Title

Leptospira in the coastal California ecosystem: Challenges and solutions for analyzing complex wildlife disease data

Author

Mummah, Riley Olivia

Publication Date

2021

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Leptospira in the coastal California ecosystem: Challenges and solutions for analyzing complex wildlife disease data

A dissertation submitted in partial satisfaction

of the requirements for the degree

Doctor of Philosophy in Biology

by

Riley Olivia Mummah

2021

ABSTRACT OF THE THESIS

Leptospira in the coastal California ecosystem:

Challenges and solutions for analyzing

complex wildlife disease data

by

Riley Olivia Mummah

Doctor of Philosophy in Biology

University of California, Los Angeles, 2021

Professor James O. Lloyd-Smith, Chair

A major frontier in disease ecology is understanding the transmission dynamics of generalist pathogens, where multiple host species are involved in ongoing circulation of a single pathogen. These dynamics violate assumptions of the simplest epidemic models and lead to challenges in characterizing the determinants of transmission because of the interwoven contributions of different host species. Since many pathogens (including all zoonoses, by definition) can infect multiple hosts, learning about complex multi-host systems from partial evidence is a pervasive problem in disease ecology. Understanding transmission dynamics of a

generalist pathogen in a community of wildlife host species faces the additional challenges of analyzing sparse wildlife data. Biological and logistical sampling constraints in wildlife systems can create temporally and spatially coarse data, which can be difficult to analyze by conventional statistical methods. Thus, analyzing and integrating these data streams requires the development of novel methods.

The coastal California ecosystem provides an opportunity to explore the transmission dynamics of a generalist pathogen in a multi-host community in the presence of complex wildlife data challenges. My primary focus is on the reintroduced Channel Island fox (*Urocyon littoralis*) population on Santa Rosa Island off the coast of southern California. After a population crash in the late 1990s and subsequent captive breeding program in the early 2000s, the Santa Rosa Island fox population was reintroduced to the wild gradually from 2003 to 2009. Shortly thereafter, an outbreak of *Leptospira*, a pathogenic bacterium known to infect most mammals, was detected in the reintroduced population. Within the broader coastal California ecosystem, *Leptospira* was known to circulate endemically in the California sea lion (*Zalophus californianus*) population since at least the 1980s. The context of the *Leptospira* outbreak on Santa Rosa Island prompted questions regarding its origin on the island and determinants of transmission in this wildlife community. In this dissertation, I present three studies analyzing the ample cross-sectional and longitudinal ecological data on the fox population, which has been monitored by the National Park Service for two decades, and pathogen genetic data isolated from four potential host species to understand transmission dynamics of a multi-host, generalist pathogen in this wildlife community.

Ecological data offers only one perspective of *Leptospira* transmission in the coastal California ecosystem. In chapter one, I use bacterial genomic data from four host species to provide qualitatively different evidence addressing two particular questions: (i) What was the source of the pathogen introduction in the reintroduced island fox population? (ii) Did pathogen fadeout occur in California sea lions during a period when other evidence showed an unprecedented pause in *Leptospira* transmission? To address both questions, I construct Bayesian time-calibrated phylogenies and use the topology to infer epidemiological linkages between hosts. For the former question, I show that bacterial isolates from Santa Rosa Island form a distinct cluster in the tree, ruling out sea lions as the direct source of the pathogen introduction to the reintroduced island fox population. For the latter question, I show that isolates obtained after the suspected pathogen fadeout period are not descended from those isolated before, suggesting that pathogen fadeout did occur in the California sea lion population, and that the post-fadeout circulation must have arisen via pathogen introduction from an external reservoir. These results are consistent with and important corroboration for other lines of evidence. This work demonstrates the utility of whole genome sequencing as a component of a multi-disciplinary, multi-data source study to untangle the complexity of wildlife disease systems.

