DATA INTEGRATION IN POPULATION AND COMMUNITY ECOLOGY USING HIERARCHICAL MODELING

By

Matthew T. Farr

A DISSERTATION

Submitted to Michigan State University in partial fulfillment of the requirements for the degree of

Zoology-Doctor of Philosophy Ecology, Evolutionary Biology and Behavior-Dual Major

ABSTRACT

DATA INTEGRATION IN POPULATION AND COMMUNITY ECOLOGY USING HIERARCHICAL MODELING

By

Matthew T. Farr

In this dissertation, I develop and apply methods for data integration using hierarchical modeling to estimate the status, trends, and demography of wildlife populations and communities. I use multi-level statistical and mathematical models to explicitly link observed data to latent ecological processes. By separately modeling observational and ecological processes, I can integrate multiple disparate data sources into a unified framework to estimate ecologically relevant population and community parameters, often in the context of wildlife conservation. In Chapter One, I apply a multispecies hierarchical distance sampling model to assess the effect of management actions on a carnivore community in the Masai Mara National Reserve, Kenya. I assess variation in species-level responses to passive management, resulting in human disturbance and apex predator declines. In Chapter Two, I develop an integrated distribution model that uses distance sampling and presence-only data to jointly estimate species abundance. I apply this model to a case study on black-backed jackals (Canis mesomelas) to evaluate the effects of anthropogenic disturbance on the distribution of jackals across the Masai Mara National Reserve. In Chapter Three, I evaluate status and trends of species in a forest dwelling duiker community using detection-nondetection data. I develop a multispecies dynamic *N*-occupancy model to estimate species-level abundance, demographic parameters, and quasiextinction probabilities. In Chapter Four, I create a spatiotemporal integrated model to estimate the effects of weather conditions on monarch butterflies (Danaus plexippus) during spring migration. Each chapter illustrates a unique application of data integration in wildlife ecology,

either by combining data on multiple species to estimate population and community-level parameters or by combining disparate data sources on a single species to estimate demography and other population-level parameters. Data integration is a powerful framework that leverages all available information to address pressing conservation challenges.

ACKNOWLEDGEMENTS

The quantity and content of words within a dissertation do not accurately reflect the experiences of obtaining a PhD. The path to completing a PhD is strenuous, and behind any successful graduate is a community of supporting individuals. Since my matriculation at Michigan State University in 2015, I have developed numerous relationships that have fostered both my academic prowess and growth as an individual. I feel privileged to have received the amount of support in both my professional and personal lives while pursuing my Doctor of Philosophy degree.

My coming to MSU was encouraged by my undergraduate advisors, Dr. Patrick Zollner and Dr. Robert Swihart. I owe them a huge debt of gratitude for recommending me to my graduate advisor, Dr. Elise Zipkin. To Elise, thank you for continuing to be such an exceptional advisor. Your commitment to helping and improving everyone and everything around you is inspiring. The Zipkin Lab and greater MSU community has truly been an outstanding group of scientists that have promoted my growth as a researcher. Beyond formal collaboration, the access to each lab member and colleague has been a great source of intellect and creativity in dealing with the idiosyncrasies of ecology and graduate school. A special thanks to Dr. Sonja Christensen, Dr. Savvas Constantinou, Kayla Davis, Dr. Andrew Dennhardt, Dr. David Green, Erika LaPlante, Dr. Connie Rojas, Dr. Sarah Saunders, Dr. Klara Scharnagl, Alli Sussman, Dr. Alex Wright, and Dr. Erin Zylstra. I am positive that our friendships will carry on. An additional thanks to my committee members, Dr. Andrew Finley, Dr. Kay Holekamp, and Dr. Gary Roloff for your commitment to improving my research. A specific thanks to Kay for the access to her outstanding long-term monitoring data and for my tenure at her field sites in Kenya. I also want to thank my family. My parents, Tom and Jane, and my sister, Rachel, have been constant in their support of me. I feel very grateful to them. I want to thank my friends as well for always keeping me grounded, focused, and motivated on pursing my passions. A specific thanks to Brandon Zinman and Cole Bleke for providing me an outlet to enjoy wildlife and the outdoors outside of academia. I cannot wait for our upcoming adventures.

Lastly, to the love of my life, Chase. My appreciation for you is never fully expressed. We have been through so many trials and tribulations in the last six years, and I am forever grateful we were able to overcome them together. As we finally end this chapter, I cannot wait for what is next for both of us and revel in the uncertainty and endless possibilities of our next chapter.

Matthew T. Farr

TABLE OF CONTENTS

| LIST OF TABLES | viii |
|--|----------------|
| LIST OF FIGURES | ix |
| INTRODUCTION | 1 |
| CHAPTER 1: Multispecies hierarchical modeling reveals variable responses of African | |
| carnivores to management alternatives | 7 |
| Abstract | 7 |
| CHAPTER 2. Integrating distance sampling and presence-only data to estimate species | |
| abundance | 0 |
| A betract | رو 0 |
| AUStract | ····· <i>)</i> |
| CHAPTER 3: Quantifying the conservation status and trends of wildlife communities us | ing |
| detection-nondetection data | 11 |
| Abstract | 11 |
| Introduction | 12 |
| Model | 15 |
| Observation process | 15 |
| Biological process | 17 |
| Community component | 18 |
| Application | 19 |
| Study area and data collection | 19 |
| Data analysis | 21 |
| Bayesian population viability analysis | 23 |
| Results | 24 |
| Population abundance and growth | 24 |
| Demographic rates | 24 |
| Population viability | 25 |
| Discussion | 26 |
| CHAPTER 4: Overcoming data gaps using integrated modeling to estimate population | |
| abundance of a migratory species | 34 |
| Abstract | 34 |
| Introduction | 34 |
| The Data | 37 |
| The Model | 38 |
| Spatial heterogeneity in population density | 39 |
| Temporal change in population density | 40 |
| Structured count data | 40 |
| Unstructured presence-only data | 41 |

| Spatiotemporal data integration | 42 |
|--|-----|
| Joint likelihood | 43 |
| Simulation Study | 43 |
| Scenario | 43 |
| Results | 45 |
| Case Study | 46 |
| Population monitoring | 46 |
| Spatiotemporal stages | 47 |
| Datasets | 48 |
| Environmental covariates | 49 |
| Observation covariates | 49 |
| Data analysis | 50 |
| Results | 51 |
| Discussion | 52 |
| | |
| APPENDICES | 61 |
| APPENDIX A: Additional application details | 62 |
| APPENDIX B: Nimble model code | 65 |
| APPENDIX C: Model output | 71 |
| APPENDIX D: Simulation code, and output | 80 |
| APPENDIX E: Case study details, code, and output | 100 |
| | |
| BIBLIOGRAPHY | 116 |
| | |

LIST OF TABLES

| Table 3.1 National parks and basic data information used in my multi-species dynamic N- occupancy analysis of the antelope community including the number of years sampled, species |
|--|
| observed, number of sampling sites, and the number of recorded presences across species. |
| Species are numerically identified as follows: (1) Cephalophus callipygus (2) Cephalophus dorsalis (3) Cephalophus harvevi (4) Cephalophus leucogaster (5) Philantomba monticola (6) |
| Nesotragus moschatus (7) Cephalophus nigrifrons (8) Cephalophus ogilbyi (9) Tragelaphus scriptus (10) Cephalophus spadix (11) Tragelaphus spekii (12) Cephalophus silvicultor |
| Table C.1 Mean annual population growth rates (i.e., geometric mean) with 95% credibleintervals of each population (species-park combination) during the study period |
| Table C.2 Mean annual site-level abundance with 95% credible intervals of each population |

| across the study period (including only the years sampled for each population/park combination) |
|--|
| Table C.3 Mean annual survival probability and gains with 95% credible intervals for each species across all parks. 77 |
| Table C.4 Park-level mean annual survival probability with 95% credible intervals78 |
| Table C.5 Quasi-extinction probability (for each of the 18 populations coming from parks with at least five years of data) for the projected window of 2020-2030 |
| Table D.1 The root mean squared error and percent relative bias (median, lower and upper quantiles) of each estimated parameter compared to the true simulated value for both the integrated (I) and presence-only (PO) analyses |
| Table E.1 Estimates (mean and 95% confidence intervals) of monarch population density per 100-m ² for each year and stage |
| Table E.2 Estimates (mean and 95% confidence intervals) of the effect of NDVI and averaged |

LIST OF FIGURES

INTRODUCTION

Human population growth and resulting pervasiveness of anthropogenic disturbance have left no wilderness or natural place unscathed (Foley et al. 2005, Watson et al. 2018). Anthropogenic-induced declines of many wildlife species have led to extirpations and extinctions, and consequential losses of biodiversity can have dramatic effects on the remaining wildlife (Ceballos et al. 2015, Zipkin et al. 2020, IUCN 2021). The goal of conservation science and wildlife management is to reverse these alarming trends through strategic protection of threatened species and their critical habitats (Watson & Venter 2017, CBD 2020). Numerous obstacles obstruct biodiversity conservation, but one of the daunting hurdles limiting such global-scale efforts has been extensive gaps in data availability on species status and trends (Meyer et al. 2015, Conde et al. 2019). Throughout both terrestrial and aquatic biomes, locations of data collection are often biased because of socio-economic constraints (e.g., financial resources for data collection are tied to developed countries), creating spatial gaps in data availability (Meyer et al. 2015). The temporal extent of data is also restricted as long-term ecological monitoring programs have existed for less than a century (Magnuson 1990). Additionally, data are often collected for only a subset of species (e.g., charismatic, indicator, umbrella) within ecological communities while many threatened species are categorized as data deficient by the International Union for Conservation of Nature (Troudet et al. 2017, IUCN 2021). Spatial, temporal, and species data gaps limit characterization of the status and trends of threatened populations (Conde et al. 2019).

The growing threat to global biodiversity has prompted the establishment of many largescale monitoring programs to overcome data gaps in estimating the status and trends of threatened species (Magnuson 1990). Data digitalization and online repositories (e.g., iNaturalist,

GBIF, Map of Life) have exponentially increased data availability for addressing both basic ecological questions and those focused on conservation (Jetz et al. 2012, Chandler et al. 2017). However, variation in the species, ecosystems, and motivations in data being collected by monitoring programs and a lack of standardization in monitoring have led to a plethora of disjointed datasets, often containing different types (e.g., presence-only, presence-absence, count, capture-recapture), which can make analysis challenging (Miller et al. 2019, Saunders et al. 2019a, Isaac et al. 2020). Data quality, or the information content of a given data type, largely depends on how the data were collected and whether the observation process was designed to account for errors such as, imperfect detection or location biases (Miller et al. 2019). A tradeoff often exists between information content and data quantity where information-rich data has a high logistical cost (e.g., financial, effort, infrastructure). Lower information content data, such as that collected via opportunistic or *ad libitum* sampling, can be conducted with relative ease as there is no protocol or design.

Quantitative ecologists have rapidly been developing methods to strategically integrate multiple data sources into a unified analysis to estimate population- and community-level parameters (Dorazio & Royle 2005, Dorazio et al. 2006, Zipkin et al. 2009, Zipkin et al. 2010, Rossman et al. 2016, Pacifici et al. 2017, Zipkin & Saunders 2018, Fletcher et al. 2019, Miller et al. 2019, Isaac et al. 2020). Data integration, which I broadly define as any quantitative analysis framework that combines multiple data sources to estimate a shared ecological process (e.g., abundance, demographic rates, covariate effects), can improve accuracy and precision of estimates over independent analyses of each data source (Zipkin et al. 2009, Zipkin & Saunders 2018). Data integration may also enable estimation of parameters that would be unidentifiable in

an analysis of a single data source (Schaub & Abadi 2011, Dorazio et al. 2014) and/or resolve discrepancies among inferences based on disparate data sources (Saunders et al. 2019a).

In my dissertation, I focus on a subset of data integration methods that use a hierarchical modeling framework to differentiate variation coming from the observation processes of one or more data sources and variation attributable to underlying ecological processes. This framework links the ecological parameter(s) of interest (e.g., abundance, survival) to the data by accounting for observation error, which may vary in structure or magnitude, among data sources (Pacifici et al. 2017, Zipkin & Saunders 2018, Fletcher et al. 2019, Miller et al. 2019, Isaac et al. 2020). I explore the utility of data integration within both community and population ecology by combining data from different species (i.e., multispecies models) and combining data of different types (single-species integrated models). Multispecies modeling uses a hierarchical framework to model the dynamics of multiple species simultaneously within a unified community analysis. Data from each species is of the same type, but observations and ecological parameters associated with each species are nested hierarchically within community-level components, allowing for variation among species within a common community (Dorazio & Royle 2005, Dorazio et al. 2006, Kéry & Royle 2009, Zipkin et al. 2009). Similar in its use of hierarchical statistical models, single-species integrated modeling combines data of different types within a unified analysis, specifying a separate observation process for each data type (Schaub & Abadi 2011, Dorazio et al. 2014, Zipkin & Saunders 2018). Because different data types inform at least one common ecological parameter, the ecological process (e.g., population abundance, population dynamics) can be jointly estimated. Each chapter in my dissertation is a separate development and application of a hierarchical multispecies or single-species integrated model to make inferences about different taxa and ecological systems.

Modeling applications presented in Chapters One and Two occur within the Masai Mara National Reserve, Kenya. Data used for these chapters was collected by the Michigan State University Mara Hyena Project, which has been monitoring spotted hyena (Crocuta crocuta) in the Reserve for over 30 years. During this long-term study, the Mara Hyena Project has surveyed the larger carnivore community in the Reserve using a variety of techniques (Green 2015). In Chapter One, I develop a multispecies hierarchical model to evaluate the effect of management alternatives and resulting human disturbances on a carnivore community within the Reserve. By integrating data for multiple carnivore species, I find that passive management and high levels of human disturbance have variable effects on carnivore abundance. African lions (Leo panthera) are adversely affected, while other species like spotted hyena and black-backed jackals (Canis mesomelas) benefit from disturbance. Variation in species' responses to disturbance could be driven by numerous processes, including direct mortality, competitive release, and habitat degradation. Although my goal was not to pinpoint explicit processes, my results indicate that passive management does not affect carnivores uniformly. By using a multispecies hierarchical model, I can both estimate abundance of rare and cryptic species that may have been discarded in analyses focused on single species, and characterize variation in the responses of multiple carnivores to human disturbance.

In Chapter Two, I develop an integrated distribution model that combines distance sampling (higher information content) and presence-only (lower information content) data to generate spatially explicit estimates of species abundance. I use a simulation study to assess the data requirements of each type and verify that the model returns unbiased and precise estimates. I apply this model to a case study on black-backed jackals in the Reserve and find that blackbacked jackals are thriving in more disturbed regions near high human density. Despite its lower

information content and potential biases, presence-only data can improve the precision of estimating ecological patterns when used in conjunction with structured data.

Chapter Three characterizes the regional distribution and dynamics of forest antelopes across Central and East Africa using detection-nondetection data from The Tropical Ecology Assessment and Monitoring Network. Recent methodological developments have expanded the capabilities of detection-nondetection data (e.g., camera trap data) to estimate both abundance and dynamics of wildlife populations (Rossman et al. 2016). I extended a single-species framework (dynamic *N*-occupancy model) for analyzing detection-nondetection data to simultaneously estimate demographic rates for multiple species in a community. This approach is particularly valuable in that it can characterize population dynamics for cryptic or rare species when detection-nondetection data may be the only available data type. Although many antelope species in this region are experiencing declines from habitat loss and poaching, my results suggest that the antelope populations were mostly stable and had low probabilities of extirpation over the next ten years.

In Chapter Four, I explore spring migration patterns of the eastern monarch butterfly (*Danaus plexippus*), which ranges from Central Mexico to the Midwestern United States and Southeast Ontario, Canada. While population dynamics of this species are affected by a range of seasonal factors, recent studies suggest that environmental conditions during spring migration may be particularly important (Zipkin et al. 2012, Saunders et al. 2016, 2018b, Zylstra et al. 2021). However, data deficiencies in spring limit our understanding of the processes affecting abundance during this critical stage of the migratory cycle. I developed a hierarchical model to integrate multiple types of unstructured opportunistic sightings of monarchs in the spring with structured data in early summer to characterize spatiotemporal variation in abundance on the

spring breeding grounds and assess the effects of spring climatic drivers. Using simulations, I assessed the ability of non-overlapping data to estimate abundance across spatiotemporal domains. When applying this integrated model to the eastern population of monarchs, I found that the effect of spring weather conditions on monarch abundance varies across its spring migration and early summer breeding grounds. These results demonstrate the potential for integrated models to circumvent data gaps during a species annual cycle.

The variety of systems, taxa, data types, and corresponding analyses covered in my dissertation illustrates the utility of data integration in population and community ecology. The frameworks I develop have a broad range of applications because of their ability to produce unbiased and precise estimates by accounting for variation between species or data types. Whether it is a multispecies model or an integrated model, the hierarchy of observation and ecological process models allows for a seamless statistical and mathematical connection between whichever data sources are available and the target parameters of interest (e.g., abundance). Data integration is a powerful tool that can harness the wide range of data now available to help overcome knowledge gaps in promoting conservation of wildlife in the face of ongoing anthropogenic threats.

CHAPTER 1: Multispecies hierarchical modeling reveals variable responses of African carnivores to management alternatives

Abstract

Carnivore communities face unprecedented threats from humans. Yet, management regimes have variable effects on carnivores, where species may persist or decline in response to direct or indirect changes to the ecosystem. Using a hierarchical multispecies modeling approach, I examined the effects of alternative management regimes (i.e., active vs. passive enforcement of regulations) on carnivore abundances and group sizes at both species and community levels in the Masai Mara National Reserve, Kenya. Alternative management regimes have created a dichotomy in ecosystem conditions within the Reserve, where active enforcement of regulations maintains low levels of human disturbance in the Mara Triangle and passive enforcement of regulations in the Talek region permits multiple forms of human disturbance. My results demonstrate that these alternative management regimes have variable effects on 11 observed carnivore species. As predicted, some species, such as African lions and bat-eared foxes, have higher population densities in the Mara Triangle, where regulations are actively enforced. Yet, other species, including black-backed jackals and spotted hyenas, have higher population densities in the Talek region where enforcement is passive. Multiple underlying mechanisms, including behavioral plasticity and competitive release, are likely causing higher black-backed jackals and spotted hyena densities in the disturbed Talek region. My multispecies modeling framework reveals that carnivores do not react to management regimes uniformly, shaping carnivore communities by differentially producing winning and losing species. Some carnivore species require active enforcement of regulations for effective conservation, while

others more readily adapt (and in some instances thrive in response) to lax management enforcement and resulting anthropogenic disturbance. Yet, high levels of human disturbance appear to be negatively affecting the majority of carnivores, with potential consequences that may permeate throughout the rest of the ecosystem. Community approaches to monitoring carnivores should be adopted as single species monitoring may overlook important intracommunity variability.

Material from: Farr, M. T., Green, D. S., Holekamp, K. E., Roloff, G. J., & Zipkin, E. F. (2019). Multispecies hierarchical modeling reveals variable responses of African carnivores to management alternatives. *Ecological Applications*, e01845.

For full text of this work, please go to: https://doi.org/10.1002/eap.1845

CHAPTER 2: Integrating distance sampling and presence-only data to estimate species abundance

Abstract

Integrated models combine multiple data types within a unified analysis to estimate species abundance and covariate effects. By sharing biological parameters, integrated models improve the accuracy and precision of estimates compared to separate analyses of individual data sets. I developed an integrated point process model to combine presence-only and distance sampling data for estimation of spatially explicit abundance patterns. Simulations across a range of parameter values demonstrate that my model can recover estimates of biological covariates, but parameter accuracy and precision varied with the quantity of each data type. I applied my model to a case study of black-backed jackals in the Masai Mara National Reserve, Kenya, to examine effects of spatially varying covariates on jackal abundance patterns. The model revealed that jackals were positively affected by anthropogenic disturbance on the landscape, with highest abundance estimated along the Reserve border near human activity. I found minimal effects of landscape cover, lion density, and distance to water source, suggesting that human use of the Reserve may be the biggest driver of jackal abundance patterns. My integrated model expands the scope of ecological inference by taking advantage of widely available presence-only data, while simultaneously leveraging richer, but typically limited, distance sampling data.

Material from: Farr, M. T., Green, D. S., Holekamp, K. E., & Zipkin, E. F. (2021). Integrating distance sampling and presence-only data to estimate abundance. *Ecology*, e03204.

For full text of this work, please go to: <u>https://doi.org/10.1002/ecy.3204</u>

CHAPTER 3: Quantifying the conservation status and trends of wildlife communities using detection-nondetection data

Abstract

Effective conservation often requires understanding species' abundance patterns and demographic rates across space and time. Ideally, such knowledge should be available for whole communities, as variations in individual species' demographic rates and responses to environmental factors provides critical information on optimal approaches to mitigate accelerating biodiversity loss. However, collecting the data necessary to simultaneously estimate abundance and demographic rates is often prohibitively time-intensive and expensive for communities of species. I developed a "multi-species dynamic N-occupancy model" that requires only detection-nondetection data (e.g., repeated presence-absence surveys) to estimate both the abundance and demographic rates of individual species over time as well as composite metrics of the whole community. Using data from a network of camera traps across tropical equatorial African national parks, I use my model to evaluate the statuses and trends of a forest-dwelling antelope community by estimating each species' local population abundance, rates of recruitment (i.e., reproduction and immigration), and apparent survival probabilities. I paired my model's results with a Bayesian population viability analysis to make mechanistic projections of species' population sizes and quasi-extinction rates within each park out to 2030. Both retrospective and prospective analyses indicate that the antelope community is fairly stable in this region (although 17% of populations/species-park combinations declined during the study period), with variations in population statuses and trends linked more closely to differences among national parks rather than differences in individual species' life histories. While certain assumptions must be met that may prevent implementation on highly mobile or socially grouping species, in many situations

my modeling framework can estimate abundance, demographic rates, and extinction risks for entire communities using commonly collected detection-nondetection data. Such frameworks can be valuable to conservation efforts seeking to understand and abate biodiversity loss.

Introduction

Knowledge of the population size and demographic rates (e.g., survival, recruitment) of species in wildlife communities is often needed to quantify, and ultimately address, threats leading to biodiversity loss (Conde et al. 2019). Traditionally, estimates of species abundance and demographic rates have relied upon 'marked data' (e.g. capture-recapture data), in which individuals are identified and followed via tags, bands, genotypes, and/or phenotypes (Pollock et al. 1990). The expensive and labor-intensive monitoring needed to generate marked data often preclude collection beyond single species. Though single-species analyses based on marked data can provide robust inferences, they are often restricted to common or charismatic species (Troudet et al. 2017). Methods to extrapolate single-species inferences (e.g., umbrella, keystone, indicator) to unmonitored community members may miss important variations among species (Cushman et al. 2010, Runge et al. 2019). Yet, accelerating biodiversity loss demands diversified approaches to monitor multiple species simultaneously and whole communities when possible (Nicholson & Possingham 2006, Ceballos et al. 2015, Zipkin et al. 2020). Community-wide assessments can provide unique information about species and their variable responses to environmental factors, including disturbance (REF).

While the need to scale up biodiversity assessments to community-levels is clear, the data to do so remain logistically challenging to obtain. Most community-wide data collection protocols rely on 'unmarked data' (e.g., presence-only, presence-absence, detectionnondetection, count). Unmarked data do not require identification or recapture of individuals and

thus can be collected more easily than marked data for community-wide monitoring programs. Arguably the most ubiquitous unmarked data type is detection-nondetection data in which observers record the presence or absence of a species at a given time and place (MacKenzie et al. 2017). A common approach to analyzing detection-nondetection data is occupancy modelling, which makes use of replicate sampling over short time frames to estimate species distributions and occurrence patterns while accounting for imperfect observation during sampling (MacKenzie et al. 2002). The advent of multi-species occupancy models (Dorazio & Royle 2005, Dorazio et al. 2006) has allowed for estimation of community occurrence processes and trends across space and time, driving discoveries in population biology, biodiversity loss, macrosystem processes, and community ecology among other areas (Kéry & Schaub 2012, MacKenzie et al. 2017, Devarajan et al. 2020). Yet without the ability to estimate demographic rates, traditional occupancy models have been restricted in their capacity to infer changes in population sizes and the underlying mechanisms driving trends. Recent development of the 'dynamic N-occupancy model' has expanded the use of unmarked data to jointly estimates population abundance and dynamics using detection-nondetection data for a single species (Rossman et al. 2016). This is done by decomposing changes in abundance into apparent survival and populations gains via recruitment (i.e., the combination of fecundity and immigration) using the biological process model developed by Dail and Madsen (2011) and capitalizing on the link between species detection probability and abundance developed by Royle and Nichols (2003).

Here, I expand the single-species dynamic *N*-occupancy modeling framework to a multispecies context, capable of estimating abundance and demographic rates for communities of related species using only detection-nondetection data. My 'multi-species dynamic *N*-occupancy model' can estimate community-level responses to environmental covariates while also capturing

species-specific variation in demographic rates and the effects of covariates. By linking individual responses via community-level distributions (Dorazio & Royle 2005, Dorazio et al. 2006), my modeling framework provides abundance and demographic rate estimates for rarer species that otherwise would be unidentifiable using a single-species approach (Kéry & Royle 2009, Zipkin et al. 2009, 2010). Scaling the single-species dynamic *N*-occupancy model to a multi-species context fills a needed gap in conservation ecology by creating a framework to explore the underlying mechanisms of community-wide population changes and biodiversity assessments.

