examinations of evolutionary dynamics [37, 61, 62, 110], and considerations of individual animal behaviors [160, 156].

ABMs have also increasingly been used to explore disease dynamics, with individual agents often transitioning between the infectious stages normally indicative of a compartmental SIR model (e.g., susceptible, infectious, and recovered [4]). The general rule set of an ABM typically gives rise to stochasticity, and thus heterogeneity, in behaviors such as movement. This makes ABMs ideal tools for exploring the role of individual movement decisions in mediating disease dynamics. However, in the few cases that have utilized ABMs to explore disease systems, individual movement dynamics are often simplified to diffusion processes [141, 107] or highly generalized jumps between patches [47, 126, 128]. Though other models have implemented more complex mechanistic movement rules to govern agent trajectories (e.g., [13, 38, 149, 11, 112]), examples of such ABMs remain fairly limited in disease ecology, and their applicability is frequently constrained by the specific nature of pathogen transmission in the focal disease system.

Here I develop a mechanistic model of herbivore movement on an anthrax-endemic landscape parameterized using empirical movement data from Etosha National Park in Namibia. Selection

function outputs have not been deployed widely as the basis for an agent-based model, but I demonstrate the utility of the framework in creating the rule set of the model. As anthrax is transmitted primarily through environmental reservoirs [153, 151, 24], the exposure process is summarized by an estimate of the contact rate between each agent and simulated infectious reservoirs distributed on the landscape (representing concentrated anthrax spores deposited at carcass sites). Using this model, I examine the influence of underlying environmental covariates in driving the exposure process.

Study System

Anthrax, the acute disease caused by the bacterium *Bacillus anthracis*, remains a persistent threat in many wildlife populations throughout the world [150, 79]. Though a variety of animal species can contract the zoonotic disease, herbivores experience the highest mortality rates, while many carnivores and scavengers exhibit resistance or tolerance [79, 27]. In some systems, anthrax outbreaks are seasonally-driven, though there may exist inter-specific differences in the timing of the peak of infections. For example, zebra in Etosha National Park in Namibia experience peaks in infection during the warm wet season (February-April), whereas elephants are more likely to be infected during the dry months of October-November [24]. A definitive explanation for these peaks remains elusive, but a number of alternatives have been proposed, including nutritional stress, heterogeneous soil ingestion rates [153], and complex coinfection dynamics [28, 29].

B. anthracis takes the form of reproducing vegetative cells in infected hosts and endospores when in soil and ponded water environments [150, 77], although some vegetative reproduction may take place within the rhizosphere of vigorously growing grasses [137]. The spores are exceptionally resilient in the face of environmental stress, and allow the infectious agent to persist in environmental reservoirs for extended periods of sub-optimal conditions [77]

Spores can enter the host organism through cutaneous lesions, by inhalation into the pulmonary system, or via the gastrointestinal (GI) tract. Many ungulates consume substantial amounts of soil in addition to vegetation during foraging bouts, and in doing so, may inadvertently ingest the pathogen [153]. Limited evidence from necropsies suggests that GI infections are the most common route of infection in herbivores [158] and will be the primary mechanism modeled here. Anthrax is highly pathogenic in herbivores, and death may occur within a few days or up to two weeks after contact with a lethal dose of *B. anthracis* spores [24].

Anthrax is endemic in the plains herbivores of Etosha National Park, Namibia, peaking in zebra, springbok, and wildebeest during the rainy season and in elephants during the dry season [153, 91, 29]. Extensive carcass surveillance efforts in Etosha National Park, Namibia, between 1968 and 2011, conducted by The Etosha Ecological Institute [153, 9, 27], were used to inform the densities of anthrax infected carcasses considered in the simulation. In addition, empirical movement data collected from Etosha National Park, were used to inform the mechanistic movement model giving rise to the simulated trajectories. Specifically, the movement trajectories of nine zebra (Table 4.1) were used to estimate a behaviorally-conditioned step-selection function (SSF; see Chapter 3), which served as the basis for individual movement decisions.

4.3 Methods

Purpose

The simulation model consists of an agent moving across a heterogeneous landscape upon which infectious reservoirs are distributed. A contact rate between an agent and these local infectious zones (LIZs; [56]) is calculated for each individual path over the course of a single anthrax season (defined as the five month period between February and June; [153]). By altering the environmental covariate layers underlying animal movement decisions, I can examine the manner by which these movements, a major contributor to individual heterogeneity, influence exposure dynamics. The contact rate represents an epidemiologically-relevant metric that is applicable only within the scope of the simulation and should not be presumed to correspond with empirical rates in the Etosha system; even so, by examining various combinations of Greenness and Wetness inputs, I can examine how the expected contact rate shifts based on external drivers.

This is not, in its current form, an agent-based model, *per se*. The movements of the individual across the landscape depend only upon a predetermined selection map derived based on a population-level selection function and the particular set of covariate layers used as the basis for the simulation; there is no interaction between the individual and any other agents, nor is its behavior influenced by any external, time-dependent variable. Rather, this approach is meant to offer greater insight into a mean expected contact rate by reducing the influence of an anomalous path. Each of these simulated paths can be seen as samples of a Monte Carlo process exploring infinite possible alternative paths. By sampling 1000 times using each selection surface, a population-level contact rate can be estimated based on Monte Carlo simulation. Observing how this rate shifts across a set of alternative covariate layers will offer insight into the role of environmental heterogeneity in the exposure process.

An overview of the methods carried out here is presented in Figure 4.1, and the caption offers a detailed description of the visualization of the approach.

Environmental Covariate Layers

The Wetness and Greenness layers were both derived from six of the component bands of the Landsat 4-5 Thematic Mapper satellite data using the equations from the previous chapter [32]. Relatively cloudless images (i.e., $< 10\%$) from the 2009 and 2010 anthrax seasons were selected, which resulted in a total of 8 possible sets of covariate layers (Table 4.2). Separate selection surfaces were calculated for the foraging and directed movement behavioral states for each combination of environmental input layers in Table 4.3. The Surface ID refers to the set of two selection surfaces that formed the basis of the simulated movement trajectories described below.

These surfaces, which emerge from the population-wide selection functions, will differ based on the input layers. By altering the Wetness and Greenness layers, the same overall selection pattern (in terms of the coefficients of the selection function) will result in different surfaces as experienced by the animal. To explore the role of environmental heterogeneity in the contact process, a range of different Wetness and Greenness layers were obtained. Variation in the mean values of both

covariates was observed across the anthrax seasons of 2009 and 2010, during which the mean Wetness ranged between -20.9 (March 2009) and -73.46 (February 2010) and the mean Greenness ranged from -8.6 (March 2009) to -37.1 (February 2010). It should be noted that February 2010 was somewhat anomalous, as February is normally when rainfall is highest, resulting in the highest Wetness values (as observed in 2008 and 2009), and the area begins to dry out after that peak. Even so, using these extreme values of mean Wetness and Greenness as well as two more months with more moderate values (May 2009 and May 2010), enables the creation of a grid of environmental covariates layers, and the Wetness and Greenness layers could be varied independently to determine whether one of both of them influenced the contact process. The 16 combinations of Wetness and Greenness layers can be seen in Table 4.3. All of the values were normalized based on the mean and standard deviation of the Wetness and Greenness layers used to develop the population-wide selection function.

Selection Map Derivation

Population-wide selection functions were calculated for the foraging and directed movement behavioral states. This involved the aggregation of the movement trajectories from both 2009 and 2010. A conditional logistic regression approach was applied to the 'used and 'available data extracted using the methods outlined in the previous chapter. The coefficient values that emerged from this analysis were then applied across all of the sets of simulations; the only aspects that differentiated these sets of simulations were the two environmental input layers. For each of these pairs of environmental covariate layers (Greenness and Wetness) incorporated into the selection surface, the values were first normalized using the mean and standard deviation of the overall average covariate layers used to derive the coefficients. The overall average Greenness and Wetness layers were calculated by combining all of the Greenness and Wetness layers collected during the anthrax periods of 2009 and 2010.

Movement Model Parameterization

The movement model was informed by the empirical movement tracks collected in Etosha National Park in 2009 and 2010. The trajectories, which were originally collected with short-interval fixes (1 minute) in addition to the longer interval fixes (20 minute), were regularized using the `adehabitatLT` package [23] in the R statistical computing environment (version 3.2.5) [133] such that all relocations had a temporal separation of 20 minutes. After projecting the spatial data (WGS84 UTM Zone 33S), all eleven paths were combined into a single data set, using an ID column to differentiate among the individuals. The final dataset used for subsequent analyses consisted of 201,190 relocations.

These data were analyzed using a hidden Markov model (HMM), implemented through the `moveHMM` package [109]. The HMM framework has grown in popularity as a method for extracting latent behavioral states from movement data [123, 122]. The general form follows that of a basic state-space model (SSM), consisting of a time series of vector-valued observations $\mathbf{z}_1, ..., \mathbf{z}_T$ and an unobservable series of scalar-valued, discrete states $S_1, ..., S_T$ that take on values from $1, ..., N$. The movement data that forms the basis of the observations tend to take the form of a time

series of locations on a Euclidean plane $(x_1, y_1), ..., (x_T, y_T)$, and these positions are treated as being measured without error. When the primary purpose of the HMM is to identify the latent behavioral states, the bivariate series of step lengths (l_t) and turning angles (ϕ_t), each governed by a set of parameters that can be vectorized as $\boldsymbol{\theta}$, tends to serve as the observations ($\mathbf{z}_t$) for fitting. Traditionally, the step length distribution is governed by a gamma distribution and the turning angle distribution by a circular vonMises distribution. For each latent state, the HMM returns estimates of the parameters underlying these distributions: μ and σ of the gamma distributions (which can be transformed into the conventional shape and rate parameters using $\frac{\mu^2}{\sigma^2}$ and $\frac{\mu}{\sigma^2}$, respectively), and the mean and concentration of the vonMises distributions.

The state process $\mathbf{S}_t$ is dictated by a matrix of transition probabilities such that $\gamma_{i,j}^t = \Pr(S_{t+1} = i|S_t = j)$, where $i, j = 1, ..., N$. Often these probabilities are related to time-varying covariates via multinomial logit-link functions (e.g., [43, 104, 109]), but in this case, the transition probability matrix $\boldsymbol{\Gamma}$ (whose $(i,j)^{th}$ element is $\gamma_{i,j}$) is treated as constant across the time series of observations. The likelihood can be expressed in terms of the set $Z_{1:t} = \{\mathbf{z}_1, ..., \mathbf{z}_t\}$ as:

$$L(Z, \theta) = \prod_{t=0}^{T-1} f(\mathbf{z}_{t+1}|z_{1:t}, \boldsymbol{\theta})$$

which is the likelihood of observing $\mathbf{z}_{t+1}$ conditional on the occurrence of the observed sequence $Z_{1:t}$, where the Markov property implies that $f(\mathbf{z}_{t+1}|z_{1:t}, \boldsymbol{\theta}) = f(\mathbf{z}_{t+1}|\mathbf{z}_t, \boldsymbol{\theta})$. Given the first term in the likelihood $f(\mathbf{z}_1|\mathbf{z}_0, \boldsymbol{\theta})$, each subsequent value of $f(\mathbf{z}_{t+1}|\mathbf{z}_t, \boldsymbol{\theta})$ can be calculated using a recursive scheme [122]. The parameters ($\hat{\boldsymbol{\theta}}$) that minimize $-\log[L(Z, \theta)]$ are then obtained using an appropriate optimization algorithm.

This behavioral state-space model was applied to the combined set of all 11 empirical movement trajectories from Etosha National Park, irrespective of the year. The `moveHMM` package accepts the number of latent states (N) as a user-defined input, so it is common practice to test alternatives and select the model that results in the lowest AIC value. Both a two-state and three-state model were fit to the data, and the built-in AIC function was used to determine the best-fitting model. The three-state model was selected based on its AIC value. The estimated parameter values governing the step length and turning angle distributions are displayed in Table 4.4. The names assigned to each of the latent behavioral states (resting, foraging, and directed movement) were selected based on the step length and turning angle distributions that characterized each. This interpretation of the HMM results is one of many possibilities, but the names are intended only to ground the sets of parameter estimates in simplified terms. The output of the HMM also includes a transition probability matrix ($\boldsymbol{\Gamma}$; Table 4.5).

The estimates of the parameter values describing the step length and turning angle distributions for each state and the estimates of the transition matrix elements are all associated with some uncertainty. The step length distribution and transition matrix were both incorporated into the simulation model described below in an *ad hoc* fashion that stripped all estimates of their uncertainties. However, to verify that the model was replicating some of the patterns observed in the

empirical movement trajectories, hidden Markov models were applied to several simulated paths and compared to HMMs from the real zebra tracks (see Pattern-Oriented Modeling below).

Simulation Approach

The simulation at each time step t begins with the assignment of the agent to one of the three behavioral states. This behavioral state was assessed by drawing upon the transition probability matrix that emerged from the HMM described above (Table 4.5). Based on the state of the individual at time $t-1$, a set of three probabilities was extracted from the matrix $\mathbf{\Gamma}$, representing the probability of moving from the current state into each of the other states, as well as the probability of remaining in the same state. These probabilities served as the weights of a stochastic draw to determine the behavioral state at time t. The same transition probability matrix governed all of the simulations, without any variation over time.

