The Under-appreciated Roles of Spatial Scale, Individual Variation, and Nonlinearity in Spatial Ecology

David Heit

University of New Hampshire

Follow this and additional works at: https://scholars.unh.edu/thesis

Recommended Citation
https://scholars.unh.edu/thesis/1761

This Thesis is brought to you for free and open access by the Student Scholarship at University of New Hampshire Scholars' Repository. It has been accepted for inclusion in Master's Theses and Capstones by an authorized administrator of University of New Hampshire Scholars' Repository. For more information, please contact Scholarly.Communication@unh.edu.
THE UNDER-APPRECIATED ROLES OF SPATIAL SCALE, INDIVIDUAL VARIATION, AND NONLINEARITY IN SPATIAL ECOLOGY

BY

DAVID R. HEIT

B.S., Michigan State University, 2018

THESIS

Submitted to the University of New Hampshire in

Partial Fulfilment of the Requirements for the Degree of

Master of Science

in

Natural Resources:

Wildlife and Conservation Biology

September 2023
This thesis was examined and approved in partial fulfilment of the requirements for the degree of
Master of Science in Wildlife and Conservation Biology by:

Thesis Director, Dr. Remington Moll, Assistant Professor
Natural Resources and the Environment, University of New Hampshire

Dr. Rebecca Rowe, Associate Professor,
Natural Resources and the Environment, University of New Hampshire

Dr. Nathan Furey, Assistant Professor,
Biological Sciences, University of New Hampshire

On July 11, 2023

Approval signatures are on file with the University of New Hampshire Graduate School
# TABLE OF CONTENTS

## LIST OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>TABLE OF CONTENTS</td>
<td>iii</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>v</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>vi</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>ix</td>
</tr>
<tr>
<td>1. CHAPTER 1: The spatial scaling and individuality of habitat selection in a widespread ungulate</td>
<td>1</td>
</tr>
<tr>
<td>1.1 ABSTRACT</td>
<td>1</td>
</tr>
<tr>
<td>1.2 INTRODUCTION</td>
<td>2</td>
</tr>
<tr>
<td>1.3 METHODS</td>
<td>9</td>
</tr>
<tr>
<td>1.3.1 Study area</td>
<td>9</td>
</tr>
<tr>
<td>1.3.2 Deer capture and collaring</td>
<td>10</td>
</tr>
<tr>
<td>1.3.3 Integrated step-selection function and covariates</td>
<td>11</td>
</tr>
<tr>
<td>1.3.4 Modeling scale of effect</td>
<td>12</td>
</tr>
<tr>
<td>1.3.5 Intrinsic and extrinsic factors on scales of effect</td>
<td>15</td>
</tr>
<tr>
<td>1.4 RESULTS</td>
<td>16</td>
</tr>
<tr>
<td>1.5 DISCUSSION</td>
<td>18</td>
</tr>
<tr>
<td>1.6 ACKNOWLEDGEMENTS</td>
<td>24</td>
</tr>
<tr>
<td>2. CHAPTER TWO: Generalized nonlinearity in wildlife ecology: research, review, and recommendations</td>
<td>25</td>
</tr>
<tr>
<td>2.1 ABSTRACT</td>
<td>25</td>
</tr>
<tr>
<td>2.2 INTRODUCTION</td>
<td>27</td>
</tr>
<tr>
<td>2.3 LINEARITY REVIEW</td>
<td>31</td>
</tr>
<tr>
<td>2.3.1 Methods</td>
<td>31</td>
</tr>
<tr>
<td>2.3.2 Results</td>
<td>32</td>
</tr>
<tr>
<td>2.4 CASE STUDY 1: MODELING MAMMAL ABUNDANCE USING POISSON REGRESSION</td>
<td>33</td>
</tr>
<tr>
<td>2.4.1 Background and Study Design</td>
<td>33</td>
</tr>
<tr>
<td>2.4.2 Abundance modeling</td>
<td>35</td>
</tr>
<tr>
<td>2.4.3 Results</td>
<td>37</td>
</tr>
<tr>
<td>2.5 CASE STUDY 2: MODELING PASSERINE OCCUPANCY</td>
<td>38</td>
</tr>
<tr>
<td>2.5.1 Background and Study Design</td>
<td>38</td>
</tr>
</tbody>
</table>
2.5.2 The Box Tidwell test .................................................................39
2.5.3 Occupancy Modeling ...............................................................40
2.5.4 Results ..................................................................................41
2.6 DISCUSSION ............................................................................42
2.7 ACKNOWLEDGEMENTS ............................................................46
LIST OF REFERENCES: ....................................................................48
LIST OF TABLES

Table 1. Covariates used in analysis of habitat selection of white-tailed deer (*Odocoileus virginianus*) in Missouri, USA from 2015 to 2019. Predicted relationships with scale of effect (SOE) are listed where positive indicates that scale of effect will increase with the covariate. References to home ranges in the table are abbreviated to ‘HR’. ................................................................. 7

Table 2. Comparisons between generalized linear model (GLM) and generalize additive model (GAM) specifications of coyotes (*Canis latrans*), and red foxes (*Vulpes vulpes*) from a study of the usage and applications of nonlinear methods in wildlife ecology. Data for the models was obtained from camera trap arrays in New Hampshire, USA. Independent variables for both model types are: Road density within 1 km, distance to nearest water source, and distance to nearest agricultural area. Significant p-values (<0.05) are shown in bold. The Akaike’s Information Criteria (AIC) is included for both the GLMs and GAMs. GAM models also were based upon RoadDens, DistW, and DistA, though their coefficients and their p-values are not shown due to space considerations. ................................................................................................. 36

Table 3. Passerine bird species that failed a Box-Tidwell test for linearity and their respective number of occurrences and variables for which the test failed. Results are part of a study examining the dynamics of nonlinearity in wildlife modeling using two case studies from Southeastern New Hampshire and the prevalence of nonlinear techniques among wildlife studies using a review of literature. ........................................................................................................... 40
LIST OF FIGURES

Figure 1. Three examples of how measured values of habitat variables can change with spatial scale due to aggregation issues associated with the modifiable areal unit problem. Scales are represented by black squares. ................................................................. 4

Figure 2. Spatial bounds of GPS-collared white-tailed deer (*Odocoileus virginianus*) in three study regions in Northern, Southern, and Southeastern Missouri, USA. Boundary polygons are overlaid on maps of roads (all types, black lines) and forest cover (NLCD classes 42, 43, 44, green areas). ................................................................. 10

Figure 3. Plots of selection of forest cover, calculated at 65 different spatial scales (blue circles) for six individual white-tailed deer in the North study area of our analysis (*Odocoileus virginianus*). Parameter coefficients of zero are indicated by orange dashed lines, and the spatial scale that resulted in the lowest AIC (scale of effect) is indicated for each individual by the large orange circles. Filled circles denote scales that resulted in significant p-values, while open circles denote insignificant p-values. Units on the x-axis represent the side lengths of buffer squares (i.e. square in which mean of forest cover values was taken). Y-axis units represent parameter coefficients ranging from negative one to one. Deer sex, age and study area are indicated in each panel...................................................................................................................17

Figure 4. Linear model coefficients for scales of effect (spatial scales that resulted in lowest AIC in integrated step selection models) modeled as a function of intrinsic (relating to the internal state of animals) and extrinsic factors (relating to exterior environment). Variables with significant p-values are marked with an asterisk.......................................................... 18
**Figure 5.** Conceptual representations of nonlinear relationship in linear regression across three link functions: identity, logit, and logarithmic. Relationships shown are based on simulated relationships, where the underlying data have a nonlinear relationship between the dependent and independent variables. Raw data points are shown as purple circles, and linear and nonlinear relationships as blue and magenta lines respectively. Raw data for GLMs are the same on both the natural and link scale (binary for logistic regression, counts for Poisson regression) as link functions relate dependent variables to independent variables without directly transforming them.

........................................................................................................................................................................ 28

**Figure 6.** Results of a literature review of studies published between 2018 and 2022 in 5 leading ecological journals showing the prevalence of various methods to account for nonlinearity in linear models. The pie chart shows the proportion of studies (out of a total of 162) that included transformations of independent variables, multi-coefficient models (abbreviated “multi-coef,” in figure; these included models with more than one coefficient for an independent variable), both, or neither. The bar chart shows how often a study also reported a test for linearity within the proportions from the pie chart so that proportions sum to 1.0. ................................................................. 32

**Figure 7.** Study area for two case studies in Southeastern New Hampshire, USA in a study examining the dynamics of nonlinearity in wildlife modeling. The first case study (blue and orange circles) used data from an array of 109 camera traps deployed in summer of 2021. The second case study (purple and orange circles) used point-count survey data from passerines on lands managed by the University of New Hampshire in the summers of 2021 and 2022.

........................................................................................................................................................................ 34
Figure 8. Generalized linear models (GLMs) and generalized additive models (GAMs) of relative abundance fit to camera trap data of coyotes (*Canis latrans*), and red foxes (*Vulpes vulpes*) collected in the summer of 2021 in New Hampshire, USA. The y-axis represents the relative abundance of a species (days with at least one detection) as a function of an independent variable, distance to agriculture, or road density. In the right-hand panel, coyote GLMs and GAMs are denoted with a solid line, and red fox GLMs and GAMs with a dashed line.