I apply the multi-species dynamic *N*-occupancy model to a case study of forest dwelling antelope species in tropical equatorial Africa. Despite growing concern for human induced biodiversity loss in the tropics (Bradshaw et al. 2009), tropical communities contain a disproportionate amount of data gaps worldwide (Collen et al. 2008, Meyer et al. 2015). Of the available data collection methods for communities, unmarked methods are typically the most feasible and as a result, are widely used in tropical regions (O'Brien 2008, Tobler et al. 2008). My target antelope community provides an ideal case study to demonstrate the untapped capabilities of unmarked data collected via remote camera trap surveys. Camera trap surveys provide one of the only suitable methods to monitor remote, forest dwelling terrestrial mammals. Limited information on demography is available for most of the species included in my analysis, despite intense environmental pressures. To evaluate the status, trends, and potential trajectory for this community across Central and East Africa, I apply the multi-species dynamic Noccupancy model to estimate species-specific abundance and demographic rates. I then project each populations' quasi-extinction (i.e., local extinction/extirpation) probability (i.e., likelihood that a population size falls below a viable threshold where extinction/extirpation is inevitable)

using Bayesian population viability analysis (BPVA; Kéry & Schaub 2012, Saunders et al. 2018a, Saunders et al. 2021). Projecting species' viabilities under potential future scenarios provides a method to quantify uncertainty in species extirpation in response to environmental drivers and is a useful metric to evaluate potential conservation actions targeted at both species and communities. My multi-species dynamic *N*-occupancy model thus provides a multifaceted approach to address a knowledge gap (abundance and vital rate estimates, species-specific and community-level estimates, population projections, and quasi-extinction probabilities) that occurs persistently throughout wildlife communities.

Model

The multi-species dynamic *N*-occupancy model uses detection-nondetection data to estimate abundance and demographic rates at both species and community levels by combining the *N*-occupancy modeling framework (Rossman et al. 2016) with the multi-species occupancy modeling framework (Dorazio & Royle 2005). The model estimates latent biological processes (i.e., abundance, apparent survival, reproduction/immigration) for individual species and accounts for imperfect detection during data collection via an observation process component. Species' biological and observation processes are then linked with a hierarchical statistical structure (i.e. through shared distributions) to estimate community-level parameters in addition to the species-level parameters. This approach leverages information across species to improve precision of species-level parameters, especially for those species that were observed less frequently due to being either rare or elusive (Zipkin et al. 2009).

Observation process

To estimate latent abundance, $N_{i,j,t}$, of species *i* at a sampling site *j* during year *t*, detectionnondetection data, $y_{i,j,k,t}$, are collected during k = 1, 2, ... K sampling replicates. I assume that

species abundance at a site *j* is closed to changes within year *t*. Thus, the K > 1 sampling replicates within a year allow us to estimate the probability that species *i* was detected at site *j* during sampling replicate k ($y_{i,j,k,t} = 1$). I model the detection-nondetection data using a Bernoulli process:

$$y_{i,j,k,t} \sim Bernoulli(p_{i,j,k,t}),$$

where $p_{i,j,k,t}$ is the detection probability of species *i* at site *j* during replicate visit *k* in year *t*. A nondetection of species *i* can result from two separate processes: the species is truly absent at the site (i.e., latent abundance of species *i* at site *j* in year *t* is zero, $N_{i,j,t} = 0$) or because the species was present ($N_{i,j,t} > 0$) but no individuals were detected during sampling. Thus, $p_{i,j,k,t}$ can be defined as the probability that at least one of the $N_{i,j,t}$ individuals at the site was detected during the k^{th} sampling event (Rossman et al. 2016, Royle & Nichols 2003):

$$p_{i,j,k,t} = 1 - \left(1 - \theta_{i,j,k,t}\right)^{N_{i,j,t}}$$

where $\theta_{i,j,k,t}$ is the detection probability of an *individual* of species *i* at site *j* during replicate visit *k* in year *t*. If there are no individuals at the site ($N_{i,j,t} = 0$), then the detection probability collapses to zero. Likewise, as latent abundance, $N_{i,j,t}$, increases, the overall detection probability, $p_{i,j,k,t}$, of the species also increases because each individual has an independent probability of being detected, $\theta_{i,j,k,t}$. Covariates can be added to $\theta_{i,j,k,t}$ to account for variation in detection by species, site, replicate visit, and/or year using a logit-link function:

$$logit(\theta_{i,j,k,t}) = \alpha_{0,i} + \alpha_{v,i} \cdot x_{v,j,k,t}.$$

Here, $\alpha_{0,i}$ is the intercept for species *i*, or average detection probability of individuals on the logit scale while $\alpha_{v,i}$ is a vector (1,2, ..., *V*) of parameter effects ($\alpha_{1,i}, \alpha_{2,i}, ..., \alpha_{V,i}$) for species *i*

of corresponding standardized covariates $x_{\nu,j,k,t}$, which may change by sampling site, replicate visit, and/or year.

Biological process

The biological process model focuses on estimating abundance, $N_{i,j,t}$, for species *i* across all *j* sites in year *t* using the approach developed by Dail and Madsen (2011). I assume that species abundance changes between years (i.e., from t - 1 to *t*) through processes of survival and recruitment. In the first year for which data are available (i.e., t = 1), I estimate abundance, $N_{i,j,1}$, using a Poisson distribution:

$$N_{i,j,1} \sim Poisson(\lambda_{i,j}),$$

where $\lambda_{i,j}$ is the expected abundance of species *i* at site *j* in the first year of sampling (Dail & Madsen 2011). I can model heterogeneity in initial abundance by adding covariates using a log-link function:

$$log(\lambda_{i,j}) = \beta_{0,i} + \beta_{v,i} \cdot w_{v,j},$$

where $\beta_{0,i}$ is the intercept (i.e., average initial abundance on the log scale) for species *i* and $\beta_{v,i}$ is a vector (1,2, ..., *V*) of parameter effects ($\beta_{1,i}, \beta_{2,i}, ..., \beta_{V,i}$) for standardized covariates $w_{v,j}$.

In subsequent years (t > 1), I assume that changes to latent abundance of species i at site j occur via births-deaths and immigration-emigration processes (Dail & Madsen 2011) and are dependent on the population size during the previous year, t - 1. I break this process into two components: $S_{i,j,t-1}$, the number of individuals of species i that survive from year t - 1 to t and remain at site j, and $G_{i,j,t-1}$, the number of new individuals of species i that are gained to site j via recruitment (reproduction and/or immigration) from year t - 1 to t (Dail & Madsen 2011, Rossman et al. 2016). Thus, total abundance in year t > 1 is:

$$N_{i,j,t} = S_{i,j,t-1} + G_{i,j,t-1}.$$

I model the number of surviving individuals between t - 1 and t using a binomial distribution:

$$S_{i,j,t-1} \sim Binomial(N_{i,j,t-1}, \omega_{i,j,t-1}),$$

where $\omega_{i,j,t-1}$ is the apparent survival probability of each individual of species *i* at site *j* between t - 1 and *t*. Apparent survival is the product of true survival and site fidelity (i.e., the inverse of permanent emigration). I model the number of individuals of species *i* gained into the population at site *j* with a Poisson distribution:

$$G_{i,j,t-1} \sim Poisson(\gamma_{i,j,t-1}),$$

where $\gamma_{i,j,t-1}$ is the expected number of individuals gained at each site from reproduction and immigration. Provided there is sufficient available data, variation in apparent survival probability $(\omega_{i,j,t-1})$ and the expected number of individuals gained to a site $(\gamma_{i,j,t-1})$ can be modeled with covariates that change by site and/or year using a logit-link function and a log-link function, respectively.

Community component

To link the species models at a community level, I assume that the species-level parameters (i.e., intercept and effect parameters on either logit- or log-link scales) in both the observation and biological process models are random effects drawn from a parameter-specific, community-level distribution shared across all species (Dorazio & Royle 2005, Dorazio et al. 2006, Zipkin et al. 2009, Zipkin et al. 2010). For example, the intercept parameter for initial species abundance, $\beta_{0,i}$, is assumed to come from a normal distribution:

$$\beta_{0,i} \sim Normal(\mu_{\beta_0}, \sigma_{\beta_0}^2),$$

with a hyper-mean μ_{β_0} (i.e., representing average expected abundance across all species in the community) and hyper-variance $\sigma_{\beta_0}^2$ (i.e., representing the variation in initial expected abundance

across species). The random effects structure allows for information sharing across species within the community, improving parameter identifiability for species with low levels of data and increasing parameter precision for most other species (Kéry & Royle 2009, Zipkin et al. 2009). In addition to estimating species-level biological and observational process parameters, the hierarchical structure of the model also produces estimates of community-level mean (e.g., μ_{β_0}) and variance (e.g., $\sigma_{\beta_0}^2$), which provide useful metrics for summarizing community dynamics.

Application

Study area and data collection

My study encompasses a network of six national parks (Udzungwa Mountains National Park [UDZ], Tanzania; Volcanoes National Park [VNP] Rwanda; Bwindi Impenetrable Forest [BIF], Uganda; Nouabale-Ndoki National Park [NNNP], Republic of Congo; Korup National Park [KRP], Cameroon; and Nyungwe Forest National Park [NFNP], Rwanda) in the equatorial region of Central and East Africa between 2009 and 2019 (Table 3.1). Equatorial Africa has a tropical climate (i.e., wet and humid) and pronounced wet and dry seasons with annual rainfall in parks ranging from 1500 to over 5000 mm (Kingdon 2015, O'Brien et al. 2020). Vegetation in the region is determined by rainfall and elevation with the predominant habitats being lowland and montane rainforest that extend into forest mosaics further from the equator (Kingdon 2015). Elevation varies across parks with the lowest elevation near sea-level in KRP (92 - 463 m) and the highest elevation in VNP (2509 - 3884 m; O'Brien et al. 2020). Parks range in area from 342 - 4000 km² and are unfenced with varying human densities (0.5 - 386 humans/km²) along the edges.

I focused on a metacommunity of forest dwelling antelopes across the network of parks (Table 3.1; Johnston & Anthony 2012). I included 12 closely-related species in my analysis: suni (Nesotragus moschatus), bushbuck (Tragelaphus scriptus), sitatunga (Tragelaphus spekii), and 9 species of duikers (Subfamily Cephalophinae; listed in Table 3.1). Predominately a browsing community, diet items for these species include fruit, flowers, and foliage with some species also eating grass (Kingdon 2015). This antelope metacommunity is geographically distributed across Sub-Saharan Africa, living in multiple forest types from lowland to alpine (Kingdon 2015), with each species range limited to only a subset of the parks in my study (Table 3.1). Forest fragmentation and resulting refugia during Pleistocene climate change is hypothesized to be responsible for the modern distribution of this community (Johnston & Anthony 2012). In the Anthropocene, humans are likely having the largest influence over the distribution and abundance patterns of this community. Antelopes are facing pervasive anthropogenic pressures including deforestation and poaching (Newing 2001), and the health, productivity, and persistence of their tropical rainforest habitat is threatened by climate change (Phillips et al. 2009, Sullivan et al. 2020). Recent assessments of this antelope community conflict on species estimated stability (O'Brien et al. 2020), and minimal information on species abundance and demographic rates have prevented conclusive inference on the current status and future trajectory of this vulnerable community.

To estimate community wide abundance and demographic rates, I collated data from the Tropical Ecology Assessment and Monitoring (TEAM) network. The TEAM network was developed for monitoring tropical species worldwide using a standardized camera trapping protocol (see TEAM Network 2011 a, b for detailed protocols). Camera traps were deployed in each park, and images were taken whenever an animal triggered a camera's motion sensor. Each

park contained camera traps across 60 sites (except NFNP, which had 97 sites) that were sampled once per year for 30 consecutive days. The available years of data varied for parks, ranging from three for VNP to 11 for UDZ (Table 3.1). In post-processing of images, individual species were identified and aggregated into detection-nondetection histories based on 5-day sampling periods (replicates) for a maximum of six replicates per year. In cases where camera traps malfunctioned, we used the amount of time that the camera was functional as a covariate within my observation model to account for variation in detection due to decreased effort. Species were only evaluated at parks within their range; thus, I estimated the abundance and dynamics of 23 populations (i.e., species-park combinations; Table 3.1).

Data analysis

I described the detection probability of individuals, $\theta_{i,j,r,k,t}$, of each species *i* at site *j* in park *r* during replicate *k* in year *t* using a logit-link:

$$logit(\theta_{i,j,r,k,t}) = \alpha_{0,i} + \alpha_1 \cdot days_{j,r,k,t},$$

where $\alpha_{0,i}$ is the species-specific intercept or detection probability of an individual during a sampling replicate when cameras were functional for the average amount of time (4.5 out of 5 days). I added the covariate $days_{j,r,k,t}$ (standardized to a mean of 0 and standard deviation of 1) to incorporate variation in the amount of time that a camera was functional at site *j* at park *r* during replicate *k* in year *t*, an effect that did not vary by species.

For the biological process model, I used a log-link function to model species' initial abundances across sites within parks:

$$log(\lambda_{i,r}) = \lambda_{0,i} + \varepsilon_{\lambda,r}.$$

The intercept, $\lambda_{0,i}$, varies by species to account for differences in baseline abundance and a parklevel random effect ($\varepsilon_{\lambda,r}$) captures variation between parks. I similarly modeled species annual survival probabilities, $\omega_{i,r}$, as:

$$logit(\omega_{i,r}) = \omega_{0,i} + \varepsilon_{\omega,r}$$

where $\omega_{0,i}$ is the baseline species-specific survival probability and $\varepsilon_{\omega,r}$ is a park-level random effect. During model development, I attempted to include environmental covariates in the models of initial abundance and the demographic parameters but large variations in covariate values between parks prevented meaningful inference and led to overly complex statistical structures (Appendix A). Further, many of the site-level (e.g., elevation, temperature, distance to edge) and park-level covariates were co-linear such that it was difficult to determine the important factors influencing species. As such, I settled on including random effects in my estimates of initial abundance and survival to account for species-specific and community-level comparisons between parks without subscribing improper mechanism to estimated differences. I modeled the expected number of individuals gained to a site, γ_i as:

$$log(\gamma_i) = \gamma_{0,i}$$

using only a species-specific intercept, $\gamma_{0,i}$, as I did not expect residual variation in gains at the park-level. I explored adding a park-level random effect but removed it due to poor convergence and lack of biological support. There is no evidence for differences in birth rates or sex ratios between parks, and immigration into sites is largely dictated by species' territorial behavior (such that there is high site fidelity across all species, Kingdon 2015), which I assumed did not vary by park.

To link the species models, I drew each species-specific parameter $(\alpha_{0,i}, \lambda_{0,i}, \omega_{0,i}, \gamma_{0,i})$ from separate community-level normal distributions with corresponding hyper-means and hypervariances. I used the mean community-level estimate for apparent annual survival across all parks (μ_{ω_0}) in combination with the park-level random effects on survival ($\varepsilon_{\omega,r}$) to derive community-level apparent annual survival for each park (i.e., $\mu_{\omega_0} + \varepsilon_{\omega,r}$). I also derived an index of annual population-level (i.e., species-park combination) abundance by summing across sites surveyed in a year. I report average abundance per site, $\widehat{N}_{i,r,t} = \frac{\sum_{j=1}^{J_{r,t}} N_{i,j,r,t}}{J_{r,t}}$, rather than total abundance within a park to account for variations in sampling effort (i.e., number of sites surveyed per year, $J_{r,t}$; Table 3.1). I calculated annual population growth rates for each population as: $\widehat{N}_{i,r,t}/\widehat{N}_{i,r,t-1}$, and summarized across years by taking the geometric mean.

I estimated parameters using a Bayesian framework via Nimble and R (de Valpine et al. 2017, R Core Team 2020 Version 4.0.2). All hyper-parameters and other fixed effect parameters were given vague priors (Appendix B). I ran 5 MCMC chains each for 100,000 iterations with a burn-in of 75,000 and thinning of 25 providing 5,000 samples from the posterior distribution for each parameter. I assessed convergence using the Gelman Rubin diagnostic (Rhat < 1.2) in addition to visually examining the chains.

Bayesian population viability analysis

I projected the population of each species to 2030 using the estimated parameter values from my multi-species dynamic *N*-occupancy model and calculated quasi-extinction probabilities for each population using a BPVA. I selected 2030 as the target year as this projection timeframe captures the next window for the Convention on Biological Diversity's updated Strategic Plan for Biodiversity (2021 - 2030; CBD 2020). I used a 50% decline in population abundance from 2020 to 2030 as the threshold for quasi-extinction as this is the criteria for International Union for Conservation of Nature's endangered species status (IUCN 2021): *Threshold* = $\hat{N}_{i,r,t=2020}$. 0.5 where $\hat{N}_{i,r,t=2020}$ is the posterior mean of park-level abundance in 2020. I calculate the quasiextinction probability as the proportion of MCMC iterations where the population (i.e., speciespark combination) abundance in 2030 was below the threshold: $QSE_{i,r} = Pr(\hat{N}_{i,r,t=2030} < Threshold)$ in which I defined extinction to be inevitable with probability > 0.95 (i.e., no overlap of the $\hat{N}_{i,r,t=2030}$ 95% CI with the threshold). I report population projections and quasiextinction probabilities only for populations with at least five years of data as estimates with a smaller temporal extent may be unreliable (Zipkin et al. 2014). I calculated the population projection estimates for the BPVA from a separate model run than the retrospective analysis (Appendix B), again using both Nimble and R (de Valpine et al. 2017, R Core Team 2020 Version 4.0.2) with identical priors and MCMC settings.

Results

Population abundance and growth

While estimates of population abundance and growth reveal that most of the antelope community was fairly stable over the study period (Figure 3.1, Appendix C Table C.1, C.2), four of the 23 populations (~17%) declined in abundance over the 11-year time frame (95% credible interval [CI] for growth rates were less than one). The four populations that had negative growth rates over the study period are: *C. callipygus*, *C. dorsalis*, *P. monticola*, *C. silvicultor*, all within the Nouabale-Ndoki National Park (NNNP). Eleven populations had stable population growth rates (95% CI for growth rates overlapped one) and eight populations increased in abundance (i.e., 95% CI above one).

Demographic rates

Though mean community-level annual apparent survival (μ_{ω_0} ; Figure 3.2A) across parks was estimated as 0.72 (CI 0.28, 0.96), there was large variation across species ($\sigma_{\omega_0} = 0.85$ [CI

0.48, 1.52], logit-scale, Appendix C Table C3) and parks ($\sigma_{\varepsilon_{r,\omega}} = 2.72$ [CI 1.14, 6.50], logit scale, Appendix C Table C.4). The larger variation in survival across parks than across species (i.e., $\sigma_{\varepsilon_{r,\omega}} > \sigma_{\omega_0}$) suggests that environmental and/or anthropogenic factors at the park-level may be contributing more to annual survival than species life history processes or species-specific responses to particular environmental factors within a park. Park-level estimates of survival (Figure 3.2A) for VNP (0.99 [CI 0.97, 0.99]), BIF (0.81 [CI 0.69, 0.91]), NFNP (0.77 [CI 0.56, 0.92]), and UDZ (0.77 [CI 0.56, 0.89]) were higher than the community-level average, likely attributable to active and effective management enforcement (Oberosler et al 2020a, 2020b, O'Brien et al. 2020). Despite estimates of stable population growth rates of its three duiker populations (Figure 3.1), low survival within KRP (0.31 [CI 0.11, 0.52]; Figure 3.2A) may be attributable to hunting pressure as local extinction of other antelope species has likely already occurred there (Viquerat et al. 2012, O'Brien et al. 2020). However, C. silvicultor is at the edge of its geographic range in KRP (Kingdon 2015), which may also be contributing to its low survival in that park (0.54 [CI 0.27, 0.72]). Across species, C. dorsalis had the lowest average annual apparent survival probability (0.56 [CI 0.12, 0.92]) while C. silvicultor had the highest (0.90 [CI 0.51, 0.99]; Figure 3.2A). The mean number of individuals gained per species annually at sites across parks ($\mu_{\gamma 0}$) was 0.24 (CI 0.10 - 0.56) with *P. monticola* having the highest estimated site-level recruitment (i.e., sum of immigration and fecundity; (1.48 [CI 1.11 - 1.87]) and *T. spekii* having the lowest (0.02 [CI 0.01 - 0.05]) (Figure 3.2B).

Population viability

None of the 18 populations with sufficient time-series for projections (i.e., four of six parks had at least five years of data) reached a quasi-extinction threshold severe enough to trigger IUCN endangered species status (i.e., 50% loss with nonoverlapping 95% CI), with all

populations having quasi-extinction probabilities of less than 0.25 (Figure 3.1; Appendix C Table C.5). Of the four populations within NNNP that were estimated to be declining, *C. silvicultor* and *T. spekii* had the highest quasi-extinction probabilities at 0.23 and 0.24, respectively (Appendix C Table C.5). While all local populations were at a low risk of extirpation over the next decade (assuming that current conditions hold), several populations in NNNP and BIF were projected to decline from 2021 to 2030 (bottom panels of Figure 3.3), suggesting that close monitoring is warranted and that management interventions may be needed to ensure persistence in the longer term.

Discussion

Achieving conservation targets for biodiversity requires quantifiable measures of the status, trends, and dynamics of populations at both species and community levels (Nicholson & Possingham 2006). Using an extensive continental-scale camera trapping network in Central and East Africa, my multi-species dynamic *N*-occupancy model estimated (1) stable population growth of the antelope community across national parks during my study period (2009-2019), (2) varying abundance and vital rates across community members and between populations of the same species, (3) differences in annual apparent survival of community members between national parks, and (4) low probabilities of quasi-extinction for the majority of populations over the next decade. These results resemble patterns seen in protected areas across equatorial Africa where effective conservation of species communities depend on the enforcement of regulations to prevent or mitigate disturbance (Farr et al. 2019, Oberosler et al 2020a, 2020b, O'Brien et al. 2020). Here, I demonstrate the upper potential of detection-nondetection data to assess communities of species by developing an analytical framework to estimate population abundance and demographic rates of each community member within the parks they inhabit. The potential
of my modeling framework to support biodiversity conservation is underpinned by detectionnondetection being the most ubiquitous data type for monitoring multiple species and the ability to quantify demographic rates and extinction risk for multiple species simultaneously.

Contrary to the perception that unmarked data provides limited information relative to marked data, I was able to estimate community-wide population abundance and demographic rates by combining multi-species models (Dorazio & Royle 2005, Dorazio et al. 2006) with the single-species dynamic *N*-occupancy (Rossman et al. 2016). While there is great potential for this approach, multiple limitations may restrict my model's application to certain data scenarios and species. Dynamic unmarked models need at least 3-5 years (i.e., time periods) of data for parameters to be identifiable but simulation results reveal that these models typically do not perform well with less than 5-10 years of sampling (Dail & Madsen 2011, Zipkin et al. 2014). Precision of estimates also depend on the number of sites sampled where too few survey locations can even lead to inaccuracies (Rossman et al. 2016). In my application of the multispecies dynamic N-occupancy model, limited time series for certain populations prevented estimation of site-level variation using covariates or multi-scale processes across parks (Appendix A). Additionally, I was unable to propagate environmental stochasticity into my BPVA (such as in Saunders et al. 2018a) because random temporal effects were unidentifiable, likely leading to underestimation of uncertainty in my future projections. For some species life histories, the basic structure of the multi-species dynamic N-occupancy model may not be feasible. For example, highly mobile and non-territorial species may violate the geographic closure assumption (i.e., no immigration or emigration out of the site during replicate visits). Other assumptions related to demographic closure (i.e., no births or deaths between replicate visits within a year), independence of sites, independence of individual detections, and species

identification may lead to the necessary exclusion of some taxonomic groups (Royle & Nichols 2003, Devarajan et al. 2020). Difficulty with parameter estimation may also occur for species that are very rarely detected. Though the multi-species framework theoretically allows for estimation of rare and elusive species (Zipkin et al. 2009), few observations of a species can lead to imprecise or unidentifiable estimates of species-specific parameters, especially in the context of a model that aims to estimate demography as well as abundance.

My case study provided an initial demonstration of the multi-species dynamic N-occupancy model and is only a prelude to this framework's full potential for conservation biology. Further applications of this approach can help explain variation in species statuses and trends by linking environmental drivers (covariates) to demographic rates. Such approaches can help elucidate the mechanistic reasons behind biodiversity loss and species declines by partitioning the effects of specific environmental factors on species' survival and/or recruitment. Coupled with a Bayesian population viability analysis, extinction risk can be linked to future environmental scenarios related to projected climate changes and management interventions (Saunders et al. 2018a, Saunders et al. 2021). Evaluating extinction risks at community levels can provide a useful metric for potential biodiversity loss while explicitly incorporating uncertainty. Quantifying extinction uncertainty across species will improve the utility of global biodiversity indices for setting conservation targets (Watermeyer et al. 2020). The comparatively lower cost of collecting unmarked data, such as detection-nondetection data, has made these data types globally available for conservation assessments. Thus, models that can use such data to estimate not only metrics of population trends, but also demographic rates, are exceptionally valuable to inform global conservation priorities and ultimately, the effectiveness of conservation interventions.