Once the behavioral state (S_t) was selected, a set of perceptual range radii were extracted from a matrix guided by the step length distribution parameters estimated in the HMM (Table 4.6). Using a randomly generated number from a uniform distribution (u_t), it was determined whether the perceptual range would be small, medium, or large for the step at time t. Because the standard deviation of the gamma distribution is not easily interpretable, an alternative means of dividing the step length distributions into three classes was needed. The use of several classes (as opposed to the use of the same large perceptual range value for all steps within a particular behavioral state) served to reduce the computational cost of each movement decision. The use of a uniform distribution would result in approximately 33% of steps falling within each of the step size classes, but the use of a normal distribution further reduces the probability of selecting the largest step size class, thereby conferring computational benefits beyond the uniform breakpoints. Thus, boundaries were delimited between the step size classes using the proportion of the normal distribution probability density function (PDF) falling within one, two, and three standard deviations of the mean. These correspond to 68%, 95%, and 99.7% of the probability density, respectively, and the radii associated with the small, medium, and large steps are determined using the gamma distribution PDF at each of those percentile values. In this way, 68% of the gamma PDF lies to the left of the small radius value, 95% lies to the left of the medium radius, and 99.7% lies to the left of the large radius (Figure 4.2). In practice, this means that the small perceptual range would be selected 68% of the time (i.e., $0 <= u_t < 0.68$), the medium range would be selected 27% of the time (i.e., $0.68 <= u_t < 0.95$) and a rare large range would be selected only 5% of the time (i.e., $u_t > 0.95$).

The process of determining the next location of the agent at time t involved the use of the selection surfaces created at the initialization of the model. These surfaces did not change throughout the course of the simulation, but each simulation scenario was associated with a unique set of selection surfaces based on the Wetness and Greenness layers that served as inputs to the step selection function. For each step, the perceptual range served as a buffer around the current position of the individual; all of the raster cells whose center points fell within the perceptual range were potential destinations of the individual. Let $\boldsymbol{C_r}$ represent the set of M cell center points that fall within the circular perceptual range with radius r. The movement process consisted of selecting one of the points contained in $\boldsymbol{C_r} = [m = 1, ..., M]$ after weighting each of the m points by the

suitability value (w_m) extracted from the appropriate selection surface (Figure 4.3). In this way, the cell that was theoretically most suitable based on the underlying environmental features (i.e., w^{max}) would have the highest probability of being selected, but the individual was not guaranteed to select that cell. This was because the selection process was stochastic rather than deterministic, and all of the cells in $\boldsymbol{C_r}$ had non-zero w unless they fell outside of the region of interest. In the case of the resting state, during which the animal was not making a concerted movement decision, a randomized selection map (as opposed to those generated using the step-selection function coefficients) was used to guide the probabilistic selection of its next step. In most cases, the perceptual range during the resting state was so small that only the current cell center would fall within the boundary, resulting in the animal remaining in its current cell. In all cases, small error terms (two values drawn from uniform distributions ranging between -14.99 and 14.99) were added to the x and y coordinates of the destination cell so that the individual did not move from cell center to cell center.

Infectious Reservoir Distribution

Locally infectious zones (LIZs), centered around the point locations at which an animal succumbs to its infection, are the critical infectious components of the anthrax system [24, 153, 151]. Due to the resilience of the *B. anthracis* spores, the area immediately surrounding an infected carcass can contain infectious material for extended periods of time (on the order of multiple years) [10]. Subsequent visits by grazers to these LIZs may result in their infection when spores in the soil are incidentally ingested along with vegetation [153]. During surveillance efforts in Etosha National Park, the species of the carcass is recorded in addition to several other attributes. Based on this information, the average yearly rate of carcass mortality, and the density of carcasses across the landscape, a set of simulated LIZ sites was created for use in the model. Thus, the distribution of LIZs proceeded according to the following steps: 1) map locations of carcass sites that exhibited anthrax-positive soil samples in the two years after deposition; 2) construct a niche model of anthrax risk based on these locations and a set of environmental covariate layers that did not serve as inputs to the step-selection model (see previous chapter for more details); 3) use the resulting surface as a set of weights to probabilistically distribute a set number of LIZs across the landscape for use in the simulation approach.

Each simulated LIZ was identified in the model as either small (representing a springbok-sized carcass), medium (representing a zebra or wildebeest carcass), or large (representing the occasional elephant-sized carcass). The rates at which each of these occurred were based on the observed proportions in the empirical carcass data from Etosha National Park (14.8% small, 82.7% medium, and 2.5% large). Each LIZ was then assigned an initial mass (associated with the individual at the time of its death) based on the size category to which it was designated. In addition, the age of the LIZ (the number of years prior to the initialization of the model, up to three) was assigned, such that the number of the LIZs on the landscape deposited each year was approximately equal and reflected empirical observations in Etosha National Park (and considered with the error associated with surveillance efforts; [9]). The carcass surveillance data enabled the initialization of the LIZ layer by informing the likely density of carcasses (0.0135 carcasses/km^2/year).

The anthrax risk layer underlying the step-selection function, which was built based on the empirical carcass data, also served as the basis for the placement of the simulated LIZs $k = \{1, ..., z\}$. In this way, the distribution of the LIZs was not random, but was weighted by the probability layer constructed using a MaxEnt niche modeling approach (see Chapter 3 methods). The initial mass of the carcass (B_k) and the time since death (A_k) were used to create a buffer around the central point associated with the LIZ. The size of the buffer (S) of each LIZ was determined by the following equation:

$$S_k = \frac{\log(B_k)}{\log(350)} * \chi, \qquad \chi \sim \mathcal{N}\left(\mu, \sigma^2\right)$$

$$\text{where} \quad \mu = 3(4 - A_k) \quad \text{and} \quad \sigma = \frac{3(4 - A_k)}{8}$$

These polygons served as the synthesized LIZ layer for the calculation of the contact rate, described below.

Though the anthrax season is likely long enough for some individuals who encountered infectious carcass sites to succumb to the disease, such mortality events were not incorporated here. This was deemed reasonable because freshly deposited carcasses have been reported to have repulsive effects on live individuals [151], meaning that they would be unlikely to influence the subsequent contact rates during the observation period.

Contact Rate Calculation

The two components used for the calculation of the contact rate across each simulated path were the circular buffer polygons created at the locations of the LIZs and a set of linear buffers along each of the steps during which the animal was in the foraging state. A 15 meter buffer was created along the length of each step to account for some of the uncertainty that exists regarding the actual path traversed between fixes. In addition, a buffer of this size can help incorporate some of the existing information about ungulate foraging behavior at finer scales than the step-selection function analysis. Using this buffer to count probable contacts implies that an animal that gets close to a LIZ will probably go slightly off its most efficient trajectory in order to forage in the immediate vicinity, where vegetation is likely to be of higher quality than the surrounding area. Because of the uncertainty regarding the dose required to cause infection or death in a wild ungulate, I considered only the probability of a transmission-enabling contact rather than the actual probability of transmission itself.

A contact rate (C) between the agent and the LIZs distributed across the landscape can be calculated simply as:

$$C = \frac{\text{number of times the two sets of buffers overlap}}{\text{total number of steps throughout anthrax season}}$$

Despite using all of the steps during the time period of interest in the denominator, an effective contact is only achieved when the animal forages at the site, so only those steps are used in calculating the numerator (Figure 4.4).

Pattern-Oriented Modeling

Given that the movement of the individual across the landscape is based upon general rules extracted from empirical paths, it would be helpful to verify that the simulation is maintaining some semblance of reality. The pattern-oriented modeling (POM; [65]) approach offers one means of verifying that certain important aspects of the simulation match measured patterns. Certainly, there are multiple mechanisms that might give rise to similar patterns, but the general principle behind POM is that a model that is able to replicate empirical patterns should be considered as theoretically plausible when more concrete means of validation are unavailable. Here, I compared the mean step lengths emerging from a hidden Markov model conducted on eleven randomly selected individual paths from the set of simulated paths on each surface to the mean step lengths emerging from HMMs on the eleven empirical zebra paths. For each of the sixteen sets of simulated paths, the means of these mean step length distribution were compared to the mean of the mean step length distribution of the observed zebra using a two-tailed t-test.

4.4 Results

Selection Map Derivation

The population-wide step-selection functions created for the foraging and directed movement states are displayed in Table 4.7. These coefficients served as the basis for all sixteen of the different surfaces applied across the simulations, with only the input layers themselves being altered. These distinct selection maps gave rise to theoretically unique simulated trajectories.

Infectious Reservoir Distribution

Based on the density calculated based on observed carcasses in Etosha National Park, a total of 282 LIZs were dispersed across the region of interest (Figure 4.5). The same set of LIZs were used across all of the simulations. The LIZs in this simulation were relatively small, with infectious radii ranging from about 1.2 meters (for the smallest carcasses, akin to a springbok) to nearly 13.3 meters (for the rare large carcasses, akin to elephants). These infectious sites occupied a total area of approximately 32,351 m^2, or about 0.032 km^2. This means that the LIZs composed approximately 0.00046% of the 6972 km^2 region of interest. Due to their probabilistic placement according to the anthrax risk map, their distribution throughout this area was neither random nor uniform. One means of measuring the distribution of points in space is the use of the measure J [90], which is evaluated over a range of radii (r). J is derived based on G (the nearest neighbor distance distribution function) and F (the empty space function of the process). $J(r)$ values equal to 1 indicate a Poisson process at the scale of r, whereas values of $J(r)$ smaller or larger than 1 indicate clustering or regularity, respectively. In the case of the LIZs distributed across the landscape, all

of the $J(r)$ values were less than 1, signifying substantial clustering in the point pattern (beyond the scale of 150 meters; see Supplementary Figure 4.8).

Pattern-Oriented Modeling

To test for similarity between the simulated paths and the empirical paths, the outputs of hidden Markov models conducted using individual paths were compared. For each path (eleven empirical and eleven randomly selected simulated for each surface), the mean step sizes from the foraging and directed movement states were extracted. Because the same empirical paths were used throughout all of the comparisons, a distribution was built based on the mean and standard deviation across all of the empirical trajectories. For the foraging state, this distribution had a mean of 142.2 meters and a standard deviation of 50.6 meters. For the directed movement state, the distribution had a mean of 682.7 meters and a standard deviation of 184.4 meters. In Figure 4.6, 1000 random samples were drawn from these distributions to create the gray histogram in the background of both plots. The means evaluated for the eleven simulated paths from all sixteen surfaces were plotted atop these histograms to compare the distributions visually. In addition, t-tests were conducted to compare the means from the eleven empirical tracks to the eleven simulated tracks for each of the surfaces. In all sixteen cases, the means from both the foraging steps and the directed movement steps were not significantly different from the means during these two states as extracted from the empirical paths (i.e., all of the p values were > 0.05; Table 4.8). These results indicate that the simulated paths are likely being drawn from step length distributions that are similar to those that underly the true step lengths observed in zebra in Etosha National Park.

Interpretation of Simulated Results

An ordinary linear regression model was built to determine whether any consistent trends emerged regarding the role of the environmental covariates (Greenness and Wetness) in the exposure process. The results indicate that neither variable is a significant predictor of mean contact rate across the 1000 simulations at the $\alpha = 0.05$ level (Table 4.10) and the fit of the model is negligible (Adjusted $R^2 = -0.01$). The inclusion of an interaction term between Greenness and Wetness resulted in a model with similarly low explanatory power. When quadratic terms of both Wetness and Greenness are added to the regression, the model fits the simulated mean contact rates much more closely (Adjusted $R^2 = 0.62$; Figure 4.7), though only the Wetness components are significant predictors of mean contact rate (Table 4.11). An additional set of carcass sites were distributed across the landscape and the contact rates recalculated to evaluate the robustness of the results to alternative LIZ distributions. The regression coefficients were nearly identical (Table 4.13) and the explanatory power of the model was even greater than the model describing the contacting rates on the initial set of LIZs (Adjusted $R^2 = 0.71$).

4.5 Discussion

A regression model containing quadratic terms proved to be highly explanatory of the pattern of contact rates calculated across the simulations. This model revealed that moderate levels of

Wetness were associated with the highest contact rates. The emergent pattern, whereby intermediate Wetness layers are associated with the highest simulated contact rates, matches the observed pattern of anthrax-related mortalities (Figure 4.7). Though there may be some lag time between exposure and death (and an additional lag before a carcass is observed), the time between infection and death is estimated at approximately a week [24]. Thus, the date of observation of a carcass is unlikely to be substantially different from the date of contact and exposure. Early in the anthrax season, multi-year rainfall data suggests that relatively high mean Wetness is expected, and in turn, relatively low contact rates. By the middle of the anthrax season, Wetness values achieve moderate levels and give rise to relatively higher contact rates. Finally, as the anthrax season nears its end, the landscape experiences very little rainfall, further reducing mean Wetness and resulting, once again, in diminished contact rates.

By calculating the mean contact rates that emerge from an exhaustive set of combinations of Wetness and Greenness surfaces, the results are removed from the actual progression of the anthrax season from high rainfall (February and March) to low rainfall (April through June). In other words, artificial scenarios in which high Wetness values were combined with low Greenness values can be tested to determine whether one environmental component dominates the relationship with contact rates. Despite these artificial pairings, the signal of changing Wetness on contact rates is evident. Assuming that higher mean contact rates are correlated with higher rates of infection and, in turn, mortality, it seems that the Wetness layer may be a reasonable predictor of anthrax-related deaths throughout the anthrax season.

The results presented here provide a compelling link between environmental factors and epidemiological processes in the form of animal movement. By altering the attractiveness of different portions of the landscape, certain combinations of Wetness and Greenness give rise to space use patterns that result in higher contact rates between a host and a set of infectious reservoirs on the landscape. Without the mechanistic movement model serving as the intermediary by which environmental factors are reflected in epidemiological processes, the observed pattern of mortality in Etosha National Park could remain a mystery. In this way, movement mediates the relationship between environmental covariates (in this case, Wetness, which tends to vary in a linear fashion over the course of the anthrax season) and epidemic dynamics (here, the peaks in both simulated contact rate and observed anthrax-related mortalities in the middle of the anthrax season).