Figure 9. Linear and nonlinear specifications of occupancy models for passerine bird species from point count data collected in the summers of 2021 and 2022 in New Hampshire, USA. Included in the plot are those that failed a Box-Tidwell test for linearity. Confidence intervals were omitted from plots for clarity. Y-axis units on the natural scale are the probability of occupancy (occ. prob.), and log of the odds ratio for occupancy on the link scale (logit(occ. prob.)).
ABSTRACT

Spatial ecology is a central component of ecological inquiry. In an ever-changing world facing threats including climate change, human sprawl, and novel zoonotic diseases, understanding how animals use space and make habitat decisions can be invaluable to research, management, and conservation of animal species. The techniques and technology used in spatial ecology have advanced continuously over time to provide increasingly detailed data and analyses and mitigate potential statistical biases. However, there remain many potential biases that warrant further attention in the discipline of spatial ecology. This is especially true for conceptual biases, or biases inherent to the ways in which data are viewed, processed, and analyzed, and whose effects are often not conspicuous during the modeling process. In this thesis, I explored three such biases that are seemingly underappreciated within the spatial ecology discipline—spatial scale, individual variation, and nonlinearity. In Chapter One, I used a dataset of white-tailed deer (*Odocoileus virginianus*) to examine the impacts of spatial scale and individual variation on models of habitat selection, and how these biases may work synergistically with each other. In Chapter Two, I reviewed contemporary literature to investigate how frequently ecologists were exploring and addressing nonlinearity in their research. I then illustrated the impacts of nonlinearity with two case studies using occupancy models of passerine birds and abundance models of mesocarnivores. Here, my coauthors and I found that the biases of spatial scale, individual variation, and nonlinearity are not only influential to the numerical outputs of spatial modeling, but when unaddressed can lead to differing and potentially inaccurate conclusions. The techniques we used in this thesis to address these biases also allow for new insights into animal ecology to be drawn that were obscured by the more conventional approaches.
CHAPTER 1: The spatial scaling and individuality of habitat selection in a widespread ungulate

1.1 ABSTRACT

Animal-habitat relationships tend to manifest at specific spatial scales. Accurately identifying these scales and accounting for the variance in habitat selection across them is crucial for linking habitat selection patterns to the ecological processes giving rise to them. Although this fundamental issue has long been recognized, it has been seldom addressed empirically in habitat selection studies. In this study, we investigated how spatial scale influences the outputs of habitat selection analyses. Furthermore, we examined whether the effect of spatial scale varies among individual animals and whether these effects could be predicted via intrinsic or extrinsic factors. We used a dataset collected from 485 GPS-collared white-tailed deer (Odocoileus virginianus) across three study sites in Missouri, USA to model habitat selection at 65 spatial scales from 900 m\(^2\) to 15 km\(^2\) using integrated step selection functions. To investigate potential drivers of spatial scaling we used multiple linear regression to model how scale of effect, defined as the spatial scale at which model AIC was minimized, could be predicted by intrinsic (age, sex, and home range size) and extrinsic factors (study site, season, mean percentage forest in home range, mean distance to nearest road in home range). Scale of effect varied substantially among individuals, and individual variation in scale of effect was predicted by home range size, study site, and proportion of forest within a home range. In contrast, other intrinsic and extrinsic factors had little to no relationship with scale of effect. Parameter coefficients for forest cover and distance to nearest road varied strongly with opposing directionality of responses across spatial scales, revealing that spatial scale may bias habitat selection analyses. Coefficients were both positive and negative at different scales for an average of 63.2\% individuals, and no single
spatial scale resulted in the scale of effect more than 9.0% of the time. Our study demonstrates that spatial scale can strongly influence model parameter coefficients, thereby raising questions about the conventional interpretation of habitat selection analyses. We discuss outstanding issues regarding the comparability of results across study sites and the future of multi-scale habitat selection analyses.

1.2 INTRODUCTION

Spatial scale is fundamental to animal ecology research (Levin, 1992; Turner, 1989; Wiens, 1989). Many ecological processes are highly scale-dependent, meaning they manifest most strongly at a particular scale (Bastille-Rousseau et al., 2018; Jackson & Fahrig, 2012). Accordingly, as the scale evaluated in a given study changes, so too can the inferences about ecological relationships (Holland et al. 2004, Jackson and Faring 2012, Moll et al., 2020). This variation can lead researchers to reach one conclusion at one scale and a different or even opposing conclusion at another scale (Gallo et al., 2018; Holland et al., 2004; Laforge et al., 2016; Orians & Wittenberger, 1991). This issue of scale-dependence is relevant across numerous ecological processes ranging from population dynamics to the anthropogenic effects on animal activity and space use (Benhamou, 2014; Mayor et al., 2009; Moll et al., 2020). Nonetheless, studies of animal ecology are often conducted at only one or a few scales, whether expressly chosen by the researcher, or implicitly determined by the scale of the associated data (Jackson & Fahrig, 2015; Wheatley & Johnson, 2009). Consequently, empirical tests of the ways ecological processes manifest at varying scales remain relatively scarce (Miguet et al., 2016).
Although spatial scale is relevant to many ecological processes, it can be especially influential in studies of animal habitat selection (Laforge et al., 2016; McGarigal et al., 2016). Here, we use the term *spatial scale* to indicate the extent at which a particular habitat variable is aggregated and quantified for the purposes of analyzing third-order habitat selection (Johnson, 1980), and the term *scale of effect* to indicate the spatial scale at which the effect of a habitat variable upon a biological response is most supported by the data. Habitat or resource selection analyses are typically used to infer selection behavior using observed animal locations and locations designated as “available” to animals (Avgar et al., 2016; Boyce et al., 2002).

The differences between spatial scale and scale of effect are crucial to consider because variation in habitat features is typically aggregated to one value, usually the mean of a value within a mapping unit (e.g., pixels, grains, or raster cells) containing an animal’s location. Without explicitly aggregating values across a different scale, the resolution of these mapping units acts as the assumed spatial scale of effect (Fig. 1). However, an animal might select for a location not only because of the resources in the immediate vicinity but also because of landscape features surrounding the location at broader spatial scales. For example, at a small spatial scale, an animal that spends considerable time in a grassland may appear to select against forest cover, but at a larger scale it is revealed that they favor grasslands positioned within a matrix of forest patches (Dunning et al., 1992). Moreover, it is now well-recognized that many animals have spatial memory and working “mental maps” of landscapes (Fagan et al., 2013; Jesmer et al., 2018). Thus, selection of a particular patch could be influenced by its relative importance within an individual’s contextual understanding of the broader landscape.
Figure 1. Three examples of how measured values of habitat variables can change with spatial scale due to aggregation issues associated with the modifiable areal unit problem. Scales are represented by black squares.

The issue of data aggregation across spatial scales constitutes a variation of the modifiable areal unit problem, or the phenomenon where spatial data aggregation can alter the outcomes of analyses (Gehlke & Biehl, 1934; Jelinski & Wu, 1996). Accordingly, mismatches between the scale of effect and the spatial scale used in analyses can result in both Type I and Type II statistical errors and bias inference (Holland et al., 2004; Mayor et al., 2009; Montgomery et al., 2011). The scale of effect for a habitat variable is not usually known to researchers a priori, thus it often has to be identified empirically (Miguet et al., 2016; Wheatley & Johnson, 2009). To accomplish this, analyses can be performed at multiple spatial scales to identify the scale of effect (Jackson & Fahrig, 2015; Wheatley & Johnson, 2009). Multi-scale approaches are an improvement upon single scale studies, but such efforts often only examine a small number of scales (Wheatley & Johnson, 2009). For example, a recent review found that the most influential scale (i.e., the scale of effect) in multi-scale analyses was often the smallest or the largest among
those considered (Jackson & Fahrig, 2015), indicating that the true scale of effect was likely outside of those examined and therefore unidentified (Miguet et al., 2016).

An additional and often overlooked issue related to the scaling of animal-habitat relationships is how scale of effect is influenced by intrinsic and extrinsic factors that vary across individual animals (Montgomery et al., 2018). Intrinsic factors are those that characterize the internal state and characteristics of individual animals (e.g. age, body size, diet, Nathan et al. 2008). Such intrinsic factors have each been shown to influence home range size and patterns of habitat selection (Mysterud et al., 2001). Extrinsic factors are those that characterize the landscapes in which animals reside (e.g. phenology, topography, or climate) and can influence animal-habitat relationships via the distribution and configuration of essential resources (Farhadinia et al., 2019; Hiller et al., 2009; MacNearney et al., 2016). These factors are influential because animals rely on their perception of the spatial configuration of habitat resources and locations of specific attributes to acquire resources, mitigate predation risk, and maintain home ranges (Fagan et al., 2013; Van Moorter et al., 2009). As individuals gain experience, information can be stored for future use (i.e., memory). Fagan et al. (2013) describes two types of memory: memory of general habitat composition (spatial memory), and memory of specific habitat features (attribute memory). Both types of memory can be influential for scaling of habitat selection. For instance, selection for forest cover may depend on spatial memory of general forest structure and composition (Merkle et al., 2014), while selection for anthropogenic features like roads would depend on the attribute memory for where specific roads or dwellings are located (Parsons et al., 2020; Zeller et al., 2021). We use the term spatial memory hereafter to refer to both types of memory. Without spatial memory, individuals can only respond to distant landscape features if they possess non-visual sensory ability, evolutionarily embedded behaviors,
or can emulate conspecifics (Jesmer et al., 2018; Nathan et al., 2008). As an individual’s spatial memory increases, so too could their ability to respond to local landscape features in the context of larger spatial scales. Both intrinsic and extrinsic factors could influence scales of effect across individuals by informing their spatial memory of the surrounding landscape, but such factors have not been examined in relation to third order habitat selection.

Here we investigate the spatial scaling of habitat selection in a widespread ungulate, the white-tailed deer (*Odocoileus virginianus*). White-tailed deer (hereafter, deer) are the most abundant and wide-ranging ungulate in North America (Hewitt, 2011) and thereby serve as a useful model species for evaluating the effects of spatial scale on habitat selection. Deer inhabit a variety of landscapes including boreal forest, desert scrubland, cropland, and cities (DeNicola et al., 2000; Mandujano-Rodriguez & Hernandez, 2019; Routh & Nielsen, 2021; C. A. Wright et al., 2019). Their ability to occupy a wide range of landscapes and levels of anthropogenic development leads to high degrees of individuality in resource selection (Haus et al., 2020). Selection behavior of individual deer also varies by sex, age, and season because dietary and shelter requirements are not constant across life history stages or the course of the year (Kie & Bowyer, 1999; Lesage et al., 2000; Michel et al., 2020). Furthermore, improving the spatial modeling of deer habitat selection is a management priority because they have high economic value to state and provincial game agencies, and face threats from numerous diseases and animal-borne pathogens (Allen et al., 2019; Belsare et al., 2021; Southwick, 2009). Prior studies have shown the effects of spatial scale on deer habitat selection (Laforge et al., 2015), but the role of intrinsic and extrinsic factors on scaling relationships remains underexplored.