Table 3.1

National parks and basic data information used in my multi-species dynamic *N*-occupancy analysis of the antelope community including the number of years sampled, species observed, number of sampling sites, and the number of recorded presences across species. Species are numerically identified as follows: (1) *Cephalophus callipygus* (2) *Cephalophus dorsalis* (3) *Cephalophus harveyi* (4) *Cephalophus leucogaster* (5) *Philantomba monticola* (6) *Nesotragus moschatus* (7) *Cephalophus nigrifrons* (8) *Cephalophus ogilbyi* (9) *Tragelaphus scriptus* (10) *Cephalophus spadix* (11) *Tragelaphus spekii* (12) *Cephalophus silvicultor*.

| Park | Country | Years | Species | Sites | Presences |
|-----------------------------|----------------------|---------------------------------|-----------------------------|-------|-----------|
| Korup NP (KRP) | Cameroon | 2011 - 2015 (<i>n</i> = 5) | 5,8,12 (<i>n</i> = 3) | 60 | 1100 |
| Nouabale-Ndoki NP (NNNP) | Republic of Congo | 2010 - 2016 (<i>n</i> = 7) | 1,2,4,5,7,11,12 (n = 7) | 60 | 5502 |
| Udzungwa NP (UDZ) | Tanzania | 2009 - 2019 (<i>n</i> = 11) | 3,6,9,10 (<i>n</i> = 4) | 60 | 2970 |
| Bwindi NP (BIF) | Uganda | 2010 - 2017 (<i>n</i> = 8) | 7,9,11,12 (n = 4) | 60 | 1551 |
| Nyungwe Forest NP (NFNP) | Rwanda | 2014 - 2017 (<i>n</i> = 4) | 7,9,12 (<i>n</i> = 3) | 97 | 284 |
| Volcanoes NP (VNP) | Rwanda | 2014 - 2016 (<i>n</i> = 3) | 7,9 (<i>n</i> = 2) | 60 | 1432 |

Figure 3.1

Population density (i.e., abundance per site) of each species at each park across all years that a park was surveyed. Points show the mean abundance while the shaded area is the 95% credible interval around the mean. Estimates are only shown for parks and years during which sampling occurred and for species that were observed in each park. Each panel shows the trends for a specific park (labeled with the parks' abbreviations) and species are each represented with a unique color.



Figure 3.2

(A) Annual apparent survival probabilities for each species in the community across the network of parks ($\omega_{0,i}$), mean survival probability at the community-level (μ_{ω_0} ; black coloring), and mean park-level survival probabilities for the entire community of species that reside in the park ($\mu_{\omega_0} + \varepsilon_{\omega,r}$; unique coloring). (B) Species-specific and community-level annual number of individuals gained ($\gamma_{0,i}, \mu_{\gamma_0}$). (A,B) Boxplots show the mean, 50% credible intervals, and 95% credible intervals. The dashed-line separates the species-specific estimates from the communityand park-level estimates.



Figure 3.3

Retrospective population density (i.e., abundance per site during study period) and projected population density (up to 2030) for each species within the four parks that were sampled for at least five years. Lines show mean estimated (solid) and projected (dashed) density while the shaded areas depict the 95% credible intervals. Estimates are shown from the initial sampling year for a given park until 2030 (with missing years interpolated for parks when data were unavailable). Each panel shows a different park, labeled with the parks' abbreviations. Species are depicted with a unique color.

Figure 3.3 (cont'd)



CHAPTER 4: Overcoming data gaps using integrated modeling to estimate population abundance of a migratory species

Abstract

Effects of environmental or anthropogenic factors on migratory populations can carry over from one season and/or spatial location to another. However, identifying the mechanistic processes driving cross-seasonal effects on migratory species is difficult due to extensive spatiotemporal gaps in monitoring data. The continuous collection of opportunistic data by volunteer-based networks provides a solution to addressing these data gaps within migratory cycles. To estimate population abundance at broad spatiotemporal scales, I develop an integrated model that leverages unstructured data during seasons and spatial locations when structured monitoring is absent. I validate my approach through simulation and then apply the framework to estimate the population density of monarch butterflies throughout their spring migratory pathway. While climate conditions during the spring breeding seasons have been identified as a key driver of annual monarch population dynamics, this portion of the annual migration lacks structured survey data. Thus, I use my integrated model to combine all available data sources and reveal that vegetation and temperature conditions in eastern Texas affected the density of monarchs during their spring migration between 2016 and 2018. Not only do these results corroborate the hypothesized importance of spring migration conditions for monarchs, but they reveal a dynamic relationship between spring conditions and their effect on fine-scale population density. More broadly, this framework highlights the power of data integration for overcoming spatiotemporal gaps during the annual cycle of migratory species.

Introduction

During the annual cycle of a species, their population may occupy only a subset of their range, which may vary across seasons (e.g., breeding, nonbreeding, migrating). Many threats to species (e.g., climate change, habitat loss) occur across broad spatiotemporal scales, but crossseasonal connections or carry-over effects, in which the cause of an observed effect at one season is linked to another (Norris 2005, Harrison et al. 2011), can complicate species conservation (Martin et al. 2007). Migratory species are among the most threatened species groups because of the challenges in enacting conservation at a given place and time when cross-seasonal connections regularly manifest between breeding grounds, nonbreeding grounds, and/or migratory pathways (Webster et al. 2002, Saunders et al. 2021, Zylstra et al. 2021). Parsing cross-seasonal connections into mechanistic processes requires data that span the spatiotemporal extent of a species annual cycle; however, structured data, which come from standardized monitoring programs, are often limited to spatiotemporal subsets of the migratory cycle (e.g., breeding grounds, non-breeding grounds) because of logistical constraints and elusive or cryptic behaviors that hamper data collection (Marra et al. 2015, Rushing et al. 2016). The continuous collection of unstructured data via volunteer-based networks (e.g., iNaturalist, eBird, GBIF, Map of Life) provides a means to fill in gaps from traditional data collection methods (Jetz et al 2012, Chandler et al. 2017). However, biases in unstructured data can impede the ability to develop meaningful inferences due to the lack of survey design (Dorazio 2014, Fithian et al. 2015).

Integrated modeling, in which multiple data sources and types are analyzed in a single, unified framework, can be a useful tool to address these issues and model dynamics of migratory species (Pacifici et al. 2017, Zipkin & Saunders 2018, Fletcher et al. 2019, Miller et al. 2019, Isaac et al. 2020). This framework can overcome the limitations of structured (i.e., limited extent

and quantity) and unstructured (i.e., limited information content) data types to estimate abundance (Dorazio 2014, Farr et al. 2021). Integrated modeling has great potential for modeling migratory species abundance across their full annual cycle (Hostetler et al. 2015). Beyond estimating abundance, this framework can resolve discrepancies in estimated population trends from separate monitoring programs by linking seasonal dynamics within a species annual cycle (Saunders et al. 2019a) or be used to estimate 'missing' parameters that would be inestimable from independent analysis of each data source (Saunders et al. 2018a, Zipkin & Saunders 2018). Integrated models have also been used to evaluate the effects of environmental factors hypothesized to drive population trends across a species full annual cycle (Rushing et al. 2017, Rushing et al. 2021, Zylstra et al. 2021).

Here, I develop an integrated model for a migratory species when both spatial and temporal data gaps exist within the annual cycle. I develop a point process model to describe spatiotemporal variation in abundance and use single-visit count data to identify baseline abundance. I then include spatiotemporal covariates and integrate opportunistic presence-only data to extend the analysis and estimate abundance during temporal portions of the annual cycle when single-visit count data are unavailable. I demonstrate the utility of this framework using a simulation study where abundance is estimated across space and time, despite limited structured sampling. I then apply our modeling framework to a case study of monarch butterflies, in which data integration over a species annual cycle may be the only means to address a pressing ecological question.

The monarch butterfly has declined across its eastern North American range since the mid-1990s (Brower et al. 2012, Agrawal & Inamine 2018), and unfavorable climatic conditions are projected to further limit monarch population abundance (Zylstra et al. 2021). The eastern

North American population of monarchs has a unique, multi-generational migratory cycle. At the beginning of the year (January - February), adult monarchs that emerged during the previous year are residing in their overwintering grounds in central Mexico. In spring (March - April), monarchs migrate north to eastern Texas and Oklahoma where they locate milkweed host plants to deposit eggs. Individuals that successfully metamorphose become the first generation of the year and continue the migration north to the summer breeding grounds across the Midwestern U.S. and Ontario, Canada. Two to three more generations are produced during the summer (May - August) until the final generation of the year enters reproductive diapause and migrates south in the fall (September - October), back to the same overwintering grounds in central Mexico (arriving by mid-December).

Multiple hypotheses have been proposed to explain the decline of monarchs in eastern North America (Thogmartin et al. 2017, Agrawal & Inamine 2018). Recent retrospective analyses demonstrate the importance of weather conditions during the spring breeding season and migratory pathway (Zipkin et al. 2012, Saunders et al. 2016, 2018b, Zylstra et al. 2021), highlighting a critical spatiotemporal window where population growth could either be inhibited or bolstered. Spring temperature and precipitation affect the recruitment of monarchs by impacting development and host plant resources (Zalucki 1982). Despite the importance of the spring migratory pathway for monarchs, structured survey data are only available in winter (nonbreeding) and summer breeding portions of their full annual cycle. I use my spatiotemporal integrated model to estimate monarch abundance and associated effects of environmental conditions during their spring migration, allowing for inference during an understudied stage of their migratory cycle.

The Data

Structured data are those resulting from surveys that were designed with the parameter of interest (i.e., abundance) and target species in mind, while unstructured data are collected without a planned survey design or monitoring protocol. Structured data typically have high information content as they minimize sampling biases and are collected in ways that allow for the explicit estimation of observation error. However, significant logistical constraints often limit the quantity of structured data that can be collected. Without a planned design, unstructured data can be obtained more easily at large scales and thus tend to be of higher volume. However, unstructured data contain sampling biases and involve data collection methods that make it difficult to account for observation errors in analyses, leading to lower information content than structured data.

Many survey designs (i.e., protocols to sample space and time) exist for structured count data. For simplicity, I consider single-visit surveys where the numbers of individuals are counted once within multiple spatial units (i.e., sites). Without replicate samples, single-visit surveys cannot account for imperfect detection leading to estimates of relative rather than absolute abundance (Knape & Korner-Nievergelt 2016). Presence-only data are the outcome of opportunistic, unstructured sampling where presences of the target species are recorded; however, the absence of the species at a given location is not recorded. Without this information, presence-only data cannot be used to estimate abundance without integration of structured data (Dorazio 2014, Farr et al. 2021). Presence-only data are also subject to sampling biases including nonrandom sampling and imperfect detection.

The Model

The purpose of our integrated model is to estimate population abundance during one or more seasons (henceforth notated as stage) of the annual migratory cycle when structured data

are limited. To estimate abundance, I use a point process to describe the distribution of individuals within a spatial domain S_t (e.g., breeding range, non-breeding range, migratory pathways) occupied by the species during each stage t = 1, ..., T within its annual cycle. As the spatial domain of the population may change over its annual cycle, I allow S_t to vary at each stage. Point process models estimate the density of individuals at stage t and point location s $(s \in S_t)$ using an intensity (i.e., number of points per unit area) function, $\lambda_t(s)$, where t is discrete and s is continuous. Population abundance at a given stage, N_t , is described as a random variable arising from a Poisson process, $N_t \sim Pois(\Lambda_t)$. Expected abundance, Λ_t , is calculated as the integral over the continuous domain S_t : $\Lambda_t = \int_{S_t} \lambda_t(s) ds$.

Spatial heterogeneity in population density

To identify an estimate of abundance, I begin by modeling population density during a stage of the annual cycle where both structured and unstructured data exist; I refer to this as baseline population density and assume for simplicity, that this occurs at t = 1. As population density varies over S_t due to environmental or anthropogenic factors, I can model baseline population density, $\lambda_{t=1}(s)$, with spatial covariates using an inhomogeneous point process, specifically a log-Gaussian Cox process (Møller et al. 1998):

$$\log(\lambda_{t=1}(s)) = \log(\lambda_0) + \beta' X_{t=1}(s) + \omega(s), \qquad \text{eqn 1}$$

where $X_{t=1}(s)$ is a vector of covariates at location *s* during the baseline stage, β' is the vector of corresponding coefficients, and λ_0 is the baseline population density when covariates *X* are at their mean values (standardized to mean of zero and standard deviation of 1). Both $\lambda_{t=1}(s)$ and λ_0 are on the log scale as their support ranges from 0 to ∞ while the support of covariates are \mathbb{R} .

Spatial heterogeneity that is unaccounted for by covariates may be captured with a spatial random effect, $\omega(s)$. A Gaussian random field (GRF) is used to describe $\omega(s)$ within the log-

Gaussian Cox process (Møller et al. 1998). The GRF is described using a multivariate normal distribution, $\omega \sim MVN(0, \tau Q)$ and implemented using a stochastic partial differential equation approximation via a triangulated spatial mesh (which may vary by stage) of $k = 1, ..., K_t$ nodes across S_t (Simpson et al. 2016, Krainski et al. 2018). Here, τ is the precision hyperparameter, and Q is a matrix that describes the correlation structure across S_t . I specify correlations between locations in S_t using a Matérn correlation function with scale parameter κ and fixed/constant smoothness parameter v (Simpson et al. 2016, Krainski et al. 2018), which indicates that correlation among $\lambda_t(s)$ values is driven by spatial proximity, with locations near one another having similar intensities.

Temporal change in population density

During subsequent stages of the annual cycle, t = 2, ..., T, population density will change because of demographic processes (i.e., mortality, recruitment, immigration, emigration) that are subject to spatiotemporally varying external factors. I extend eqn 1 to implicitly account for population change:

$$\log(\lambda_t(s)) = \log(\lambda_0) + \beta' X_t(s) + \delta_t + \omega(s).$$
 eqn 2

Here $X_t(s)$ is a vector of covariates for each subsequent stage. Population change not captured by fluctuations in covariate values is captured by δ_t , which is the change (i.e., combined gains and losses) in mean population density between stages across S_t .

Structured count data

I assume counts, C_{tj} , occur at set of sites (i.e., spatial locations) j = 1, ..., J with an associated area D_j where $D_{1,...,J} \subset S_t$. Count data are also described with a log-Gaussian Cox process: $C_{tj} \sim Pois\left(\int_{D_j} \lambda_t(s) ds\right)$. Count data and most structured data types cannot be used as direct measures of true latent abundance as there are naturally errors within the data collection

process. When counts are replicated, an observation process model can be included to account for imperfect detection (Royle 2004). However, for simplicity and similar to many structured monitoring programs (e.g., Breeding Bird Survey, Christmas Bird Count), I do not explicitly account for observation error, and use a single-visit count model which assumes constant detection probability over space and time to estimate relative abundance. As many count survey designs aggregate individual observations within a site (i.e., they do not include point location, s_i , for each individual *i*), a change-of-support problem must be addressed (Pacifici et al. 2019). Given that the expected sum of Poisson random variables is equivalent to the sum of their means, I specify an area offset and approximate the integral over each site j as: $D_j \bar{\lambda}_{tj} \approx \int_{D_j} \lambda_t(s) ds$. This assumes a homogenous process within each site (i.e., $\bar{\lambda}_{tj}$ is the mean population density for D_j) and a linear relationship between area and intensity (i.e., no density dependence). Counts at each stage t and site j can be modeled as: $C_{tj} \sim Pois(D_j \bar{\lambda}_{tj})$, where $\log(\bar{\lambda}_{tj}) = \log(\lambda_0) + \log(\bar{\lambda}_{tj})$ $\beta' X_{tj} + \delta_t + \omega_j$. Covariates, X_{tj} , are summarized across each D_j , and a projection matrix, A_{kj} , is needed to interpolate the random effects, ω_i , from each node k in the spatial mesh to each site j (Krainski et al. 2018). The log-likelihood, $\ell_c(\lambda_0, \beta, \delta_t, \omega)$, of the count data follows the classical Poisson log-likelihood.

Unstructured presence-only data

Presence-only data resulting from opportunistic sampling are modeled using a thinned log-Gaussian Cox process, $Y_t \sim Pois\left(\int_{S_t} \lambda_t(s) p_t(s) ds\right)$, where Y_t is the number of presence-only observations for stage t within spatial domain S_t (Dorazio 2014, Farr et al. 2021). The thinning function, $p_t(s)$, represents the observation process that relates true population density at location s, $\lambda_t(s)$, to the presence-only data (i.e., the number of detections at location s). Imperfect detection and sampling bias can be modeled within a linear predictor of $p_t(s)$ using a logit-link function, intercept (p_0) , covariates $(W_t(s))$ and their effects (α) that capture each of these observation processes (Dorazio 2014, Farr et al. 2021).

The log-likelihood function for presence-only data is:

$$\ell_{PO}(\lambda_{0},\beta,\delta_{t},\omega,p_{0},\alpha) = \sum_{t=1}^{T} \left(-\int_{S_{t}} \frac{\lambda_{0} \cdot \exp\left(\beta' X_{t}(s) + \delta_{t} + \omega(s) + p_{0} + \alpha' W_{t}(s)\right)}{\exp(p_{0} + \alpha' W_{t}(s)) + 1} ds + \sum_{i=1}^{Y_{t}} \left(\log(\lambda_{0}) + \delta_{t} + \beta' X_{t}(s_{i}) + \omega(s_{i}) + p_{0} + \alpha' W_{t}(s_{i}) - \log\left(\exp\left(p_{0} + \alpha' W_{t}(s_{i})\right) + 1\right) \right) \right) ds$$

Within a single stage t, the log-likelihood can be decomposed in two parts. The first is the integral across domain S_t for the thinned process. To approximate the integral over S_t , the thinned process is estimated at each of the mesh nodes k specified above, and a weighted sum is calculated where the nodes are provided a spatial (i.e., area) weight w_k :

$$\int_{S_t} \lambda(s) p(s) ds \approx \sum_{k=1}^K w_k \lambda_k p_k.$$
eqn 4

The second part of the log-likelihood is the contribution of all the presence-only observations $i = 1 \dots, Y_t$ at stage t for each of the locations s_i . A projection matrix, A_{ki} is also needed to interpolate the random effects ω_i from each node k to each presence-only observation i at location s_i . The covariates, X_t and W_t need to be extracted for each node k in addition to the presence-only locations s_i .

Spatiotemporal data integration

Within our framework, structured data provide an anchor point for estimating the baseline population density, λ_0 , which is unidentifiable in a standalone presence-only data analysis (Dorazio 2014, Farr et al. 2021). I assume that structured data exist at ≥ 1 stages t within a species annual cycle, whereas unstructured data exist throughout the cycle, t = 1, ..., T (i.e., structured data are more limited temporally than unstructured data). During a stage when both

data types simultaneously exist, the structured data can be used to estimate baseline population density, λ_0 . Thus, data integration creates a link between unstructured data and population density, which can be used to estimate population density during subsequent stages, t = 2, ..., T, when structured data are unavailable. If structured data exist during other portions of the annual cycle, they can also be integrated during an additional stage t. The implicit assumptions of this framework are that changes in population density are explained by covariates or temporal deviations (i.e., main effects of each stage, δ_t) from baseline density. Since unstructured data are used to estimate δ_t , the relationship between unstructured data and population density is assumed to be constant over space and time, with the thinning rate accurately capturing changes in observation error.

Joint likelihood

To create the integrated model, a joint likelihood is formed by the product of the likelihoods of structured (e.g., count) and unstructured (e.g., presence-only) data: $L_{IDM}(\lambda_0, \beta, \delta_t, \omega, p_0, \alpha) = L_C(\lambda_0, \beta, \delta_t, \omega) \cdot L_{PO}(\lambda_0, \beta, \delta_t, \omega, p_0, \alpha)$ eqn 5 The joint likelihood of the integrated distribution model assumes independence between data types. However, violations of this assumption are likely to have minimal impact on precision and accuracy of parameter estimation and associated uncertainty (Abadi et al. 2010, Fletcher et al. 2019).

Simulation Study

Scenario

The objective of our simulation study is to evaluate whether it is reasonable to assume a shared baseline population density between structured and unstructured data types during multiple spatiotemporal stages of abundance for a hypothetical target species. I devised a

scenario that was applicable to migratory species, like the eastern monarch butterfly population, which occupies non-overlapping geographic regions during different stages (e.g., breeding, nonbreeding, migrating) of their annual cycle. I established two stages, where each stage contained a spatially nonoverlapping domain, S_1 or S_2 , of distinct size and shape (Figures 4.1A, B), with $S_1 \approx$ 36% larger than S_2 . Both unstructured and structured data were available during the first stage, while only unstructured data were available during the second stage. I simulated presence-only observations for the unstructured data and single-visit counts for the structured data.

I specified λ_0 to yield an expected baseline mean abundance across both S_1 and S_2 . To create a realistic scenario, I created an inhomogeneous point process via a single environmental covariate across both domains, $X_t(s)$, and unmeasured spatial heterogeneity, $\omega(s)$. I specified a single covariate effect β that was constant between stages with unique covariate values, $X_t(s)$ for each stage (Figures 4.1A, B; Appendix D). The spatial random effect, $\omega(s)$, was simulated using a GRF with a Matérn correlation function and precision (τ), scale (κ), and smoothness, (v; fixed at v = 1; Appendix D) hyperparameters. I then simulated the location and number of individuals for each stage using a Poisson point process and the corresponding intensity function.

I simulated presence-only data (i.e., observations of individuals) for both stages (Figures 4.1C, D) where the thinning process incorporated effect α and a single covariate, W(s), representing sampling bias (simulated from a GRF). I standardized the covariate by subtracting the minimum value and dividing by the mean, which forced areas with low sampling intensity (i.e., lower covariate values) towards the intercept, p_0 . I set $p_0 = \text{logit}(0.01)$; thus, locations at the lower end of the covariate range represent low or no sampling effort (i.e., $p_t \approx 0$). The log-likelihood was estimated using eqns 3 and 4. I simulated single-visit counts at 25 sites for stage 1 (Figure 4.1C). I incorporated imperfect detection to mimic a realistic scenario. For each

individual within a given site, I simulated their detection probability using a random Bernoulli trial with a spatially homogenous probability of 0.90. Without repeated surveys, count data cannot be used to estimate absolute abundance; thus, in this simulation, our model could only estimate relative abundance.

I simulated 1000 datasets and estimated population density along with relative abundance to evaluate our ability to leverage structured data over space and time. Fixed effect parameters $(\lambda_0, \beta, \delta, \kappa, \tau, p_0, \alpha)$, along with covariates $X_t(s)$ and W(s), were constant across simulations. The spatial random effects, $\omega(s)$, and individual locations changed between simulations. Parameters were estimated using R-INLA, Template Model Builder, and R (Lindgren & Rue 2015, Kristensen et al. 2016, R Core Team 2020). To assess convergence (i.e., negative log likelihood is at a minimum), I checked that the absolute values of the final gradient for each parameter were near 0 and checked that the Hessian matrix was positive definite (Skaug & Fournier 2006). To assess performance, I calculated root-mean-squared error, bias, and percent relative bias for the mean estimate of each parameter. I also derived estimates of abundance and associated uncertainty to compare with true abundance. Finally, I estimated abundance for each of the 1000 simulated datasets based only on the presence-only data to compare with estimates from our integrated framework to assess the improvements in precision and bias as a result of integrating structured and unstructured datasets.

Results

Our integrated framework successfully identified baseline population density, λ_0 , with lower bias (-2% relative bias) than the presence-only analysis (231% relative bias; Figure 4.1E; see Table D.1 Appendix D for full results). The slight error in our integrated analysis likely reflects that I cannot adjust for imperfect detection with single-visit counts. When structured data

were unavailable, it was impossible to identify the intercept parameter of an intensity function, as has been demonstrated previously (Dorazio 2014, Farr et al. 2021). Estimates of covariate effect, β , were accurate using both frameworks, though more so using the integrated analysis (6% relative bias) relative to the presence-only analysis (10% relative bias; Figure 4.1E). These results were similar to those from other studies comparing integrated and presence-only analyses (Farr et al. 2021), and the small error in the integrated analysis was likely due to each domain, S_1 and S_2 , only capturing a portion of the covariate range (Figure 4.1).

Unmeasured spatial variability, $\omega(s)$, was similarly estimated between the integrated and presence-only analyses. Hyperparameters, τ and κ , were likely biased as S_1 and S_2 only captured a portion of the simulation GRF for $\omega(s)$ (Figure 4.1). Unmeasured population change, δ , was estimated with slight error for both the integrated (28% relative bias) and presence-only analyses (32% error; Figure 4.1E). Using the estimated parameters, I predicted abundance across each stage of the annual cycle. The integrated model was also able to estimate abundance for stage 2, when structured data were unavailable, with low bias (3% relative bias; Figure 4.1F). Abundance was also estimated with low bias in stage 1 (-9% relative bias; Figure 4.1F). Overall, abundances were underestimated because single-visit counts do not account for imperfect detection. Alternatively, abundance was overestimated in each stage (both $\approx 1 \times 10^8$ % relative bias) when using only the presence-only data due to its inability to estimate λ_0 .

Case Study

Population monitoring

The vast geographic range that monarchs inhabit during their annual migratory cycle make them difficult to monitor throughout the year. Structured monitoring occurs annually at the overwintering grounds in central Mexico and across the summer breeding grounds in the

Midwestern U.S. and Ontario, Canada. However, structured monitoring does not occur during the spring breeding season nor during the spring or fall migrations. With spring conditions potentially becoming less favorable for monarch recruitment (Zylstra et al. 2021), a deeper understanding of the mechanistic links between climate and monarch abundance may aid decision-making for this population of conservation concern. Fortunately, multiple volunteerbased monitoring programs collect observations of adult monarchs in the late spring and early summer through unstructured, opportunistic sampling. Using our integrated framework, I leveraged these data to estimate abundance of monarchs during spring migration and assess the effects of weather conditions during this critical stage.

Spatiotemporal stages

Our analysis encompassed the spring migration of monarchs during 2016 - 2018. I focused on these three years because the availability of unstructured data during the spring migration has increased markedly in recent years (see www.journeynorth.org). Within a given year, I partitioned spring migration into three distinct spatiotemporal stages. I indexed years with r (R = 3) and use S_t (T = 3) to denote each spatiotemporal stage during spring migration in a given year. The spatiotemporal stages were constant across years and defined similarly to previous studies (Saunders et al. 2019b, Zylstra et al. 2021). Our first stage, S_1 , encompassed eastern Texas and Oklahoma (93.5° W to 100° W, 25.8° N to 37° N) between 8 March and 4 April. This stage was specified to encompass adult monarchs arriving from Mexico as they lay eggs that become the first generation of the year. The next stage, S_2 , covered the same spatial extent but occurred between 19 April and 2 May, which included the sightings of adults from the first generation of the year post-metamorphosis. The final stage, S_3 , captured these adults after they arrived on the summer breeding grounds, an area that spans the Midwestern U.S. and

Ontario, Canada (74.3° W to 97.2° W, 39.8° N to 49.4° N), during the early part of the summer breeding season (3 May to 6 June). Although monarchs disperse outside of this area during summer, I focused on the Midwest because the majority of individuals that arrive on the overwintering grounds in Mexico originate from this region (Flockhart et al. 2017). This spatiotemporal stage is also important as it contains both structured and unstructured data, while S_1 and S_2 contain very few to no structured sampling data.