Another interesting pattern that may be supported by the results of this simulation exercise are rates of sub-lethal exposure in ungulates in Etosha based on field investigations. Using serum samples from 154 zebra revealed that 52%-87% exhibited some level of anti-anthrax antibody titres [27], though the rate at which titres may wane over time is not known. Across all sixteen of the simulated surfaces, the probability of encountering at least one LIZ over the course of the anthrax season was approximately 51.1% (ranging from 44.4% to 56.2%). If one assumes that titres are detectable for about a year after encountering a non-infectious dose, these contact rates are quite close to the lower end of the observed rate. If, however, one assumes that titres last for approximately three years, the proportion of animals in the population likely to exhibit anti-anthrax antibodies would reflect the upper bound of the estimate ($\approx 88.3\%$).

The general purpose underlying the simulation of movement trajectories is to extract general

rules from empirical tracks and expand upon the limited sample size frequently associated with GPS tracking studies. An alternative approach would involve simulating a large number of different sets of LIZ distributions on the landscape and simply using the eleven empirical tracks to calculate a distribution of contact rates. However, I decided to simulate both the movement trajectories and LIZ sites in order to distance the results appropriately from the real life system so that the contact rates calculated here would not be mistaken for true estimates of contact between zebra and anthrax spores. In most wildlife disease systems, precise contact rates, whether generalized over an entire host population or recorded for each individual separately, are only very rarely calculated directly. The effort involved in the near-continuous monitoring of individuals normally precludes such calculations (but see fairly extensive studies regarding tuberculosis transmission among badgers and livestock; e.g., [134, 168]).

The goal of this exercise was explore the theoretical relationships between environmental covariates and epidemiological processes using a mechanistic movement model parameterized, in part, on empirical movement trajectories. Simulation enabled estimates of the rates at which individuals were likely to come into contact with an indirectly-transmitted pathogen within an environmental reservoir across a range of environmental scenarios. The outputs of the simulation included a set of movement trajectories (covering the five-month anthrax period at a temporal resolution of 20 minutes per fix) and the locations and sizes of LIZs across the landscape. Due to the limitations of most field surveillance efforts, it is unlikely that a researcher would have the ability to map out the exact locations of every infected carcass on the landscape (estimates from the Etosha system based on a hierarchical model of distance sampling place the rate of detection at approximately 25%; [9]), so the simulation framework offered an alternative approach that enabled full knowledge of the distribution of risk across the landscape.

Several of the complexities of infection dynamics, including considerations of heterogeneity in dosages and immune responses, were excluded from this model. Instead, the emphasis was on potential transmission events and a more readily measurable metric: the contact rate between a host and an infectious reservoir represented by LIZ sites. Several additional variables and components would be needed to consider more complete infection dynamics. Additional state variables would include parameters defining agent immune systems and changing bacterial densities at LIZs. Infection and immune response components would also be required, which could account for altered movement patterns in infected hosts and disease-induced mortality during and after the anthrax season. Due to the lack of empirical data for parameterizing such variables and processes, the model presented here was not extended in this manner.

4.6 Tables

Table 4.1: Serial location data (collected every 20 minutes from 9 zebra) that were used for parameterizing movement in the simulation model. Two of the individuals had movement tracks that extended across two anthrax seasons (defined as the period from February through June), so 11 paths were extracted from this data set.

Individual ID	Number of Points	Start Date	End Date
AG059	20,769	2009-04-20	2010-08-29
AG061	17,196	2009-04-20	2010-03-03
AG062	12,794	2009-04-20	2010-01-30
AG063	25,721	2009-04-20	2010-04-30
AG068	32,661	2009-04-20	2010-08-29
AG252	23,450	2009-10-06	2010-08-29
AG253	21,676	2009-10-06	2010-12-17
AG255	23,470	2009-10-06	2010-08-29
AG256	23,519	2009-10-06	2010-08-29

Table 4.2: All Landsat images with less than 10% cloud cover were treated as potential environmental covariate layers. The eight dates displayed here indicate all of the dates during the anthrax seasons of 2009 and 2010 that met that criterion. The four dates with asterisks were selected because they offered the best range of Greenness and Wetness values to explore the contact rate dynamics. The two extreme sets of values were selected as well as two more moderate layers.

Layer Dates	Greenness Value	Wetness Value
2009-03-22*	-8.62	-20.92
2009-04-23	-18.97	-21.98
2009-05-09	-19.82	-32.49
2009-05-25*	-18.53	-33.98
2010-02-05*	-37.14	-73.46
2010-04-10	-16.77	-22.82
2010-05-12	-23.25	-45.67
2010-05-28*	-24.93	-43.82

Table 4.3: Mean Greenness and Wetness values for each of the sets of input layers used in the simulations. The values are calculated across the region of interest in Etosha National Park, and include the portion of the Etosha Pan included in that region. The sixteen sets of layers form the basis of the grid of covariate layers over which the probabilities of contact are compared to determine the role of these environmental factors in the exposure process. In addition to the combinations that were used in the simulations, the overall mean values used in the population-level selection function is presented.

Surface ID	Greenness Value	Greenness Date	Wetness Value	Wetness Date
1	-37.14	2010-02-05	-73.46	2010-02-05
2	-37.14	2010-02-05	-43.82	2010-05-28
3	-37.14	2010-02-05	-33.98	2009-05-25
4	-37.14	2010-02-05	-20.92	2009-03-22
5	-24.93	2010-05-28	-73.46	2010-02-05
6	-24.93	2010-05-28	-43.82	2010-05-28
7	-24.93	2010-05-28	-33.98	2009-05-25
8	-24.93	2010-05-28	-20.92	2009-03-22
9	-18.53	2009-05-25	-73.46	2010-02-05
10	-18.53	2009-05-25	-43.82	2010-05-28
11	-18.53	2009-05-25	-33.98	2009-05-25
12	-18.53	2009-05-25	-20.92	2009-03-22
13	-8.62	2009-03-22	-73.46	2010-02-05
14	-8.62	2009-03-22	-43.82	2010-05-28
15	-8.62	2009-03-22	-33.98	2009-05-25
16	-8.62	2009-03-22	-20.92	2009-03-22
Mean	-21.00		-36.89	

Table 4.4: Parameter estimates from the Hidden Markov Model.

State	**Step Lengths (Gamma)**			**Turning Angles (vonMises)**	
	mu	sigma	zero-mass	mean	concentration
Resting	20.40	22.92	0.001	2.917	0.017
Foraging	154.7	131.1	0.001	0.002	1.389
Directed	739.5	262.9	0.0	-0.002	4.079

Table 4.5: Transition Probability Matrix ($\mathbf{\Gamma}$) estimated from the combined dataset of all 11 empirical zebra tracks collected during the anthrax season in 2009 and 2010. This matrix was used as the static basis for state transitions in the simulation.

	State 1 $(t+1)$	State 2 $(t+1)$	State 3 $(t+1)$
State 1 (t)	0.78	0.22	2.6e-10
State 2 (t)	0.16	0.75	0.08
State 3 (t)	0.01	0.23	0.76

Table 4.6: The distance (in meters) used as the radius of the perceptual range of the simulant based on their current behavioral state and whether they would be taking a small, medium, or large step.

	State 1	State 2	State 3
Small	23	181	858
Medium	66	413	1420
Large	136	747	2121

Table 4.7: Results of the conditional logistic mixed effects models as applied to foraging points (n = 39,631) and directed movement points (n = 15,867) 20-minute fix intervals.

Foraging	coef	exp(coef)	se(coef)	z	p	sig
Wetness (β_{W})	-0.15	0.86	0.025	-5.76	8.4e-09	***
Greenness (β_{G})	0.22	1.25	0.030	7.47	8.1e-14	***
Road Density (β_{RD})	-0.00	1.00	0.005	-0.83	0.4	
Anthrax Risk (β_{AR})	-0.08	0.93	0.009	-8.32	<2e-16	***

Directed	coef	exp(coef)	se(coef)	z	p	sig
Wetness (β_{W})	-2.18	0.113	0.066	-33.21	<2e-16	***
Greenness (β_{G})	-0.22	0.80	0.076	-2.91	0.004	**
Road Density (β_{RD})	0.06	1.06	0.007	8.83	<2e-16	***
Anthrax Risk (β_{AR})	0.10	1.11	0.021	5.08	3.8e-07	***

Table 4.8: The results of the individual-level hidden Markov models applied to eleven randomly selected simulated paths from each selection surface. The mean step sizes from both the foraging and directed movement states are presented, as well as the p values from a two-tailed t-test between these simulated paths and the eleven empirical trajectories.

Surface ID	Mean (Foraging)	p (Foraging)	Mean (Directed)	p (Directed)
1	138	0.83	630	0.42
2	142	0.98	632	0.45
3	149	0.73	661	0.73
4	150	0.66	659	0.71
5	145	0.89	648	0.59
6	142	0.52	635	0.17
7	143	0.95	643	0.54
8	144	0.93	641	0.53
9	141	0.96	631	0.45
10	135	0.70	613	0.31
11	132	0.60	604	0.24
12	145	0.89	645	0.57
13	144	0.94	641	0.52
14	130	0.53	592	0.20
15	139	0.86	630	0.42
16	150	0.68	665	0.78

Table 4.9: The mean number of contacts and contact rates emerging from the 1000 simulated trajectories on each of the sixteen alternative selection surfaces. Also presented are the standard deviation of the number of contacts, the number of simulations (out of 1000) that resulted in zero contacts, and the maximum number of contact events observed in a single simulated trajectory.

Surface ID	Mean	Rate	St.Dev	Zeros	Max
1	2.17	2.01e-04	3.30	486	22
2	2.34	2.17e-04	3.37	464	21
3	2.17	2.01e-04	3.02	470	16
4	2.00	1.85e-04	3.38	556	22
5	2.08	1.92e-04	3.05	505	16
6	2.29	2.12e-04	3.18	459	21
7	2.29	2.12e-04	3.35	479	19
8	1.99	1.85e-04	3.23	541	20
9	2.14	1.98e-04	3.23	499	20
10	2.32	2.15e-04	3.30	452	22
11	2.50	2.18e-04	3.32	458	20
12	2.02	1.87e-04	3.24	523	20
13	2.29	2.12e-04	3.35	470	24
14	2.51	2.33e-04	3.42	438	20
15	2.12	1.97e-04	3.20	498	23
16	2.22	2.06e-04	3.61	529	29

Table 4.10: Ordinary Linear Regression of Wetness and Greenness onto mean contact rate ($R^2 = 0.12$; Adjusted $R^2 = -0.0101$).

	Estimate	Std. Error	t value	p	sig
(Intercept)	2.24	0.120	18.75	8.59e-11	***
Greenness	0.004	0.004	1.16	0.27	
Wetness	-0.001	0.002	-0.71	0.49	

Table 4.11: Non-linear regression including the quadratic terms for both Wetness and Greenness ($R^2 = 0.72$; Adjusted $R^2 = 0.6181$)

	Estimate	Std. Error	t value	p	sig
(Intercept)	1.65	0.193	8.58	3.33e-06	***
Wetness	-3.59e-02	7.42e-03	-4.84	5.2e-04	***
Wetness^2	-3.56e-04	7.54e-05	-4.71	6.4e-04	***
Greenness	1.58e-02	1.11e-02	1.43	0.18	
Greenness^2	2.54e-04	2.35e-04	1.08	0.30	

4.7 Figures

Figure 4.1: **Methods Overview:** The first component of the approach is the derivation of the selection maps (A). There were sixteen alternative environmental scenarios, each with its own unique combination of Wetness and Greenness layers. For each scenario, these two layers were combined with a mean anthrax risk map and a static road density map, and the coefficients from the behaviorally-contingent step selection function (see Chapter 3) were used to create surfaces for the foraging (purple squares) and directed movement (gold squares) behavioral states. A random raster was used as the selection surface during the resting state. (Continued on next page.)