We hypothesized that deer habitat selection patterns would change significantly as a function of spatial scale, and that varying scales of effect (SOE) could be predicted by intrinsic
or extrinsic factors (Table 1). We specifically tested six hypotheses, three intrinsic (I1-I3) and three extrinsic (E1-E3), as articulated here and summarized in Table 1 with supporting references. We tested these hypotheses in the context of deer ecology, but similar mechanisms and predictions are likely relevant to many other species.

**I1.** SOE will increase with age because older individuals will possess greater spatial memory and a broader “mental map” of resource locations.

**I2.** SOE will be larger for males than females because of males’ more extensive movements and broader space use.

**I3.** SOE will increase with home range size because individuals with larger home ranges use resources at broader scales.

**E1.** SOE will vary by study areas because of landscape-level variation in landscape configuration.

**E2.** SOE for forest will increase with decreasing forest cover because individuals must traverse broader areas to maintain access to cover.

**E3.** SOE for roads will decrease with decreasing distance to roads because individuals living near roads will select against them at fine scales to avoid road mortality.
Table 1. Covariates used in analysis of habitat selection of white-tailed deer (*Odocoileus virginianus*) in Missouri, USA from 2015 to 2019. Predicted relationships with scale of effect (SOE) are listed where positive indicates that scale of effect will increase with the covariate. References to home ranges in the table are abbreviated to ‘HR’.

### Intrinsic Covariate

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Source</th>
<th>Justification</th>
<th>Prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>Tooth wear (Severinghaus, 1949)</td>
<td>(Foley et al., 2015; E. S. Long et al., 2008; McGovern et al., 2020; C. A. Wright et al., 2019)</td>
<td>All SOE increase with age</td>
</tr>
<tr>
<td>Sex</td>
<td>Visual identification</td>
<td>(Nixon et al., 2007; C. A. Wright et al., 2019)</td>
<td>Males have larger SOE</td>
</tr>
<tr>
<td>HR size</td>
<td>GPS collar data</td>
<td>(Hasapes &amp; Comer, 2017; Hygnstrom &amp; Vercauteren, 2000)</td>
<td>All SOE increase with home range size</td>
</tr>
</tbody>
</table>

### Extrinsic Covariate

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Source</th>
<th>Justification</th>
<th>Prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study site</td>
<td>GPS collar data</td>
<td>(Duparc et al., 2019; Moll et al., 2021)</td>
<td>Larger SOE in North</td>
</tr>
<tr>
<td>Mean percentage forest cover in HR</td>
<td>NLCD cover types 41 (deciduous), 42 (evergreen), 43 (mixed)</td>
<td>(Grovenburg et al., 2011; E. S. Long et al., 2005)</td>
<td>Negative with forest SOE</td>
</tr>
<tr>
<td>Mean distance to nearest road in HR</td>
<td>TIGER county road shapefiles (U.S. Census Bureau, 2019).</td>
<td>(Jacobs et al., 2021; Tucker et al., 2018)</td>
<td>Positive with distance to road SOE</td>
</tr>
<tr>
<td>Season</td>
<td>GPS collar data</td>
<td>(Beier &amp; McCullough, 1990; Hygnstrom et al., 2011; E. S. Long et al., 2013)</td>
<td>Larger forest SOE in winter</td>
</tr>
</tbody>
</table>

To test these hypotheses, we constructed integrated step selection functions (iSSFs) at the individual level for 485 white-tailed deer across a wide range of spatial scales in three distinct regions in Missouri, USA. Our results reveal the influential nature of spatial scale on habitat selection analyses and shed light upon whether intrinsic and extrinsic factors can explain or
predict patterns of spatial scaling. These findings connect emerging theory regarding spatial scaling to quantitative analysis of animal-habitat relationships and help inform the management and conservation of wildlife and the resources that support them.

1.3 METHODS

1.3.1 Study area

We conducted our study across three study regions in northwest, southern, and southeastern Missouri, USA (hereafter North, South, and Southeast sites, Fig. 2). The North site comprises 5,779 km$^2$ characterized by fragmented forest patches in a largely agricultural landscape with undulating topography and a predominantly uniform gridded road system. Land cover in the North site is 2.5% impervious surface, 44.5% cultivated crops, 12.3% forest, 37.7% grassland, and 3.0% other land types (C. A. Wright et al., 2019). The South site comprises 9,332 km$^2$ and has more contiguous forest cover and less regular road patterns than the North, as roads tend to track with topography associated with riverine systems. Land cover in the South site is 2.1% impervious surface, 1.7% cultivated crops, 52.6% forest, 43.0% grassland, and 0.6% other land types. The Southeast site comprises 7,285 km$^2$ of a transitionary area between the Ozark Mountains and more fertile plains to the east. As such it has patterns of forest cover and roads that are intermediate between the North and South. Habitat cover in the Southeast site is 5.0% impervious surface, 29.1% cultivated crops, 42.9% forest, 18.1% grassland, 4.8% other. The climate of Missouri is characterized by warm, humid summers with a mean maximum July temp of 31°C, mean temperature of 24°C, and long, dry winters with mean minimum January temperature of -8°C and mean annual snowfall of 40 cm (Decker, 2021).
Figure 2. Spatial bounds of GPS-collared white-tailed deer (*Odocoileus virginianus*) in three study regions in Northern, Southern, and Southeastern Missouri, USA. Boundary polygons are overlaid on maps of roads (all types, black lines) and forest cover (NLCD classes 42, 43, 44, green areas).

1.3.2 Deer capture and collaring

We captured 664 deer (2015 to 2019, January to March for North and South deer, June to September for Southeast deer) using modified clover traps (VerCauteren et al. 1999) and rocket nets (Haulton et al. 2001), and chemically immobilized individuals using established methods (Kreeger 1996, Kilpatrick and Spohr 1999, see Wright et al. 2019 for details). Deer capture and chemical immobilization procedures were approved by the Institutional Animal Care and Use Committee at the University of Missouri (protocol number 8216). We estimated age by tooth wear and replacement (Severinghaus 1949), and divided deer into three age classes at the time of capture: fawns (6 months), yearling (1.5 yr), and adult (> 1.5 yr). We fit each captured deer with an 825-g Iridium global positioning system (GPS) radio-collar (model G2110E; Advanced Telemetry Systems, Isanti, Minnesota, USA) that recorded a location every five hours thereafter. Locational error for the GPS collars was estimated to be 5.92 meters during a preliminary field
accuracy test (C. Wright, 2018). Over the course of deployment, collar settings and fix rates were adjusted for various purposes including locating neonates. For our analysis we retained from the initial 664 deer those which had more than 100 relocations and could be subsampled to conform to a 5-hour fix rate (n=485, 219 in North, 248 in South, 18 in Southeast) and 1,137 sub-individuals (490 in North, 591 in South, and 56 in Southeast), corresponding to 597,854 GPS locations.

1.3.3 Integrated step-selection function and covariates

We modeled deer habitat selection across scales using integrated step selection functions (iSSFs) within the ‘amt’ package in the R programming environment (R Core Team, 2020; Signer et al., 2019). Like many other methods of habitat selection analyses, iSSFs model a binary input of observed and available locations as a function of habitat features. However, instead of sampling available locations within a larger area, iSSFs sample available locations from distributions of the distance traveled (step length) and change in bearing (turning angle) between successive points (Avgar et al., 2016, Fieberg et al., 2020). Thus, iSSFs are inherently scale dependent in the way that available points are delineated. The ‘integrated’ aspect of iSSFs denotes that the distributions of the movement parameters are based on theoretical distributions from a movement model, rather than empirical distributions of pooled locations (Avgar et al., 2016). We sampled available locations from gamma distributions of step lengths and von Mises distributions of turning angles for each animal. The median 5-hour step length in our study was 217.5 meters (mean = 311.9, sd = 323.8). We sampled locations at a ratio of five available locations for each used location, which is sufficient for conditional approaches like iSSFs where available locations are directly paired with used locations (Thurfjell et al., 2014). We used iSSFs to quantify selection with respect to two habitat covariates: forest cover and distance to nearest
road. We chose these two covariates because both have been shown in previous studies to be important for deer; forest cover provides food resources as well as thermal and hiding cover, and roads influence deer movement patterns and are a source of vehicle mortality risk (Kammerle et al., 2017; E. S. Long et al., 2010; Trombulak & Frissell, 2000). These covariates also did not covary with each other (correlation < 0.6). Additionally, recent research (Laforge et al., 2015) examining a sample of adult deer \((n = 14)\) found that selection of forest cover and road density varied with spatial scale, thus these covariates serve as useful habitat characteristics for examinations of spatial scaling in habitat selection.

We used rasterized maps of forest cover and distance to nearest road to evaluate the spatial scaling of selection. We developed the forest cover raster using the National Landcover Database (NLCD) by combining the NLCD forest classifications of deciduous, mixed, and evergreen forest (Homer et al., 2015). We developed the distance to nearest road raster using QGIS to convert U.S. Census TIGER county-level road shapefiles (U.S. Census Bureau, 2019) into a raster where each cell corresponds to the distance to the nearest road. We initially explored whether separating primary, secondary, and tertiary road types would improve model fit over all road types combined, however, we determined differences to be negligible and therefore used distance to any road in subsequent analyses.

1.3.4 Modeling scale of effect

To represent forest cover and distance to nearest road across a suite of spatial scales, we updated the value of each 30 x 30 m raster cell using the mean of a moving window with dimensions varying from 1 x 1 cells to 129 x 129 cells and every possible odd-numbered window dimension in between (see Froese et al. 2015). By this process, we evaluated 65 different spatial scales ranging from 900 m\(^2\) to ~15 km\(^2\). This largest scale approximates the maximum size of
deer home ranges in our study. The purpose of using a home range-sized scale was not to investigate first- or second-order selection, but rather because this scale serves as a logical upper bound for studying third-order habitat selection. Given that iSSFs sample available points from the area surrounding used points at distances based on a distribution of an individual’s step lengths, selection manifests as individual movement decisions at each step. The scale of effect then signifies the amount of spatial information an animal takes into account in habitat selection decisions. Individuals might employ different amounts of spatial information (i.e. at different spatial scales) for each decision. Thus, the scale of effect as we refer to it here represents the spatial scale of aggregated habitat characteristics that an individual selects for across all of its movement decisions at a particular temporal scale (in this case, 5 hours). The key point is that selection of habitat at a given step could be informed by spatial characteristics at larger scales, but this is rarely considered or known *a priori* in habitat selection studies.