Datasets

I collated presence-only data on monarchs during spring and early summer from five unstructured monitoring programs: North American Butterfly Association's Butterflies I've Seen program (BIS), iNaturalist, eButterfly, Butterflies and Moths of North America, and Journey North. These programs allow nonscientists to report butterfly sightings (e.g., date, location, species, count) to an online database that is then made available to scientists. From each program, I extracted monarch sightings (location and date) during our study period and specified them as individual points (i.e., a single presence).

For structured monitoring data, I integrated single-visit counts from five separate programs. Unlike unstructured monitoring, variation in protocols exist between programs. The first program I used was North American Butterfly Association's Fourth of July (NFJ) surveys. These structured surveys are conducted annually in summer, with optimal timing depending on local conditions, at specified sites throughout the summer breeding range. Volunteers survey areas within a 25-km diameter circle and record the number of adults butterflies observed, by species (Oberhauser et al. 2015). I summed monarch observations among volunteers during each survey, resulting in a single count for each site and year C_{tj} . The other four monitoring programs were 'Pollard walk' surveys from state-level butterfly monitoring networks (BMNs) in Illinois,

Ohio, Iowa, and Michigan. Volunteers walked fixed transects and counted the number of adult butterflies observed. Unlike NFJ surveys, volunteers surveyed Pollard transects multiple times within a year; however, I only used the first replicate of the year to avoid pseudo-replication. As with NFJ surveys, I summarized monarch counts during each visit to a site. The majority of structured sampling existed during early summer, S_3 , with less than 5 structured counts occurring during late spring, S_2 , each year.

Environmental covariates

I included environmental covariates that are likely to influence monarch recruitment and abundance. Monarch larvae are obligate feeders on milkweed (Asclepias spp.; Pleasants & Oberhauser 2013); thus, milkweed availability is likely to drive monarch recruitment. Because data on milkweed distribution and abundance are lacking, I used normalized difference vegetation index (NDVI), a measure of landscape greenness, as a proxy for milkweed distribution during the spring and early summer (Flockhart et al. 2013, Lemoine 2015). Temperature is also vital to the recruitment of monarchs as it influences the rates of development and survival from eggs to adults (Zalucki 1982). I used the number of growing degree days (GDD) to describe thermal conditions across the spring migratory pathway (Zipkin et al. 2012, Saunders et al. 2016, 2018, Zylstra et al. 2021). GDD is the accumulation of temperature within a specific range (McMaster & Wilhelm 1997) that allows for monarch development. GDD values accumulated from a separate start date for each stage until the date each observation or survey occurred. To standardize for differences in days accumulated, I summarized GDD as the average daily GDD. For details about how NDVI and GDD variables were calculated for each observation, survey location, and mesh node, see Appendix E.

Observation covariates

The amount of effort expended searching for monarchs on each structured survey was likely to affect the number of individuals observed. Because the area searched during each structured survey (e.g., length of each BMN transect or area of each NFJ survey) was not reported, I converted the number of person hours spent surveying into area sampled based on the average human walking speed (5-km/hr; Browning et al. 2006) and assuming that butterflies within 1-m of the observer were detected. For the presence-only data, I used two covariates in the thinning function to account for spatiotemporal variation in sampling intensity. First, I used the spatiotemporal density of non-monarch butterfly (superfamily *Papilionoidea*) observations in iNaturalist, assuming that areas of high-frequency butterfly observations correlate with high sampling intensity. I used human population density as another proxy for sampling intensity because the number of observers relates to the number of people in the surrounding area (Geldmann et al. 2016, Appendix E).

Data analysis

I used the integrated modeling framework described above with a few case-specific modifications. I selected a 100-m² resolution for presence-only data, and converted the area offset to reflect this baseline resolution. In addition to the fixed effects describing population change between domains within a single annual cycle (δ_t), I included a fixed effect to capture population change between years (γ_r). I included an interaction between each environmental covariate and stage to capture any nonstationarity in climatic effects (Rollinson et al. 2021). I also specified the effect of observation covariates to vary by stage and year to account for changes in their effects on sampling effort. I added quadratic effects for both NDVI and averaged daily GDD to account for peaks in optimal climatic conditions. For simplicity, I show the updated eqn 2 here and specify the entire likelihood within TMB in Appendix E:

$$\log(\lambda_{tr}(s)) = \log(\lambda_0) + \beta_{1,t} \cdot NDVI_{tr}(s) + \beta_{2,t} \cdot NDVI_{tr}(s)^2 + \beta_{3,t} \cdot GDD_{tr}(s) + \beta_{4,t} \cdot GDD_{tr}(s)^2 + \delta_t + \gamma_r + \omega(s).$$
 eqn 6

I use location *s* above as a generic description, but I estimate population density within the likelihood at each s_i , node *k*, and site *j* (Appendix E).

Results

Using our integrated model, I estimated monarch population density (i.e., relative density) during spring migration from 2016 - 2018 when structured data were limited. In each year, monarch density increased across each migratory stage respectively (Figure 4.2). Climatic conditions along the spring migratory pathway, particularly early in migration in Texas, had significant effects on the density of monarchs (Figure 4.3A). I identified nonlinear effects of both NDVI and averaged daily GDD during the spring and early summer (Figures 4.3B, C). The relationship between covariates and peak monarch density differed between the spring and summer, shifting from higher densities under moderately warm and highly vegetated conditions to higher densities under cooler and less vegetated conditions (Figures 4.3D-F). Densities were highest when conditions were near their spatiotemporal means of a given stage (during 2016 – 2018) of NDVI and averaged daily GDD, except for early spring when expected density peaked above the mean NDVI (Figure 4.3D-F). The influence of vegetation and climatic conditions was also weaker during the early summer compared to spring (Figure 4.3A), with monarchs likely to be present over a wide range of conditions (Figures 4.2C, F, I, 4.3B, C, F).

I assessed seasonal and yearly changes in monarch densities, rather than total abundance, as single-visit counts do not account for imperfect detection. Monarch densities increased over the course of spring migration each year, with the lowest densities during early spring migration (Figure 4.2). Increases in densities from spring to summer (Figures 4.2A, B, D, E, G, H) reflects spring breeding and recruitment of individuals as monarchs peak in abundance during the summer (Zylstra et al. 2021). Abundance trends across years reflect similar patterns seen in the overwintering grounds during 2016 - 2018, with densities considerably higher in 2018 than in the previous two years (Table E.1 Appendix E).

Discussion

In this study, I developed an integrated model using a log-Gaussian Cox process that can estimate abundance during multiple seasons of a species' annual cycle when structured data are sparse. The continuous collection of unstructured data in combination with the comparatively higher information content inherent to structured data improves inferences on migratory populations, as compared to independent analyses of either data source. By addressing data gaps through data integration across multiple stages of a species' migratory or annual cycle, our modeling framework can be used to elucidate cross-seasonal connections between environmental factors and observed patterns in population abundance and/or dynamics.

I demonstrate both the validity and merit of my framework through a simulation study and by directly estimating monarch population density throughout the spring migration when structured data are minimal. The spring migration has been identified as a critical stage for the eastern monarch population because of a potential cross-seasonal effect of spring climate on the long-term trends and annual variation in summer monarch abundance (Saunders et al. 2016, Saunders et al. 2018b, Zylstra et al. 2021). Our model corroborates these long-term analyses, despite the use of a different spatiotemporal resolution than has been used in previous work. The demonstrated effects of spring NDVI and averaged daily GDD on monarch densities suggest that spring conditions may drive recruitment rates of monarchs in spring, either directly (by influencing growth rates) or indirectly (by influencing the availability of milkweed). Similar to

previous studies (Saunders et al. 2016, Zylstra et al. 2021), monarch densities were highest when conditions were mild and near the spatiotemporal (space and year) average of a given stage.

One of the novel findings from our analysis relates to nonstationarity of the effects of environmental conditions on monarch population densities (i.e., when mean effects or their covariance are not constant across space or time; Rollinson et al. 2021) in spring vs early summer, with a shift from peak densities under warmer and greener conditions vs. cooler and less vegetated conditions, respectively (Figure 3D-F). The initiative to understand cross-seasonal processes has unveiled the frequency of nonstationary effects over space and time within ecology (Rollinson et al. 2021). In our case study, I explicitly captured nonstationarity over time using a temporal interaction for the covariates and implicitly captured it over space as the spring and summer domains were spatially distinct. Given the point process foundation of our model, it can be adapted to situations when stationarity is violated by specifying more complex statistical structures (Jarzyna et al. 2014, Ingebrigtsen et al. 2014, Lindgren & Rue 2015, Krainski et al. 2018). The ability to incorporate nonstationarity and spatial random effects into our framework helps maintain the assumption of a baseline population density by increasing the model's flexibility in explaining spatiotemporal variation and dependences.

Our integrated framework, which uses unstructured data to estimate population size during stages when structured data are unavailable, relies on one critical assumption: that the relationship between unstructured data and population density is constant over space and time, with the thinning rate accurately capturing changes in observation error. Thus, valid inferences rely on a model that properly accounts for the observation processes through which the data were collected. These observational covariates are critical to properly estimating both population density and the effect of the environmental covariates, and should be carefully examined.

Unstructured data use a thinned point process with relevant covariates; however, relevant covariates may be hard to identify or measure (Kéry & Royle 2020). The use of covariates to correct for sampling bias of presence-only observations is the subject of much debate (Troudet et al. 2017) and warrants further study (Kelling et al. 2019). I selected observation covariates following other presence-only analyses (Geldmann et al. 2016, Momeni-Dehaghi et al. 2021), but improper selection can lead to biased inferences (Ries et al. 2019, Wepprich 2019). Log-Gaussian Cox processes may be a solution, as they are powerful tools to estimate spatiotemporal patterns in both ecological and observational processes and can be specified to improve estimation where spatial biases in sampling occur (Tang et al. 2021). Preferably, structured data should formally incorporate methods to account for error, such as imperfect detection.

In addition to selecting relevant covariates to capture variation in observational and biological patterns, other choices made in the modeling process can affect inferences. Covariates must be summarized over space and time, and the scale at which summarizing occurs could influence our understanding of the effects of climate or other environmental factors on species populations. Emerging evidence from macrosystems ecology indicates that multi-scale connections may drive population patterns, and thus, the role of scale is critical in capturing cryptic yet important ecological processes (Fei et al. 2015). Inferences will also be affected by the unit of measurement $(1-km^2 vs 1 \times 6-m^2)$ for the area offset, which accounts for effort in terms of the area sampled by structured surveys. The weight (log likelihood values) of the presence only data source is determined by the spatial resolution of the model, which is indirectly set by the area offset unit of measurement. Finally, data weighting between structured and unstructured sources can be a concern when structured surveys are unable to anchor baseline density estimates by incorporating imperfect detection. I encourage future practitioners to conduct sensitivity

analyses to ensure that inference is invariant to these choices or choices are made with biological support.

My spatiotemporal integrated model offers a powerful approach for estimating densities of migratory species that range over vast extents, as well as covariate effects on those densities. The spatial point process framework also allows for easy incorporation of techniques for addressing nonstationary processes, which is a growing issue in macrosystems research. Crossseasonal processes and data gaps are not only a challenge for migratory species. Thus, my modeling framework can be applied to other species or systems that would benefit from data integration to provide a complete picture of ecological processes operating across broad spatial ranges and the full annual cycle.

Figure 4.1

Figure 1. (A-D) Visualization of my simulation study for the spatiotemporal integrated model with stage, domain S_1 on the left (A, C) and stage, domain S_2 on the right (B, D) with dashed lines indicating the extent of each domain. The top row (A, B) depicts true population abundance and distribution with black dots representing observations of individuals. The middle row (C, D) shows the data for each stage within each domain. (C) S_1 contains presence-only data (white dots) and 25 sites (dashed circles) with single-visit counts (magnitude of counts are within dashed circles). (D) S_2 contains only presence-only data. (A-D) Background gradient indicates the expected population density (individuals per unit area) for each domain and stage (A, C are the same [stage 1] and B, D are the same [stage 2]). (E, F) Simulation results showing bias (estimated minus true value) from each of the integrated and presence-only models. Bottom left (E) shows results for biases in estimates of the baseline population density (λ_0), covariate effect (β), and population change (δ). Bottom right (F) shows results for biases in estimates of abundance for each domain and stage.

Figure 4.1 (cont'd)



Figure 4.2

Predicted population density of monarchs from 2016 to 2018 during their spring migration. Density values were calculated using parameter estimates and values of NDVI and daily GDD averaged across each stage at a given location and year. Population density was estimated and reported at a resolution of 100-m². Rows and columns correspond to years and stages, respectively.



Figure 4.3

Effects of NDVI and averaged daily GDD on monarch population density. Top left (A) is the parameter estimates (mean and 95% confidence intervals) for the linear and quadratic effects of NDVI and averaged daily GDD at each stage. Middle left (B, C) is the estimated marginal effects (mean and 95% confidence intervals) of NDVI (B) and averaged daily GDD (C) on monarch population density at each stage, respectively. Right column (D-F) is the marginal effect (mean) of both NDVI and averaged daily GDD on monarch population density for early spring (D), late spring (E), and early summer (F). Crosses indicate observed values between 2016 and 2018 and dashed lines indicate the observed mean values during 2016 - 2018.



APPENDICES

APPENDIX A: Additional application details

For the multi-species dynamic N-occupancy model application, I assessed multiple anthropogenic and environmental factors hypothesized to affect the antelope community's abundance and demographic rates. Though the final model did not contain covariates on the biological process, I describe the various covariates that I considered including and an explanation of why I refrained from including covariates in the final model.

Anthropogenic covariates

Humans may play a role in the spatial distributions and survival of antelope species within my target community as antelopes are often poached as part of the bushmeat trade (Newing 2001). I summarized multiple covariates as a proxy for hunting pressure because direct metrics were not available. I used human population density outside of parks and distance to edge as proxies for anthropogenic disturbance. Because human population density was either high (> 300 humans/km2) or low (< 35 humans/km2) at each park, I specified human density as a binary variable (i.e., 0 for low density [UDZ, NNNP, KRP] and 1 for high density [VNP, BIF, NFNP]). Distance to edge was used as a proxy for hunting pressure and other human disturbance. Distances were recorded from each camera trap site to the nearest accessible edge (e.g., public road, navigable river, park boundary adjacent to human land use).

Environmental covariates

Though previous research on tropical rainforest dwelling antelopes suggests neutral responses to weather conditions (O'Brien et al. 2020), ongoing climate change continues to be a concern for tropical rainforest (Phillips et al. 2009, Sullivan et al. 2020). I characterized local weather (i.e., rainfall) at the site-level to assess the potential influence of climate change on the antelope community. I hypothesized that climate may indirectly affect the population status and
trends of antelopes by altering forest productivity and resulting food availability (e.g., foliage, fruit). I used CHIRPS (Climate Hazards group Infrared Precipitation with Stations) annual precipitation values (mm), which combines satellite and station data to interpolate localized (i.e., 0.05° resolution) rainfall (Funk et al. 2015). Antelope species within this community vary in their diet composition (i.e., percent of foliage, fruit, grass); thus, each species is tied to a different forest types and corresponding elevation range (Kingdon 2015). To capture this variation, I also evaluated a site-level elevation covariate.

Model building

To interpret multiple covariate effects, I standardized each covariate to have a mean of zero and standard deviation of one. For site-level covariates I standardized across all parks as standardizing within each park prevented meaningful inference. During model building I first constructed a null model. To determine a final model, I attempted to add covariates to abundance, survival, and gains using a forward selection approach (Saunders et al. 2019b, 2019c, Saunders et al. 2021). However, parameter unidentifiability prevented this approach from working as convergence (Rhat < 1.2) for many species-specific effects was not achievable. I hypothesize that large variation in covariate values across parks prevented accurate estimation of effects. Ideally, population-specific (i.e., interaction between a given covariate, species, and park) effects could account for this issue, but data volumes for many parks were constrained by short temporal extents (Table 3.1). I also ruled out scaling up site-level covariates (i.e., mean of a park) to the park-level as the low number of parks (i.e., six in total) did not allow covariates to accurately represent (i.e., proxy for true relationship) the anthropogenic or environmental process of interest. Ultimately, I settled on using park-level random effects on initial abundance and apparent survival. The high similarities between sites within a park and the large variations that I

observed with covariate values among parks led us to determine that park-level random effects were the most practical solution.

APPENDIX B: Nimble model code

Nimble code to fit the 'multi-species dynamic N-occupancy model' for the antelope community across the network of TEAM sites within national parks between 2009-2019. Both the retrospective and projected analysis code are shown below.

Definitions of constants

nparks: number of national parks (6).
Nspecs: number of species (12).
parkS & parkE: indicates the parks where a species was observed (varies by species).
nsite: number of sites within a park (varies by park).
nstart & nend: indicates the years when a park was surveyed (varies by parks).
nyears: number of years for projection corresponding to 2030 (22).
nreps: number of replicates (6).

Retrospective model code

```
model.code <- nimbleCode({</pre>
#-Priors-#
# Detection hyperparameters
mu.a0 ~ dunif(0, 1) # Community-level intercept on probability scale
mu.a0L <- logit(mu.a0) # Community-level intercept on logit scale</pre>
tau.a0 ~ dgamma(0.1, 0.1) # Community-level precision
# Effect of effort (days sampled)
alpha1 \sim dnorm(0, 0.1)
# Initial abundance hyperparameters
mu.b0 ~ dnorm(0, 0.1) # Community-level intercept on log scale
tau.b0 ~ dgamma(0.1, 0.1) # Community-level precision
# Apparent survival hyperparameters
mu.o0 ~ dunif(0, 1) # Community-level intercept on probability scale
mu.oOL <- logit(mu.o0) # Community-level intercept on logit scale</pre>
tau.o0 ~ dgamma(0.1, 0.1) # Community-level precision
# Gains hyperparameters
mu.g0 ~ dnorm(0, 0.1) # Community-level intercept on log scale
tau.g0 ~ dgamma(0.1, 0.1) # Community-level precision
# Precison of park-level random effect on initial abundance
tau.eps.l \sim dgamma(0.1, 0.1)
# Precison of park-level random effect on apparent survival
tau.eps.o ~ dgamma(0.1, 0.1)
```

```
for(r in 1:nparks) {
# Park-level random effect on initial abundance
eps.l[r] ~ dnorm(0, tau.eps.l)
# Park-level random effect on apparent survival
eps.o[r] ~ dnorm(0, tau.eps.o)
# Predicted community-level apparent survival for each park
logit(park.surv[r]) <- mu.oOL + eps.o[r]</pre>
} # End r
for(i in 1:nspecs) {
# Species-specific intercept on detection
alpha0[i] ~ dnorm(mu.a0L, tau.a0)
# Species-specific intercept on initial abundance
beta0[i] ~ dnorm(mu.b0, tau.b0)
# Species-specific intercept on apparent survival
omega0[i] ~ dnorm(mu.o0L, tau.o0)
# Species-specific intercept on gains
gamma0[i] ~ dnorm(mu.g0, tau.g0)
for(r in parkS[i]:parkE[i]) {
# Predicted population-level apparent survival
logit(pop.surv[r,i]) <- omega0[i] + eps.o[r]</pre>
for(j in 1:nsite[r]){
# Linear predictor for initial abundance
log(lambda[j,r,i]) <- beta0[i] + eps.l[r]</pre>
# Initial abundance
N[nstart[r],j,r,i] ~ dpois(lambda[j,r,i])
for(k in 1:nreps) {
# Linear predictor for detection probability (in year 1)
logit(r[k,nstart[r],j,r,i]) <- alpha0[i] + alpha1 * days[k,nstart[r],j,i]</pre>
# Observation process (in year 1)
y[k,nstart[r],j,r,i] ~ dbern(p[k,nstart[r],j,r,i])
# Site level detection (N-occupancy parameterization) (in year 1)
p[k,nstart[r],j,r,i] <- 1 - pow((1 - r[k,nstart[r],j,r,i]),</pre>
N[nstart[r],j,r,i])
} # End k
for(t in (nstart[r]+1):nend[r]) {
for(k in 1:nreps){
```

```
# Linear predictor for detection probability (in year t+1)
logit(r[k,t,j,r,i]) <- alpha0[i] + alpha1 * days[k,t,j,i]</pre>
# Observation process (in year t+1)
y[k,t,j,r,i] \sim dbern(p[k,t,j,r,i])
# Observation process (in year t+1)
p[k,t,j,r,i] <- 1 - pow((1 - r[k,t,j,r,i]), N[t,j,r,i])</pre>
} # End k
# Linear predictor for apparent survival
logit(omega[t-1,j,r,i]) <- omega0[i] + eps.o[r]</pre>
# Biological process
N[t,j,r,i] \leq S[t-1,j,r,i] + G[t-1,j,r,i]
# Apparent survival
S[t-1,j,r,i] ~ dbin(omega[t-1,j,r,i], N[t-1,j,r,i])
# Gains
G[t-1,j,r,i] ~ dpois(gamma[t-1,j,r,i])
# Linear predictor for gains
log(gamma[t-1,j,r,i]) <- gamma0[i]</pre>
} # End t
} # End j
for(t in nstart[r]:nend[r]) {
# Population (species-park) abundance per year
Nhat[t,r,i] <- sum(N[t,1:nsite[r],r,i])</pre>
} # End t
} # End r
} # End s
})
```

Prospective model code

```
model.code <- nimbleCode({
    #-Priors-#
    # Detection hyperparameters
    mu.a0 ~ dunif(0, 1) # Community-level intercept on probability scale
    mu.a0L <- logit(mu.a0) # Community-level intercept on logit scale
    tau.a0 ~ dgamma(0.1, 0.1) # Community-level precision</pre>
```

```
# Effect of effort (days sampled)
alphal \sim dnorm(0, 0.1)
# Initial abundance hyperparameters
mu.b0 \sim dnorm(0, 0.1) # Community-level intercept on log scale
tau.b0 ~ dgamma(0.1, 0.1) # Community-level precision
# Apparent survival hyperparameters
mu.o0 ~ dunif(0, 1) # Community-level intercept on probability scale
mu.oOL <- logit(mu.oO) # Community-level intercept on logit scale</pre>
tau.o0 ~ dgamma(0.1, 0.1) # Community-level precision
# Gains hyperparameters
mu.g0 ~ dnorm(0, 0.1) # Community-level intercept on log scale
tau.g0 ~ dgamma(0.1, 0.1) # Community-level precision
# Precison of park-level random effect on initial abundance
tau.eps.l \sim dgamma(0.1, 0.1)
# Precison of park-level random effect on apparent survival
tau.eps.o ~ dgamma(0.1, 0.1)
for(r in 1:nparks) {
# Park-level random effect on initial abundance
eps.l[r] ~ dnorm(0, tau.eps.l)
# Park-level random effect on apparent survival
eps.o[r] ~ dnorm(0, tau.eps.o)
# Predicted community-level apparent survival for each park
logit(park.surv[r]) <- mu.oOL + eps.o[r]</pre>
} # End r
for(i in 1:nspecs) {
# Species-specific intercept on detection
alpha0[i] ~ dnorm(mu.a0L, tau.a0)
# Species-specific intercept on initial abundance
beta0[i] ~ dnorm(mu.b0, tau.b0)
# Species-specific intercept on apparent survival
omega0[i] ~ dnorm(mu.o0L, tau.o0)
# Species-specific intercept on gains
gamma0[i] ~ dnorm(mu.g0, tau.g0)
for(r in parkS[i]:parkE[i]){
# Predicted population-level apparent survival
logit(pop.surv[r,i]) <- omega0[i] + eps.o[r]</pre>
for(j in 1:nsite[r]){
# Linear predictor for initial abundance
```

```
log(lambda[j,r,i]) <- beta0[i] + eps.l[r]</pre>
# Initial abundance
N[nstart[r],j,r,i] ~ dpois(lambda[j,r,i])
for(k in 1:nreps) {
# Linear predictor for detection probability (in year 1)
logit(r[k,nstart[r],j,r,i]) <- alpha0[i] + alpha1 * days[k,nstart[r],j,i]</pre>
# Observation process (in year 1)
y[k,nstart[r],j,r,i] ~ dbern(p[k,nstart[r],j,r,i])
# Site level detection (N-occupancy parameterization) (in year 1)
p[k,nstart[r],j,r,i] <- 1 - pow((1 - r[k,nstart[r],j,r,i]),</pre>
N[nstart[r],j,r,i])
} # End k
for(t in (nstart[r]+1):nend[r]) {
for(k in 1:nreps){
# Linear predictor for detection probability (in year t+1)
logit(r[k,t,j,r,i]) <- alpha0[i] + alpha1 * days[k,t,j,i]</pre>
# Observation process (in year t+1)
y[k,t,j,r,i] \sim dbern(p[k,t,j,r,i])
# Observation process (in year t+1)
p[k,t,j,r,i] <- 1 - pow((1 - r[k,t,j,r,i]), N[t,j,r,i])</pre>
} # End k
# Linear predictor for apparent survival
logit(omega[t-1,j,r,i]) <- omega0[i] + eps.o[r]</pre>
# Biological process
N[t,j,r,i] <- S[t-1,j,r,i] + G[t-1,j,r,i]
# Apparent survival
S[t-1,j,r,i] ~ dbin(omega[t-1,j,r,i], N[t-1,j,r,i])
# Gains
G[t-1,j,r,i] ~ dpois(gamma[t-1,j,r,i])
# Linear predictor for gains
log(gamma[t-1,j,r,i]) <- gamma0[i]</pre>
} # End t
for(t in (nend[r]+1):nyears) {
# Linear predictor for apparent survival
logit(omega[t-1,j,r,i]) <- omega0[i] + eps.o[r]</pre>
# Biological process
```

```
N[t,j,r,i] <- S[t-1,j,r,i] + G[t-1,j,r,i]
# Apparent survival
S[t-1,j,r,i] ~ dbin(omega[t-1,j,r,i], N[t-1,j,r,i])
# Gains
G[t-1,j,r,i] ~ dpois(gamma[t-1,j,r,i])
# Linear predictor for gains
log(gamma[t-1,j,r,i]) <- gamma0[i]</pre>
} # End t
} # End j
for(t in nstart[r]:nyears){
# Population (species-park) abundance per year
Nhat[t,r,i] <- sum(N[t,1:nsite[r],r,i])</pre>
} # End t
} # End r
} # End s
})
```

APPENDIX C: Model output

Output from the multi-species dynamic N-occupancy model. For the retrospective analysis I provide estimates of annual population growth and population abundance during the study period. I also report species-specific, community-level, and park-level demographic rates. For the population projection and Bayesian viability analysis, I provide estimate of the quasiextinction probability of each population.