In chapter two, I focus on the *Leptospira* outbreak in reintroduced island foxes to identify and quantify risk factors for infection while addressing two challenges that frequently arise in wildlife data: interval censoring and time-varying covariates. I first addressed the challenge of extensive interval censoring by bootstrapping a set of datasets with an imputed time of infection for individual foxes using a quantitative serology model. The resulting

synthetic datasets lacked interval censoring, which allowed the use of a Cox proportional hazards model and enabled me to account for time-varying covariates through a counting process formulation. I find that higher 24-month cumulative precipitation increases the risk of infection and suggests that long-term fluctuations in the water table significantly influence transmission of *Leptospira*. I also show that risk of infection decreases with increasing fox abundance, contrary to conventional expectation, which may be due to the stabilization of fox social structure on the island as the population increased post-reintroduction. This chapter highlights the need for intensive and sustained data collection and new methodologies for analysis in wildlife systems and lays a foundation for future studies to investigate transmission risk to inform prevention and control strategies in wildlife populations.

Collecting movement data is a resource-limited endeavor, and many studies monitor the locations of individuals at a coarser scale for purposes other than movement analyses. In chapter three, through the integration of a novel spatial data type, obtained through field notes and expert interpretation, and innovative methodology, I propose a method to construct wildlife movement trajectories from location data of varying resolution, again using the Santa Rosa Island fox population as a case study. I integrate unconventional expert-drawn polygons with traditional GPS data by resampling locations on the date of observation for every individual fox. I then fit smoothing splines through the coordinate directions to interpolate each fox's location for every day in its observation window. By pairing the movement estimates with time of infection estimates estimated via a quantitative serology model, I reconstruct the spatiotemporal origin of the *Leptospira* outbreak and estimate transmission to most likely have begun in mid-to-late 2005 along the northern shore of the island. Our approach lays the

groundwork to reap the full benefit of rich, long-term monitoring datasets, which could provide vital insights into a species' movement ecology and better inform conservation and management.

The studies presented in these three chapters apply a breadth of analytic methods to tackle challenges of wildlife data and illustrate how methodological developments and integration of different data streams can be utilized to describe the transmission dynamics of a multi-host generalist pathogen. This work lays the foundation to capitalize on ample lower quality data collected from long-term monitoring programs and to address fundamental challenges in studying wildlife disease ecology.

The dissertation of Riley Olivia Mummah is approved.

Nandita Garud

Van Maurice Savage

Priyanga A. Amarasekare

James O. Lloyd-Smith, Committee Chair

University of California, Los Angeles

2021

To the people who have shown up for me

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SUPPLEMENTARY FIGURES

LIST OF SUPPLEMENTARY TABLES

LIST OF ACRONYMS

CPH Cox proportional hazards regression

CSL California sea lion (*Zalophus californicanus*)

ES Northern elephant seal (*Mirounga angustirostris*)

NPS National Park Service

SRI Santa Rosa Island

VHF Very high frequency

ACKNOWLEDGMENTS

Thank you…

To my advisor, Jamie Lloyd-Smith, for his patience and guidance over the past five years. It was a lot to ask to complete two degrees during my time at UCLA, but I thank you for supporting me through it and making it possible to do both. Thank you for offering me a chance to work on the fox project and the very cool data that came with it, but most of all, thank you for giving me the chance to work with the very smart and capable members of the Lloyd-Smith Lab.

To my committee members, Priyanga Amarasekare, Van Savage, and Nandita Garud, for their thoughtful comments and guidance on my work and their incredible support and patience in my sprint to the finish. I also want to thank Marc Suchard for his helpful insights.

To the Lloyd-Smith lab, for teaching me a lot about myself and the kind of scientist and person I want to be (and sitting through MANY lab meetings with me).

To Katie Gostic and Monique Ambrose. Thank you both for leading the way and setting the stage for other graduate students in the lab. To Monique, thank you for never judging me when I questioned why I was still in the program. To Katie, for telling me to just do it.

To Celine Snedden and Ana Gomez, for your encouragement over the past year and a half, predominantly in the form of “Writing Group,” (which was definitely more of a support group).

To Katie Prager, for the hours that you have listened to me and heard me out, for the incredible support you have given me over the years, and your judgment-free attitude when I brought you my concerns. I always enjoyed our calls during your runs and your occasional

podcast or book recommendation. I will sorely miss working with such a supportive and positive person.