The moving window approach we used here also has subtly different interpretations than definitions of spatial scale used in other studies. In such analyses, the grain (minimum mapping unit) and extent (maximum mapping unit) are often manipulated separately or concurrently to investigate different aspects of landscape ecology (Dunning et al., 1992; Milne, 1997). As noted earlier, we define spatial scale as the extent in which a habitat variable is aggregated. In using moving windows, the grain (resolution) remains unchanged for each raster and a new value is calculated for each cell (Fig. 1). We used this approach as we were most interested in investigating how landscape composition interacts with spatial scale, but we acknowledge that the shape and distribution of patches across a landscape might also be important to investigate in future work (Thompson & McGarigal, 2002). After preparing rasters for forest cover and distance to nearest road across all scales, we partitioned deer locations by individual, year, and
season to compare effects of spatial scale across ages and seasons. We chose season thresholds for summer and winter that would allow us to confidently exclude dispersal movements. These thresholds were: June 15th – September 15th (summer), and December 15th – March 15th (winter) (Moll et al., 2021). Accordingly, we hereafter use the term *sub-individual* to refer to individuals in a given year and season. We excluded locations collected during spring and autumn, as they coincide with periods of natal and mating dispersal (Benhamou, 2014; E. S. Long et al., 2008; Nixon et al., 2007). We also screened locations for outliers caused by GPS location errors using the process described by Bjørneraas et al. (2010). We then fit iSSFs to both covariates for each sub-individual and each scale. For multiple sub-individuals, iSSFs are designed to be fit in as individual models, rather than population-level models with individual random effects (Scrafford et al., 2018; Signer et al., 2019; Stuber & Gruber, 2020). Models contained either proportion forest or distance to nearest road and, to reduce bias in parameter coefficients, the log of step length (Forester et al., 2009; Signer et al., 2019). Given the volume of the dataset and broad suite of scales, we did not investigate additional covariates or combinations of our focal covariates.

The above process resulted in 147,810 total iSSF models. For each model, we recorded the Akaike’s Information Criterion (AIC) and determined the scale of effect for each sub-individual to be the spatial scale at which AIC was minimized (D. R. Anderson & Burnham, 2002). Researchers sometimes define the scale of effect as the scale corresponding to the largest selection coefficient. We expect that AIC-defined scales of effect will frequently correspond to scales with the strongest selection coefficients, but AIC carries the additional benefit of explicitly quantifying model uncertainty between competing spatial scales. To quantify this uncertainty, we recorded the AIC weights associated with each scale of effect (Alfaro and Huelsenbeck 2006). Recording these weights allowed us to account for the fact that, while a scale of effect can
always be identified, the covariate might have an overall weak effect even at that scale. Such cases would have low AIC weights, and thus would minimally impact overall inference.

1.3.5 Intrinsic and extrinsic factors on scales of effect

The iSSF analysis above resulted in a scale of effect for each sub-individual and each habitat covariate, as well as an AIC weight that quantified uncertainty. Using the scale of effect results and their associated uncertainties from that analysis, we then evaluated the hypotheses described above and summarized in Table 1. To do so, we constructed two linear models – one for forest cover, one for distance to nearest road – with the scale of effect for each sub-individual modeled as a function of the intrinsic and extrinsic factors, both models weighted by the AIC weight of the best performing scale from the iSSFs. The intrinsic factors we used were sex, age, and home range size; and the extrinsic factors were study site, season, mean forest cover in home range, and mean distance to nearest road in home range. We determined the intrinsic factors of sex and age at the time of capture and updated the ages for sub-individuals that were tracked in multiple years. Since we partitioned individuals by year and season, one individual may correspond to multiple scales of effect for our linear modeling dataset. Thus, we also included a random effect for each sub-individual in the linear models (Kuznetsova et al., 2020). We estimated seasonal home range size for each sub-individual using kernel density estimation (95% adaptive kernel, reference bandwidth) via the ‘amt’ package in R (Fleming & Calabrese, 2017; Signer et al., 2019). To account for large differences between deer home range sizes, we logarithmically transformed home range sizes for use in the iSSF models (Kerkhoff & Enquist, 2009). In our preliminary exploration, we found that males home ranges (n = 503, mean = 7.09km²) were larger on average than females (n = 634, mean = 5.95km²). To incorporate the uncertainty of scales of effect into this second analysis step, we used a weighted regression
where the weights were the AIC weights from the first analysis step (Alfaro & Huelsenbeck, 2006).

1.4 RESULTS

Partitioning our 485 deer into sub-individuals, we had 1,137 sub-individuals to use for the two phases of our analysis. In phase one of our analysis, scale of effect ranged from 900 m$^2$ to 15 km$^2$. Parameter coefficients at the scale of effect varied from -1.42 to 2.51 (mean = 0.36, sd = 0.56) for forest, and -1.75 to 2.33 (mean = 0.14, sd = 0.74) for distance to nearest road. The smallest scale (900 m$^2$) was the scale of effect for 68 sub-individuals (6.0%) in the forest models, and 93 sub-individuals (8.2%) in the distance to nearest road models. The largest scale (15km$^2$) was the scale of effect for 41 sub-individuals in the forest models (3.6%) and 92 sub-individuals in the distance to nearest road models (8.0%). Each of the 65 spatial scales tested resulted in the scale of effect for at least two sub-individuals in both the forest and distance to nearest road models. Although every spatial scale tested resulted in the scale of effect at least part of the time, no single spatial scale corresponded to more than 9.0% of resulting scales of effect. AIC weights for scales of effect ranged from 0.02 to 1.00 in both the forest (mean = 0.47, sd = 0.27) and distance to nearest road models (mean = 0.33, sd = 0.23). Across sub-individuals, we observed many patterns of spatial scaling. For example, parameter coefficients varied with scale in straightforward linear trends (Fig. 3a), as well as in parabolic (Fig. 3b), cyclical (Fig. 3d), and chaotic patterns (Fig. 3e). While in many cases, the scale with the lowest AIC corresponded to the scale with the strongest selection coefficient, we encountered instances in which this was not the case (Fig. 3a, 3e, and 3f). Unsurprisingly, we also found that scales that resulted in selection coefficients near 0 were not often statistically significant (Fig. 3, open circles). Across spatial
scales, we observed many instances in which a sub-individual had both positive and negative selection coefficients that were significant depending on the spatial scale (e.g. Fig. 3b.). In total this occurred for 57.1% of sub-individuals for the forest models and in 69.2% of sub-individuals for the road models.

Figure 3. Plots of selection of forest cover, calculated at 65 different spatial scales (blue circles) for six individual white-tailed deer in the North study area of our analysis (Odocoileus virginianus). Parameter coefficients of zero are indicated by orange dashed lines, and the spatial scale that resulted in the lowest AIC (scale of effect) is indicated for each individual by the large orange circles. Filled circles denote scales that resulted in significant p-values, while open circles denote insignificant p-values. Units on the x-axis represent the side lengths of buffer squares (i.e. square in which mean of forest cover values was taken). Y-axis units represent parameter coefficients ranging from negative one to one. Deer sex, age and study area are indicated in each panel.

In the linear models in the second phase of our analysis, most intrinsic and extrinsic factors had little association with scale of effect (Fig. 4). Compared with the intercept (North site) the South study site had a significant negative association with scale of effect in the forest
model (coef. = -281.1, \(p = 0.014\)), and the Southeast site had a significant negative effect in the
distance to nearest road model (coef. = -561.2, \(p = 0.029\)). Home range size had a similar
significant negative association in both the forest (coef. = -267.9, \(p < 0.001\)) and distance to
nearest road models (coef. = -275.3, \(p < 0.001\)). Finally, proportion forest cover within the
home range had a significant positive effect on forest scale of effect (coef. = 240.6, \(p < 0.001\)).

### Figure 4

Linear model coefficients for scales of effect (spatial scales that resulted in lowest
AIC in integrated step selection models) modeled as a function of intrinsic (relating to the
internal state of animals) and extrinsic factors (relating to exterior environment). Variables with
significant p-values are marked with an asterisk.

#### 1.5 DISCUSSION

We found that habitat selection models were highly sensitive to the spatial scales at
which covariates were aggregated. The relationships between habitat selection and spatial scale
varied greatly amongst sub-individuals, and patterns of scaling relationships were highly varied
(Fig. 3). Along with the diversity of observed scales of effect, this variability underscores the
importance of considering multiple spatial scales across a range of values into analyses of habitat selection. Evaluating habitat selection using a small subset of scales (i.e., low, medium, high) would likely fail to capture the scale of effect, as even spatial scales similar in size to the scale of effect in our analysis frequently corresponded to very different parameter coefficients (Fig. 3). Given that parameter coefficients reversed direction across scales over 50% of the time, our results suggest that inference can be strongly influenced by researcher-defined spatial scales of variable aggregation. Furthermore, several intrinsic and extrinsic factors influenced the spatial scaling of animal-habitat relationships.