Retrospective analysis

Table C.1

Mean annual population growth rates (i.e., geometric mean) with 95% credible intervals of each population (species-park combination) during the study period.

| National park | Species | Annual population growth |
|---------------|----------------|--------------------------|
| | - | rate |
| UDZ | C. harveyi | 1.04 (CI 1.02, 1.06) |
| UDZ | N. moschatus | 1.00 (CI 0.98, 1.02) |
| UDZ | T. scriptus | 1.19 (CI 1.09, 1.26) |
| UDZ | C. spadix | 1.01 (CI 0.98, 1.04) |
| VNP | C. nigrifrons | 1.07 (CI 1.03, 1.10) |
| VNP | T. scriptus | 1.01 (CI 0.99, 1.03) |
| NFNP | C. nigrifrons | 1.41 (CI 1.29, 1.54) |
| NFNP | T. scriptus | 1.41 (CI 0.91, 2.08) |
| NFNP | C. silvicultor | 1.54 (CI 1.21, 1.94) |
| BIF | C. nigrifrons | 1.08 (CI 1.06, 1.11) |
| BIF | T. scriptus | 1.12 (CI 0.96, 1.29) |
| BIF | T. spekii | 1.16 (CI 0.97, 1.36) |
| BIF | C. silvicultor | 1.06 (CI 1.03, 1.09) |
| NNNP | C. callipygus | 0.93 (CI 0.90, 0.95) |
| NNNP | C. dorsalis | 0.96 (CI 0.93, 1.00) |
| NNNP | C. leucogaster | 1.49 (CI 1.33, 1.69) |
| NNNP | P. monticola | 0.92 (CI 0.90, 0.95) |
| NNNP | C. nigrifrons | 1.07 (CI 0.89, 1.26) |
| NNNP | T. spekii | 1.02 (CI 0.81, 1.23) |
| NNNP | C. silvicultor | 0.94 (CI 0.91, 0.97) |
| KRP | P. monticola | 1.03 (CI 0.98, 1.08) |
| KRP | C. ogilbyi | 1.00 (CI 0.93, 1.07) |

Table C1 (cont'd)

KRP

C. silvicultor

0.92 (CI 0.67, 1.19)

Table C.2

Mean annual site-level abundance with 95% credible intervals of each population across the

| National park | Species | Year | Site-level abundance |
|---------------|----------------|------|----------------------|
| KRP | P. monticola | 2011 | 1.49 (CI 1.28, 1.73) |
| KRP | P. monticola | 2012 | 1.59 (CI 1.37, 1.83) |
| KRP | P. monticola | 2013 | 2.02 (CI 1.77, 2.32) |
| KRP | P. monticola | 2014 | 1.90 (CI 1.65, 2.18) |
| KRP | P. monticola | 2015 | 1.68 (CI 1.43, 1.95) |
| KRP | C. ogilbyi | 2011 | 1.24 (CI 0.97, 1.60) |
| KRP | C. ogilbyi | 2012 | 1.17 (CI 0.92, 1.50) |
| KRP | C. ogilbyi | 2013 | 1.30 (CI 1.05, 1.62) |
| KRP | C. ogilbyi | 2014 | 1.23 (CI 0.97, 1.57) |
| KRP | C. ogilbyi | 2015 | 1.23 (CI 0.95, 1.58) |
| KRP | C. silvicultor | 2011 | 0.06 (CI 0.03, 0.12) |
| KRP | C. silvicultor | 2012 | 0.06 (CI 0.05, 0.10) |
| KRP | C. silvicultor | 2013 | 0.06 (CI 0.03, 0.10) |
| KRP | C. silvicultor | 2014 | 0.05 (CI 0.03, 0.08) |
| KRP | C. silvicultor | 2015 | 0.04 (CI 0.02, 0.10) |
| NNNP | C. callipygus | 2010 | 4.69 (CI 3.87, 5.65) |
| NNNP | C. callipygus | 2011 | 3.57 (CI 3.02, 4.23) |
| NNNP | C. callipygus | 2012 | 2.92 (CI 2.47, 3.43) |
| NNNP | C. callipygus | 2013 | 2.97 (CI 2.52, 3.48) |
| NNNP | C. callipygus | 2014 | 3.05 (CI 2.57, 3.58) |
| NNNP | C. callipygus | 2015 | 2.61 (CI 2.17, 3.10) |
| NNNP | C. callipygus | 2016 | 2.94 (CI 2.45, 3.48) |
| NNNP | C. dorsalis | 2010 | 2.10 (CI 1.75, 2.52) |
| NNNP | C. dorsalis | 2011 | 1.69 (CI 1.42, 2.02) |
| NNNP | C. dorsalis | 2012 | 1.75 (CI 1.47, 2.08) |
| NNNP | C. dorsalis | 2013 | 1.90 (CI 1.63, 2.22) |
| NNNP | C. dorsalis | 2014 | 1.71 (CI 1.45, 2.02) |
| NNNP | C. dorsalis | 2015 | 1.57 (CI 1.30, 1.88) |
| NNNP | C. dorsalis | 2016 | 1.69 (CI 1.40, 2.02) |
| NNNP | C. leucogaster | 2010 | 0.01 (CI 0.00, 0.03) |
| NNNP | C. leucogaster | 2011 | 0.04 (CI 0.00, 0.12) |
| NNNP | C. leucogaster | 2012 | 0.14 (CI 0.07, 0.23) |
| NNNP | C. leucogaster | 2013 | 0.25 (CI 0.18, 0.35) |
| NNNP | C. leucogaster | 2014 | 0.28 (CI 0.20, 0.40) |
| NNNP | C. leucogaster | 2015 | 0.23 (CI 0.13, 0.37) |
| NNNP | C. leucogaster | 2016 | 0.23 (CI 0.13, 0.38) |

study period (including only the years sampled for each population/park combination).

Table C2 (cont'd)

| NNNP | P. monticola | 2010 | 6.21 (CI 5.32, 7.18) |
|------|----------------|------|----------------------|
| NNNP | P. monticola | 2011 | 4.70 (CI 4.10, 5.37) |
| NNNP | P. monticola | 2012 | 4.19 (CI 3.67, 4.80) |
| NNNP | P. monticola | 2013 | 4.19 (CI 3.68, 4.77) |
| NNNP | P. monticola | 2014 | 4.19 (CI 3.68, 4.75) |
| NNNP | P. monticola | 2015 | 3.59 (CI 3.10, 4.15) |
| NNNP | P. monticola | 2016 | 3.78 (CI 3.28, 4.33) |
| NNNP | C. nigrifrons | 2010 | 0.05 (CI 0.03, 0.08) |
| NNNP | C. nigrifrons | 2011 | 0.01 (CI 0.00, 0.05) |
| NNNP | C. nigrifrons | 2012 | 0.08 (CI 0.07, 0.10) |
| NNNP | C. nigrifrons | 2013 | 0.01 (CI 0.00, 0.05) |
| NNNP | C. nigrifrons | 2014 | 0.02 (CI 0.00, 0.05) |
| NNNP | C. nigrifrons | 2015 | 0.05 (CI 0.02, 0.10) |
| NNNP | C. nigrifrons | 2016 | 0.07 (CI 0.02, 0.15) |
| NNNP | T. spekii | 2010 | 0.07 (CI 0.03, 0.18) |
| NNNP | T. spekii | 2011 | 0.06 (CI 0.00, 0.15) |
| NNNP | T. spekii | 2012 | 0.07 (CI 0.03, 0.15) |
| NNNP | T. spekii | 2013 | 0.05 (CI 0.00, 0.15) |
| NNNP | T. spekii | 2014 | 0.06 (CI 0.00, 0.15) |
| NNNP | T. spekii | 2015 | 0.09 (CI 0.05, 0.18) |
| NNNP | T. spekii | 2016 | 0.08 (CI 0.02, 0.18) |
| NNNP | C. silvicultor | 2010 | 2.17 (CI 1.87, 2.52) |
| NNNP | C. silvicultor | 2011 | 2.08 (CI 1.83, 2.37) |
| NNNP | C. silvicultor | 2012 | 1.99 (CI 1.77, 2.23) |
| NNNP | C. silvicultor | 2013 | 1.84 (CI 1.62, 2.08) |
| NNNP | C. silvicultor | 2014 | 1.72 (CI 1.52, 1.97) |
| NNNP | C. silvicultor | 2015 | 1.54 (CI 1.33, 1.77) |
| NNNP | C. silvicultor | 2016 | 1.50 (CI 1.28, 1.73) |
| UDZ | C. harveyi | 2009 | 1.96 (CI 1.68, 2.28) |
| UDZ | C. harveyi | 2010 | 1.99 (CI 1.73, 2.30) |
| UDZ | C. harveyi | 2011 | 2.12 (CI 1.85, 2.43) |
| UDZ | C. harveyi | 2012 | 2.19 (CI 1.92, 2.52) |
| UDZ | C. harveyi | 2013 | 2.31 (CI 2.03, 2.65) |
| UDZ | C. harveyi | 2014 | 2.47 (CI 2.17, 2.83) |
| UDZ | C. harveyi | 2015 | 2.54 (CI 2.22, 2.92) |
| UDZ | C. harveyi | 2016 | 2.65 (CI 2.32, 3.03) |
| UDZ | C. harveyi | 2017 | 2.77 (CI 2.43, 3.17) |
| UDZ | C. harveyi | 2018 | 2.80 (CI 2.43, 3.23) |
| UDZ | C. harveyi | 2019 | 2.91 (CI 2.52, 3.35) |
| UDZ | N. moschatus | 2009 | 0.91 (CI 0.73, 1.12) |
| UDZ | N. moschatus | 2010 | 0.88 (CI 0.72, 1.07) |

Table C2 (cont'd)

| UDZ | N. moschatus | 2011 | 0.86 (CI 0.70, 1.03) |
|-----|---------------|------|----------------------|
| UDZ | N. moschatus | 2012 | 0.89 (CI 0.73, 1.07) |
| UDZ | N. moschatus | 2013 | 1.02 (CI 0.88, 1.18) |
| UDZ | N. moschatus | 2014 | 1.05 (CI 0.90, 1.22) |
| UDZ | N. moschatus | 2015 | 1.02 (CI 0.87, 1.18) |
| UDZ | N. moschatus | 2016 | 0.96 (CI 0.82, 1.12) |
| UDZ | N. moschatus | 2017 | 0.95 (CI 0.80, 1.10) |
| UDZ | N. moschatus | 2018 | 0.90 (CI 0.75, 1.07) |
| UDZ | N. moschatus | 2019 | 0.91 (CI 0.75, 1.08) |
| UDZ | T. scriptus | 2009 | 0.01 (CI 0.00, 0.05) |
| UDZ | T. scriptus | 2010 | 0.06 (CI 0.05, 0.08) |
| UDZ | T. scriptus | 2011 | 0.06 (CI 0.03, 0.10) |
| UDZ | T. scriptus | 2012 | 0.13 (CI 0.12, 0.15) |
| UDZ | T. scriptus | 2013 | 0.08 (CI 0.03, 0.13) |
| UDZ | T. scriptus | 2014 | 0.11 (CI 0.07, 0.15) |
| UDZ | T. scriptus | 2015 | 0.13 (CI 0.10, 0.17) |
| UDZ | T. scriptus | 2016 | 0.12 (CI 0.08, 0.17) |
| UDZ | T. scriptus | 2017 | 0.11 (CI 0.08, 0.15) |
| UDZ | T. scriptus | 2018 | 0.10 (CI 0.07, 0.15) |
| UDZ | T. scriptus | 2019 | 0.13 (CI 0.10, 0.17) |
| UDZ | C. spadix | 2009 | 1.27 (CI 0.93, 1.67) |
| UDZ | C. spadix | 2010 | 1.25 (CI 0.93, 1.63) |
| UDZ | C. spadix | 2011 | 1.25 (CI 0.93, 1.63) |
| UDZ | C. spadix | 2012 | 1.20 (CI 0.85, 1.60) |
| UDZ | C. spadix | 2013 | 1.24 (CI 0.92, 1.63) |
| UDZ | C. spadix | 2014 | 1.26 (CI 0.93, 1.67) |
| UDZ | C. spadix | 2015 | 1.29 (CI 0.93, 1.70) |
| UDZ | C. spadix | 2016 | 1.38 (CI 1.07, 1.78) |
| UDZ | C. spadix | 2017 | 1.40 (CI 1.05, 1.80) |
| UDZ | C. spadix | 2018 | 1.38 (CI 1.02, 1.82) |
| UDZ | C. spadix | 2019 | 1.38 (CI 1.00, 1.83) |
| BIF | C. nigrifrons | 2010 | 0.75 (CI 0.67, 0.85) |
| BIF | C. nigrifrons | 2011 | 0.82 (CI 0.72, 0.93) |
| BIF | C. nigrifrons | 2012 | 0.95 (CI 0.85, 1.08) |
| BIF | C. nigrifrons | 2013 | 0.72 (CI 0.60, 0.85) |
| BIF | C. nigrifrons | 2014 | 0.83 (CI 0.72, 0.95) |
| BIF | C. nigrifrons | 2015 | 0.98 (CI 0.85, 1.12) |
| BIF | C. nigrifrons | 2016 | 1.42 (CI 1.28, 1.57) |
| BIF | C. nigrifrons | 2017 | 1.33 (CI 1.17, 1.50) |
| BIF | T. scriptus | 2010 | 0.02 (CI 0.00, 0.07) |
| BIF | T. scriptus | 2011 | 0.01 (CI 0.00, 0.03) |

Table C2 (cont'd)

| BIF | T. scriptus | 2012 | 0.01 (CI 0.00, 0.03) |
|------|----------------|------|----------------------|
| BIF | T. scriptus | 2013 | 0.04 (CI 0.03, 0.07) |
| BIF | T. scriptus | 2014 | 0.04 (CI 0.03, 0.07) |
| BIF | T. scriptus | 2015 | 0.04 (CI 0.02, 0.07) |
| BIF | T. scriptus | 2016 | 0.06 (CI 0.05, 0.08) |
| BIF | T. scriptus | 2017 | 0.07 (CI 0.05, 0.10) |
| BIF | T. spekii | 2010 | 0.04 (CI 0.02, 0.10) |
| BIF | T. spekii | 2011 | 0.04 (CI 0.02, 0.12) |
| BIF | T. spekii | 2012 | 0.05 (CI 0.02, 0.12) |
| BIF | T. spekii | 2013 | 0.05 (CI 0.02, 0.13) |
| BIF | T. spekii | 2014 | 0.06 (CI 0.02, 0.15) |
| BIF | T. spekii | 2015 | 0.08 (CI 0.05, 0.17) |
| BIF | T. spekii | 2016 | 0.09 (CI 0.05, 0.17) |
| BIF | T. spekii | 2017 | 0.10 (CI 0.05, 0.20) |
| BIF | C. silvicultor | 2010 | 1.02 (CI 0.88, 1.20) |
| BIF | C. silvicultor | 2011 | 1.03 (CI 0.90, 1.18) |
| BIF | C. silvicultor | 2012 | 1.06 (CI 0.92, 1.23) |
| BIF | C. silvicultor | 2013 | 1.22 (CI 1.07, 1.38) |
| BIF | C. silvicultor | 2014 | 1.42 (CI 1.27, 1.58) |
| BIF | C. silvicultor | 2015 | 1.35 (CI 1.17, 1.53) |
| BIF | C. silvicultor | 2016 | 1.44 (CI 1.27, 1.63) |
| BIF | C. silvicultor | 2017 | 1.53 (CI 1.35, 1.73) |
| NFNP | C. nigrifrons | 2014 | 0.29 (CI 0.25, 0.34) |
| NFNP | C. nigrifrons | 2015 | 0.53 (CI 0.42, 0.64) |
| NFNP | C. nigrifrons | 2016 | 0.65 (CI 0.52, 0.79) |
| NFNP | C. nigrifrons | 2017 | 0.81 (CI 0.65, 0.99) |
| NFNP | T. scriptus | 2014 | 0.02 (CI 0.00, 0.06) |
| NFNP | T. scriptus | 2015 | 0.04 (CI 0.01, 0.08) |
| NFNP | T. scriptus | 2016 | 0.05 (CI 0.02, 0.09) |
| NFNP | T. scriptus | 2017 | 0.06 (CI 0.02, 0.10) |
| NFNP | C. silvicultor | 2014 | 0.07 (CI 0.03, 0.11) |
| NFNP | C. silvicultor | 2015 | 0.11 (CI 0.05, 0.18) |
| NFNP | C. silvicultor | 2016 | 0.15 (CI 0.08, 0.24) |
| NFNP | C. silvicultor | 2017 | 0.24 (CI 0.15, 0.34) |
| VNP | C. nigrifrons | 2014 | 3.22 (CI 2.88, 3.62) |
| VNP | C. nigrifrons | 2015 | 3.42 (CI 3.07, 3.83) |
| VNP | C. nigrifrons | 2016 | 3.67 (CI 3.28, 4.08) |
| VNP | T. scriptus | 2014 | 2.88 (CI 2.48, 3.32) |
| VNP | T. scriptus | 2015 | 2.91 (CI 2.52, 3.32) |
| VNP | T. scriptus | 2016 | 2.95 (CI 2.55, 3.37) |

Table C.3

Mean annual survival probability and gains with 95% credible intervals for each species across

all parks.

| Species | Annual apparent survival | Gains |
|-----------------|--------------------------|----------------------|
| C. callipygus | 0.61 (CI 0.15, 0.94) | 1.34 (CI 0.88, 1.95) |
| C. dorsalis | 0.56 (CI 0.12, 0.92) | 0.92 (CI 0.67, 1.21) |
| C. harveyi | 0.88 (CI 0.43, 0.98) | 0.36 (CI 0.24, 0.51) |
| C. leucogaster | 0.73 (CI 0.24, 0.96) | 0.09 (CI 0.06, 0.15) |
| P. monticola | 0.67 (CI 0.18, 0.95) | 1.48 (CI 1.11, 1.87) |
| N. moschatus | 0.83 (CI 0.32, 0.98) | 0.14 (CI 0.09, 0.20) |
| C. nigrifrons | 0.66 (CI 0.17, 0.95) | 0.22 (CI 0.18, 0.27) |
| C. ogilbyi | 0.89 (CI 0.44, 0.99) | 0.57 (CI 0.36, 0.85) |
| T. scriptus | 0.66 (CI 0.16, 0.95) | 0.03 (CI 0.02, 0.05) |
| C. spadix | 0.82 (CI 0.31, 0.98) | 0.22 (CI 0.11, 0.37) |
| T. spekii | 0.73 (CI 0.19, 0.97) | 0.02 (CI 0.01, 0.05) |
| C. silvicultor | 0.90 (CI 0.51, 0.99) | 0.12 (CI 0.09, 0.16) |
| Community-level | 0.72 (CI 0.28, 0.96) | 0.24 (CI 0.10, 0.56) |

Table C.4

| National park | Annual apparent survival |
|---------------|--------------------------|
| UDZ | 0.77 (CI 0.56, 0.89) |
| VNP | 0.99 (CI 0.97, 1.00) |
| NFNP | 0.77 (CI 0.56, 0.92) |
| BIF | 0.81 (CI 0.69, 0.91) |
| NNNP | 0.66 (CI 0.51, 0.80) |
| KRP | 0.31 (CI 0.11, 0.52) |

Park-level mean annual survival probability with 95% credible intervals.

Population projection

Table C.5

Quasi-extinction probability (for each of the 18 populations coming from parks with at least five years of data) for the projected window of 2020-2030.

| National park | Species | Quasi-extinction probability |
|---------------|----------------|------------------------------|
| KRP | P. monticola | 0.00 |
| KRP | C. ogilbyi | 0.00 |
| KRP | C. silvicultor | 0.05 |
| NNNP | C. callipygus | 0.00 |
| NNNP | C. dorsalis | 0.00 |
| NNNP | C. leucogaster | 0.07 |
| NNNP | P. monticola | 0.00 |
| NNNP | C. nigrifrons | 0.00 |
| NNNP | T. spekii | 0.24 |
| NNNP | C. silvicultor | 0.23 |
| UDZ | C. harveyi | 0.00 |
| UDZ | N. moschatus | 0.01 |
| UDZ | T. scriptus | 0.24 |
| UDZ | C. spadix | 0.01 |
| BIF | C. nigrifrons | 0.04 |
| BIF | T. scriptus | 0.03 |
| BIF | T. spekii | 0.23 |
| BIF | C. silvicultor | 0.01 |

APPENDIX D: Simulation code, and output

Simulation study script

Attached here is the R script to run the simulation study.

#-----# #-Libraries-# #-----#

library(spatstat)

library(RandomFields)

library(INLA)

library(rgeos)

library(TMB)

library(gridExtra)

#-----# #-Functions-# #-----#

#Function to create dual mesh (original code from Krainski et al. 2018)
source("rDualMesh.R")

```
#Logit function
logit <- function(pp)
{
    log(pp) - log(1-pp)
}
#Inverse logit
expit <- function(eta)
{
    1/(1+exp(-eta))
```

}

```
#-----#
#-Spatio-temporal domains-#
#-----#
#Full spatio-temporal extent
win <- owin(c(0, 3), c(0, 3)) #Window
loc.d <- 3 * cbind(c(0, 1, 1, 0, 0), c(0, 0, 1, 1, 0)) #Coordinates
#Domain 1
win1 <- owin(c(1, 2.75), c(0.25, 2))
#Domain 2
win2 <- owin(c(0, 3), c(0, 3), poly =
list(x=c(0.25,0.75,0.75,2.75,2.75,0.25), y=c(0.25,0.25,2.25,2.25,2.75,2.75)))
#Number of pixels
npix <- 300
spatstat.options(npixel = npix)
#Create triangulated mesh over domain
mesh <- inla.mesh.2d(loc.domain = loc.d,</pre>
                    offset = c(0.3, 1),
                    max.edge = c(0.3, 0.7),
                     cutoff = 0.05)
# mesh1 <- inla.mesh.2d(boundary = inla.sp2segment(as(win1,</pre>
"SpatialPolygons")), max.edge = 0.3, cutoff = 0.05)
#
# mesh2 <- inla.mesh.2d(boundary = inla.sp2segment(as(win2,</pre>
"SpatialPolygons")), max.edge = 0.3, cutoff = 0.05)
#X range of mesh
x0 <- seq(min(mesh$loc[, 1]), max(mesh$loc[, 1]), length = npix)</pre>
```

```
#Y range of mesh
y0 <- seq(min(mesh$loc[, 2]), max(mesh$loc[, 2]), length = npix)</pre>
#Mesh window
Mwin <- owin(c(min(x0), max(x0)), c(min(y0), max(y0)))
#----#
#-Simulate LGCP-#
#----#
#Parameters of intensity function
beta <- c(6, 0.5)
#Change in intensity
delta <- c(0.5)
#Parameters of thinning function
alpha <- c(logit(0.01), 0.85)
#Ecological covariate
set.seed(100)
RandomFields::RFoptions(spConform=FALSE)
xcov1 <- RFsimulate(model = RMgauss(scale = 1.25), x = x0, y = y0, grid =</pre>
TRUE)
xcov2 <- xcov1 + RFsimulate(model = RMgauss(var = 0.5, scale = 9.5), x = x0,</pre>
y = y0, grid = TRUE)
x1 <- (xcov1 - mean(c(xcov1, xcov2))/sd(c(xcov1, xcov2)))</pre>
x2 <- (xcov2 - mean(c(xcov1, xcov2))/sd(c(xcov1, xcov2)))</pre>
#Thinning covariate
pcov <- RFsimulate(model = RMgauss(scale = 1), x = x0, y = y0, grid = TRUE)
pcov <- (pcov - min(pcov))/sd(pcov)</pre>
set.seed(NULL)
```

```
#Range of Matern covaraince
range <- 1.2
#Scale of Matern covariance
kappa <- sqrt(8)/range</pre>
#Variance of Matern covaraince
sigma2 <- 0.2
#Precision of Matern covariance
tau <- sqrt(4*pi*kappa*kappa*sigma2)</pre>
#Smoothness of Matern covaraince
nu <- 1
#Simulate spatial covariance
zcov <- RFsimulate(model = RMmatern(var = sigma2, scale = kappa, nu = nu), x</pre>
= x0, y = y0, grid = TRUE)
#Intensity function
X1 <- as.im(x1, W = Mwin)[win]
X2 < -as.im(x2, W = Mwin)[win]
z <- as.im(zcov, W = Mwin)[win]</pre>
lambda1 <- exp(beta[1] + beta[2] * X1 + z)
lambda2 <- exp(beta[1] + delta[1] + beta[2] * X2 + z)</pre>
#Format intensity function
lambda1 <- as.im(lambda1, W=win)</pre>
lambda2 <- as.im(lambda2, W=win)</pre>
#Simulate LGCP
PP1 <- rpoispp(lambda1)[win]</pre>
PP2 <- rpoispp(lambda2)[win]</pre>
#-----#
#-Simulate opportunistic sampling-#
#-----#
```

```
83
```

```
#Thinning function
thin <- alpha[1] + alpha[2] * pcov</pre>
thin <- expit(thin)</pre>
#Format thinning function
thin <- as.im(thin, W=Mwin)[win]</pre>
#Latent observations
Z1 <- rbinom(n = PP1$n, size = 1, prob = thin[PP1])</pre>
PO1 <- PP1[Z1==1]
PO1 <- PO1[win1]
Z2 <- rbinom(n = PP2$n, size = 1, prob = thin[PP2])
PO2 <- PP2[Z2==1]
PO2 <- PO2[win2]
#----#
#-Simulate counts-#
#----#
#Number of counts
ncount <- 25
#Count locations
u.loc <- expand.grid(seq(1 + 0.175, 2.75 - 0.175, 0.35), seq(0.25 + 0.175, 2
- 0.175, 0.35))
#Latent abundance
N <- NULL
#Counts
count <- rep(0, ncount)</pre>
```

```
#Covariate value @ location
Cov <- NULL
#Unit area
unit.A <- NULL
for(j in 1:ncount) {
  unit <- disc(radius = 0.15, centre = as.numeric(u.loc[j,]))</pre>
  count[j] <- PP1[unit]$n</pre>
  count[j] <- rbinom(1, count[j], 0.9)</pre>
  Cov[j] <- mean(as.im(x1, W=Mwin)[unit])</pre>
  unit.A[j] <- area(unit)</pre>
}
#Count dataframe
countdf <- data.frame(count, Cov, unit.A, u.loc)</pre>
#----#
#-SPDE approach-#
#----#
#Update mesh
mesh <- inla.mesh.2d(boundary = loc.d, max.edge = 0.3, cutoff = 0.05)</pre>
#Create SPDE objects (Matern covariance)
spde <- inla.spde2.matern(mesh)</pre>
#Create dual mesh for weights
dmesh <- book.mesh.dual(mesh)</pre>
#Format location domain
domain.polys <- Polygons(list(Polygon(loc.d)), '0')</pre>
domainSP <- SpatialPolygons(list(domain.polys))</pre>
```

```
85
```

```
#Weights of each node (area of dual mesh)
weight <- sapply(1:length(dmesh), function(i) {
    if (gIntersects(dmesh[i, ], domainSP))
        return(gArea(gIntersection(dmesh[i, ], domainSP)))
    else return(0)</pre>
```