To Angie Ghanem, for being someone to lean on, for hours of listening to me complain, and for always making an effort to be my friend and stay in touch. Talking and catching up with you was always something I could look forward to.

To Lauren Smith. Thank you for getting me through grad school. For surviving with me through our first few years and blossoming together into seasoned grad students. Thank you for always sharing your discerning mind and offering advice. You are always someone I can trust to guide me through a tough situation.

To Sarah Helman & Mason Gamble. Your friendship is one of my most cherished things in my life, and I am so lucky that you've both been by my side during this time. You are two of the most generous and supportive people I've ever known. Thank you for always being a light in my life and being proud members of #TeamRiley.

To Amanda Tokuyama, for reflecting my own advice back to me. For reading SO many drafts. For helping to write and edit. The list of things you have done to support me is seemingly endless. You have grown into one of my closest friends, and it is one of my most valued experiences over the last five years. Thank you for our #RockHardFriendship, Thursday night dinners, and for finishing our degrees together.

To Alex and Bonnie Crubaugh ("The Parents"), for being on my team. I always appreciated your interest and questions to better understand my work and what I was going through.

To my mum, thank you for showing up for me: now, before, and for the times yet to come. Your support does not go unnoticed. Thank you for always having my back.

To my dad, you once told me that if it was humanly possible, I could do it, and if I couldn't do it, then it wasn't humanly possible. Your confidence in me got me through some of the hardest days.

To Grammie. You have showed up to games, band concerts, dance recitals, and graduations. Your weekly phone calls, visits to California, and unending support has pushed me through some dark days. Thank you for always being there for me.

To Bryant. Words cannot express the level of support you have shown me over the throughout this degree. I quite literally could not have done this without you; this accomplishment is something we share. Thank you for putting up with late nights, grumpy mornings, my (your) frustration, and for always believing I could do it (especially when I didn't believe in myself).

What everyone says it true; it does take a village to get anywhere in life. I could not have asked for a better group of supporters: Anita Smith; Bruce Brumbaugh; Nancy and Doug Brackbill; Kris and Tim Lloyd; Royce and Andi Book; Joyce Eckhart; Kenny, Tawny, Haley, Harrison and Olivia Mummah; Lisa Harling; Lance Gilliland; Pappy; Duck and Betty Jean Gilliland; Dr. James Farrell; Alex Smith; Sara Makanani; Sabrina Simmons; Kathy Cannon; Oshiomah Oyageshio; Jeff and Allison Arnold; Dana Williams; Mark Juhn; Marshal and Victoria Rittenger; Mary Chiu; Kelli Sio and Steve Gustafson; Ross Mummah.

My dissertation work was made possible and supported by a Department of Defense Strategic Environmental Research Development Program (SERDP) grant (RC-2635). Travel funds

from the UCLA Graduate Division and the UCLA Department of Ecology and Evolutionary Biology have facilitated collaborations and dissemination of my results at national and international conferences. The contents of this work are solely the responsibility of the authors and do not necessarily represent the views of the US Government. The funders had no role in data analysis, study design or interpretation of results.

This work was done in collaboration with the "Lepto Team" within the Lloyd-Smith lab group. The team consists of five graduate students (Angela Guglielmino, Ana Gomez, Sarah Helman, Amanda Tokuyama, and myself), a postdoc (Benny Borremans), a research associate (Katherine Prager), and James Lloyd-Smith. Many thank go out to all members of the team for their contributions to this project and my dissertation.

I thank Angela Guglielmino who led the field work for the fox system and offered crucial field- and system-specific knowledge. Angela was also responsible for drawing the thousands of telemetry polygons featured in chapter three, which was many, many hours of work. Thank you for all of your efforts.

I thank Benny Borremans who developed the quantitative serology model featured in chapters two and three. Benny also offered lots of wisdom and insights into the analytics of the project.

I thank Katie Prager for her helpful insights into the system, edits and feedback on project deliverables, and overall support of this project's success.

I thank Jamie Lloyd-Smith for bringing this team together to do this novel, interesting work and for your ideas, which push the bounds on what can be achieved.