We hypothesized that variation in home range size among sub-individuals would exhibit a positive relationship with scale of effect (Hypothesis I3); however, we found the inverse – as home range size increased, scale of effect decreased. A plausible explanation for this finding is that in maintaining larger home ranges, deer are naturally required to spend less time overall in any given location, thus they may have more difficulty in maintaining spatial memory for this larger area and instead respond to changes in both forest cover and distance to nearest road at smaller, more local scales. Our findings also did not support our hypothesis that forest cover within home ranges would have a negative relationship with scale of effect of our forest models (Hypothesis E3). Forest cover is important to deer for many reasons including food procurement, concealment, thermoregulation, and protection from severe weather between foraging excursions and dispersals (DelGiudice et al., 2013; E. S. Long et al., 2005; Moll et al., 2021; Wiemers et al., 2014). A positive relationship between forest cover and scale of effect suggests that, for deer in areas with low forest cover (e.g., patchy or fragmented areas), forest selection decisions may arise from more exploratory movement behavior and proximate cues of habitat variation (e.g., simple visual cues).
In support of our initial hypothesis, we found that study area was a significant predictor of scale of effect (Hypothesis E1). For forest cover, differences among the study areas may have been driven by the same mechanisms described above, namely larger contiguous forests in the South and Southeast sites compared to the North site (Fig. 1). However, this effect was not observed for distance to nearest road, which was associated with significant differences in scale of effect across all three study sites (Fig. 4). This result is relevant for spatial analysis more generally because it suggests that white-tailed deer and other animals may exhibit selection behavior at different scales simply by existing in different regions.

Similar effects on habitat selection have been observed for migratory species like songbirds, which respond to the intersecting regional effects of climate, anthropogenic disturbance, landscape fragmentation, and invasive species (Herse et al., 2017; Oguchi et al., 2018; Packett & Dunning, 2009). Thus, our results suggest such landscape-level effects might also be present for resident large mammalian species and influence the scale at which they select for habitat resources. The effect of study site on scale of effect also has important implications for the comparability of studies of deer and other species in distinct study sites across regions and globally. Habitat selection across study sites that are spatially distant, either within a single research effort or compared across studies, may need to be interpreted under the consideration that scaling relationships likely vary across sites.

More generally, our results underscore how spatial scale can strongly affect the outcome of habitat selection analyses conducted at an individual animal level. Population-level studies of habitat selection are often most tractable, but may mask variation from individual scales of effect (Laforge et al., 2016; Montgomery et al., 2018). In our study, scaling relationships exhibited high variation across sub-individuals in the direction and magnitude of estimated parameter
coefficients. Thus, depending on the spatial scale employed, one could come to a different – or even opposite – inference regarding the habitat selection of individuals (Fig. 3). If unaccounted for, the impacts of spatial scale could impair the accuracy of spatial analyses and lead to both Type I and Type II statistical errors (Holland et al., 2004; Hurlbert & Jetz, 2007; Mayor et al., 2009). This is especially meaningful for threatened and endangered species for which spatial research provides invaluable insights and means for conservation planning (Dietz et al., 2020; Ripple et al., 2014; Wilson et al., 2011). For deer, accurate information regarding habitat relationships is vital for predicting future movement patterns and mitigating the spread of pathogens such as chronic wasting disease (D. P. Anderson et al., 2013; Belsare et al., 2021).

Ever-increasing capabilities of telemetry and remote sensing technology can allow and even encourage researchers to use the finest scale possible (Benson, 2010; Mayor et al., 2009). The 900 m² scale, the smallest in our study, is currently the smallest available for many commonly used data sources in North American studies like the NLCD (Homer et al., 2015). However, this scale rarely corresponded to the scale of effect for a given habitat variable. In fact, we observed instances where a 900 m² scale would result in the opposite inference when compared with selection quantified at the scale of effect (Fig. 3e) or support the inference that there is no selection at all (Fig. 3f). This discrepancy indicates that while the decision to use the finest scale of data available may be conducted in hopes of improving statistical accuracy in spatial analyses, it may bias habitat selection if the scale of effect is larger.

Although ecologists have recognized that spatial scale and individuality matter in many contexts (Levin, 1992), the synergy between these concepts has not been widely incorporated into studies of animal habitat selection (Laforge et al., 2015, 2016). We found that spatial scaling of habitat selection is influenced by both intrinsic and extrinsic factors, and that scaling
relationships can be highly variable among sub-individuals. Given that age, sex, and season were not significant predictors of scale of effect, the variable patterns of spatial scaling observed across individuals may be driven by other characteristics like home range size. Other studies have shown that habitat selection behavior can vary across individuals at a single scale (Haus et al., 2020; Montgomery et al., 2018), however our study indicates that individuals also select for resources at different scales. The driving force of such individuality is likely a combination of personality and behavioral plasticity, both of which can influence the pattern and scale of habitat selection separately and interactively, but are difficult to tease apart (T. A. Anderson & Johnson, 2014; Dingemanse et al., 2010; Haus et al., 2020). In our study, patterns of spatial scaling were also not predicted by sex, age, or season. These dynamics combined underscore the need to consider spatial scale more explicitly in resource selection analyses, perhaps by combining scaling of habitat selection with recent advancements in the study of individual spatial behaviors (Osipova et al., 2019; Stuber et al., 2022).

Continued advancement in quantitative techniques has allowed studies like ours to demonstrate the effects of spatial scale on habitat selection analyses. However, we recognize that modeling an exhaustive range of spatial scales as we have done here may be computationally intractable in many applications. Indeed, we also had to make tradeoffs in our analytical approach to accommodate computational needs, including fitting relatively simple models and partitioning individuals by season and year, which could elevate Type I error. We navigated the tradeoff between the complexity of modeling efforts and acceptable levels of potential bias by analyzing a large sample size of animals and using only clearly justified habitat coefficients, thereby reducing the influence of Type I and II errors (di Stephano, 2003). We fit univariate step selection functions for the purpose of computational tractability. As a result, we were only able
to compare forest models to other forest models, and road models to road models via AIC. Many real-world applications will require that models contain multiple relevant covariates in the same model. This and the other strategic decisions we employed here will likely not be possible for many studies, as such a more exploratory approach and tentative interpretations of findings might be warranted. Furthermore, the relationships we have described here between spatial scale and intrinsic and extrinsic factors may not persist across other study areas, regions, or species. We found that the aggregation effect (size of aggregations) of the modifiable areal unit problem is impactful for selection studies, and future work could expand upon this to investigate the zoning effect (shape of aggregations). Such applications could combine multi-scale analysis with shape-varying approaches like path-selection functions (Zeller et al., 2016) or kernel-based selection analyses (Millspaugh et al., 2006).

This study demonstrated that estimated selection coefficients vary with spatial scale and that this scale-dependent selection varies by individual. These dynamics combined underscore the need to consider spatial scale more explicitly in resource selection analyses. Additional work needs to be done to identify the appropriate range of scales that will enable identification of the scale of effect for a given species and context. Our recommendation is that researchers carefully consider the range of scales that could be relevant to the species and question of interest and to evaluate as many of these scales as resources and time allow. We further suggest that modeling frameworks should consider individual-level scales of effect. While there remain many questions about the proper implementation of spatial scale into habitat selection analysis, the matter of spatial scale also presents exciting opportunities to improve and refine our understanding of habitat selection and its interplay with individuality and environmental variation. Ultimately, our
work supports the general conclusion that spatial scale is an important aspect of habitat selection that deserves critical attention in future research

1.6 ACKNOWLEDGEMENTS

Funding for this work was provided, in part, by the National Science Foundation Graduate Research Fellowship Program, the University of New Hampshire, and the New Hampshire Agricultural Experiment Station. Reproduced with permission from Springer Nature.
2. CHAPTER TWO: Generalized nonlinearity in wildlife ecology: research, review, and recommendations

2.1 ABSTRACT

Generalized linear models (GLMs) are an integral tool for wildlife ecologists. Like simple linear regression, GLMs entail several important assumptions, one of the most influential being the assumption of linearity. Ecological relationships are frequently nonlinear and while relationships on the natural scale of GLMs are inherently nonlinear, GLMs still assume linearity on the link scale. Here we clarify nonlinearity in GLMs and explore how it is investigated and addressed among ecologists. We first review recent ecological literature to quantify how often the linearity assumption is tested and the frequency of nonlinear GLM methods in wildlife research. We then present two case studies that confront the assumption of linearity using two common GLMs (a Poisson regression, and a hierarchical logistic regression) fit to empirical data from field studies. The case studies illustrate the potential effects of nonlinearity in these common GLM applications and explore methods of addressing such nonlinearity. In the first case study we compare GLMs to generalized additive models (GAMs) using relative abundance of mammal species from camera trap data. In the second case study we test for nonlinearity in occupancy-detection models using point-count data of passerines. We reviewed 162 studies published in the last five years in five leading ecology journals and found that less than 15% of studies reported testing for linearity. These studies used techniques like transformations and GAMs more often than they reported a linearity test. In our first case study, we found that GAMs out-performed GLMs in modeling relative abundance, and GAMs helped uncover a potential human shield effect related to nonlinear responses of competing carnivore species to landscape development. In our second case study, we found that 14% of passerine bird species failed a
statistical test (the Box-Tidwell test) for linearity. When we applied a nonlinear model to these species, we found that differences between the linear and nonlinear model predictions were similar for some species but changed substantially for others, which has implications for both ecological inference and management and conservation decision-making. We illustrate in this paper that reporting a test for linearity is not commonplace in recent ecological studies employing GLMs, despite the fact that nonlinearity is likely prevalent in the ecological relationships being examined. We offer the recommendations that future ecology studies conduct and report tests for linearity and, in cases where nonlinearity is identified, we suggest researchers use formal methods to assess nonlinearity in GLMs.
2.2 INTRODUCTION

Ecological systems are complex, and this complexity often presents a formidable statistical challenge to researchers trying to disentangle many interacting variables that rarely relate to one another in a straightforward, linear manner (Anand et al., 2010; Bolker et al., 2013; Maurer, 1999; Moll et al., 2016). Consequently, the statistical methods used in ecology have rapidly developed over the last several decades, and quantitative expertise is often highly valued among ecologists, sometimes to the point of controversy (Ellison & Dennis, 2010; McGill, 2012; Millspaugh & Gitzen, 2010). As quantitative methods continue to advance and complex models become commonplace, it can become increasingly easy to overlook foundational assumption checks (R. G. Long, 2008; Nimon, 2012; Osborne & Waters, 2019). Yet regardless of how sophisticated statistical methods become, understanding and reporting foundational assumption checks remains critically important (Zuur et al., 2010). One such fundamental assumption in regression models is that of linearity, which assumes all independent variables in a model have a linear relationship with the dependent variable, such that a constant change in an independent variable leads to a constant change in a dependent variable. In linear regression, this assumption is especially influential and is often listed as the most important assumption for this method (e.g., Gelman et al., 2020). When unaddressed, nonlinearity can cause inaccurate parameter estimation, and a reduction in statistical power (Gelman et al., 2020).