})

rm(dmesh)

#-----# #-Compile Data-# #-----#

#Node index
nodes <- mesh\$n</pre>

#Observation index
nobs <- sum(PO1\$n, PO2\$n)
ncount <- dim(countdf)[1]</pre>

#Period index
t_i <- rep(0:1, c(PO1\$n, PO2\$n))
t n <- 2</pre>

#Counts
counts <- countdf\$count</pre>

#Area @ each count location
Area <- countdf\$unit.A</pre>

```
#Covaraites @ nodes
nloc <- as.ppp(mesh$loc[,1:2], W = Mwin)</pre>
```

```
nCov <- matrix(NA, nrow = nodes, ncol = t n)</pre>
nCov[,1] \leq as.im(x1, W = Mwin)[nloc]
nCov[,2] <- as.im(x2, W = Mwin)[nloc]</pre>
nBias <- as.im(pcov, W = Mwin)[nloc]</pre>
#Covariates @ obs
Cov <- c(as.im(x1, W = Mwin)[PO1])
         as.im(x2, W = Mwin)[PO2],
         countdf$Cov)
Bias <- c(as.im(pcov, W = Mwin)[PO1],
          as.im(pcov, W = Mwin)[PO2])
#Location of obsevations
locxy <- rbind(cbind(PO1$x, PO1$y)[,2:1],</pre>
               cbind(PO2$x, PO2$y)[,2:1],
               as.matrix(countdf[,4:5]))
#Projection matrix of observations
A <- as(inla.spde.make.A(mesh, locxy), "dgTMatrix")</pre>
#----#
#-Prediction-#
#----#
Grid <- as.matrix(expand.grid(seq(0.01,2.99,0.01), seq(0.01,2.99,0.01)))
npred <- dim(Grid)[1]</pre>
Apred <- as(inla.spde.make.A(mesh, Grid), "dgTMatrix")</pre>
predX <- matrix(NA, ncol = t n, nrow = npred)</pre>
predX[,1] <- interp.im(X1, x = Grid[,1], y = Grid[,2])
predX[,2] <- interp.im(X2, x = Grid[,1], y = Grid[,2])</pre>
```

#-----#

```
#-Integrated Model-#
#----#
#Compile TMB code (only once)
compile("Simulation.cpp")
#Load TMB code
dyn.load(dynlib("Simulation"))
#Compile data
data <- list("nodes" = nodes, "nobs" = nobs, "ncount" = ncount,</pre>
             "t i" = t i, "t n" = t n,
             "weight" = weight, "area" = Area, "A" = A, "counts" = counts,
             "nBias" = nBias, "nCov" = nCov, "Bias" = Bias, "Cov" = Cov,
             "spde" = spde$param.inla[c("M0","M1","M2")],
             "npred" = npred, "Apred" = Apred, "predX" = predX)
#PHASE 1: Fit fixed effects
#Parameters to estimate
params <- list("beta0" = 0, "beta1" = 1, "delta1" = 1,</pre>
               "alpha0" = logit(0.01), "alpha1" = 0,
               "log_kappa" = log(sqrt(8)/range), "log_tau_0" = 1,
               "omega" = rep(0, mesh$n))
#Define random effects
random = c("omega")
#Map
map <- list(</pre>
  "log_kappa" = as.factor(NA),
  "log_tau_0" = as.factor(NA),
  "omega" = factor(rep(NA, mesh$n)))
```

```
88
```

```
#Make AD objective function
obj1 <- MakeADFun(data = data, parameters = params, random = random, map =
map, DLL="Simulation")
#Trace parameters
obj1$env$tracepar <- TRUE</pre>
#Minimize objective function
opt1 <- nlminb(obj1$par, obj1$fn, obj1$gr)</pre>
#Calculate standard deviations
out1 <- sdreport(obj1)</pre>
#PHASE 2: Fit random effects
params <- list("beta0" = opt1$par[1], "beta1" = opt1$par[2], "delta1" =</pre>
opt1$par[3],
                "alpha0" = opt1$par[4], "alpha1" = opt1$par[5],
                "log kappa" = log(kappa), "log tau O" = 1,
                "omega" = rep(0, mesh$n))
#Map
map <- list(</pre>
  "beta0" = as.factor(NA),
  "beta1" = as.factor(NA),
  "delta1" = as.factor(NA),
  "alpha0" = as.factor(NA),
  "alpha1" = as.factor(NA))
```

```
obj2 <- MakeADFun(data = data, parameters = params, random = random, map =
map, DLL="Simulation")</pre>
```

#Trace parameters

#Make AD objective function

```
obj2$env$tracepar <- TRUE
#Minimize objective function
opt2 <- nlminb(obj2$par, obj2$fn, obj2$gr)</pre>
#Calculate standard deviations
out2 <- sdreport(obj2)</pre>
#Store output
output <- as.data.frame(round(cbind(Truth = c(beta, delta, alpha, PP1$n,
PP2$n, kappa, tau),
rbind(summary(out1)[!grepl("omega|sigma O|tau O|kappa",
rownames(summary(out1))),],
                                          summary(out2)[grepl("kappa|tau 0",
rownames(summary(out2))),])),
                 digits = 2))
colnames(output)[2:3] <- c("IM Estimate", "IM Std. Error")</pre>
#----#
#-Presence-only model-#
#----#
#Compile TMB code (only once)
compile("Simulation PO.cpp")
#Load TMB code
dyn.load(dynlib("Simulation PO"))
#Compile data
data <- list("nodes" = nodes, "nobs" = nobs, "t i" = t i, "t n" = t n,</pre>
             "weight" = weight, "A" = A[1:nobs,],
             "nBias" = nBias, "nCov" = nCov, "Bias" = Bias, "Cov" =
Cov[1:nobs],
             "spde" = spde$param.inla[c("M0", "M1", "M2")],
```

```
"npred" = npred, "Apred" = Apred, "predX" = predX)
#PHASE 1: Fit fixed effects
#Parameters to estimate
params <- list("beta0" = 0, "beta1" = 1, "delta1" = 1,</pre>
               "alpha0" = logit(0.01), "alpha1" = 0,
               "log kappa" = log(sqrt(8)/range), "log tau 0" = 1,
               "omega" = rep(0, mesh$n))
#Map
map <- list(</pre>
  "log_kappa" = as.factor(NA),
  "log tau O" = as.factor(NA),
  "omega" = factor(rep(NA, mesh$n)))
#Make AD objective function
obj3 <- MakeADFun(data = data, parameters = params, random = random, map =
map, DLL="Simulation PO")
#Trace parameters
obj3$env$tracepar <- TRUE
#Minimize objective function
opt3 <- nlminb(obj3$par, obj3$fn, obj3$gr)</pre>
#Calculate standard deviations
out3 <- sdreport(obj3)</pre>
params <- list("beta0" = opt1$par[1], "beta1" = opt1$par[2], "delta1" =</pre>
opt1$par[3],
               "alpha0" = opt1$par[4], "alpha1" = opt1$par[5],
               "log kappa" = log(sqrt(8)/range), "log tau O" = 1,
```

```
91
```

"omega" = rep(0, mesh\$n))

```
#Map
map <- list(</pre>
  "beta0" = as.factor(NA),
  "beta1" = as.factor(NA),
  "delta1" = as.factor(NA),
  "alpha0" = as.factor(NA),
  "alpha1" = as.factor(NA))
#Define random effects
random = c("omega")
#Make AD objective function
obj4 <- MakeADFun(data = data, parameters = params, map = map, random =
random, DLL="Simulation PO")
#Trace parameters
obj4$env$tracepar <- TRUE
#Minimize objective function
opt4 <- nlminb(obj4$par, obj4$fn, obj4$gr)</pre>
#Calculate standard deviations
out4 <- sdreport(obj4)</pre>
#Store output
output <- merge(output,</pre>
as.data.frame(round(rbind(summary(out3)[!grepl("omega|sigma_0|tau_0|kappa",
rownames(summary(out3))),],
                                            summary(out4)[grepl("kappa|tau O",
rownames(summary(out4))),]), digits = 2)),
                by='row.names', all=TRUE)
```

```
92
```

colnames(output)[5:6] <- c("PO Estimate", "PO Std. Error")</pre>

```
output$Row.names[7:8] <- c("N[1]", "N[2]")</pre>
rownames(output) <- output$Row.names</pre>
output <- output[,-1]</pre>
convergence <- data.frame("pdHess" = c(out1$pdHess, out2$pdHess, out3$pdHess,</pre>
out4$pdHess),
                            "grd.size" = c(all(abs(out1$gradient.fixed) <
0.01),
                                            all(abs(out2$gradient.fixed) <</pre>
0.01),
                                            all(abs(out3$gradient.fixed) <</pre>
0.01),
                                            all(abs(out4$gradient.fixed) <</pre>
0.01)))
rownames(convergence) <- c("IM_fixed", "IM_random", "PO_fixed", "PO_random")</pre>
out <- list(output, convergence)</pre>
message("finished")
#----#
#-Save file-#
#----#
ID <- length(list.files("./Output/")) + 1</pre>
save(out, file = paste("./Output/output", ID, ".Rds", sep=""))
tmp <- list(summary(out1), summary(out2), summary(out3), summary(out4))</pre>
save(tmp, file = "tmp out.Rds")
```

```
Template Model Builder code for simulation
```

```
#include <TMB.hpp>
// Integrated Log Gaussian Cox Process
template <class Type>
Type objective_function<Type>::operator() ()
{
```

```
// objective function -- joint negative log-likelihood
using namespace R_inla;
using namespace density;
using namespace Eigen;
```

// DATA //

```
// Indices
DATA_INTEGER( nodes ); // Number of nodes
DATA_INTEGER( nobs ); // Number of presence-only observations
DATA_INTEGER( ncount ); // Number of structured sampling sites
DATA_INTEGER( t_n ); // Number of stages
DATA_IVECTOR( t_i ); // Stage ID for data
```

```
// Projection & weight data
DATA_VECTOR( weight ); // Node weight
DATA_VECTOR( area ); // Area sweept for counts
DATA SPARSE MATRIX( A ); // Projection matrix
```

// Count dataset
DATA VECTOR(counts); // Number of counts per site

```
// Covaraite dataset
DATA_VECTOR( nBias ); //Sampling bias @ nodes
DATA MATRIX( nCov ); //Environmental covariate @ nodes
```

```
DATA VECTOR( Bias ); //Sampling bias @ observations
DATA VECTOR( Cov ); //Environmental covariate @ data
// SPDE objects
DATA STRUCT(spde, spde t); // Sparse matrix for Matern covariance structure
// Prediction
DATA_INTEGER( npred ); // Number of pixels for prediction
DATA MATRIX( predX );
                           // Environmental covariate for prediction
DATA SPARSE MATRIX( Apred ); // Projection matrix for prediction
// PARAMETERS //
// Fixed effects
PARAMETER( beta0 ); // Baseline population density
PARAMETER ( betal ); // Effect of environmental covariate
PARAMETER( delta1 ); // Effect of stage (population change)
PARAMETER( alpha0 ); // Thinning function intercept
PARAMETER( alpha1 ); // Effect of sampling bias
PARAMETER( log kappa ); // Scale parameter of Matern covariance
PARAMETER( log tau 0 ); // Precision parameter of Matern covariance
// Random effects
PARAMETER VECTOR( omega ); // Spatial random effect
// Population density at each stage
vector<Type> beta(t n);
beta(0) = beta0;
beta(1) = beta0 + delta1;
// Derived parameters
Type kappa = exp(log kappa);
```

```
Type tau O = \exp(\log tau O);
```

95

```
Type range = sqrt(8)/kappa;
  Type sigma O = 1/sqrt(4*PI*tau O*tau O*kappa*kappa);
  vector<Type> jnll_comp(4);
  jnll comp.setZero();
  // Probability of random effects
  SparseMatrix<Type> Q = Q spde(spde,kappa); // Matern covariance (see R inla
namespace)
  jnll comp(0) += GMRF(Q)( omega );
  // Holding values
  vector<Type> Omega(nodes);
  vector<Type> omg(nobs + ncount);
  // Transform GMRFs
  for(int k=0; k<nodes; k++) {</pre>
    Omega(k) = omega(k) / tau O;
  }
  // Project GMRFs
  omg = A * Omega; // Project omega to points
  // Intensity function @ nodes
  for(int t=0; t<t n; t++) {</pre>
    for(int k=0; k < nodes; k++) {
      jnll comp(1) += weight(k) * exp(alpha0 + alpha1 * nBias(k) + beta(t) +
beta1 * nCov(k,t) + Omega(k)) / (exp(alpha0 + alpha1 * nBias(k)) + 1); //
Integration nodes
    }
  }
  // Intensity function @ presence-only data
  for(int i=0; i<nobs; i++) {</pre>
```

```
jnll comp(2) -= alpha0 + alpha1 * Bias(i) + beta(t i(i)) + beta1 * Cov(i)
+ omg(i) - log(exp(alpha0 + alpha1 * Bias(i)) + 1); // Observation points
  }
  // Intensity function @ counts
  vector<Type> lambda(ncount);
  for(int j=0; j<ncount; j++) {</pre>
    lambda(j) = area(j) * exp(beta0 + beta1 * Cov(j+nobs) + omg(j+nobs));
    jnll comp(3) -= dpois(counts(j), lambda(j), true);
  }
  // Prediction
 vector<Type> predO(npred);
 pred0 = Apred * Omega;
 vector<Type> pred(t n);
  vector<Type> Npred(t_n);
  for(int t=0; t<t_n; t++) {</pre>
    for(int g=0; g<npred; g++) {</pre>
      pred(t) += exp(beta(t) + beta1 * predX(g,t) + predO(g));
    }
   Npred(t) = pred(t)/npred * 9;
  }
  // Joint NLL
  Type jnll = jnll_comp.sum();
  // Reporting
  REPORT ( kappa );
  REPORT( tau_0 );
  REPORT ( sigma O );
  REPORT ( range );
```

```
REPORT( Npred );
ADREPORT( Npred );
ADREPORT( kappa );
ADREPORT( tau_0 );
ADREPORT( sigma_0 );
REPORT( jnll_comp );
REPORT( jnll );
return jnll;
```

```
}
```
Simulation output

Table D.1

The root mean squared error and percent relative bias (median, lower and upper quantiles) of each estimated parameter compared to the true simulated value for both the integrated (I) and presence-only (PO) analyses.

| Parameters | Model | RMSE | Percent relative bias |
|-----------------------|-------|----------|--|
| p_0 | Ι | 1.50 | 30.22 (22.39, 38.04) |
| | PO | 13.57 | 332.72 (67.39, 353.97) |
| α | Ι | 0.16 | 3.53 (-9.41, 16.47) |
| | PO | 0.16 | 2.35 (-10.59, 14.12) |
| λ_0 | Ι | 0.45 | -2.08 (-7.00, 2.67) |
| | PO | 12.25 | 231.00 (28.54, 244.50) |
| β | Ι | 0.22 | 6.00 (-24.00, 36.00) |
| | PO | 0.20 | 10.00 (-12.00, 36.00) |
| δ | Ι | 0.34 | 28.00 (-10.00, 68.00) |
| | PO | 0.35 | 32.00 (-10.00, 72.00) |
| log (κ) | Ι | 0.20 | 19.77 (9.30, 27.91) |
| | PO | 0.16 | -15.12 (-23.26, -6.98) |
| σ | Ι | 0.61 | 300.00 (260.00, 340.00) |
| | PO | 0.81 | 400.00 (365.00, 440.00) |
| N_1 | Ι | 697.47 | -9.40 (-15.18, 3.40) |
| | PO | 1.41E+10 | $1.22 \times 10^8 (382.41, 2.61 \times 10^8)$ |
| <i>N</i> ₂ | Ι | 1662.19 | 3.41 (-8.29, 18.27) |
| | PO | 2.47E+10 | 1.35×10^8 (384.12, 3.23×10^8) |

APPENDIX E: Case study details, code, and output

Covariate descriptions

I provide some details on the summary of covariates for the case study. I acquired NDVI values from Terra Moderate Resolution Imaging Spectroradiometer (MODIS v.006) Vegetation Indices (MOD13A2) at a spatial resolution of 1-km² and a temporal resolution of 16 days (Didan 2015). At each presence-only location, s_i , and at each single-visit site, D_j , I extracted the mean NDVI value within a 10-km radius circle centered on each location during the appropriate 16-day window.

Temperature is also vital to the recruitment of monarchs as it influences the rates of development and survival from eggs to adults. I used the number of growing degree days (GDD) to describe the thermal conditions across the spring migratory pathway. GDD is the accumulation of temperature within a physiological threshold for development (McMaster & Wilhelm 1997). The GDD value accumulated for a given location is approximated as:

$$GDD = \sum_{n=1}^{N} \begin{cases} T_{min} & \text{if } \overline{T}_n < T_{min} \\ \overline{T}_n - T_{min} & \text{if } T_{min} \le \overline{T}_n \le T_{max} \\ T_{max} & \text{if } T_{max} < \overline{T}_n \end{cases}$$

where \overline{T}_n is the average temperature of a day n, T_{min} is the minimum temperature threshold (i.e., 11.5°C for monarchs), T_{max} is the maximum temperature threshold (i.e., 33°C for monarchs), and N is the number of days accumulated since n = 1. For my purposes I set n = 1 to be 7 days prior to the start of each stage. Average daily temperatures were obtained from Daymet (Thornton et al. 2020), which interpolates temperature values across North America at a 1 km² resolution. Since observations could have different number of days accumulated, N, depending on when they were seen during a given stage, I averaged GDD values to be averaged daily GDD by dividing by N.

I used two covariates to correct for sampling biases. I used the number of other butterfly observations within the superfamily *Papilionoidea* amalgamated from the various presence-only data sources used within the study. Other butterfly observations are correlated with where butterfly observers sample in space. I also used population density (CIESIN 2018) to correct for sampling bias as high human abundance is correlated with higher number of observers (Geldmann et al. 2016). This metric has also been used to correct for presence-only monarch data in another study (Momeni-Dehaghi et al. 2021).

Case study script

Attached here is the R script to run the case study analysis on monarch butterflies.

```
#----#
#-Libraries-#
#----#
library(tidyverse)
library(sf)
library(INLA)
library(TMB)
#Function to create dual mesh (original code from Krainski et al. 2018)
source("~/Monarchs/DataAnalysis/rDualMesh.R")
#----#
#-Load data-#
#----#
#Spring domain
domain sp <-
st read("~/Monarchs/DataFormatting/BaseData/Domain/Domain Spring.shp")
#Summer domain
domain su <-
st read("~/Monarchs/DataFormatting/BaseData/Domain/Domain Summer.shp")
#Monarch data
load(file = "~/Monarchs/DataFormatting/FormattedData.Rdata")
#Mesh node data
load(file = '~/Monarchs/DataFormatting/FormattedNodeData.Rdata')
#----#
#-Format data-#
#----#
#Geographic projection
```

```
prj <- "+proj=aea +lat 1=29.5 +lat 2=45.5 +lat 0=37.5 +lon 0=-96 +x 0=0
+y 0=0 +ellps=GRS80 +datum=NAD83 +units=km +no defs"
#Set projection
domain sp <- st transform(domain sp, crs = st crs(prj))</pre>
domain su <- st transform(domain su, crs = st crs(prj))</pre>
#Spring domain boundary
domain sp seg <- inla.sp2segment(as(domain sp, "Spatial"))</pre>
#Summer domain boundary
domain su seg <- inla.sp2segment(as(domain su, "Spatial"))</pre>
#Construct triangulated mesh
mesh <- list()</pre>
mesh[[1]] <- inla.mesh.2d(boundary=domain sp seg,</pre>
                            max.edge = 75,
                            cutoff=30)
mesh[[2]] <- inla.mesh.2d(boundary=domain sp seg,</pre>
                            max.edge = 75,
                            cutoff=30)
mesh[[3]] <- inla.mesh.2d(boundary=domain su seq,</pre>
                            max.edge = 75,
                            cutoff = 30)
#Create SPDE objects (Matern covariance structure)
spde1 <- inla.spde2.matern(mesh[[1]])</pre>
spde2 <- inla.spde2.matern(mesh[[2]])</pre>
spde3 <- inla.spde2.matern(mesh[[3]])</pre>
#Create dual mesh for weights (see Krainski et al. 2018)
dmesh1 <- book.mesh.dual(mesh[[1]])</pre>
dmesh1 <- st as sf(dmesh1)</pre>
st crs(dmesh1) <- st_crs(domain_sp)</pre>
dmesh2 <- book.mesh.dual(mesh[[2]])</pre>
dmesh2 <- st as sf(dmesh2)</pre>
st crs(dmesh2) <- st crs(domain sp)</pre>
dmesh3 <- book.mesh.dual(mesh[[3]])</pre>
dmesh3 <- st as sf(dmesh3)</pre>
st crs(dmesh3) <- st crs(domain su)</pre>
#Weights of each node (area of dual mesh)
weight1 <- sapply(1:dim(dmesh1)[1], function(i) {</pre>
  if (st intersects(dmesh1[i,], domain sp, sparse = FALSE))
    return(st area(st intersection(dmesh1[i,], domain sp)))
  else return(0)
})
weight2 <- sapply(1:dim(dmesh2)[1], function(i) {</pre>
  if (st intersects(dmesh2[i,], domain sp, sparse = FALSE))
    return(st area(st intersection(dmesh2[i,], domain sp)))
```

```
else return(0)
})
weight3 <- sapply(1:dim(dmesh3)[1], function(i) {</pre>
  if (st intersects(dmesh3[i,], domain su, sparse = FALSE))
    return(st area(st intersection(dmesh3[i,], domain su)))
  else return(0)
})
#Projection matrix of observations to nodes
A1 <- as(inla.spde.make.A(mesh[[1]], st coordinates(Data %>% filter(period ==
1))), "dgTMatrix")
A2 <- as(inla.spde.make.A(mesh[[2]], st coordinates(Data %>% filter(period ==
2))), "dgTMatrix")
A3 <- as(inla.spde.make.A(mesh[[3]], st coordinates(Data %>% filter(period ==
3))), "dgTMatrix")
#----#
#-Compile data-#
#----#
#Number of years
t n <- Data %>% summarize(out = n distinct(yr)) %>% select(out) %>% .$out
#Number of stages
p n <- Data %>% summarize(out = n distinct(period)) %>% select(out) %>% .$out
#Number of nodes in early spring
nodes1 <- mesh[[1]]$n</pre>
#Number of nodes in late spring
nodes2 <- mesh[[2]]$n</pre>
#Number of nodes in early summer
nodes3 <- mesh[[3]]$n</pre>
#Number of nodes across stages and years
nodes <- as.integer(t n*(nodes1 + nodes2 + nodes3))</pre>
#MOVE TO NODE FORMATTING
NodeDF$period <- as.factor(rep(1:3, c(nodes1 * 3, nodes2 * 3, nodes3 * 3)))
#Year-stage identifier for nodes
tp k <- as.integer(fct cross(NodeDF$period, NodeDF$Year)) - 1</pre>
#Stage identifier for nodes
p k <- as.integer(NodeDF$period) - 1</pre>
#Number of presence-only observations
nobs <- dim(Data %>% filter(type == "obs"))[1]
#Number of presence-only observations in early spring
nobs1 <- dim(Data %>% filter(type == "obs" & period == 1))[1]
#Number of presence-only observations in late spring
nobs2 <- dim(Data %>% filter(type == "obs" & period == 2))[1]
```