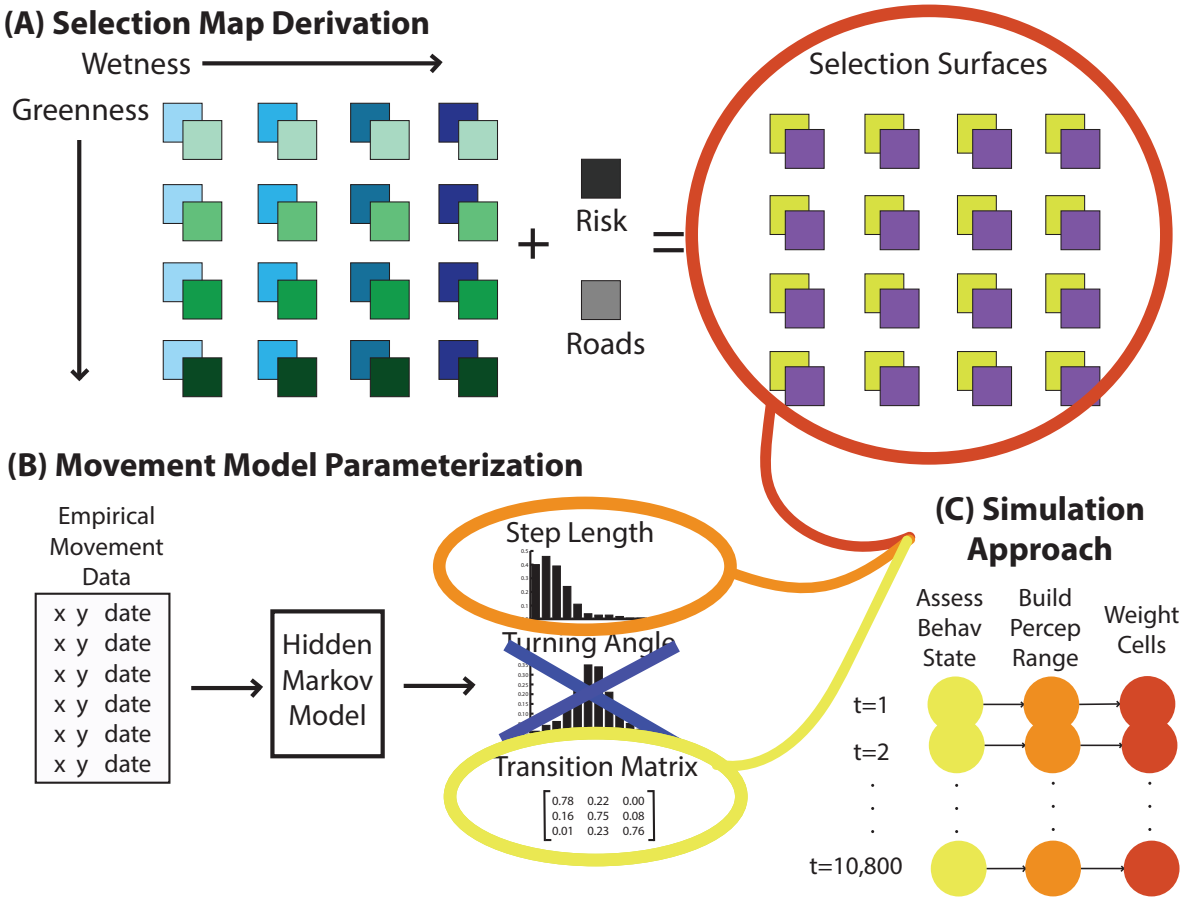


Figure 4.1: (Continued from previous page.) The next component was the movement model parameterization component (B) wherein a hidden Markov modeling (HMM) framework was applied to the empirical movement trajectories. The HMM resulted in a set of parameter estimates for the distributions underlying the step lengths (gamma) and turning angles (von-Mises) during each of the behavioral states, as well as estimates of transition probabilities between the states. Of these four outputs (selection surfaces, step length distribution parameter estimates, turning angle distribution parameter estimates, and estimated transition probability matrix), all but the turning angle parameters were used in the simulation approach. This simulation approach (C) consisted of three primary actions, carried out in sequence and then repeated over the course of the anthrax season. The first of these actions was the assessment of the behavioral state (yellow circles). This action drew upon the transition matrix as the basis for a stochastic process in which the state of the agent at time t was selected based on its state at time $t-1$. The next action was the perceptual range construction (orange circles). This action drew upon the parameters estimates of the step length distribution emerging from the HMM to determine the size of the radius over which the agent would search for its next location. The final action was the weighting of the cells within the perceptual range (red circles). This action drew upon the selection surfaces derived at the start of the simulation for the scenario in question. Cells within the radius were assigned weights based on the relative selection probabilities. The actual location of the next step was determined using another stochastic selection procedure in which the probability of cell being selected corresponded directly to the weight assigned to it. The behavioral state and the location of the animal were recorded at each of the 10,800 time steps, and this process was repeated 1000 times for each of the sixteen sets of selection surfaces.

Figure 4.2: **Step Size Distributions:** A schematic figure of the small, medium, and large perceptual radii as dictated by the gamma distributions underlying the step lengths for each behavioral state (Resting, Foraging, and Directed Movement). In each panel, a set of randomly generated steps were drawn from the gamma distributions and subsequently colored by the thresholds defined as the small, medium, and large perceptual range radii. These distributions indicate the relative probabilities of selecting each of the different perceptual range sizes based on the underlying step size distributions.

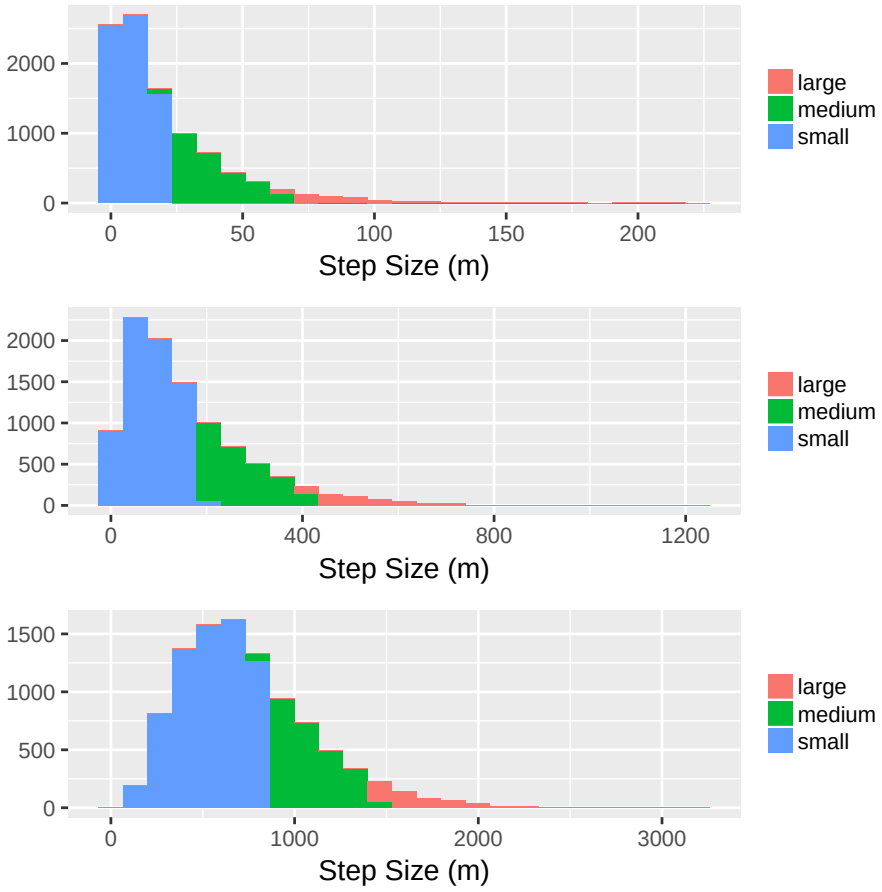


Figure 4.3: **Perceptual Range Schematic:** An example graphic that demonstrates the mechanistic movement process. Following the selection of a perceptual range radius, a buffer (black circle) was constructed around the current position of the agent (red point). To determine the next position, all of the cells whose center fell within the perceptual range (black points) were considered, and weighted by the relative suitability as dictated by the selection surface underlying the current behavioral state of the agent. Cells with larger suitability values were therefore more likely to be selected for the next step, but were not selected in a deterministic fashion.

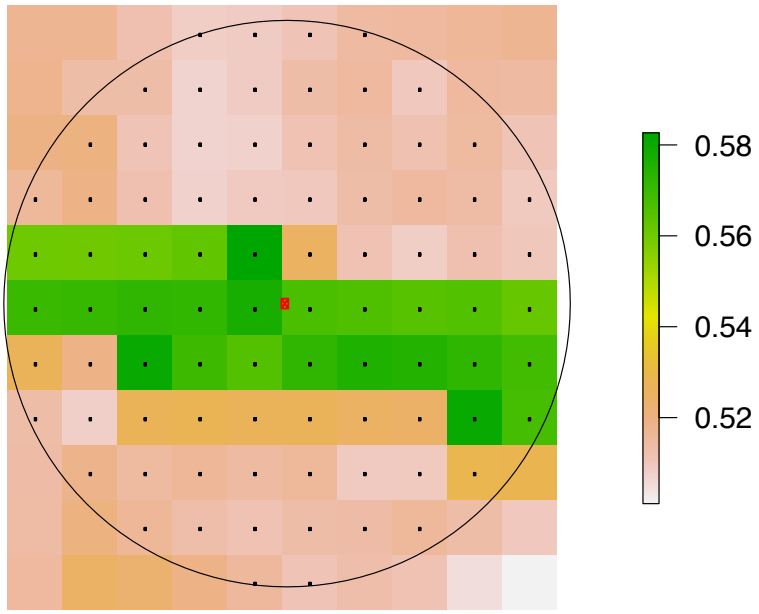


Figure 4.4: **Contact Calculation Schematic:** A schematic diagram of a single contact event between a host and a LIZ. The line represents the inferred path of the individual moving across the landscape, with each point along the path representing recorded positional fixes. When the animal is in the foraging state, a buffer of 15 meters is constructed along the linear path between consecutive points (red polygons). Note that the lines traversed during periods in which the animal is not in the foraging state do not have buffers around them. If the animal encounters a LIZ (blue circle) during a foraging period, indicated by an overlap of the two polygons, it counts as a single contact event. In this case, one LIZ site is centered along the path, but even a partial overlap would be counted as a contact because the animal is assumed to leave its path slightly to investigate the high quality vegetation associated with LIZ sites. Another LIZ was unexplored by the host animal, indicated by a lack of overlap between the red foraging polygons and the blue circle nearer to the bottom left, so no contacts were counted between them.

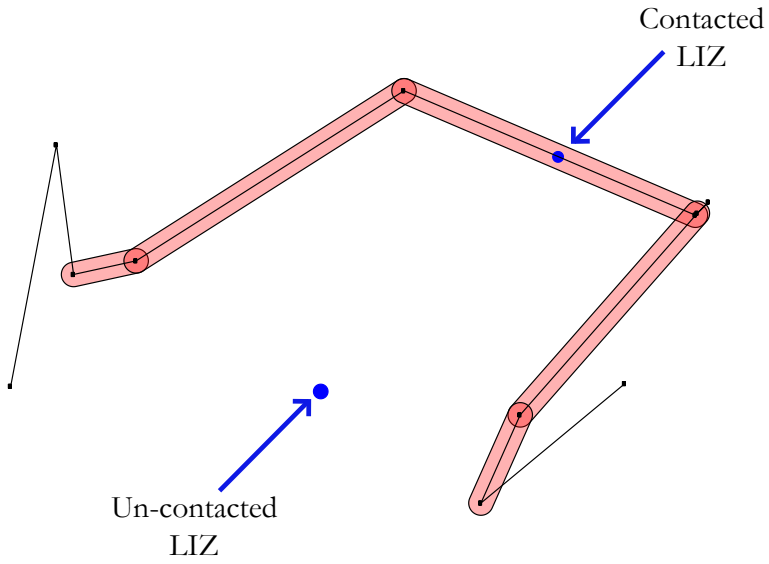


Figure 4.5: **Distribution of Local Infectious Zones (LIZs):** A map of the region of interest in Etosha National Park with the positions of the simulated local infectious zones mapped and colored by the species of the carcass (green for small-bodied animals like springbok, red for medium-sized animals like zebra, and blue for the relatively uncommon large-bodied animal akin to an elephant.

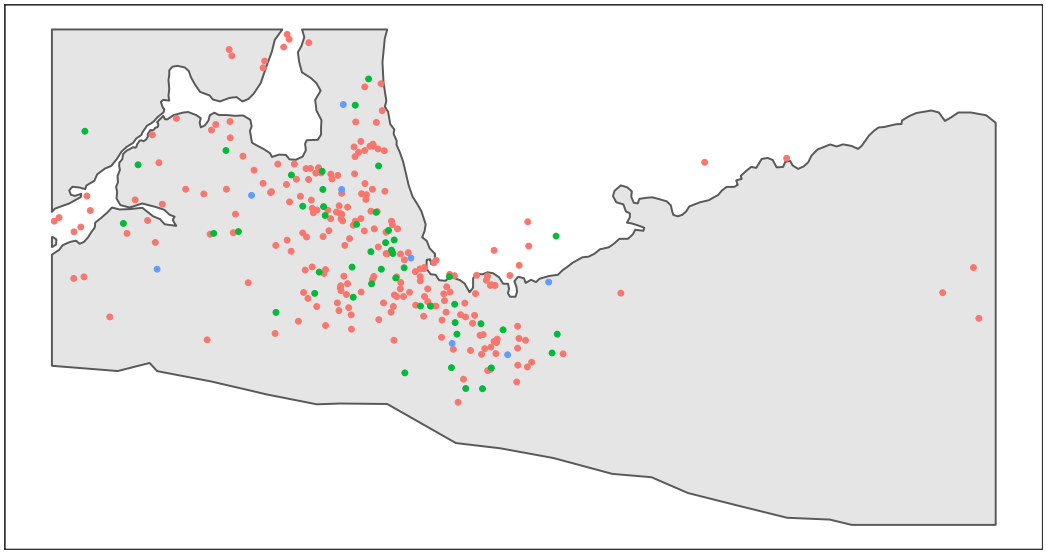


Figure 4.6: **Foraging Step Size Distribution Comparison:** As a means of comparing the simulated tracks to the empirical tracks, the mean step lengths of both the foraging and directed movement states were calculated using individual HMMs for the 11 empirical tracks (red bar) and 11 randomly selected simulated paths from each of the 16 selection surfaces. The results are illustrated here in the form of a box plot, where the thick black line within each box represents the mean of the mean step lengths, the box itself extends from -1 standard deviation to +1 standard deviation, and the additional whiskers extend to the minimum and maximum values (if they are not contained by the box). The points represent the actual calculated mean step length values for each individual with a small horizontal offset to aid in visualizing the distribution.