Addressing the linearity assumption is crucial in ecological research because linear regression is one of the most fundamental and commonly used tools in the ecologist's toolkit (Gelman et al., 2020; Gotelli & Ellison, 2004; Schabenberger & Gotway, 2017). Importantly, this assumption applies to both simple linear regression models and generalized linear regression models (GLMs). GLMs are linear models that have been generalized to fit dependent variables
that are not normally-distributed and for which simple linear regression is ill-suited, such as, binary data (e.g., resulting from a binomial or Bernoulli distribution) or count data (e.g., resulting from a Poisson distribution, Nelder & Wedderburn, 1972). GLMs are widespread in ecology because such data types are commonplace in the natural world. GLMs relate independent and dependent variables to each other via a mathematical function, that allows for the assumptions of linear regression to be met. This mathematical function for a given data type is known as the link function, and the scale on which it is the independent variable is known as the link scale, as opposed to the natural scale which is the scale of the original data (Fox et al., 2015; Gotelli & Ellison, 2004). For example, Poisson regression, which is often used to model count data, typically employs a natural logarithm (hereafter simply log) link function to mathematically relate the mean of the Poisson distribution ($\lambda$) to a linear combination of dependent variables in a way that realistically represents the data (e.g., does not allow the model to predict negative counts).

Interpretation of the results of GLMs for ecological purposes typically occurs on the natural scale rather than the link scale, and relationships between independent and dependent variables on the natural scale are often nonlinear. Thus, conceptualization of nonlinearity on the link scale may not come readily for many ecologists (Boldina & Beninger, 2016; Cohen et al., 2013). In simple linear regression, nonlinearity can be directly visualized using scatter plots in an exploratory probe of one’s data (Fig. 5). However, the nature of GLMs and their link functions make visualizing and conceptualizing nonlinearity in a GLM framework less straight-forward (Fig. 5). Thus, for the purposes of this paper, we use the term nonlinearity to specifically reference nonlinearity on the link scale in accordance with this assumption.
Figure 5. Conceptual representations of nonlinear relationship in linear regression across three link functions: identity, logit, and logarithmic. Relationships shown are based on simulated relationships, where the underlying data have a nonlinear relationship between the dependent and independent variables. Raw data points are shown as purple circles, and linear and nonlinear relationships as blue and magenta lines respectively. Raw data for GLMs are the same on both the natural and link scale (binary for logistic regression, counts for Poisson regression) as link functions relate dependent variables to independent variables without directly transforming them.

The specific reasons why one might expect a nonlinear relationship vary based on the independent variable in question, but they mostly stem from the effect of an independent variable not being constant across all values of a dependent variable. In other words, there is often not a one-to-one relationship between a change in the independent variable and the dependent variable as is assumed in simple linear regression. For some variables this nonlinearity can be obvious, such as the effect of distance variables on a dependent variable. For example, a landscape feature
like distance to forest cover may only take effect on an animal’s movement behavior after the threshold of an animal’s perception range has been passed (Zollner, 2000).

Once nonlinearity is identified, there are multiple ways to address it. The first and simplest is transformation (Zuur et al., 2009). Transforming a variable is to apply a mathematical function (e.g., logarithm, exponent) to a specific independent or dependent variable to meet the assumptions of linear regression. Transforming a dependent variable is often done to meet the assumption of normality, while transformations of independent variables are often used to meet the linearity assumption. For example, if an independent variable varies by orders of magnitude, for instance body mass, one might apply a log-transformation to it to linearize its effect. If this transformation is performed in a GLM, there would now be a nonlinear relationship for body mass on the link scale, but a linear relationship with log of body mass. An alternative to transformation is to use generalized additive models (GAMs) that allow for even more complex relationships by fitting multiple nonlinear models to a variable and combining them (Wood, 2017). We describe GAMs in more detail in our first case study below.

The importance of nonlinearity is not a new concept (Sheppard, 1914). Seventy-five years ago, H.O. Hartley (1948) commented on the need for further inquiry into nonlinear relationships by the academic community. The reasons he cited for researchers’ reluctance to use nonlinear methods were a.) The complication in the computational procedure when estimating nonlinear parameters, b.) The lack of exactness in goodness-of-fit tests, c.) The fact that any transformations usually involve certain unknown parameters which must be estimated from the sample, and d.) The difficulty of deciding which of the many possible nonlinear regressions is suggested by theory. Although the analytic and computation tools we possess to address these issues are vastly superior now than when these observations were made, each of these difficulties
remain present in statistical modeling today. To confront these issues, we explore nonlinearity in generalized linear models in wildlife ecology. Specifically, we seek to: 1) determine the frequency at which wildlife ecologists are using or reporting testing for linearity or using nonlinear techniques, 2) explore the dynamics of nonlinearity and consequences of not addressing it in common statistical methods in wildlife ecology using two case studies, and 3) provide recommendations for best practices for ecologists to address nonlinearity in the future. It is not our goal to provide a comprehensive “how-to” for fitting all types of nonlinear GLMs to ecological data, which is beyond the scope of this paper. Instead, in this paper we seek to bring this perennial issue into a fresh light by providing examples and general recommendations of addressing nonlinearity in wildlife research, with the ultimate goal of stimulating additional work, clarifying ecological inference, and improving management and conservation decision-making.

2.3 LINEARITY REVIEW

2.3.1 Methods

We conducted a formal literature review to determine the degree to which researchers reported testing for nonlinearity or used nonlinear modeling approaches. We queried all studies published during a 5-year period (2018 – 2022) in 5 leading ecology journals (Methods in Ecology and Evolution, Ecography, Ecology Letters, Ecology, and the Journal of Animal Ecology). We searched for studies that contained in their title or abstract the terms “model” or “regression”, and the terms “wildlife” or “animal”. From these, we retained studies for analysis that we could confirm 1) were the result of original research; and 2) analyzed data using a linear
regression, GLM, or GAM. This process excluded papers that introduced new R packages, described conceptual models or models made for one specific purpose that were not regressions, or used ‘black box’ methods for which model specifications were seldom described, such as hidden Markov models and machine learning algorithms. For each relevant study, we extracted the following information from the manuscript text and/or supplemental information, when applicable: 1) if the study reported a test for nonlinearity either before or during modeling; 2) if the study used a transformation of an independent variable (e.g., a logarithmic transformation); 3) if the study used a multi-coefficient model, defined as models that had more than one coefficient associated with an independent variable (e.g., a quadratic model or GAM); and 4) the link function(s) used in the study (identity, logit, or log).

2.3.2 Results

We reviewed 273 studies, of which 162 met the criteria described above for analysis. We found that 14.2% (n = 23) of studies reported testing for nonlinearity, 25.9% (n = 42) of studies transformed an independent variable, 20.4% (n = 33) of studies fit a multi-coefficient model, 4.9% (n = 8) of studies used both a transformation and a multi-coefficient model, and 48.8% (n = 79) used neither. Additionally, 33.3% (n = 54) of studies used an identity link function, 40.1% (n = 65) used a logit link function, and 34.6% (n = 56) used a log link function (note that some studies used multiple link functions, thus proportions exceed 100%). Within each of the methods the proportion of studies that reported testing for nonlinearity were: 19.1% (n = 8 of 42) of studies using transformations, 30.3% (n=10 of 33) of studies using multi-coefficient models, 50% (n = 4 of 8) of studies using both, and 1.3% (n = 1 of 79) of studies using neither.
Figure 6. Results of a literature review of studies published between 2018 and 2022 in 5 leading ecological journals showing the prevalence of various methods to account for nonlinearity in linear models. The pie chart shows the proportion of studies (out of a total of 162) that included transformations of independent variables, multi-coefficient models (abbreviated “multi-coef,” in figure; these included models with more than one coefficient for an independent variable), both, or neither. The bar chart shows how often a study also reported a test for linearity within the proportions from the pie chart so that proportions sum to 1.0.

2.4 CASE STUDY 1: MODELING MAMMAL ABUNDANCE USING POISSON REGRESSION

2.4.1 Background and Study Design
This case study is focused on the relative abundance modeling of coyotes (Canis latrans) and red foxes (Vulpes vulpes, hereafter ‘foxes’). These two canid species are competitors, and coyotes are dominant over smaller foxes and often reduce their abundance through interference competition and intraguild predation (Gosselink et al., 2007; Levi & Wilmers, 2012; Moll et al., 2018). Prior research on these species suggests that their competitive interactions might be mediated by human development of the landscape (Moll et al., 2018, 2023). Specifically, a “human shield” effect might reduce the interference competition experienced by foxes from coyotes, which are more human averse and therefore less abundant in more developed locations that foxes readily inhabit (Moll et al., 2018). However, this relationship does not always cleanly manifest (Cervantes et al., 2023; Mueller et al., 2018), suggesting that nonlinear dynamics among these variables could be present. Thus, these species act as a useful case study to explore nonlinearity in a GLM framework.

To quantify relative abundance, we deployed camera traps across 109 sites in the 3,200 km² Southeast Management Region in New Hampshire (New Hampshire Fish and Game, 2022), 34 of which coincided with locations also used for bird point counts in case study 2 (Fig. 7). The study area consists of approximately 49% forested land, 25% anthropogenic development, 14% wetland, 4% open water. The remaining 8% is comprised of a combination of agriculture, grasslands, shrublands, and sand beaches (Dewitz & U.S. Geological Survey, 2021). We selected 75 of the camera sites using a generalized random tessellation design, stratified to not be within 1 km of each other (Stevens Jr & Olsen, 2003), the remaining 33 sites were selected with a fully random design within properties managed by the University of New Hampshire and were separated by at least 250 m (Moll et al., 2023; Poisson et al., 2023). We selected all sites using QGIS (QGIS Association, 2021), then made minor adjustments in the field to locate a suitable
location to affix the camera on a tree at approximately knee height (~ 50 cm), facing north (>270° or <90°), and not conspicuously placed to avoid theft or tampering. We programmed cameras to take three images with each motion trigger, and then not re-trigger for five minutes (Lepard et al., 2018). For this case study, we used images captured during summer 2022 (Jun. 15 – Sep. 15). We obtained road data from the TIGER county-level road database (U.S. Census Bureau, 2019). We derived all other variables using the National Land Cover Database (Dewitz & U.S. Geological Survey, 2021).