#Number of presence-only observations in early summer nobs3 <- dim(Data %>% filter(type == "obs" & period == 3))[1] #Number of single-visit sites ncount <- dim(Data %>% filter(type == "count"))[1] #Number of single-visit sites in early spring ncount1 <- dim(Data %>% filter(type == "count" & period == 1))[1] #Number of single-visit sites in late spring ncount2 <- dim(Data %>% filter(type == "count" & period == 2))[1] #Number of single-visit sites in early summer ncount3 <- dim(Data %>% filter(type == "count" & period == 3))[1] #Single-visit counts @ each site counts <- as.numeric(Data %>% filter(type == "count") %>% select(count) %>% .\$count) #Year-stage identifier for data tp i <- as.integer(fct cross(as.factor(Data\$period), as.factor(Data\$yr))) - 1</pre> #CHECK THIS #Stage identifier for data p i <- as.integer(as.factor(Data\$period)) - 1</pre> #Weight of each node @ 100 m^2 weight <- rep(c(weight1, weight2, weight3), t n) * 1e4</pre> #Area/effort offset area <- as.numeric(Data %>% filter(type == "count") %>% select(effort) %>% .\$effort) #Covert to 100 m^2 area <- area*6*1000*1e-2 #Observation covariates Data <- Data %>% mutate(daysaccum = as.numeric(as.Date(paste0(yr, "-", mo day)) as.Date(paste0(yr, ifelse(period == 1, "-03-01", ifelse(period == 2, "-03-22", "-04-26"))))) + 1, gdd.avg = gdd2/daysaccum)#Normalized Difference Vegetation Index @ data NDVI <- Data\$NDVI #Averaged daily growing degree days @ data GDD <- Data\$gdd.avg #Number of other butterfly observations @ data Bias <- as.numeric(Data %>% filter(type == "obs") %>% select(Bias) %>% .\$Bias) #Population density @ data

```
PopD <- as.numeric(Data %>% filter(type == "obs") %>% select(PopD) %>%
.$PopD)
#Node covariates
NodeDF <- NodeDF %>% mutate(daysaccum = ifelse(period == 1, 35, 42),
                             gdd.avg = gdd2/daysaccum)
#Normalized Difference Vegetation Index @ nodes
nNDVI <- NodeDF$NDVI
#Averaged daily growing degree days @ nodes
nGDD <- NodeDF$gdd.avg
#Number of other butterfly observations @ nodes
nBias <- NodeDF$Bias
#Population density @ nodes
nPopD <- NodeDF$PopD
#Scale covaraites for both data & nodes
tmp1 <- (NDVI - mean(c(NDVI, nNDVI)))/sd(c(NDVI, nNDVI))</pre>
tmp2 <- (GDD - mean(c(GDD, nGDD)))/sd(c(GDD, nGDD))</pre>
tmp3 <- (Bias - mean(c(Bias, nBias)))/sd(c(Bias, nBias))</pre>
tmp4 <- (PopD - mean(c(PopD, nPopD)))/sd(c(PopD, nPopD))</pre>
tmp5 <- (nNDVI - mean(c(NDVI, nNDVI)))/sd(c(NDVI, nNDVI))</pre>
tmp6 <- (nGDD - mean(c(GDD, nGDD)))/sd(c(GDD, nGDD))</pre>
tmp7 <- (nBias - mean(c(Bias, nBias)))/sd(c(Bias, nBias))</pre>
tmp8 <- (nPopD - mean(c(PopD, nPopD)))/sd(c(PopD, nPopD))</pre>
NDVI <- tmp1
GDD <- tmp2
Bias <- tmp3
PopD <- tmp4
nNDVI <- tmp5
nGDD <- tmp6
nBias <- tmp7
nPopD <- tmp8
#----#
#-Optimize nll-#
#----#
#Compile TMB code (only once)
# compile("./DataAnalysis/IntegratedModel Final.cpp")
#Load TMB code
dyn.load(dynlib("./DataAnalysis/IntegratedModel Final"))
#PHASE 1: Fit fixed effects
#Compile data for TMB
data <- list("nodes1" = nodes1, "nodes2" = nodes2, "nodes3" = nodes3,</pre>
             "nobs1" = nobs1, "nobs2" = nobs2, "nobs3" = nobs3,
```

```
"ncount" = ncount, "ncount1" = ncount1, "ncount2" = ncount2,
"ncount3" = ncount3,
             "counts" = counts,
             "t n" = t n, "p n" = p n, "tp k" = tp k, "tp i" = tp i, "p i" =
pi, "pk" = p\overline{k},
             "weight" = weight, "area" = area,
             "A1" = A1, "spde1" = spde1$param.inla[c("M0", "M1", "M2")],
             "A2" = A2, "spde2" = spde2$param.inla[c("M0", "M1", "M2")],
             "A3" = A3, "spde3" = spde3$param.inla[c("M0", "M1", "M2")],
             "nBias" = nBias, "nPopD" = nPopD,
             "nNDVI" = nNDVI, "nGDD" = nGDD,
             "Bias" = Bias, "PopD" = PopD,
             "NDVI" = NDVI, "GDD" = GDD)
#Parameters to estimate
params <- list("beta0" = 0,
               "beta1" = rep(0,3), "beta2" = rep(0,3),
               "beta3" = rep(0,3), "beta4" = rep(0,3),
               "delta1" = 0, "delta2" = 0,
               "gamma1" = 0, "gamma2" = 0,
               "alpha0" = 0, "alpha1" = rep(0,9), "alpha2" = rep(0,9),
               "log kappa" = 0, "log tau" = 1,
               "omegal" = rep(0, nodes1),
               "omega2" = rep(0, nodes2),
               "omega3" = rep(0, nodes3))
#Hold random effects constant
map <- list(</pre>
  "log kappa" = as.factor(NA),
  "log tau" = as.factor(NA),
  "omega1" = factor(rep(NA, nodes1)),
  "omega2" = factor(rep(NA, nodes2)),
  "omega3" = factor(rep(NA, nodes3))
)
#Random effects
random <- c("omega1", "omega2", "omega3")</pre>
#Make AD objective function
obj1 <- MakeADFun(data = data, parameters = params, random = random, map =
map, DLL="IntegratedModel Final")
#Trace parameters
obj1$env$tracepar <- TRUE
#Minimize objective function
opt1 <- nlminb(obj1$par, obj1$fn, obj1$gr)</pre>
#AIC
2 * length(obj1$par) - 2 * (-1 * sum(obj1$report()$jnll comp[3:12]))
#Output
out1 <- sdreport(obj1)</pre>
summary(out1)
#Extract parameter values
beta0 <- out1$par.fixed[1]</pre>
```

```
beta1 <- out1$par.fixed[2:4]</pre>
beta2 <- out1$par.fixed[5:7]</pre>
beta3 <- out1$par.fixed[8:10]</pre>
beta4 <- out1$par.fixed[11:13]</pre>
delta1 <- out1$par.fixed[14]</pre>
delta2 <- out1$par.fixed[15]</pre>
gamma1 <- out1$par.fixed[16]</pre>
gamma2 <- out1$par.fixed[17]</pre>
#Expected population densities per stage and year
beta <- NULL
#2016 stage 1 (early spring)
beta[1] <- beta0 + gamma1 + gamma2 + delta1 + delta2</pre>
#2016 stage 2 (late spring)
beta[2] <- beta0 + gamma1 + delta1 + delta2</pre>
#2016 stage 3 (early summer)
beta[3] <- beta0 + delta1 + delta2</pre>
#2017 stage 1 (early spring)
beta[4] <- beta0 + gamma1 + gamma2 + delta1</pre>
#2017 stage 2 (late spring)
beta[5] <- beta0 + gamma1 + delta1</pre>
#2017 stage 3 (early summer)
beta[6] <- beta0 + delta1</pre>
#2018 stage 1 (early spring)
beta[7] <- beta0 + gamma1 + gamma2</pre>
#2018 stage 2 (late spring)
beta[8] <- beta0 + gamma1</pre>
#2018 stage 3 (early summer)
beta[9] <- beta0</pre>
#Compile output
Output <- data.frame(year = factor(rep(2016:2018, each = 3)),</pre>
                       period = factor(rep(1:3, 3)),
                       mean.den = beta)
Output$lower.den <- NA
Output$upper.den <- NA
#Profile confidence intervals
Output[1,4:5] <- confint(tmbprofile(obj1, lincomb =</pre>
as.numeric(names(obj1$par) %in%
c("beta0", "gamma1", "gamma2", "delta1", "delta2"))))
Output[2,4:5] <- confint(tmbprofile(obj1, lincomb =</pre>
as.numeric(names(obj1$par) %in% c("beta0","gamma1","delta1","delta2"))))
Output[3,4:5] <- confint(tmbprofile(obj1, lincomb =</pre>
as.numeric(names(obj1$par) %in% c("beta0","delta1","delta2"))))
Output[4,4:5] <- confint(tmbprofile(obj1, lincomb =</pre>
as.numeric(names(obj1$par) %in% c("beta0","gamma1","gamma2","delta1"))))
```

```
Output[5,4:5] <- confint(tmbprofile(obj1, lincomb =</pre>
as.numeric(names(obj1$par) %in% c("beta0","gamma1","delta1"))))
Output[6,4:5] <- confint(tmbprofile(obj1, lincomb =</pre>
as.numeric(names(obj1$par) %in% c("beta0","delta1"))))
Output[7,4:5] <- confint(tmbprofile(obj1, lincomb =</pre>
as.numeric(names(obj1$par) %in% c("beta0","gamma1","gamma2"))))
Output[8,4:5] <- confint(tmbprofile(obj1, lincomb =</pre>
as.numeric(names(obj1$par) %in% c("beta0","gamma1"))))
Output[9,4:5] <- confint(tmbprofile(obj1, lincomb =</pre>
as.numeric(names(obj1$par) %in% c("beta0"))))
#PHASE 2: Fit random effects
#Parameters to estimate
params2 <- list("beta0" = beta0, "beta1" = beta1,</pre>
                 "beta2" = beta2, "beta3" = beta3,
                 "beta4" = beta4,
                 "delta1" = delta1, "delta2" = delta2,
                 "gamma1" = gamma1, "gamma2" = gamma2,
                 "alpha0" = out1$par.fixed[18],
                 "alpha1" = out1$par.fixed[19:27],
                 "alpha2" = out1$par.fixed[28:36],
                 "log_kappa" = 0, "log_tau" = 1,
"omega1" = rep(0, nodes1), "omega2" = rep(0, nodes2),
"omega3" = rep(0, nodes3))
#Hold fixed effects constant
map2 <- list(</pre>
  "alpha0" = as.factor(NA),
  "alpha1" = factor(rep(NA, 9)),
  "alpha2" = factor(rep(NA, 9)),
  "beta0" = as.factor(NA),
  "beta1" = factor(rep(NA, 3)),
  "beta2" = factor(rep(NA, 3)),
  "beta3" = factor(rep(NA, 3)),
  "beta4" = factor(rep(NA, 3)),
  "delta1" = as.factor(NA),
  "delta2" = as.factor(NA),
  "gamma1" = as.factor(NA),
  "gamma2" = as.factor(NA)
)
#Make AD objective function
obj2 <- MakeADFun(data = data, parameters = params2, map = map2, random =
random, DLL="IntegratedModel Final")
#Trace parameters
obj2$env$tracepar <- TRUE
#Minimize objective function
opt2 <- nlminb(obj2$par, obj2$fn, obj2$gr)</pre>
#Output
out2 <- sdreport(obj2)</pre>
#Estimated parameter values
```

```
Effect <- data.frame(param = rep(c("beta1", "beta2", "beta3", "beta4"), each
= 3),
                      period = factor(rep(1:3, 4)),
                      mean.den = c(beta1, beta2, beta3, beta4))
Effect$lower <- NA
Effect$upper <- NA
Effect[1,4:5] <- confint(tmbprofile(obj1, lincomb = c(0,1,rep(0,34))))</pre>
Effect[2,4:5] <- confint(tmbprofile(obj1, lincomb = c(0,0,1,rep(0,33))))</pre>
Effect[3,4:5] <- confint(tmbprofile(obj1, lincomb = c(0,0,0,1,rep(0,32))))
Effect[4,4:5] <- confint(tmbprofile(obj1, lincomb = c(0,0,0,0,1,rep(0,31))))</pre>
Effect[5,4:5] <- confint(tmbprofile(obj1, lincomb =</pre>
c(0,0,0,0,0,1,rep(0,30))))
Effect[6,4:5] <- confint(tmbprofile(obj1, lincomb = c(rep(0,6),1,rep(0,29))))
Effect[7,4:5] <- confint(tmbprofile(obj1, lincomb = c(rep(0,7), 1, rep(0,28))))
Effect[8,4:5] <- confint(tmbprofile(obj1, lincomb = c(rep(0,8),1,rep(0,27))))</pre>
Effect[9,4:5] <- confint(tmbprofile(obj1, lincomb = c(rep(0,9),1,rep(0,26))))
Effect[10,4:5] <- confint(tmbprofile(obj1, lincomb =</pre>
c(rep(0,10),1,rep(0,25))))
Effect[11,4:5] <- confint(tmbprofile(obj1, lincomb =</pre>
c(rep(0,11), 1, rep(0,24))))
Effect[12,4:5] <- confint(tmbprofile(obj1, lincomb =</pre>
c(rep(0, 12), 1, rep(0, 23))))
```

Template Model Builder code for case study

The TMB code to generate the objective function containing the negative log likelihood for

optimization.

```
#include <TMB.hpp>
// Integrated Log Gaussian Cox Process
template <class Type>
Type objective function<Type>::operator() ()
{
  // Objective function -- joint negative log-likelihood
  using namespace R inla;
  using namespace density;
  using namespace Eigen;
  // Indices
  DATA INTEGER( nodes1 ); // Number of nodes in early spring
  DATA INTEGER( nodes2 ); // Number of nodes in late spring
  DATA INTEGER( nodes3 ); // Number of nodes in early summer
  DATA INTEGER( nobs1 ); // Number of presence-only observations in early
spring
                          // Number of presence-only observations in late
  DATA INTEGER ( nobs2 );
spring
                          // Number of presence-only observations in early
 DATA INTEGER( nobs3 );
summer
  DATA INTEGER( ncount ); // Number of structured sampling sites
  DATA INTEGER ( ncount1 ); // Number of structured sampling sites in early
spring
```

DATA INTEGER(ncount2); // Number of structured sampling sites in late spring DATA INTEGER (ncount3); // Number of structured sampling sites in early summer DATA INTEGER(t n); // Number of years DATA INTEGER (p n); // Number of stages DATA IVECTOR(tp i); // Index for year and stage combination for the data // Index for year and stage combination for the DATA IVECTOR (tp k); nodes DATA IVECTOR(p i); // Stage ID for the data DATA IVECTOR (p k); // Stage ID for the nodes // Projection & weight data DATA VECTOR (weight); // Node weight DATA VECTOR(area); // Area sweept for counts DATA SPARSE MATRIX(A1); // Projection matrix early spring DATA SPARSE MATRIX(A2); // Projection matrix late spring DATA SPARSE MATRIX(A3); // Projection matrix early summer //Count data DATA VECTOR(counts); // Number of counts per site // Covaraite dataset DATA VECTOR(nBias); // Number of other butterfly observations @ nodes DATA VECTOR(nPopD); // Human population density @ nodes DATA VECTOR (nNDVI); // NDVI @ nodes DATA VECTOR (nGDD); // GDD @ nodes DATA VECTOR(Bias); // Number of other butterfly observations @ data DATA VECTOR (PopD); // Human population density @ data DATA VECTOR (NDVI); // NDVI @ data DATA VECTOR (GDD); // GDD @ data // SPDE objects DATA STRUCT(spde1, spde t); // Sparse matrix for Matern covariance structure in early spring DATA STRUCT(spde2, spde t); // Sparse matrix for Matern covariance structure in late spring DATA STRUCT(spde3, spde t); // Sparse matrix for Matern covariance structure in early summer // Fixed effects // Baseline population density PARAMETER(beta0); PARAMETER VECTOR(betal); // Effect of NDVI PARAMETER_VECTOR(beta2); // Quadratic effect of NDVI PARAMETER VECTOR(beta3); // Effect of GDD PARAMETER VECTOR(beta4); // Quadratic effect of GDD // Effect of 2016 PARAMETER(delta1); // Effect of 2017 PARAMETER(delta2); // Effect of early spring PARAMETER (gamma1); // Effect of late spring PARAMETER(gamma2); // Thinning function intercept PARAMETER(alpha0); PARAMETER VECTOR(alpha1); // Effect of number of other butterfly observations PARAMETER VECTOR(alpha2); // Effect of human population density PARAMETER (log kappa); // Scale parameter of Matern covariance

```
PARAMETER( log tau ); // Precision parameter of Matern covariance
// Random effects
PARAMETER VECTOR( omegal ); // Spatial random effect in early spring
PARAMETER VECTOR( omega2 ); // Spatial random effect in late spring
PARAMETER VECTOR( omega3 ); // Spatial random effect in early summer
// Population density at each stage and year
vector<Type> beta(t n * p n);
beta(0) = beta0 + gamma1 + gamma2 + delta1 + delta2; //2016 early spring
beta(1) = beta0 + gamma1 + delta1 + delta2; //2016 late spring
beta(2) = beta0 + delta1 + delta2;
                                                   //2016 early summer
                                                //2017 early spring
beta(3) = beta0 + gamma1 + gamma2 + delta1;
beta(4) = beta0 + gamma1 + delta1;
                                                    //2017 late spring
beta(5) = beta0 + delta1;
                                                    //2017 early summer
beta(6) = beta0 + gamma1 + gamma2;
                                                    //2018 early spring
beta(7) = beta0 + gamma1;
                                                     //2018 late spring
beta(8) = beta0;
                                                    //2018 early summer
// Derived parameters for computational purposes
Type kappa = exp(log kappa);
Type tau = exp(log tau);
Type range = sqrt(8)/kappa;
Type sigma = 1/sqrt(4*PI*tau*tau*kappa*kappa);
vector<Type> jnll comp(12);
jnll comp.setZero();
// Probability of random effects
SparseMatrix<Type> Q1 = Q_spde(spde1,kappa);
SparseMatrix<Type> Q2 = Q_spde(spde2,kappa);
SparseMatrix<Type> Q3 = Q_spde(spde3, kappa);
jnll comp(0) += GMRF(Q1)( omega1 );
jnll comp(1) += GMRF(Q2)( omega2 );
jnll comp(2) += GMRF(Q3)( omega3 );
// Holding values
vector<Type> Omega1(nodes1);
vector<Type> Omega2(nodes2);
vector<Type> Omega3(nodes3);
vector<Type> omg1(nobs1 + ncount1);
vector<Type> omg2(nobs2 + ncount2);
vector<Type> omg3(nobs3 + ncount3);
// Transform GMRFs
//Early spring
for(int k=0; k<nodes1; k++) {</pre>
  Omega1(k) = omega1(k) / tau;
//Late spring
for(int k=0; k<nodes2; k++) {</pre>
  Omega2(k) = omega2(k) / tau;
}
//Summer
for(int k=0; k < nodes3; k++) {
```

```
111
```

```
Omega3(k) = omega3(k) / tau;
  }
  // Project GMRFs
  omg1 = A1 * Omega1;
  omg2 = A2 * Omega2;
  omg3 = A3 * Omega3;
  // Intensity function @ nodes
  for(int t=0, j=0, i=nodes1*t n, g=(nodes1+nodes2)*t n; t<t n; t++) {</pre>
    for(int k=0; k<nodes1; k++, j++) {</pre>
      jnll comp(3) += weight(j) * exp(alpha0 + alpha1(tp k(j)) * nBias(j) +
alpha2(tp k(j)) * nPopD(j) + beta(tp k(j)) + beta1(p k(j)) * nNDVI(j) +
beta2(p k(j)) * nNDVI(j) * nNDVI(j) + beta3(p k(j)) * nGDD(j) + beta4(p k(j))
* nGDD(j) * nGDD(j) + Omegal(k)) / (exp(alpha0 + alpha1(tp k(j)) * nBias(j) +
alpha2(tp k(j)) * nPopD(j)) + 1);
    }
    for(int k=0; k<nodes2; k++, i++) {</pre>
      jnll comp(4) += weight(i) * exp(alpha0 + alpha1(tp k(i)) * nBias(i) +
alpha2(tp k(i)) * nPopD(i) + beta(tp k(i)) + beta1(p k(i)) * nNDVI(i) +
beta2(p k(i)) * nNDVI(i) * nNDVI(i) + beta3(p k(i)) * nGDD(i) + beta4(p k(i))
* nGDD(i) * nGDD(i) + Omega2(k)) / (exp(alpha0 + alpha1(tp k(i)) * nBias(i) +
alpha2(tp_k(i)) * nPopD(i)) + 1);
    }
    for(int k=0; k<nodes3; k++, g++) {</pre>
      jnll comp(5) += weight(g) * exp(alpha0 + alpha1(tp k(g)) * nBias(g) +
alpha2(tp k(g)) * nPopD(g) + beta(tp k(g)) + beta1(p k(g)) * nNDVI(g) +
beta2(p k(g)) * nNDVI(g) * nNDVI(g) + beta3(p k(g)) * nGDD(g) + beta4(p k(g))
* nGDD(g) * nGDD(g) + Omega3(k)) / (exp(alpha0 + alpha1(tp k(g)) * nBias(g) +
alpha2(tp k(g)) * nPopD(g)) + 1);
    }
  }
  // Intensity function @ presence-only observations
  int h=0;
  for(int i=0; i<nobs1; i++, h++) {</pre>
    jnll_comp(6) -= alpha0 + alpha1(tp_i(h)) * Bias(h) + alpha2(tp_i(h)) *
PopD(h) + beta(tp i(h)) + beta1(p i(h)) * NDVI(h) + beta2(p i(h)) * NDVI(h) *
NDVI(h) + beta3(p i(h)) * GDD(h) + beta4(p i(h)) * GDD(h) * GDD(h) + omq1(i)
-\log(\exp(alpha0 + alpha1(tp i(h)) * Bias(h) + alpha2(tp i(h)) * PopD(h)) +
1);
 }
  for(int i=0; i<nobs2; i++, h++) {</pre>
    jnll comp(7) -= alpha0 + alpha1(tp i(h)) * Bias(h) + alpha2(tp i(h)) *
PopD(h) + beta(tp i(h)) + beta1(p i(h)) * NDVI(h) + beta2(p i(h)) * NDVI(h) *
NDVI(h) + beta3(p i(h)) * GDD(h) + beta4(p i(h)) * GDD(h) * GDD(h) + omg2(i)
- log(exp(alpha0 + alpha1(tp i(h)) * Bias(h) + alpha2(tp i(h)) * PopD(h)) +
1);
  }
  for(int i=0; i<nobs3; i++, h++) {</pre>
    jnll comp(8) = alpha0 + alpha1(tp i(h)) * Bias(h) + alpha2(tp i(h)) *
PopD(h) + beta(tp i(h)) + beta1(p i(h)) * NDVI(h) + beta2(p i(h)) * NDVI(h) *
NDVI(h) + beta3(p i(h)) * GDD(h) + beta4(p i(h)) * GDD(h) * GDD(h) + omg3(i)
```

```
- log(exp(alpha0 + alpha1(tp i(h)) * Bias(h) + alpha2(tp i(h)) * PopD(h)) +
1);
 }
  // Intensity function @ structured sampling sites
 vector<Type> lambda(ncount);
  int f=0;
  for(int i=0; i<ncount1; i++, h++, f++) {</pre>
    lambda(f) = area(f) * exp(beta(tp i(h)) + beta1(p i(h)) * NDVI(h) +
beta2(p i(h)) * NDVI(h) * NDVI(h) + beta3(p i(h)) * GDD(h) + beta4(p i(h)) *
GDD(h) * GDD(h) + omg1(i+nobs1));
    jnll comp(9) -= dpois(counts(f), lambda(f), true);
  }
  for(int i=0; i<ncount2; i++, h++, f++) {</pre>
    lambda(f) = area(f) * exp(beta(tp i(h)) + betal(p i(h)) * NDVI(h) +
beta2(p i(h)) * NDVI(h) * NDVI(h) + beta3(p i(h)) * GDD(h) + beta4(p_i(h)) *
GDD(h) * GDD(h) + omg2(i+nobs2));
    jnll comp(10) -= dpois(counts(f), lambda(f), true);
  }
  for(int i=0; i<ncount3; i++, h++, f++) {</pre>
    lambda(f) = area(f) * exp(beta(tp i(h)) + betal(p i(h)) * NDVI(h) +
beta2(p_i(h)) * NDVI(h) * NDVI(h) + beta3(p_i(h)) * GDD(h) + beta4(p_i(h)) *
GDD(h) * GDD(h) + omg3(i+nobs3));
    jnll comp(11) -= dpois(counts(f), lambda(f), true);
  }
  // Joint NLL
  Type jnll = jnll comp.sum();
  // Reporting
  REPORT( tau );
  REPORT ( kappa );
  REPORT( sigma );
  ADREPORT( tau );
  ADREPORT ( kappa );
 ADREPORT ( sigma );
 REPORT( jnll comp );
 REPORT ( jnll );
  return jnll;
}
```

Model output

Table E.1

Estimates (mean and 95% confidence intervals) of monarch population density per 100-m^2 for

each year and stage.

| Year | Stage | Population density per 100-m ² |
|------|--------------|---|
| 2016 | Early Spring | 0.001 (CI 0.001, 0.002) |
| 2016 | Late Spring | 0.002 (CI 0.001, 0.002) |
| 2016 | Early Summer | 0.003 (CI 0.003, 0.004) |
| 2017 | Early Spring | 0.004 (CI 0.003, 0.005) |
| 2017 | Late Spring | 0.004 (CI 0.004, 0.006) |
| 2017 | Early Summer | 0.009 (CI 0.008, 0.011) |
| 2018 | Early Spring | 0.010 (CI 0.008, 0.012) |
| 2018 | Late Spring | 0.012 (CI 0.010, 0.015) |
| 2018 | Early Summer | 0.022 (CI 0.020, 0.025) |

Table E.2

Estimates (mean and 95% confidence intervals) of the effect of NDVI and averaged daily GDD

| Covariate | Stage | Effect (log-scale) |
|-------------------|--------------|-------------------------|
| NDVI | Early Spring | 0.22 (CI 0.09, 0.36) |
| NDVI | Late Spring | 1.87 (CI 1.63, 2.14) |
| NDVI | Early Summer | 0.41 (CI 0.30, 0.52) |
| NDVI ² | Early Spring | 0.23 (CI 0.16, 0.29) |
| NDVI ² | Late Spring | -1.62 (CI -1.87, -1.40) |
| NDVI ² | Early Summer | -0.33 (CI -0.42, -0.24) |
| GDD | Early Spring | 0.97 (CI 0.73, 1.23) |
| GDD | Late Spring | 0.56 (CI 0.31, 0.82) |
| GDD | Early Summer | 0.02 (CI -0.13, 0.17) |
| GDD^2 | Early Spring | -1.17 (CI -0.13, 0.17) |
| GDD^2 | Late Spring | -0.89 (CI -0.94, -0.67) |
| GDD^2 | Early Summer | -0.09 (CI -0.21, 0.02) |

per each stage on monarch population density on the log-scale.