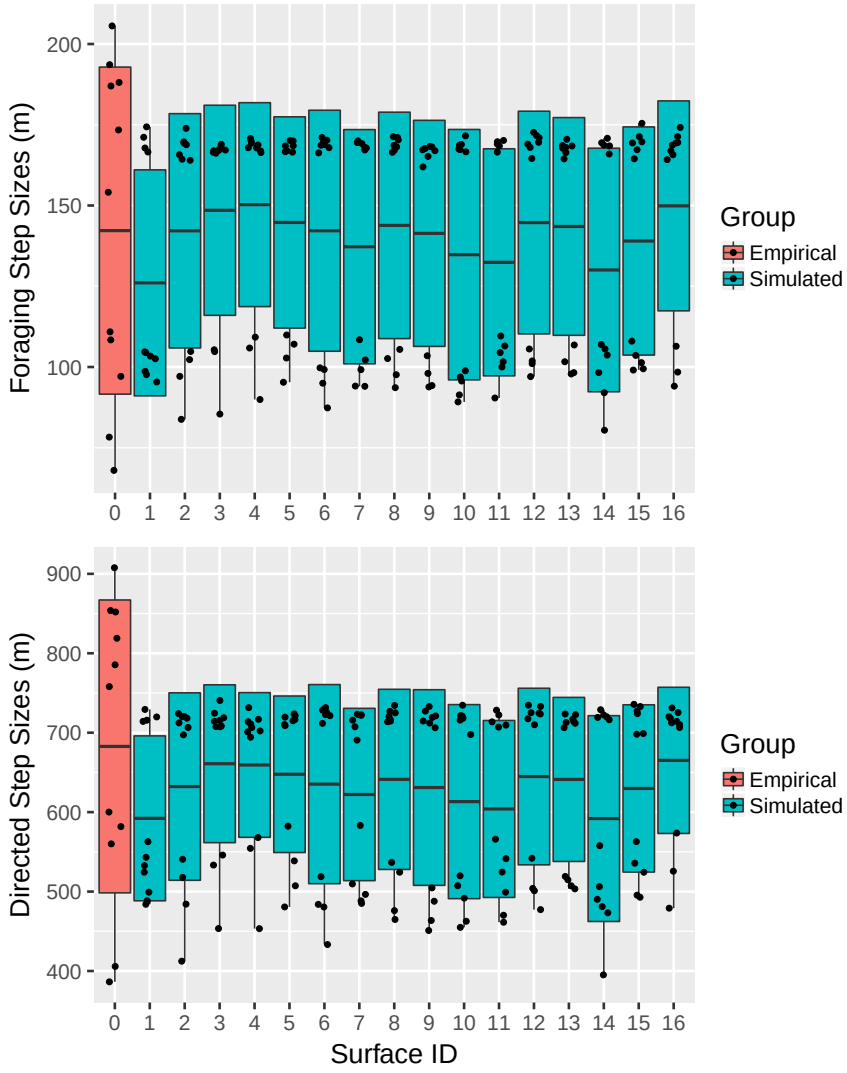
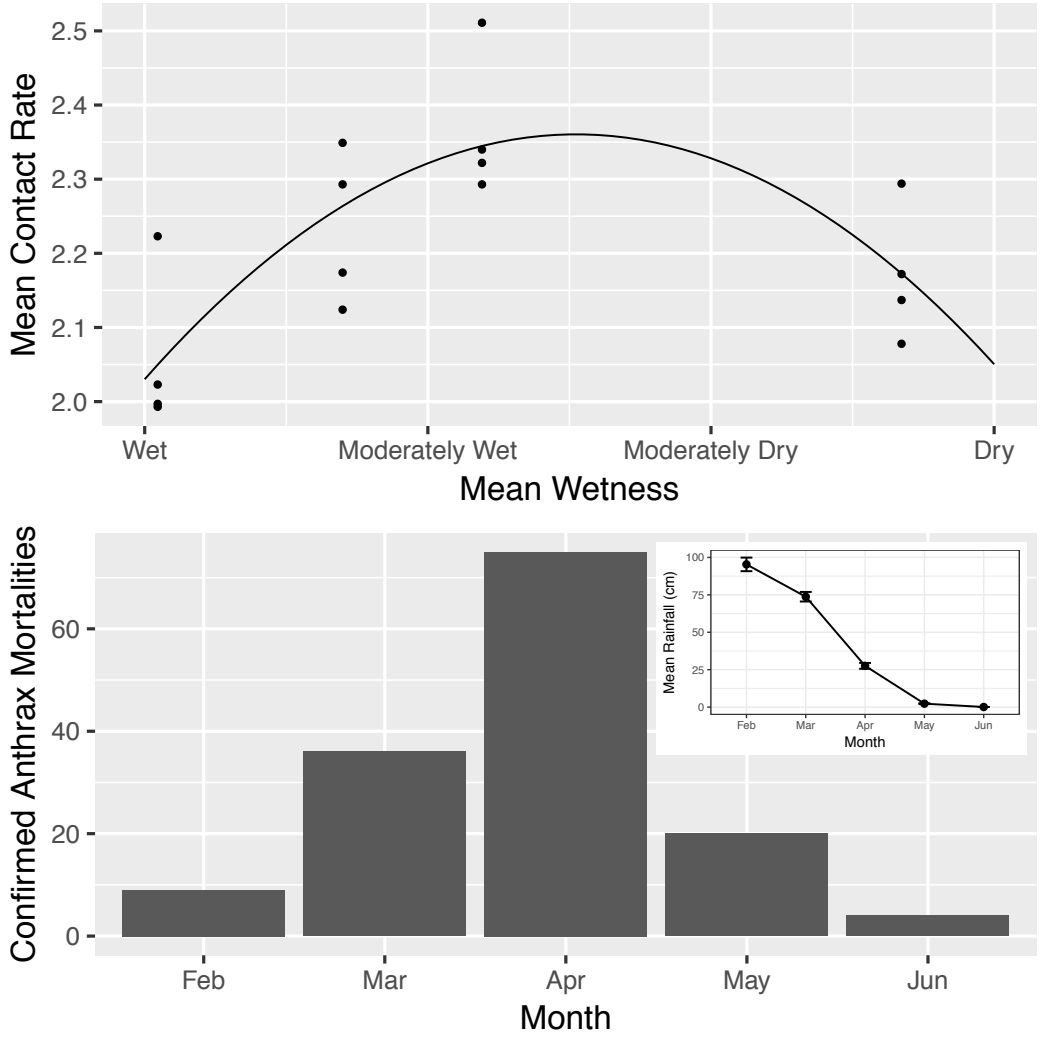


Figure 4.7: **Contact Rate Trend in Relation to Anthrax Mortality Surveillance Data:** The upper panel represents the fitted non-linear regression curve across the tested values of Wetness used in the simulations. The bottom panel presents a histogram of the number of actual anthrax-induced mortalities observed between 1996 and 2009 based on the probable month of death. The inset plot represents the mean rainfall data over the same period, indicating that the anthrax season tends to range from wet in the early months to dry in the later months.



4.8 Supplementary Materials

Figure 4.8: **Estimated J function for LIZ distribution at various scales (r):** $J(r)$ is a measure of the spatial distribution of a point process. If $J(r)$ is smaller than 1 within a radius r of a given point when averaged across a set of points in a particular space (in this case, the region of interest in Etosha National Park), then the points in this space are more clustered than points placed at random (i.e., $J(r) = 1$) at the spatial scale of r. If $J(r)$ is greater than 1, then the points are more regular in the space than points placed at random. Note a the distribution of points in a space can be random or regular at one spatial scale (here, at radii less than about 150 meters), and clustered at other spatial scales (here, at all radii greater than 150 meters).

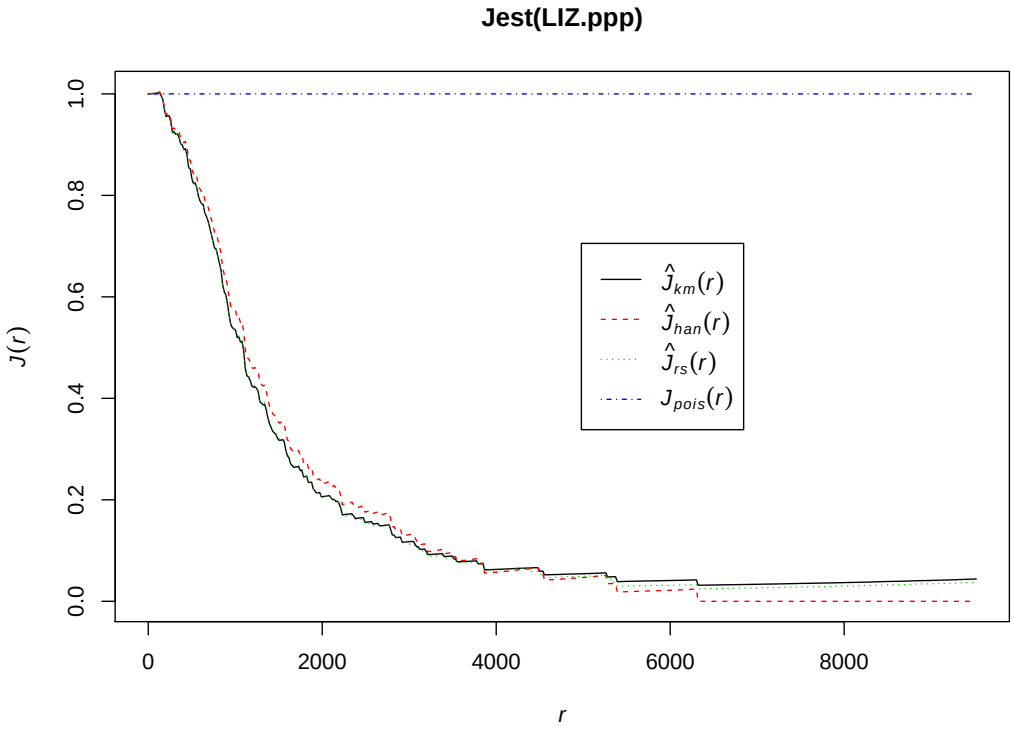


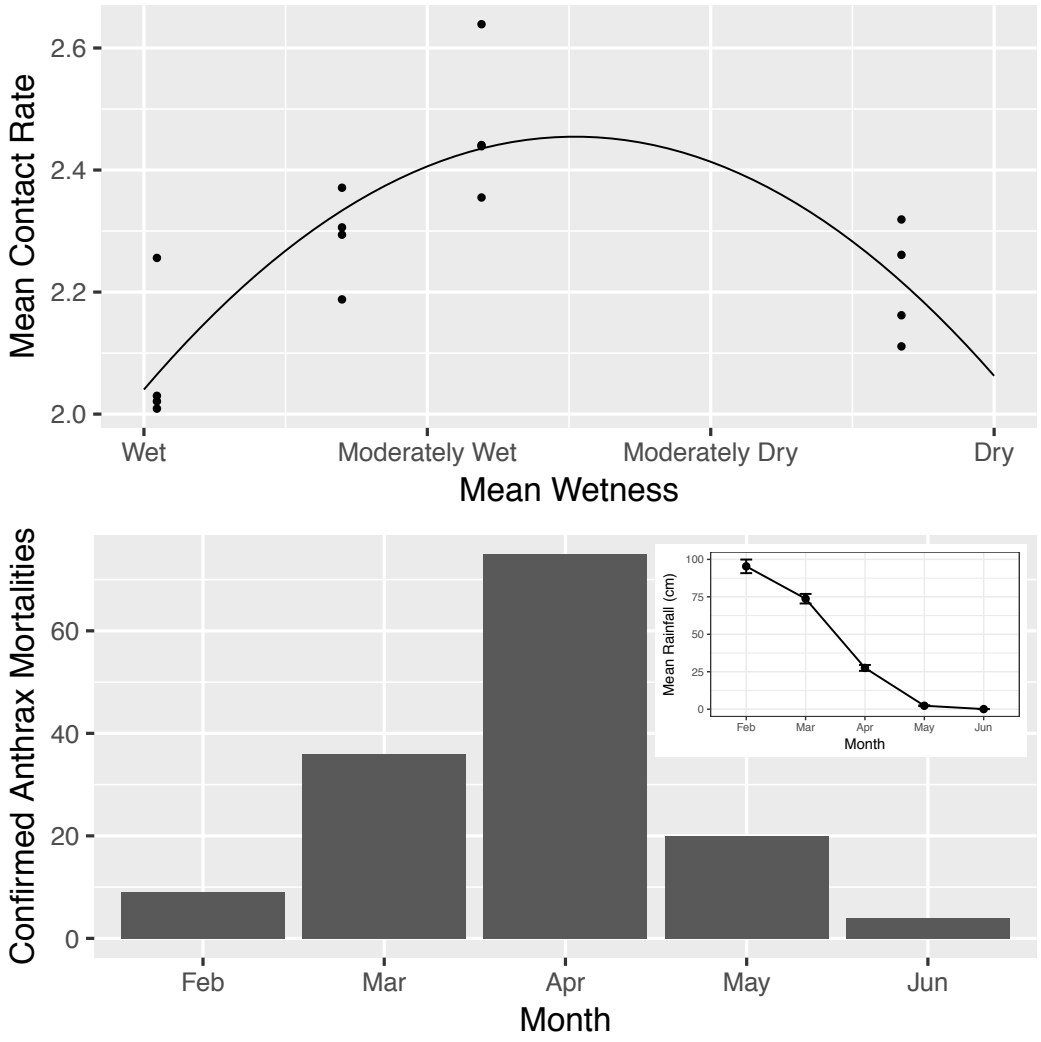
Table 4.12: The mean number of contacts and contact rates emerging from the 1000 simulated trajectories on each of the sixteen alternative selection surfaces. Also presented are the standard deviation of the number of contacts, the number of simulations (out of 1000) that resulted in zero contacts, and the maximum number of contact events observed in a single simulated trajectory.