Figure 7: Study area for two case studies in Southeastern New Hampshire, USA in a study examining the dynamics of nonlinearity in wildlife modeling. The first case study (blue and orange circles) used data from an array of 109 camera traps deployed in summer of 2021. The second case study (purple and orange circles) used point-count survey data from passerines on lands managed by the University of New Hampshire in the summers of 2021 and 2022.

2.4.2 Abundance modeling

We modeled the relative abundance of these mammal species using a Poisson regression with the number of days with at least one detection as our dependent variable instead of the
absolute number of observations, to help account for variation in detection rates across species and to create a sampling design amenable to Poisson regression. Hereafter we refer to this metric as relative abundance. For each species we fit two models: a Poisson GLM, and a Poisson GAM. GAMs are flexible models that use combinations of many additive polynomials (i.e., smoothed functions) in the linear predictor, that allow for ‘wiggliness’ in a fitted relationship that can conform to complex patterns (Pedersen et al., 2019; Wood, 2017). These individual polynomial terms are often called ‘splines’ or ‘knots’. The number of splines in a model, and the parameters of their smoothing functions determine how wiggly a model is allowed to be (Pedersen et al., 2019; Wood, 2017). A general model form for GAMs is:

\[ y = \alpha + f_1(x_1) + f_2(x_2) + \ldots + f_n(x_n) \]

where \( y \) is the dependent variable, \( \alpha \) is the \( y \)-intercept, \( x_n \) are the \( n \) independent variables, and \( f_n \) are the associated \( n \) splines.

In each GLM and GAM we included three independent variables: road density within 1 km\(^2\) (a proxy for human landscape development in this study area), distance to nearest water source, and distance to nearest agricultural area. We chose these variables based on their importance to these focal species as shown in previous work (Moll et al., 2023). In GAMs, we used restricted maximum likelihood to define our smoothing functions and thin-plate regression to determine the number of splines, but constrained the smoothing basis functions to 10 dimensions (\( k=10 \)) to encompass the estimated degrees of freedom for all species (Wood, 2017). To obtain a measure of whether the GAMs provided a better model fit than the GLMs, we compared GLMs to GAMs using the Akaike Information Criterion (AIC). AIC allows for the comparison of models, and as long as models are based on the same underlying data and probability distribution, can allow for comparison of linear and nonlinear models (Fox et al.,
For model visualizations, we fit univariate models of each independent variable and each species.

2.4.3 Results

We captured 354 images (n = 203 coyote and n = 151 foxes) between June 15 and September 15, 2022. For both species, the GAMs were heavily favored (coyote ΔAIC = 32.8, fox ΔAIC = 214.5) over GLMs via AIC (Table 2). We found that the GLMs for foxes and coyotes were approximately the inverse of each other for road density, with coyotes having a negative relationship with road density and foxes having a positive relationship. In the GAMs, however, we found this relationship to still be inversed, but in a cyclical pattern where peaks in coyote relative abundance coincided with troughs in fox abundance and vice-versa (Fig. 8).

Table 2: Comparisons between generalized linear model (GLM) and generalize additive model (GAM) specifications of coyotes (Canis latrans), and red foxes (Vulpes vulpes) from a study of the usage and applications of nonlinear methods in wildlife ecology. Data for the models was obtained from camera trap arrays in New Hampshire, USA. Independent variables for both model types are: Road density within 1 km, distance to nearest water source, and distance to nearest agricultural area. Significant p-values (<0.05) are shown in bold. The Akaike’s Information Criteria (AIC) is included for both the GLMs and GAMs. GAM models also were based upon RoadDens, DistW, and DistA, though their coefficients and their p-values are not shown due to space considerations.

| Species | GLM | Road Density | | | Distance to Water | | | Distance to Agriculture | | | AIC |
|---------|-----|--------------|---|---|------------------|---|---|--------------------------|---|---|
| Coyote  | Estimate | -0.18 | p-value | 0.064 | 0.08 | p-value | 0.299 | Estimate | -0.29 | p-value | 0.002 | 457.8 |
| Red fox | Estimate | 0.31 | p-value | <0.001 | 0.08 | p-value | 0.418 | Estimate | -0.12 | p-value | 0.0261 | 509.4 |

<table>
<thead>
<tr>
<th>Species</th>
<th>GAM</th>
<th>EDF</th>
<th>p-value</th>
<th>EDF</th>
<th>p-value</th>
<th>EDF</th>
<th>p-value</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coyote</td>
<td>EDF</td>
<td>4.583</td>
<td>p-value</td>
<td>0.006</td>
<td>4.645</td>
<td>&lt;0.001</td>
<td>1.001</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Red fox</td>
<td>EDF</td>
<td>6.733</td>
<td>p-value</td>
<td>&lt;0.001</td>
<td>6.260</td>
<td>&lt;0.001</td>
<td>5.749</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>
Figure 8. Generalized linear models (GLMs) and generalized additive models (GAMs) of relative abundance fit to camera trap data of coyotes (*Canis latrans*), and red foxes (*Vulpes vulpes*) collected in the summer of 2021 in New Hampshire, USA. The y-axis represents the relative abundance of a species (days with at least one detection) as a function of an independent variable, distance to agriculture, or road density. In the right-hand panel, coyote GLMs and GAMs are denoted with a solid line, and red fox GLMs and GAMs with a dashed line.

2.5 CASE STUDY 2: MODELING PASSERINE OCCUPANCY

2.5.1 Background and Study Design

In our second case study, we focus on a passerine community (i.e., songbirds, order *Passeriformes*) in Southeastern New Hampshire. Passerine communities often contain many more species in a given area than other groups like mammals (Benítez-López et al., 2010). These birds are highly varied, and often have very specific habitat requirements that we expect to manifest nonlinear relationships for some species, especially for landscape variables like land cover type and anthropogenic development (Benítez-López et al., 2010; Ktitorov et al., 2008;
When evaluating individual species, residuals vs. fitted plots are often used in an attempt to obtain a general sense of whether a linear relationship is appropriate. However, these plots don’t provide an objective threshold past which a relationship can be deemed nonlinear, and for speciose communities, it is not as tractable to go through this subjective analysis with each individual species. Thus, we utilize an empirical test, the Box-Tidwell test (described in further detail below) to determine which species exhibit nonlinear relationships.

We conducted independent-observer point counts (Betts et al., 2008; Frey et al., 2012) of passerine bird species across 70 sites in Southeastern New Hampshire, USA during the two consecutive summer breeding seasons of June-July during 2021 and 2022. We selected sites randomly within areas with land access permission and we separated sites by a minimum of 250 m. This study area also coincides with the camera trap surveys used and described in our first case study (Fig. 7). The point-count survey period each day began 30 minutes before sunrise and concluded four and a half hours after sunrise, as detection rates are expected to decline beyond this time frame because birds reduce or discontinue singing (De Wan et al., 2009; Kroll et al., 2007; Leu et al., 2017). Each survey lasted eight minutes, during which the observer recorded all bird species seen or heard (Leu et al., 2017). We visited most sites three times each year (range per site across years combined = 2 - 6, mean = 5.36), with a total of 379 total survey-days.

2.5.2 The Box Tidwell test

We first tested each species in our dataset for nonlinearity. The Box-Tidwell test (Box & Tidwell, 1962) identifies nonlinearity in a dataset by fitting a linear specification of a model alongside an interaction between the independent variable and the natural log of that same variable:
\[ y = \alpha + \beta_1 \times x_1 + \beta_2 \times x_1 \times \ln(x_1) \]

where \( y \) is the dependent variable, \( x_1 \) is the independent variable, \( \alpha \) is the y-intercept, \( \beta_1 \) is the linear coefficient, and \( \beta_2 \) is the interaction coefficient. If the interaction coefficient is significantly different from zero (typically using a cutoff of \( p < 0.05 \)), it is evidence of the presence of a nonlinear relationship, although the test itself does not indicate the exact specifications or polynomial that should be used to correct this nonlinearity.

### 2.5.3 Occupancy Modeling

We modeled occupancy for all species with 10 or more observations across both survey years. We fit an occupancy model to each species with several independent variables hypothesized to influence these species’ occurrence as well as an interaction between the variables and their natural logarithms. The independent variables we used were distance to nearest road, distance to nearest body of water, proportion forest cover, and proportion wetland cover as these have potential to impact passerine occurrence (Ascensão et al., 2022; Brown et al., 2014; Rahlin et al., 2022). We also fit a detection sub-model with time of day as an independent variable, which we found during preliminary testing to be the only significant variable over other variables we collected at the time of the survey: week of survey, cloud cover, and wind speed. For all variables we used a logistic prior distribution with a location of zero and a scale of one (Northrup & Gerber, 2018). For the species that failed the Box-Tidwell test, we fit univariate occupancy models with only the variable for which the species failed and compared that model to an identical one with a natural log transformation of that variable, as this is the transformation used in the Box-Tidwell test (Box & Tidwell, 1962), though as we note in the Discussion other polynomials could be used. We fit both the Box-Tidwell tests and univariate models in a Bayesian framework using JAGS via the ‘r2jags’ package in R (Version 4.2.1, Plummer, 2003; R
Core Team, 2022; Su et al., 2015). We fit all models with 8000 MCMC iterations with three un-thinned chains and a burn-in of 1000 iterations (Link & Eaton, 2012).