BIBLIOGRAPHY

BIBLIOGRAPHY

- Abadi, F., Gimenez, O., Arlettaz, R., and Schaub, M. (2010) An assessment of integrated population models: bias, accuracy, and violation of the assumption of independence. Ecology, 91: 7-14.
- Agrawal, A.A., and Inamine, H. (2018) Mechanisms behind the monarch's decline. Science, 360: 1294-1296.
- Bradshaw, C.J.A., Sodhi, N.S. & Brook, B.W. (2009) Tropical turmoil: a biodiversity tragedy in progress. Frontiers in Ecology and the Environment, 7: 79-87.
- Brower, L.P., Taylor, O.R., Williams, E.H., Slayback, D.A., Zubieta, R.R., and Ramírez, M.I. (2012) Decline of monarch butterflies overwintering in Mexico: is the migratory phenomenon at risk? Insect Conservation and Diversity, 5: 95-100.
- Browning, R.C., Baker, E.A., Herron, J.A., & Kram, R. (2006) Effects of obesity and sex on the energetic cost and preferred speed of walking. Journal of Applied Physiology, 100: 390-398.
- Ceballos, G., Ehrlich, P.R., Barnosky, A.D., García, A., Pringle, R.M. & Palmer T.M. (2015). Accelerated modern human-induced species losses: Entering the sixth mass extinction. Science Advances, 1: e1400253.
- Center for International Earth Science Information Network CIESIN Columbia University (2018) Gridded Population of the World, Version 4 (GPWv4): Population Density, Revision 11. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). DOI: 10.7927/H49C6VHW.
- Chandler, M., See, L., Copas, K., Bonde, A.M.Z., López, B.C., Danielsen, F., ... Turak, E. (2017) Contribution of citizen science towards international biodiversity monitoring. Biological Conservation, 213, 280-294.
- Collen, B., Ram, M., Zamin, T. & McRae, L. (2008) The tropical biodiversity data gap: addressing disparity in global monitoring. Tropical Conservation Science, 1: 75-88.
- Conde, D.A., Staerk, J., Colchero, F., da Silva, R., Schöley, J., Baden, H.M., ... Vaupel, J.W. (2019). Data gaps and opportunities for comparative and conservation biology.
 Proceedings of the National Academy of Sciences of the United States of America, 116: 9658-9664
- Convention on Biological Diversity. (2020) Zero draft of the post-2020 global biodiversity framework. <u>https://www.cbd.int/doc/c/efb0/1f84/a892b98d2982a829962b6371/wg2020-02-03-en.pdf</u>.

- Cushman, S.A., McKelvey, K.S., Noon, B.R. & McGarigal K. (2010) Use of abundance of one species as a surrogate for abundance of others. Conservation Biology, 24: 830-840.
- Dail D. & Madsen L. (2011) Models for estimating abundance from repeated counts of an open metapopulation. Biometrics, 67: 577-587.
- Devarajan, K., Morelli, T.L. & Tenan, S. (2020) Multi-species occupancy models: review, roadmap, and recommendations. Ecography, 43: 1612-1624.
- Didan, K. (2015) MOD13A2 MODIS/Terra Vegetation Indices 16-Day L3 Global 1km SIN Grid V006 [Data set]. NASA EOSDIS Land Processes DAAC. DOI: 10.5067/MODIS/MOD13A2.006.
- Dorazio, R.M. & Royle, J.A. (2005) Estimating size and composition of biological communities by modeling the occurrence of species. Journal of the American Statistical Association, 100: 389-398.
- Dorazio, R.M., Royle, J.A., Söderström, B. & Glimskär, A. (2006) Estimating species richness and accumulation by modeling species occurrence and detectability. Ecology, 87: 842-854.
- Farr, M.T., Green, D.S., Holekamp, K.E., Roloff, G.J. & Zipkin, E.F. (2019) Multispecies hierarchical modeling reveals variable responses of African carnivores to management alternatives. Ecological Applications, 29: e01845.
- Farr, M.T., Green, D.S., Holekamp, K.E., and Zipkin, E.F. (2021) Integrating distance sampling and presence-only data to estimate species abundance. Ecology, 102: e03204.
- Fei, S., Guo, Q., and Potter, K. (2016) Macrosystems ecology: novel methods and new understanding of multi-scale patterns and processes. Landscape Ecology, 31: 1-6.
- Fithian, W., Elith, J., Hastie, T. & Keith, D.A. (2015) Bias correction in species distribution models: Pooling survey and collection data for multiple species. Methods in Ecology and Evolution, 6, 424–438.
- Fletcher, R.J., Hefley, T.J., Robertson, E.P., Zuckerberg, B., McCleery, R.A., & Dorazio, R.M. (2019) A practical guide for combining data to model species distributions. Ecology, 100: e02710.
- Flockhart, D.T.T., Wassenaar, L.I., Martin, T.G., Hobson, K.A., Wunder, M.B., and Norris, D.R. (2013) Tracking multi-generational colonization of the breeding grounds by monarch butterflies in eastern North America. Proceedings of the Royal Society B, 280: 20131087.
- Foley, J.A., DeFries, R., Asner, G.P., Barford, C., Bonan, G., Carpenter, S.R., ... Snyder, P.K. (2005) Global consequences of land use. Science, 309: 570-574.

- Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., ... Michaelsen, J. (2015) The climate hazards infrared precipitation with stations-a new environmental record for monitoring extremes. Scientific Data, 2: 150066.
- Geldmann, J., Heilmann-Clausen, J., Holm, T.E., Levinsky, I., Markuseen, B., Olsen, K., ... Tøttrup, A.P. (2016) What determines spatial bias in citizen science? Exploring four recording schemes with different proficiency requirements. Diversity and Distributions, 22: 1139-1149.
- Green, D. S. (2015) Anthropogenic disturbance, ecological change, and wildlife conservation at the edge of the Mara-Serengeti ecosystem. Dissertation. Michigan State University, East Lansing, Michigan, USA.
- Gustafsson, K.M., Agrawal, A.A., Lewenstein, B.V., and Wolf, S.A. (2015) The monarch butterfly through time and space: the social construction of an icon. BioScience, 65: 612-622.
- Harrison, X.A., Blount, J.D., Inger, R., Norris, D.R., & Bearhop, S. (2011) Carry-over effects as drivers of fitness differences in animals. Journal of Animal Ecology, 80: 4-18.
- Hostetler, J.A., Sillett, T.S., & Marra, P.P. (2015) Full-annual-cycle population models for migratory birds. The Auk, 132: 433-449.
- Ingebrigtsen, R., Lindgren, F., & Steinsland, I. (2014) Spatial models with explanatory variables in the dependence structure. Spatial Statistics, 8: 20-38.
- Isaac, N.J.B., Jarzyna, M.A., Keil, P., Dambly, L.I., Boersch-Supan, P.H., Browning, E., ... O'Hara, R.B. (2020) Data integration for large-scale models of species distributions. Trends in Ecology & Evolution, 35: 56-67.
- IUCN. (2021) The IUCN Red List of Threatened Species. Version 2021-1. https://www.iucnredlist.org.
- Jarzyna, M.A., Finley, A.O., Porter, W.F., Maurer, B.A., Beier, C.M., & Zuckerberg, B. (2014) Accounting for the space-varying nature of the relationships between temporal community turnover and the environment. Ecography, 37: 1073-1083.
- Jetz, W., McPherson, J.M., & Guralnick, R.P. (2012) Integrating biodiversity distribution knowledge: toward a global map of life. Trends in Ecology & Evolution, 27: 151-159.
- Johnston, A.R. & Anthony, N.M. (2012) A multi-locus species phylogeny of African forest dukers in the subfamily Cephalophinae: evidence for a recent radiation in the Pleistocene. BMC Evolutionary Biology, 12: 120.

- Kelling, S., Johnston, A., Bonn, A., Fink, D., Ruiz-Gutierrez, V., Bonney, R., ... Guralnick, R. (2019) Using semistructured surveys to improve citizen science data for monitoring biodiversity. BioScience, 69: 170-179.
- Kéry, M. & Royle, J.A. (2009) Inference about species richness and community structure using species-specific occupancy models in the national Swiss breeding bird survey MHB.
 Modeling Demographic Processes in Marked Populations (eds D.L. Thomson et al.), pp. 639-656. Springer, New York.
- Kéry, M. & Schaub, M. (2012) Bayesian population analysis using WinBUGS: A hierarchical perspective. Oxford, UK: Academic Press.
- Kéry, M. & Royle, J.A. (2020) Applied hiearchcial modeling in ecology: Analysis of distribution, abundance and species richness in R and BUGS: Volume 2: Dynamic and advanced models. Oxford, UK: Academic Press.
- Kingdon, J. (2015) The Kingdon field guide to African mammals 2nd edition. London, UK: Bloomsbury Publishing.
- Knape, J. and Korner-Nievergelt, F. (2016) On assumptions behind estimates of abundance from counts at multiple sites. Methods in Ecology and Evolution, 7: 206-209.
- Krainski, E.T., Gómez-Rubio, V., Bakka, H., Lenzi, A., Castro-Camilo, D., Simpson, D., ... Rue, H. (2018) Advanced spatial modeling with stochastic partial differential equations using R and INLA. Chapman and Hall, New York.
- Kristensen, K., Nielsen, A., Berg, C.W., Skaug, H., and Bell, B.M. (2016) TMB: Automatic differentiation and Laplace approximation. Journal of Statistical Software, 70: 1-21.
- Lemoine, N.P. (2015) Climate change may alter breeding ground distributions of eastern migratory monarchs via range expansion of Asclepias host plants. PLoS ONE, 10: e0118614.
- Lindgren, F. & Rue, H. (2015) Bayesian spatial modelling with R-INLA. Journal of Statistical Software, 63: 1-25.
- MacKenzie, D.I., Nichols, J.D., Lachman, G.B., Droege, S., Royle, J.A. & Langtimm, C.A. (2002) Estimating site occupancy rates when detection probabilities are less than one. Ecology, 83: 2248-2255.
- MacKenzie, D.I., Nichols, J.D., Royle, J.A., Pollock, K.H., Bailey, L.L. & Hines, J.E. (2017) Occupancy estimation and modeling: Inferring patterns and dynamics of species occurrence. Oxford, UK: Academic Press.
- Magnuson, J.J. (1990) Long-term ecological research and the invisible present. BioScience, 40:495-501.

- Marra, P.P., Cohen, E.B., Loss, S.R., Rutter, J.E., and Tonra, C.M. (2015) A call for full annual cycle research in animal ecology. Biology Letters, 11: 20150552.
- Martin, T.G., Chadès, I, Arcese, P., Marra, P.P, Possingham, H.P., & Norris, D.R. (2007) Optimal conservation of migratory species. PLoS ONE, 2: e751.
- McMaster, G.S. & Wilhelm, W.W. (1997) Growing degree-days: one equation, two interpretations. Agricultural and Forest Meteorology, 87: 291-300.
- Meyer, C., Kreft, H., Guralnick, R. & Jetz W. (2015) Global priorities for an effective information basis of biodiversity distributions. Nature Communications, 6: 8221.
- Miller, D.A.W., Pacifici, K., Sanderlin, J.S., & Reich, B.J. (2019) The recent past and promising future for data integration methods to estimate species' distributions. Methods in Ecology and Evolution, 10, 22–37.
- Møller, J., Syversveen, A.R., and Waagepetersen, R.P. (1998) Log Gaussian Cox Processes. Scandinavian Journal of Statistics, 25: 451-482.
- Momeni-Dehaghi, I., Bennett, J.R., Mitchell, G.W., Rytwinski, T., & Fahrig, L. (2021) Mapping the premigration distribution of eastern Monarch butterflies using community science data. Ecology and Evolution, DOI: 10.1002/ece3.7912.
- Newing, H. (2001) Bushmeat hunting and management: implications of duiker ecology and interspecific competition. Biodiversity and Conservation, 10: 99-118.
- Norris, D.R. (2005) Carry-over effects and habitat quality in migratory populations. Oikos, 109: 179-186.
- Oberhauser, K., Wiederholt, R. Diffendorfer, J.E., Semmens, D., Ries, L., Thogmartin, W.E., ... Semmens, B. (2015) A trans-national monarch butterfly population model and implications for regional conservation priorities. Ecological Entomology, 42: 51-60.
- Oberosler, V., Tenan, S., Zipin, E.F. & Rovero, F. (2020a) Poor management in protected areas is associated with lowered tropical mammal diversity. Animal Conservation, 23: 171-181.
- Oberosler, V., Tenan, S., Zipin, E.F. & Rovero, F. (2020b) When parks work: Effect of anthropogenic disturbance on occupancy of tropical forest mammals. Ecology and Evolution, 10: 3881-3894.
- O'Brien, T.G. (2008) On the use of automated cameras to estimate species richness for largeand medium-sized rainforest mammals. Animal Conservation, 11: 179-181.Phillips, O.L., Aragão, L.E.O.C., Lewis, S.L., Fisher, J.B., Lloyd, J., López-González, G., ... Torres-

Lezama, A. (2009) Drought sensitivity of the Amazon rainforest. Science, 323: 1344-1347.

- O'Brien, T.G., Ahumada, J., Akampurila, E., Beaudrot, L., Boekee, K., Brncic, T., ... Strindberg, S. (2020) Camera trapping reveals trends in forest duiker populations in African National Parks. Remote Sensing in Ecology and Conservation, 6: 168-180.
- Pacifici, K., Reich, B.J., Miller, D.A.W., Gardner, B., Stauffer, G., Susheela, S., ... Collazo, J.A. (2017) Integrating multiple data sources in species distribution modeling: a framework for data fusion. Ecology, 98: 840-850.
- Pacifici, K., Reich, B.J., Miller, D.A.W., & Pease, B.S. (2019) Resolving misaligned spatiotemporal data with integrated distribution models. Ecology, 100, e02709.
- Phillips, O.L., Aragão, L.E.O.C., Lewis, S.L., Fisher, J.B., Lloyd, J., López-González, G., ... Torres-Lezama, A. (2009) Drought sensitivity of the Amazon rainforest. Science, 323: 1344–1347.
- Pleasants, J.M. & Oberhauser, K.S. (2013) Milkweed loss in agricultural fields because of herbicide use: effect on the monarch butterfly population. Insect Conservation and Diversity, 6: 135-144.
- Pollock, K.H., Nichols, J.D., Brownie, C., & Hines, J.E. (1990) Statistical inference for capturerecapture experiments. Wildlife Monographs, 107: 3–97.
- R Core Team. (2020) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
- Reis, L., Zipkin, E.F., & Guralnick, R.P. (2019) Tracking trends in monarch abundance over the 20th century is currently impossible using museum records. Proceedings of the National Academy of Sciences, 28: 13745-13748.
- Rollinson, C.R., Finley, A.O., Alexander, M.R., Banerjee S., Hamil, K.D., Koenig, L.E., ..., Zipkin, E.F. (2021) Working across space and time: nonstationarity in ecological research and application. Macrosystems Biology, 19: 66-72.
- Rossman S., Yackulic, C.B., Saunders, S.P., Reid, J., Davis, R. & Zipkin E.F. (2016) Dynamic N-occupancy models: estimating demographic rates and local abundance from detectionnondetection data. Ecology, 97: 3300–3307.
- Royle, J.A. & Nichols, J.D. (2003) Estimating abundance from repeated presence-absence data or point counts. Ecology, 84: 777–790.
- Royle, J.A. (2004) N-mixture models for estimating population size from spatially replicated counts. Biometrics, 60: 108-115.

- Runge, C.A., Withey, J.C., Naugle, D.E., Fargione, J.E., Helmstedt, K.J., Larsen, A.E., ... Tack, J.D. (2019) Single species conservation as an umbrella for management of landscape threats. PLoS ONE, 14: e0209619.
- Rushing, C.S., Ryder, T.B., & Marra, P.P. (2016) Quantifying drivers of population dynamics for a migratory bird throughout the annual cycle. Proceedings of the Royal Society B, 283, 20152846.
- Rushing, C.S., Hostetler, J.A., Sillett, T.S., Marra, P.P., Rotenberg, J.A., & Ryder, T.B. (2017) Spatial and temporal drivers of avian population dynamics across the annual cycle. Ecology, 98: 2837-2850.
- Rushing, C.S., Tatenhove, A.M.V., Sharp, A., Ruiz-Gutierrez, V., Freeman, M.C., Sykes, P.W., ... Sillett, T.S. (2021) Integrating tracking and resight data enables unbiased inferences about migratory connectivity and winter range survival from archival tags. Ornithological Applications, 123: 1-14.
- Saunders, S.P., Ries, L., Oberhauser, K.S., & Zipkin, E.F. (2016) Evaluating confidence in climate-based predictions of population change in a migratory species. Global Ecology and Biogeography, 25: 1000-1012.
- Saunders, S.P., Cuthbert, F.J., & Zipkin, E.F. (2018a) Evaluating population viability and efficacy of conservation management using integrated population models. Journal of Applied Ecology, 55: 1380–1392.
- Saunders, S.P., Ries, L., Oberhauser, K.S., Thogmartin, W.E., & Zipkin, E.F. (2018b) Local and cross-seasonal associations of climate and land use with abundance of monarch butterflies. Ecography, 41: 278-290.
- Saunders, S.P., Farr, M.T., Wright, A.D., Bahlai, C.A., Ribeiro, J.W., Rossman, S., Sussman, A.L., Arnold, T.W., & Zipkin, E.F. (2019a) Disentangling data discrepancies and deficiencies with integrated population models. Ecology, 100, e02714.
- Saunders, S.P., Ries, L., Neupane, N., Ramírez, M.I., García-Serrano, E., Rendón-Salinas, E., & Zipkin, E. F. (2019b) Multiscale seasonal factors drive the size of winter monarch colonies. Proceedings of the National Academy of Sciences, 116: 8609-8614.
- Saunders, S.P., Hall, K.A.L., Hill, N., & Michel, N.L. (2019c) Multiscale effects of wetland availability and matrix composition on wetland breeding birds in Minnesota, USA. The Condor: Ornithological Applications, 121: duz024.
- Saunders, S.P., Piper, W., Farr, M.T., Bateman, B.L., Michel, N.L., Westerkam, H., & Wilsey, C.B. (2021) Interrelated impacts of climate and land-use change on a widespread waterbird. Journal of Animal Ecology, 90:1165-1176.

- Schaub, M., and Abadi, F. (2011) Integration population models: a novel analysis framework for deeper insights into population dynamics. Journal of Orthinology, 152: 227-237.
- Simpson, D., Illian, J.B., Lindgren, F., Sørbye, S.H., & Rue, H. (2016) Going off grid: computationally efficient inference for log-Gaussian Cox processes. Biometrika, 103: 49-70.
- Skaug, H.J. & Fournier, D.A. (2006) Automatic approximation of the marginal likelihood in non-Gaussian hierarchical models. Computational Statistics & Data Analysis: 51, 699-709.
- Sullivan, M.J.P., Lewis, S.L., Affum-Baffoe, K., Castilho, C., Costa, F., Cuni Sanchez, A., ... Phillips, O.L. (2020) Long-term thermal sensitivity of Earth's tropical forests. Science, 368: 869-874.
- Tang, B., Clark, J.S., & Gelfand, A.E. (2021). Modeling spatially biased citizen science effort through the eBird database. Environmental and Ecological Statistics, 28: 609-630.
- TEAM Network. (2011a) Terrestrial Vertebrate Protocol Implementation Manual, v. 3.1. Tropical Ecology, Assessment and Monitoring Network.
- TEAM Network. (2011b) TEAM Network Sampling Design Guidelines, v. 1.0. Tropical Ecology, Assessment and Monitoring Network.
- Thogmartin, W.E., Wiederholt, R., Oberhauser, K., Drum, R.G., Diffendorfer, J.E., Altizer, S., ... Lopez-Hoffman, L. (2017) Monarch butterfly population decline in North America: identifying the threatening processes. Royal Society Open Science, 4: 170760.
- Thornton, M.M., R. Shrestha, Y. Wei, P.E. Thornton, S. Kao, & Wilson, B.E. (2020) Daymet: Daily Surface Weather Data on a 1-km Grid for North America, Version 4. ORNL DAAC, Oak Ridge, Tennessee, USA. DOI: 10.3334/ORNLDAAC/1840.
- Tobler, M.W., Carrillo-Percastegui, S.E., Pitman, R.L. & Powel G. (2008) An evaluation of camera traps for inventorying large- and medium-sized terrestrial rainforest mammals. Animal Conservation, 11: 169–178.
- Troudet, J., Grandcolas, P., Blin, A, Vignes-Lebbe, R. & Legendre F. (2017) Taxonomic bias in biodiversity data and societal preferences. Scientific Reports, 7: 9132.
- de Valpine, P., D. Turek, C. J. Paciorek, C. Anderson-Bergman, D. T. Lang & R. Bodik. (2017) Programming With Models: Writing Statistical Algorithms for General Model Structures With NIMBLE. Journal of Computational and Graphical Statistics, 26: 403–413.
- Viquerat, S.M.A., Bobo, K.S., Müller, M., Kiffner, C. & Waltert, M. (2012) Estimating forest duiker (Cephalophinae) density in Korup National Park: a case study on the performance of three line transect methods. South African Journal of Wildlife Research, 42: 1–10.

- Watermeyer, K.E., Guillera-Arroita, G., Bal, P., Burgass, M.J., Bland, L.M., Collen, B., ... Nicholson, E. (2020) Using decision science to evaluate global biodiversity indices. Conservation Biology, 35: 492–501.
- Watson, J.E.M. & Venter, O. (2017) A global plan for nature conservation. Nature, 550: 48-49.
- Watson, J.E.M., Venter, O., Lee, J., Jones, K.R., Robinson, J.G., Possingham, H.P., & Allan, J.R. (2018) Protect the last of the wild. Nature, 563: 27-30.
- Webster, M.S., Marra, P.P., Haig, S.M., Bensch, S., & Holmes, R.T. (2002) Links between worlds: unraveling migratory connectivity. Trends in Ecology & Evolution, 17: 76-83.
- Wepprich, T. (2019) Monarch butterfly trends are sensitive to unexamined changes in museum collections over time. Proceedings of the National Academy of Sciences, 28: 13742-13744.
- Zalucki, M.P. (1982) Temperature and rate of development in Danaus plexippus L. and D. chrysippus L. (Lepidoptera: Nymphalidae). Journal of the Australian Entomological Society, 21: 241-246.
- Zipkin, E.F., Dewan, A. & Royle, J.A. (2009) Impacts of forest fragmentation on species richness: A hierarchical approach to community modelling. Journal of Applied Ecology, 46: 815–822.
- Zipkin, E.F., Royle, J.A., Dawson, D.K. & Bates, S. (2010) Multi-species occurrence models to evaluate the effects of conservation and management actions. Biological Conservation, 143: 479–484.
- Zipkin, E.F., Ries, L., Reeves, R., Regetz, J., & Oberhauser, K.S. (2012) Tracking climate impacts on the migratory monarch butterfly. Global Change Biology, 18: 3039-3049.
- Zipkin, E.F., Thorson, J.T., See, K., Lynch, H.J., Grant, E.H.C., Kanno, Y., ... Royle, J.A. (2014) Modeling structured population dynamics using data from unmarked individuals. Ecology, 95: 22–29.
- Zipkin, E.F. & Saunders S.P. (2018) Synthesizing multiple data types for biological conservation using integrated population models. Biological Conservation, 217, 240-250.
- Zipkin, E.F., DiRenzo, G.V., Ray, J.M., Rossman, S., & Lips, K.R. (2020) Tropical snake diversity collapses after widespread amphibian loss. Science, 367: 814-816.
- Zylstra, E.R., Ries, L., Neupane, N., Saunders, S.P., Ramírez, M.I., Rendón-Salinas, E., ... Zipkin, E.F. (2021) Changes in climate drive recent monarch butterfly dynamics. Nature Ecology & Evolution, DOI: 0.1038/s41559-021-01504-1.