Surface ID	Mean	Rate	St.Dev	Zeros	Max
1	2.11	1.95e-04	3.21	512	22
2	2.36	2.18e-04	3.10	436	18
3	2.29	2.12e-04	3.19	479	19
4	2.02	1.87e-04	3.31	542	23
5	2.16	2.00e-04	3.27	504	20
6	2.44	2.26e-04	3.38	456	19
7	2.37	2.20e-04	3.42	478	21
8	2.03	1.88e-04	3.39	555	18
9	2.26	2.09e-04	3.26	478	20
10	2.44	2.26e-04	3.29	434	18
11	2.31	2.14e-04	3.24	473	22
12	2.01	1.86e-04	3.09	546	20
13	2.32	2.15e-04	3.33	479	21
14	2.64	2.44e-04	3.47	429	16
15	2.19	2.03e-04	3.16	513	18
16	2.26	2.09e-04	3.49	527	22

Table 4.13: Additional Non-linear regression including the quadratic terms for both Wetness and Greenness ($R^2 = 0.7883$; Adjusted $R^2 = 0.7113$) to demonstrate robustness across alternative LIZ site distributions.

	Estimate	Std. Error	t value	p	sig
(Intercept)	1.49	0.198	7.52	1.17e-05	***
Wetness	-4.52e-02	7.61e-03	-5.93	9.84e-05	***
Wetness^2	-4.48e-04	7.74e-05	-5.79	1.22e-04	***
Greenness	1.12e-02	1.14e-02	0.99	0.34	
Greenness^2	1.31e-04	2.41e-04	0.54	0.60	

Figure 4.9: **Reanalysis of Contact Rate Trend in Relation to Anthrax Mortality Surveillance Data:** The upper panel represents the fitted non-linear regression curve across the tested values of Wetness used in the simulations. The contact rates here are derived based on a second set of LIZs distributed across the landscape. The bottom panel represents a histogram of the number of anthrax-induced mortalities observed between 1996 and 2009 based on the probable month of death. The inset plot represents the mean rainfall data over the same period, indicating that the anthrax season tends to range from wet in the early months to dry in the later months.



Chapter 5

Conclusion

Movement data represent seemingly inexhaustible sources of information, ranging from basic characterizations of space usage to more detailed inferences about movement decisions and behaviors. The constituent chapters of this dissertation were built upon a set of movement data points collected nearly a decade prior to their analysis. In the intervening period, technological advancements have led to dramatic improvements in the devices used to collect data on animal positions at increasingly fine temporal scales [83]. In addition, numerous complementary datasets are commonly collected in conjunction with movement data, offering the potential for a more comprehensive understanding the motivations underling animal movement decisions [116]. The application of advanced computational and statistical methods (e.g., machine learning algorithms [101, 12, 154, 21]) to these datasets may enable the accurate classification of a more descriptive set of behavioral states, and may even expose the fundamental movement elements (FMEs; [58]) that are often subsumed into broader canonical activity modes (CAMs) when describing animal movements.

One of the fundamental limitations of the work presented here was the incomplete nature of carcass surveillance data in Etosha National Park. Though we have an estimate of the magnitude of the discrepancy between the available data and reality [9], there exists no method to approximate the locations of these unobserved carcasses. Instead, assessments of exposure risk were conducted based on entirely simulated systems or highly uncertain predictive surfaces that project statistical associations between known carcass sites and environmental covariates (resulting in the delineation of potential transmission zones; PTZs). While the first of these approaches offers 'perfect' knowledge of the locally infectious zone (LIZ) sites, it displaces the empirical zebra movement paths from reality, relegating them to rule-generators to guide agent movements across the simulated landscape. The second approach allows for the analysis of empirical movement data, but only broad scale estimates of risk can be used, potentially resulting in highly biased interpretations. Though both approaches have their advantages and disadvantages and allow researchers to pose and answer different questions, the truth of ungulate exposure to anthrax will remain elusive without more comprehensive data on pathogen presence.

An important theme throughout much of the work on animal movement and disease transmission is the role of scale. Decisions regarding the temporal scale at which data were originally

collected, the spatial scale of remotely sensed or simulated data for use as predictors of movement, and the spatiotemporal scale at which an analysis for movement data are ultimately conducted will all influence the results and subsequent interpretations. Where data are available, analyses should be conducted at multiple scales to determine the robustness of findings. Though both movement and environmental data will have a lower bound in resolution, it is not necessarily the case that using these lower bounds will result in the most accurate analyses. There are cases in which the resolution of one data source will require that another dataset be coarsened in order to match. For example, the temporal scale of the movement data used throughout this dissertation was a 20 minute fix rate, and much of the environmental data was remotely sensed at a spatial resolution of 30 meters. If, however, the movement data were even finer, say a fix interval of 1 minute, it could be worthwhile to coarsen the data for analysis if the animals tended to stay within a single 30 meter cell for many of their relocations. Depending on the question at hand, finer data might not always be the optimal data. Similarly, collecting bioclimatic data at the scale of 1 meter may not offer much more information than when it is collected at the scale of 30 meters or even 1 kilometer. Considerations of scale will be integral to any spatiotemporal analysis, and results should be examined for sensitivity to scale whenever possible [171].

Future Directions

Perhaps the most promising avenue for exploration is the role of behavior in disease dynamics. Despite the proliferation of statistical methods for extracting behavior from movement trajectories, the practice remains substantially less common than other forms of movement analysis (see Figure 1.1). The findings of Chapter 3 regarding the different selection patterns during the foraging and directed movement states hint at the potential importance of such analyses in revealing important dynamics with respect to the exposure process. It could be interesting to perform a similar analysis on the springbok data, as they tend to browse for their forage rather than grazing (like zebra). In theory, this could put them at a substantially lower risk of exposure than zebra, so any differences between the behavioral states may be less apparent than those observed here. In addition to exploring the potential selection patterns in other susceptible ungulate species, it could be interesting to explore the role of memory in zebra movement. Though memory has been linked (through mechanistic modeling) to stable home range emergence [113, 155, 49] and optimal foraging patterns [106, 139], it has not been extensively explored in terms of navigating a 'landscape of disgust' [163]. If zebra are indeed selecting against areas with high risk whilst foraging, a deeper investigation into the potential mechanisms underlying that behavior should be performed.

Another logical direction for future inquiry would be explicit inclusion of some of the complex factors underlying infection, rather than merely exposure. Given the numerous advancements in remote camera trap technology and the promise of unmanned aerial vehicles (UAVs or drones; [2]), it is not infeasible to conceive of a large-scale monitoring effort that paired the two technologies to identify individuals that ultimately succumb to anthrax. Drones have been demonstrably useful in wildlife monitoring efforts [74, 73] and offer access to areas too remote for human sampling efforts. The development of a drone-based carcass surveillance program could aid in filling the data gap implied by [9]. Once identified, remote camera traps could be set at carcass sites, as in [152]. Using novel classification methods developed to identify species [157] and individuals [26] from

camera trap images, one could potentially create a dataset linking individuals visiting known (and monitored) carcass sites to their eventual demise. This would offer the first insights into the actual infection dynamics in a natural anthrax system. Due to the decreasing cost of GPS devices, it may be feasible to track a larger portion of the zebra population as well, potentially even capturing the movement trajectory of an animal that ultimately succumbs to anthrax infection. Such data could shed light on the changes that an animal undergoes throughout the course of infection (as in [33]). With more information on the nature of transmission, including the probability of contracting an infection given an exposure of a certain length, it would be possible to augment agent-based models like the one presented in Chapter 4 to capture that important aspect of disease. The inclusion of the transmission and infection components to the model would enable the investigation of potential mechanisms underlying the pattern of anthrax endemicity observed in Etosha National Park.

Bibliography

[1] B. Abrahms et al. "Lessons from integrating behaviour and resource selection: activity-specific responses of African wild dogs to roads". In: *Animal Conservation* 19.3 (2016), pp. 247–255.

[2] Karen Anderson and Kevin J Gaston. "Lightweight unmanned aerial vehicles will revolutionize spatial ecology". In: *Frontiers in Ecology and the Environment* 11.3 (2013), pp. 138–146.

[3] Roy M Anderson, Robert M May, and B Anderson. *Infectious Diseases of Humans: Dynamics and Control*. Vol. 28. Wiley Online Library, 1992.

[4] Roy M Anderson, Robert M May, et al. "Population biology of infectious diseases: Part I". In: *Nature* 280.5721 (1979), pp. 361–367.

[5] Amanda V Bakian, Kimberly A Sullivan, and Eben H Paxton. "Elucidating spatially explicit behavioral landscapes in the Willow Flycatcher". In: *Ecological modelling* 232 (2012), pp. 119–132.

[6] Alassane S Barro et al. "Redefining the Australian anthrax belt: Modeling the ecological niche and predicting the geographic distribution of Bacillus anthracis". In: *PLoS neglected tropical diseases* 10.6 (2016), e0004689.

[7] William S Beatty et al. "Landscape effects on mallard habitat selection at multiple spatial scales during the non-breeding period". In: *Landscape ecology* 29.6 (2014), pp. 989–1000.

[8] William S Beatty et al. "The role of protected area wetlands in waterfowl habitat conservation: implications for protected area network design". In: *Biological Conservation* 176 (2014), pp. 144–152.

[9] Steve E Bellan et al. "A hierarchical distance sampling approach to estimating mortality rates from opportunistic carcass surveillance data". In: *Methods in ecology and evolution* 4.4 (2013), pp. 361–369.

[10] Steve E Bellan et al. "Effects of experimental exclusion of scavengers from carcasses of anthrax-infected herbivores on Bacillus anthracis sporulation, survival, and distribution". In: *Applied and Environmental microbiology* 79.12 (2013), pp. 3756–3761.

[11] Aniruddha V Belsare and Matthew E Gompper. "A model-based approach for investigation and mitigation of disease spillover risks to wildlife: Dogs, foxes and canine distemper in central India". In: *Ecological Modelling* 296 (2015), pp. 102–112.

[12] Owen R Bidder et al. "Love thy neighbour: automatic animal behavioural classification of acceleration data using the k-nearest neighbour algorithm". In: *PloS one* 9.2 (2014), e88609.

[13] Tyler R Bonnell et al. "An agent-based model of red colobus resources and disease dynamics implicates key resource sites as hot spots of disease transmission". In: *Ecological Modelling* 221.20 (2010), pp. 2491–2500.

[14] Frédéric Bordes et al. "Home range and parasite diversity in mammals". In: *The American Naturalist* 173.4 (2009), pp. 467–474.

[15] Luca Börger et al. "An integrated approach to identify spatiotemporal and individual-level determinants of animal home range size". In: *The American Naturalist* 168.4 (2006), pp. 471–485.

[16] François Bousquet and Christophe Le Page. "Multi-agent simulations and ecosystem management: a review". In: *Ecological modelling* 176.3 (2004), pp. 313–332.

[17] Mark S Boyce and Lyman L McDonald. "Relating populations to habitats using resource selection functions". In: *Trends in Ecology & Evolution* 14.7 (1999), pp. 268–272.

[18] Mark S Boyce et al. "Scale and heterogeneity in habitat selection by elk in Yellowstone National Park". In: *Ecoscience* 10.4 (2003), pp. 421–431.

[19] Norman E Breslow, Nicholas E Day, and Walter Davis. *Statistical methods in cancer research*. Vol. 1. International agency for research on cancer Lyon, 1980.

[20] Ryan K Brook and Stéphane M McLachlan. "Transdisciplinary habitat models for elk and cattle as a proxy for bovine tuberculosis transmission risk". In: *Preventive Veterinary Medicine* 91.2 (2009), pp. 197–208.

[21] Ella Browning et al. "Predicting animal behaviour using deep learning: GPS data alone accurately predict diving in seabirds". In: *Methods in Ecology and Evolution* 9.3 (2018), pp. 681–692.

[22] William Henry Burt. "Territoriality and home range concepts as applied to mammals". In: *Journal of mammalogy* 24.3 (1943), pp. 346–352.

[23] C. Calenge. "The package adehabitat for the R software: tool for the analysis of space and habitat use by animals". In: *Ecological Modelling* 197 (2006), p. 1035.

[24] Colin J Carlson et al. "Spores and soil from six sides: interdisciplinarity and the environmental biology of anthrax (Bacillus anthracis)". In: *bioRxiv* (2017), p. 165548.

[25] Fiona M Caryl, Christopher P Quine, and Kirsty J Park. "Martens in the matrix: the importance of nonforested habitats for forest carnivores in fragmented landscapes". In: *Journal of Mammalogy* 93.2 (2012), pp. 464–474.

[26] Gullal Singh Cheema and Saket Anand. "Automatic Detection and Recognition of Individuals in Patterned Species". In: *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer. 2017, pp. 27–38.

[27] Carrie A Cizauskas et al. "Frequent and seasonally variable sublethal anthrax infections are accompanied by short-lived immunity in an endemic system". In: *Journal of Animal Ecology* 83.5 (2014), pp. 1078–1090.

[28] Carrie A Cizauskas et al. "Gastrointestinal helminths may affect host susceptibility to anthrax through seasonal immune trade-offs". In: *BMC ecology* 14.1 (2014), p. 27.

[29] Carrie A Cizauskas et al. "Seasonal patterns of hormones, macroparasites, and microparasites in wild African ungulates: the interplay among stress, reproduction, and disease". In: *PloS one* 10.4 (2015), e0120800.

[30] Bradley W Compton, Judith M Rhymer, and Mark McCollough. "Habitat selection by wood turtles (Clemmys insculpta): an application of paired logistic regression". In: *Ecology* 83.3 (2002), pp. 833–843.