2.5.4 Results

After removing species with less than 10 observations, we ran Box-Tidwell tests for the five independent variables listed above for 42 bird species (a total of 210 tests). In total there were eight failures (3.8% of 210) in six species (14.2% of 42, Table 3). We observed at least one failure for each independent variable considered in the occupancy sub-model, but none for time of day in the detection sub-model. We also observed failures across models for species that strongly varied in commonness, as our most-observed (red-eyed vireo, n = 143 detections) and least-observed species (bobolink, n = 10 detections) both failed the Box-Tidwell test (Table 3). In most of the failures, the magnitude of the effect of the independent variable was small on the natural scale (Fig. 9). There were species that had similar trends on the natural and link scales (Fig. 9, yellow warbler) and species that differed strongly between these scales (Fig. 9, cedar waxwing). In three instances, the log-transformation yielded a positive effect where the linear case of the model was negative, or vice versa.

Table 3: Passerine bird species that failed a Box-Tidwell test for linearity and their respective number of occurrences and variables for which the test failed. Results are part of a study examining the dynamics of nonlinearity in wildlife modeling using two case studies from Southeastern New Hampshire and the prevalence of nonlinear techniques among wildlife studies using a review of literature.

<table>
<thead>
<tr>
<th>Species</th>
<th>No.</th>
<th>Box-Tidwell</th>
<th>Mean (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Red-eyed Vireo</strong> (<em>Vireo olivaceus</em>)</td>
<td>143</td>
<td>Distance to road</td>
<td>2.21 (0.19, 5.35)</td>
</tr>
<tr>
<td><strong>Song Sparrow</strong> (<em>Melospiza melodia</em>)</td>
<td>79</td>
<td>Distance to road</td>
<td>-1.49 (-2.96, -0.32)</td>
</tr>
<tr>
<td><strong>Cedar Waxwing</strong> (<em>Bombycilla cedrorum</em>)</td>
<td>42</td>
<td>Forest cover</td>
<td>1.68 (0.17, 4.00)</td>
</tr>
<tr>
<td><strong>Yellow Warbler</strong> (<em>Setophaga petechia</em>)</td>
<td>31</td>
<td>Distance to road</td>
<td>-1.55 (-3.30, -0.09)</td>
</tr>
<tr>
<td><strong>Eastern Phoebe</strong> (<em>Sayornis phoebe</em>)</td>
<td>18</td>
<td>Distance to water</td>
<td>1.23 (0.14, 3.07)</td>
</tr>
</tbody>
</table>
**Bobolink** (*Dolichonyx oryzivorous*)

<table>
<thead>
<tr>
<th>Species</th>
<th>Feature</th>
<th>Coefficient ± CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bobolink</td>
<td>Wetland cover</td>
<td>-2.63 ± (-5.97, -0.26)</td>
</tr>
<tr>
<td></td>
<td>Distance to water</td>
<td>1.05 ± (0.13, 2.20)</td>
</tr>
<tr>
<td></td>
<td>Forest cover</td>
<td>-2.10 ± (-5.29, -0.32)</td>
</tr>
</tbody>
</table>

**Figure 9.** Linear and nonlinear specifications of occupancy models for passerine bird species from point count data collected in the summers of 2021 and 2022 in New Hampshire, USA. Included in the plot are those that failed a Box-Tidwell test for linearity. Confidence intervals were omitted from plots for clarity. Y-axis units on the natural scale are the probability of occupancy (occ. prob.), and log of the odds ratio for occupancy on the link scale (logit(occ. prob.)).

**2.6 DISCUSSION**
Generalized linear models are an integral tool in wildlife ecology. Like simple linear regression, GLMs are subject to foundational assumptions, like the assumption of linearity. This assumption applies to the link-scale of the GLM, but interpretation of the outputs of GLM most often occurs on the natural scale. It is this dynamic that makes conceptualization of what a nonlinear effect on the link scale means on the natural scale difficult. We aimed in this study to investigate whether nonlinearity exists on the link scale of GLMs in real-world ecological examples, and the degree to which studies report testing for nonlinearity in their studies.

In our review of contemporary ecology literature, studies seldom reported using linearity tests before applying a nonlinear technique. For instance, more than 25% of studies used a transformation in their analysis, but less than a quarter of those studies reported testing for linearity. Overall, less than 15% of studies overall reported testing. This trend could be an indication of a few dynamics occurring separately or concurrently. Studies could be applying nonlinear methods more often than they test for nonlinearity, basing their decision to use a nonlinear method either on a priori expectations or prior research. Conversely, studies could be testing for linearity, but not including those results in the main text. Both scenarios suggest that highlighting the importance of the linearity assumption and reporting tests for nonlinear relationships is not commonplace in wildlife ecology literature. Importantly, there are specific variables in ecology research where nonlinear techniques are almost universally applied and explaining their use is likely redundant, such as log-transformations of body mass in allometry research (Kerkhoff & Enquist, 2009). Nonetheless, the nuances of transformations on the link scale of GLMs is still worth discussion even for variables like body mass, and our findings that there is a lack of attention placed on nonlinearity for virtually all other variables.
In our first case study, we found that GAMs not only outperformed GLMs in Poisson regression of relative abundance, but more importantly elucidated patterns obscured by GLMs. These patterns provide additional ecological interpretations that would not be able to be drawn from the GLMs alone. In our analysis of foxes and coyotes we found that not only did both species have a nonlinear relationship with distance to nearest road, but their relationships were the inverse of one-another in both the GLM and GAM. When inspecting only the GLM, we can deduce that coyotes and foxes might simply have differing tolerances to anthropogenic development. However, the GAM reveals a more complex cyclical pattern (Fig. 8). This could be the result of foxes interfacing with anthropogenic development primarily to avoid coyotes spatially, i.e. a “human shield effect” (Moll et al., 2018). Most importantly, this relationship was only recognizable when using a GAM, and no output of the GLM necessarily points to the necessity of fitting a GAM.

Interpreting the numerical outputs of a GAM may be less straightforward than a standard GLM that has one coefficient for each independent variable. The multiple coefficients that come out of GAMs are potentially not as easily communicable or actionable for interested parties and stakeholders like land managers. However, the utility of GAMs, as we found in this case study, is often in the visual representation itself which can elucidate dynamics that linear models or GLMs could not. Importantly, we fit relatively simplistic GAMs, and do not in this case study delve into the technical details of GAMs themselves. With GAMs flexibility comes sensitivity to input parameters like smoothing basis functions and the number of splines to use; as such, misspecifications can lead to over or underfitting (Furey et al., 2013; Pedersen et al., 2019; Wood, 2017). It is not sufficient to fit a GAM with default specifications in R without careful inspection of data properties like estimated degrees of freedom and assume that any potential
nonlinearity has been addressed. GAMs frequently require fine-tuning to achieve an appropriate fit (see Wood, 2017 for more information). As with nonlinearity itself, it is important to be thorough in preliminary evaluations of ones’ data, specify the parameters of GAMs accordingly, and justify the choices made (see Pedersen et al., 2019; Wood, 2017).

We found in our second case study that 14% of passerine bird species failed a Box-Tidwell test for linearity. Although this comprised a small number of the total tests we performed, the nonlinear models we fit for the species that failed highlight the impacts of nonlinearity in GLMs. Firstly, for three species we found that the coefficient from the log-transformed model was of the opposite direction than the linear specification (Fig 9e, 9f, 9h), meaning that unaddressed nonlinearity could lead researchers to opposite and possibly incorrect conclusions. Secondly, though often subtle, the differences between the linear and nonlinear GLMs are non-trivial for ecological interpretation. For yellow warblers (Setophaga petechia) in this case study, the nonlinear GLM predicts that the probability of occupancy would be approximately 12% less than the linear GLM at values of distance to nearest road near 200 m, and 15% greater near 800 m. These differences could constitute important information to know for a practitioner managing this species. In this case, habitat restoration efforts for yellow warblers near roads may be more or less effective depending on where they are sited. The effect of this potential model inaccuracy would be further exacerbated for threatened and endangered species or situations where financial resources are limited (Bottrill et al., 2009; Game et al., 2013). Finally, interpretations of GLMs typically occur on the natural scale. For some species in this case study, the natural scale and link scale relationships were quite similar (Fig. 9d), but for other species with effect sizes greater in magnitude, the differences between natural scale and link scale were more pronounced (Fig. 9c). Since the data in a GLM with a log-transformed
independent variable are transformed in two ways: the natural log in the linear predictor, and the link function itself, the direction and magnitude of the effect on the link scale can influence what pattern will manifest on the natural scale.

In this study, we have demonstrated that nonlinearity can be present on the link scale of GLMs, and nonlinear methods can elucidate ecological patterns obscured by standard GLMs. Addressing this nonlinearity not only improves model fit, but also provides novel information on which ecologists and stakeholders can make more informed conservation and management decisions. Also, the field of wildlife ecology would benefit from more standard practices for addressing and reporting nonlinearity in linear modeling. Thus, it is our recommendation that wildlife ecologists: 1.) test their data for nonlinearity, and when using GLMs test specifically for nonlinearity on the link scale, 2.) report the results of such and the details of any nonlinear methods used, and 3.) offer justification for the tests and nonlinear mitigation techniques used. Not only will these practices improve the efficacy and reproducibility of research methods, but we believe increased understanding of and attention to nonlinear dynamics will allow for increased and novel insights into ecological patterns and processes.

2.7 ACKNOWLEDGEMENTS

I thank E. Johnson, I. Aldrich, S. Buzzell, K. Falconer, A. Lussier, J. Parry, C. Renteria, S. Richard, L. Santos, F. Shinost, M. Shubert, and M. Twombly for their work in the field and with photo identification. Credit for animal silhouettes in figures: Gabrielle Palomo Muños via phylopic.com. Funding sources for this research include, the National Science Foundation Graduate Research Fellowship Program, the Pittman-Robertson Federal Aid in Wildlife
Restoration Act Grant No. F21AF01526 (W-112-R-1), the New Hampshire Fish and Game Department, the University of New Hampshire, and the New Hampshire Agricultural Experiment Station. This is NHAES Scientific Contribution number [Number to be generated upon paper acceptance]. This work was supported by the USDA National Institute of Food and Agriculture Hatch Project 1024128.
LIST OF REFERENCES:


QGIS Association. (2021). *QGIS geographic information system* [Computer software]. QGIS.org