[31] Daniel Cornélis et al. "Spatiotemporal dynamics of forage and water resources shape space use of West African savanna buffaloes". In: *Journal of Mammalogy* 92.6 (2011), pp. 1287–1297.

[32] Eric P Crist and Richard C Cicone. "A physically-based transformation of Thematic Mapper data—The TM Tasseled Cap". In: *IEEE Transactions on Geoscience and Remote sensing* 3 (1984), pp. 256–263.

[33] Paul C Cross et al. "Energetic costs of mange in wolves estimated from infrared thermography". In: *Ecology* 97.8 (2016), pp. 1938–1948.

[34] Valerie A Curtis. "Infection-avoidance behaviour in humans and other animals". In: *Trends in Immunology* 35.10 (2014), pp. 457–464.

[35] Benjamin D Dalziel, Juan M Morales, and John M Fryxell. "Fitting probability distributions to animal movement trajectories: using artificial neural networks to link distance, resources, and memory". In: *The American Naturalist* 172.2 (2008), pp. 248–258.

[36] Stephen Davis et al. "Spatial analyses of wildlife contact networks". In: *Journal of the Royal Society Interface* 12.102 (2015), p. 20141004.

[37] Donald L DeAngelis and Wolf M Mooij. "Individual-based modeling of ecological and evolutionary processes". In: *Annu. Rev. Ecol. Evol. Syst.* 36 (2005), pp. 147–168.

[38] Elise Dion, Louis VanSchalkwyk, and Eric F Lambin. "The landscape epidemiology of foot-and-mouth disease in South Africa: A spatially explicit multi-agent simulation". In: *Ecological Modelling* 222.13 (2011), pp. 2059–2072.

[39] Somayeh Dodge et al. "The environmental-data automated track annotation (Env-DATA) system: linking animal tracks with environmental data". In: *Movement Ecology* 1.1 (2013), p. 3.

[40] Eric R Dougherty et al. “A cross-validation-based approach for delimiting reliable home range estimates”. In: *Movement Ecology* 5.1 (2017), p. 19.

[41] Eric R Dougherty et al. “Going through the motions: incorporating movement analyses into disease research”. In: *Ecology Letters* 12.4 (2018), pp. 588–604.

[42] Joni A Downs and Mark W Horner. “Effects of point pattern shape on home-range estimates”. In: *Journal of Wildlife Management* 72.8 (2008), pp. 1813–1818.

[43] Anne-Cécile Dragon et al. “Comparative analysis of methods for inferring successful foraging areas from Argos and GPS tracking data”. In: *Marine Ecology Progress Series* 452 (2012), pp. 253–267.

[44] Thierry Duchesne, Daniel Fortin, and Nicolas Courbin. “Mixed conditional logistic regression for habitat selection studies”. In: *Journal of Animal Ecology* 79.3 (2010), pp. 548–555.

[45] Salome Dürr and Michael P Ward. “Roaming behaviour and home range estimation of domestic dogs in Aboriginal and Torres Strait Islander communities in northern Australia using four different methods”. In: *Preventive veterinary medicine* 117.2 (2014), pp. 340–357.

[46] Hendrik Edelhoff, Johannes Signer, and Niko Balkenhol. “Path segmentation for beginners: an overview of current methods for detecting changes in animal movement patterns”. In: *Movement ecology* 4.1 (2016), p. 21.

[47] Stephen Eubank et al. “Modelling disease outbreaks in realistic urban social networks”. In: *Nature* 429.6988 (2004), p. 180.

[48] Vanessa O Ezenwa. “Selective defecation and selective foraging: antiparasite behavior in wild ungulates?” In: *Ethology* 110.11 (2004), pp. 851–862.

[49] William F Fagan et al. “Spatial memory and animal movement”. In: *Ecology letters* 16.10 (2013), pp. 1316–1329.

[50] Verónica Farías and Todd K Fuller. “Native vegetation structure and persistence of endangered Tehuantepec jackrabbits in a neotropical savanna in Oaxaca, México”. In: *Biodiversity and conservation* 18.7 (2009), pp. 1963–1978.

[51] Matthew L Farnsworth et al. “Linking chronic wasting disease to mule deer movement scales: a hierarchical Bayesian approach”. In: *Ecological Applications* 16.3 (2006), pp. 1026–1036.

[52] John Fieberg and Luca Börger. “Could you please phrase “home range” as a question?” In: *Journal of Mammalogy* 93.4 (2012), pp. 890–902.

[53] Chris H Fleming et al. “Rigorous home range estimation with movement data: a new autocorrelated kernel density estimator”. In: *Ecology* 96.5 (2015), pp. 1182–1188.

[54] Bonnie Fairbanks Flint, Dana M Hawley, and Kathleen A Alexander. “Do not feed the wildlife: associations between garbage use, aggression, and disease in banded mongooses (Mungos mungo)”. In: *Ecology and evolution* 6.16 (2016), pp. 5932–5939.

[55] Daniel Fortin et al. “Wolves influence elk movements: behavior shapes a trophic cascade in Yellowstone National Park”. In: *Ecology* 86.5 (2005), pp. 1320–1330.

[56] Wayne M Getz. “Biomass transformation webs provide a unified approach to consumer–resource modelling”. In: *Ecology letters* 14.2 (2011), pp. 113–124.

[57] Wayne M Getz and John Pickering. “Epidemic models: thresholds and population regulation”. In: *The American Naturalist* 121.6 (1983), pp. 892–898.

[58] Wayne M Getz and David Saltz. “A framework for generating and analyzing movement paths on ecological landscapes”. In: *Proceedings of the National Academy of Sciences* 105.49 (2008), pp. 19066–19071.

[59] Wayne M Getz and Christopher C Wilmers. “A local nearest-neighbor convex-hull construction of home ranges and utilization distributions”. In: *Ecography* 27.4 (2004), pp. 489–505.

[60] Wayne M Getz et al. “LoCoH: nonparameteric kernel methods for constructing home ranges and utilization distributions”. In: *PLoS One* 2.2 (2007), e207.

[61] Wayne M Getz et al. “Panmictic and clonal evolution on a single patchy resource produces polymorphic foraging guilds”. In: *PloS one* 10.8 (2015), e0133732.

[62] Wayne M Getz et al. “Sympatric speciation in structureless environments”. In: *BMC evolutionary biology* 16.1 (2016), p. 50.

[63] Cameron S Gillies et al. “Application of random effects to the study of resource selection by animals”. In: *Journal of Animal Ecology* 75.4 (2006), pp. 887–898.

[64] Volker Grimm et al. “A standard protocol for describing individual-based and agent-based models”. In: *Ecological modelling* 198.1 (2006), pp. 115–126.

[65] Volker Grimm et al. “Pattern-oriented modeling of agent-based complex systems: lessons from ecology”. In: *science* 310.5750 (2005), pp. 987–991.

[66] Thomas J Habib et al. “Modelling landscape effects on density–contact rate relationships of deer in eastern Alberta: implications for chronic wasting disease”. In: *Ecological Modelling* 222.15 (2011), pp. 2722–2732.

[67] Spencer R Hall et al. “Eating yourself sick: transmission of disease as a function of foraging ecology”. In: *Ecology Letters* 10.3 (2007), pp. 207–218.

[68] S Harris et al. “Home-range analysis using radio-tracking data–a review of problems and techniques particularly as applied to the study of mammals”. In: *Mammal review* 20.2-3 (1990), pp. 97–123.

[69] Zepee Havarua, Wendy C Turner, and John KE Mfune. “Seasonal variation in foraging behaviour of plains zebra (Equus quagga) may alter contact with the anthrax bacterium (Bacillus anthracis)”. In: *Canadian Journal of Zoology* 92.4 (2014), pp. 331–337.

[70] Tomislav Hengl et al. “SoilGrids — 1km global soil information based on automated mapping”. In: *PLoS One* 9.8 (2014), e105992.

[71] Robert J. Hijmans et al. *dismo: Species Distribution Modeling*. R package version 1.1-4. 2017. URL: `https://CRAN.R-project.org/package=dismo`.

[72] Robert J Hijmans et al. “Very high resolution interpolated climate surfaces for global land areas”. In: *International journal of climatology* 25.15 (2005), pp. 1965–1978.

[73] Jarrod C Hodgson et al. “Drones count wildlife more accurately and precisely than humans”. In: *Methods in Ecology and Evolution* (2017).

[74] Jarrod C Hodgson et al. “Precision wildlife monitoring using unmanned aerial vehicles”. In: *Scientific reports* 6 (2016), p. 22574.

[75] Jon S Horne et al. “Analyzing animal movements using Brownian bridges”. In: *Ecology* 88.9 (2007), pp. 2354–2363.

[76] Martin Hugh-Jones and Jason Blackburn. “The ecology of Bacillus anthracis”. In: *Molecular aspects of medicine* 30.6 (2009), pp. 356–367.

[77] Martin Hugh-Jones and Jason Blackburn. “The ecology of Bacillus anthracis”. In: *Molecular aspects of medicine* 30.6 (2009), pp. 356–367.

[78] ME Hugh-Jones and V De Vos. “Anthrax and wildlife”. In: *Revue Scientifique et Technique-Office International des Epizooties* 21.1 (2002), pp. 359–384.

[79] ME Hugh-Jones and V De Vos. “Anthrax and wildlife”. In: *Revue Scientifique et Technique-Office International des Epizooties* 21.1 (2002), pp. 359–384.

[80] Örjan Johansson et al. “Land sharing is essential for snow leopard conservation”. In: *Biological Conservation* 203 (2016), pp. 1–7.

[81] Douglas H Johnson. “The comparison of usage and availability measurements for evaluating resource preference”. In: *Ecology* 61.1 (1980), pp. 65–71.

[82] Jonna Katajisto and Atte Moilanen. “Kernel-based home range method for data with irregular sampling intervals”. In: *Ecological Modelling* 194.4 (2006), pp. 405–413.

[83] Roland Kays et al. “Terrestrial animal tracking as an eye on life and planet”. In: *Science* 348.6240 (2015), aaa2478.

[84] Matt J Keeling and Pejman Rohani. *Modeling Infectious Diseases in Humans and Animals*. Princeton University Press, 2008.

[85] Andrew M Kittle et al. “Landscape-level movement patterns by lions in western Serengeti: comparing the influence of inter-specific competitors, habitat attributes and prey availability”. In: *Movement Ecology* 4.1 (2016), p. 17.

[86] Bart Kranstauber et al. “A dynamic Brownian bridge movement model to estimate utilization distributions for heterogeneous animal movement”. In: *Journal of Animal Ecology* 81.4 (2012), pp. 738–746.

[87] Peter N Laver and Marcella J Kelly. “A critical review of home range studies”. In: *Journal of Wildlife Management* 72.1 (2008), pp. 290–298.

[88] Clinton B Leach, Colleen T Webb, and Paul C Cross. “When environmentally persistent pathogens transform good habitat into ecological traps”. In: *Royal Society open science* 3.3 (2016), p. 160051.

[89] Martin Leclerc et al. “Quantifying consistent individual differences in habitat selection”. In: *Oecologia* 180.3 (2016), pp. 697–705.

[90] MNM van Lieshout and AJ Baddeley. “A nonparametric measure of spatial interaction in point patterns”. In: *Statistica Neerlandica* 50.3 (1996), pp. 344–361.

[91] PM Lindeque and PCB Turnbull. “Ecology and epidemiology of anthrax in the Etosha National Park, Namibia”. In: (1994).

[92] James O Lloyd-Smith et al. “Superspreading and the effect of individual variation on disease emergence”. In: *Nature* 438.7066 (2005), pp. 355–359.

[93] Jed A Long and Trisalyn A Nelson. “A review of quantitative methods for movement data”. In: *International Journal of Geographical Information Science* 27.2 (2013), pp. 292–318.

[94] Maite Louzao et al. “Conserving pelagic habitats: seascape modelling of an oceanic top predator”. In: *Journal of Applied Ecology* 48.1 (2011), pp. 121–132.

[95] Maite Louzao et al. “Coupling instantaneous energy-budget models and behavioural mode analysis to estimate optimal foraging strategy: an example with wandering albatrosses”. In: *Movement Ecology* 2.1 (2014), p. 8.

[96] Andrew J Lyons, Wendy C Turner, and Wayne M Getz. “Home range plus: a space-time characterization of movement over real landscapes”. In: *Movement Ecology* 1.1 (2013), p. 2.

[97] VA Macandza, N Owen-Smith, and JW Cain. “Habitat and resource partitioning between abundant and relatively rare grazing ungulates”. In: *Journal of Zoology* 287.3 (2012), pp. 175–185.

[98] BFL Manly et al. *Resource selection by animals: statistical design and analysis for field studies*. Kluwer, 2002.

[99] BFL Manly et al. *Resource selection by animals: statistical design and analysis for field studies*. Springer Science & Business Media, 2007.

[100] Maija K Marsh et al. “Spatial and temporal heterogeneities in the contact behaviour of rabbits”. In: *Behavioral ecology and sociobiology* 65.2 (2011), pp. 183–195.

[101] Paula Martiskainen et al. “Cow behaviour pattern recognition using a three-dimensional accelerometer and support vector machines”. In: *Applied animal behaviour science* 119.1 (2009), pp. 32–38.

[102] Hamish McCallum et al. “Breaking beta: deconstructing the parasite transmission function”. In: *Phil. Trans. R. Soc. B* 372.1719 (2017), p. 20160084.

[103] Kevin McGarigal et al. “Multi-scale habitat selection modeling: a review and outlook”. In: *Landscape ecology* 31.6 (2016), pp. 1161–1175.

[104] Ann E McKellar et al. “Using mixed hidden Markov models to examine behavioral states in a cooperatively breeding bird”. In: *Behavioral Ecology* 26.1 (2014), pp. 148–157.

[105] Adam J McLane et al. “The role of agent-based models in wildlife ecology and management”. In: *Ecological Modelling* 222.8 (2011), pp. 1544–1556.

[106] JA Merkle, D Fortin, and Juan Manuel Morales. “A memory-based foraging tactic reveals an adaptive mechanism for restricted space use”. In: *Ecology letters* 17.8 (2014), pp. 924–931.

[107] Lauren Ancel Meyers, MEJ Newman, and Babak Pourbohloul. “Predicting epidemics on directed contact networks”. In: *Journal of theoretical biology* 240.3 (2006), pp. 400–418.

[108] Lauren Ancel Meyers et al. “Network theory and SARS: predicting outbreak diversity”. In: *Journal of theoretical biology* 232.1 (2005), pp. 71–81.

[109] Théo Michelot et al. “Estimation and simulation of foraging trips in land-based marine predators”. In: *Ecology* (2016).

[110] Marcos Moleón et al. “Carnivore carcasses are avoided by carnivores”. In: *Journal of Animal Ecology* (2017).

[111] Lillian R Morris et al. “Elk resource selection and implications for anthrax management in Montana”. In: *Journal of Wildlife Management* 80.2 (2016), pp. 235–244.

[112] Aristides Moustakas and Matthew R Evans. “A big-data spatial, temporal and network analysis of bovine tuberculosis between wildlife (badgers) and cattle”. In: *Stochastic Environmental Research and Risk Assessment* 31.2 (2017), pp. 315–328.

[113] Thomas Mueller and William F Fagan. “Search and navigation in dynamic environments–from individual behaviors to population distributions”. In: *Oikos* 117.5 (2008), pp. 654–664.

[114] Maureen Murray et al. “Poor health is associated with use of anthropogenic resources in an urban carnivore”. In: *Proceedings of the Royal Society of London B: Biological Sciences* 282.1806 (2015).

[115] Ran Nathan et al. “A movement ecology paradigm for unifying organismal movement research”. In: *Proceedings of the National Academy of Sciences* 105.49 (2008), pp. 19052–19059.

[116] Ran Nathan et al. “Using tri-axial acceleration data to identify behavioral modes of free-ranging animals: general concepts and tools illustrated for griffon vultures”. In: *Journal of Experimental Biology* 215.6 (2012), pp. 986–996.

[117] Erlend B Nilsen, Simen Pedersen, and John DC Linnell. "Can minimum convex polygon home ranges be used to draw biologically meaningful conclusions?" In: *Ecological Research* 23.3 (2008), pp. 635–639.

[118] Joseph M Northrup et al. "Practical guidance on characterizing availability in resource selection functions under a use–availability design". In: *Ecology* 94.7 (2013), pp. 1456–1463.

[119] Charles L Nunn, Peter H Thrall, and Peter M Kappeler. "Shared resources and disease dynamics in spatially structured populations". In: *Ecological Modelling* 272 (2014), pp. 198–207.

[120] N Owen-Smith and V Goodall. "Coping with savanna seasonality: comparative daily activity patterns of African ungulates as revealed by GPS telemetry". In: *Journal of Zoology* 293.3 (2014), pp. 181–191.

[121] Michele B Parsons et al. "Global positioning system data-loggers: a tool to quantify fine-scale movement of domestic animals to evaluate potential for zoonotic transmission to an endangered wildlife population". In: *PloS one* 9.11 (2014), e110984.

[122] Toby A Patterson et al. "Classifying movement behaviour in relation to environmental conditions using hidden Markov models". In: *Journal of Animal Ecology* 78.6 (2009), pp. 1113–1123.

[123] Toby A Patterson et al. "State–space models of individual animal movement". In: *Trends in ecology & evolution* 23.2 (2008), pp. 87–94.

[124] Toby A Patterson et al. "Using GPS data to evaluate the accuracy of state–space methods for correction of Argos satellite telemetry error". In: *Ecology* 91.1 (2010), pp. 273–285.

[125] Sara H Paull et al. "From superspreaders to disease hotspots: linking transmission across hosts and space". In: *Frontiers in Ecology and the Environment* 10.2 (2012), pp. 75–82.

[126] Liliana Perez and Suzana Dragicevic. "An agent-based approach for modeling dynamics of contagious disease spread". In: *International journal of health geographics* 8.1 (2009), p. 50.

[127] Sarah E Perkins et al. "Comparison of social networks derived from ecological data: implications for inferring infectious disease dynamics". In: *Journal of Animal Ecology* 78.5 (2009), pp. 1015–1022.

[128] T Alex Perkins et al. "Heterogeneity, mixing, and the spatial scales of mosquito-borne pathogen transmission". In: *PLoS Computational Biology* 9.12 (2013), e1003327.

[129] Steven J Phillips and Miroslav Dudík. "Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation". In: *Ecography* 31.2 (2008), pp. 161–175.

[130] Clémence Poirotte et al. "Mandrills use olfaction to socially avoid parasitized conspecifics". In: *Science advances* 3.4 (2017), e1601721.

[131] Kelly M Proffitt et al. "Elk distribution and spatial overlap with livestock during the brucellosis transmission risk period". In: *Journal of Applied Ecology* 48.2 (2011), pp. 471–478.

[132] Mathieu Pruvot et al. "Pathogens at the livestock-wildlife interface in Western Alberta: does transmission route matter?" In: *Veterinary research* 45.1 (2014), p. 18.

[133] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria, 2017. URL: https://www.R-project.org/.

[134] Dave Ramsey et al. "The effects of reducing population density on contact rates between brushtail possums: implications for transmission of bovine tuberculosis". In: *Journal of Applied Ecology* 39.5 (2002), pp. 806–818.

[135] Carrie Lynn Roever et al. "The pitfalls of ignoring behaviour when quantifying habitat selection". In: *Diversity and Distributions* 20.3 (2014), pp. 322–333.

[136] SM Rooney, A Wolfe, and TJ Hayden. "Autocorrelated data in telemetry studies: time to independence and the problem of behavioural effects". In: *Mammal Review* 28.2 (1998), pp. 89–98.

[137] Elke Saile and Theresa M Koehler. "Bacillus anthracis multiplication, persistence, and genetic exchange in the rhizosphere of grass plants". In: *Applied and environmental microbiology* 72.5 (2006), pp. 3168–3174.

[138] Anna K Schweiger et al. "Foraging ecology of three sympatric ungulate species–Behavioural and resource maps indicate differences between chamois, ibex and red deer". In: *Movement ecology* 3.1 (2015), p. 6.

[139] Dana P Seidel and Mark S Boyce. "Patch-use dynamics by a large herbivore". In: *Movement Ecology* 3.1 (2015), p. 7.

[140] Dana P Seidel and Mark S Boyce. "Varied tastes: home range implications of foraging-patch selection". In: *Oikos* 125.1 (2016), pp. 39–49.

[141] David L Smith et al. "Predicting the spatial dynamics of rabies epidemics on heterogeneous landscapes". In: *Proceedings of the National Academy of Sciences* 99.6 (2002), pp. 3668–3672.

[142] Sandra E Smith-Aguilar, Gabriel Ramos-Fernández, and Wayne M Getz. "Seasonal changes in socio-spatial structure in a group of free-living spider monkeys (Ateles geoffroyi)". In: *PloS one* 11.6 (2016), e0157228.

[143] Orr Spiegel et al. "Moving beyond curve fitting: using complementary data to assess alternative explanations for long movements of three vulture species". In: *The American Naturalist* 185.2 (2015), E44–E54.

[144] Orr Spiegel et al. “What’s your move? Movement as a link between personality and spatial dynamics in animal populations”. In: *Ecology Letters* 20.1 (2017), pp. 3–18.

[145] John R Squires et al. “Combining resource selection and movement behavior to predict corridors for Canada lynx at their southern range periphery”. In: *Biological Conservation* 157 (2013), pp. 187–195.

[146] Pieter Johan Steenkamp, Henriette van Heerden, and Ockert Louis van Schalkwyk. “Ecological suitability modeling for anthrax in the Kruger National Park, South Africa”. In: *PloS one* 13.1 (2018), e0191704.

[147] Ariana Strandburg-Peshkin et al. “Shared decision-making drives collective movement in wild baboons”. In: *Science* 348.6241 (2015), pp. 1358–1361.

[148] Stanley M Tomkiewicz et al. “Global positioning system and associated technologies in animal behaviour and ecological research”. In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 365.1550 (2010), pp. 2163–2176.

[149] Jeff A Tracey et al. “An agent-based movement model to assess the impact of landscape fragmentation on disease transmission”. In: *Ecosphere* 5.9 (2014), pp. 1–24.

[150] PCB Turnbull. “Introduction: anthrax history, disease and ecology”. In: *Anthrax*. Springer, 2002, pp. 1–19.

[151] Wendy C Turner et al. “Fatal attraction: vegetation responses to nutrient inputs attract herbivores to infectious anthrax carcass sites”. In: *Proceedings of the Royal Society of London B: Biological Sciences* 281.1795 (2014), p. 20141785.

[152] Wendy C Turner et al. “Lethal exposure: an integrated approach to pathogen transmission via environmental reservoirs”. In: *Scientific reports* 6 (2016), p. 27311.

[153] Wendy C Turner et al. “Soil ingestion, nutrition and the seasonality of anthrax in herbivores of Etosha National Park”. In: *Ecosphere* 4.1 (2013), pp. 1–19.

[154] John Joseph Valletta et al. “Applications of machine learning in animal behaviour studies”. In: *Animal Behaviour* 124 (2017), pp. 203–220.

[155] Bram Van Moorter et al. “Identifying movement states from location data using cluster analysis”. In: *Journal of Wildlife Management* 74.3 (2010), pp. 588–594.

[156] Bram Van Moorter et al. “Memory keeps you at home: a mechanistic model for home range emergence”. In: *Oikos* 118.5 (2009), pp. 641–652.

[157] Alexander Gomez Villa, Augusto Salazar, and Francisco Vargas. “Towards automatic wild animal monitoring: Identification of animal species in camera-trap images using very deep convolutional neural networks”. In: *Ecological Informatics* 41 (2017), pp. 24–32.

[158] Mapesa Moses Wafula, Atimnedi Patrick, and Tumwesigye Charles. “Managing the 2004/05 anthrax outbreak in Queen Elizabeth and Lake Mburo National Parks, Uganda”. In: *African Journal of Ecology* 46.1 (2008), pp. 24–31.

[159] Susan F Walker et al. “Factors driving pathogenicity vs. prevalence of amphibian panzootic chytridiomycosis in Iberia”. In: *Ecology letters* 13.3 (2010), pp. 372–382.

[160] Magnus Wang and Volker Grimm. “Home range dynamics and population regulation: an individual-based model of the common shrew Sorex araneus”. In: *Ecological Modelling* 205.3 (2007), pp. 397–409.

[161] Nicola Weber et al. “Badger social networks correlate with tuberculosis infection”. In: *Current Biology* 23.20 (2013), R915–R916.

[162] Mia Wege et al. “Cross-seasonal foraging site fidelity of subantarctic fur seals: implications for marine conservation areas”. In: *Marine Ecology Progress Series* 554 (2016), pp. 225–239.

[163] Sara B Weinstein, Julia C Buck, and Hillary S Young. “A landscape of disgust”. In: *Science* 359.6381 (2018), pp. 1213–1214.

[164] David M Williams, Amy C Dechen Quinn, and William F Porter. “Informing disease models with temporal and spatial contact structure among GPS-collared individuals in wild populations”. In: *PloS One* 9.1 (2014), e84368.

[165] Christopher C Wilmers et al. “The golden age of bio-logging: how animal-borne sensors are advancing the frontiers of ecology”. In: *Ecology* 96.7 (2015), pp. 1741–1753.

[166] Ryan R Wilson, Lynne Gilbert-Norton, and Eric M Gese. “Beyond use versus availability: behaviour-explicit resource selection”. In: *Wildlife Biology* 18.4 (2012), pp. 424–430.

[167] Rosie Woodroffe and Christl A Donnelly. “Risk of contact between endangered African wild dogs Lycaon pictus and domestic dogs: opportunities for pathogen transmission”. In: *Journal of applied ecology* 48.6 (2011), pp. 1345–1354.

[168] Rosie Woodroffe et al. “Badgers prefer cattle pasture but avoid cattle: implications for bovine tuberculosis control”. In: *Ecology letters* 19.10 (2016), pp. 1201–1208.

[169] Brian J Worton. “Kernel methods for estimating the utilization distribution in home-range studies”. In: *Ecology* 70.1 (1989), pp. 164–168.

[170] IJ Yockney et al. “Comparison of ranging behaviour in a multi-species complex of free-ranging hosts of bovine tuberculosis in relation to their use as disease sentinels”. In: *Epidemiology & Infection* 141.7 (2013), pp. 1407–1416.

[171] Katherine A Zeller et al. “Sensitivity of landscape resistance estimates based on point selection functions to scale and behavioral state: pumas as a case study”. In: *Landscape Ecology* 29.3 (2014), pp. 541–557.

[172] Katherine A Zeller et al. “Sensitivity of resource selection and connectivity models to landscape definition”. In: *Landscape Ecology* 32.4 (2017), pp. 835–855.

[173] Katherine A Zeller et al. “Using step and path selection functions for estimating resistance to movement: pumas as a case study”. In: *Landscape ecology* 31.6 (2016), pp. 1319–1335.

[174] Royi Zidon et al. “Zebra migration strategies and anthrax in Etosha National Park, Namibia”. In: *Ecosphere* 8.8 (2017).