I thank Ana Gomez and Sarah Helman for their valuable contributions to my understanding of *Leptospira* ecology and modelling, which helped me advance my projects.

I thank Amanda Tokuyama for creating the map featured in chapter two. Amanda also read numerous drafts of my chapters and offered valuable edits.

I thank Ottar Bjørnstad, not only for introducing me to this field and taking me under his wing as an undergraduate, but for his valued suggestions on the movement reconstruction. I thank Kat Shea for her encouragement and wisdom throughout my career.

I thank everyone who helped me with my bioinformatics pipeline and genomic analysis. I thank Sorel FitzGibbon for providing early advice on my bioinformatics pipeline. It got me off to a great start. I also thank UCSD scientists Michael Mattias and Joshua Lefler for their assistance in *Leptospira* whole genome sequence assembly and for hosting me during my visit to San Diego. I thank Nandita Garud for taking the time to talk with me to answer my questions about genomic data.

I would like to thank all the laboratory collaborators who made this work possible. I thank Renee Galloway at the Centers for Disease Control and Prevention for her amazing efforts in performing serum MAT analyses and isolating *Leptospira* from our cultures. I would also like to thank USDA team members David Alt, Rick Hornsby and Jarlath Nally for their work on obtaining *Leptospira* isolates from the bacterial cultures sent to them, sequencing *Leptospira* isolates, and general advice regarding *Leptospira* epidemiology and microbiology.

The staff and volunteers with the National Park Service are responsible for obtaining samples from both the island foxes and the spotted skunks. I thank them for their tremendous efforts and for their willingness to obtain and share data on fox movement and locations.

Additionally, the veterinarians, veterinary technicians, researchers, and volunteers with The Marine Mammal Center in Sausalito, California collect, archive, and manage data and samples from stranded California sea lions, and I owe them many thanks for it.

VITA

2021 Food Studies Graduate Certificate
University of California Los Angeles

2020 MS Epidemiology
University of California Los Angeles

2020 Leaders in Sustainability Graduate Certificate
University of California Los Angeles

2016 Master of Applied Statistics
The Pennsylvania State University

2016 BS Statistics
The Pennsylvania State University

2016 BS Biology
The Pennsylvania State University

2016 BS Mathematics
The Pennsylvania State University

PUBLICATIONS

Griffin C, **Mummah R**, and deForest R. A Finite Population Destroys a Traveling Wave in Spatial Replicator Dynamics. Chaos, Solitons, & Fractals. *Accepted.*

Mummah RO, Hoff NA, Rimoin AW, Lloyd-Smith JO. Controlling emerging zoonoses at the animal-human interface. OneHealth Outlook. 2020.

Helman SK, **Mummah RO**, Gostic KM, Buhnerkempe MG, Prager KC, Lloyd-Smith JO. Estimating prevalence and test accuracy in disease ecology: how Bayesian latent class analysis can boost or bias imperfect test results. Ecology & Evolution. 2020.

Gostic K, Gomez AC, **Mummah RO**, Kucharski AJ, Lloyd-Smith JO. Estimated effectiveness of symptom and risk screening to prevent the spread of COVID-19. Elife. 2020.

Sura SA, Smith LL, Ambrose MR, Guerra Amorim CE, Beichman AC, Gomez ACR, Juhn M, Kandlikar GS, Miller JS, Mooney JS, **Mummah RO**, Lohmueller KE, Lloyd-Smith JO. Ten simple rules for giving an effective academic job talk. PLoS Comput Biol. 2019.

Li SL, Bjørnstad O, Ferrari M, **Mummah R**, Runge MC, Fonnesbeck C, Tildesley MJ, Probert WJM, Shea K. Essential information: Uncertainty and optimal control of Ebola outbreaks. PNAS. 2017.

1 Molecular ecology of *Leptospira interrogans* in the coastal California ecosystem

1.1 Abstract

Parsing cross-species transmission in multi-host systems is difficult because the relative contribution of each host to the transmission dynamics is unknown, but current sequencing technologies have created unprecedented opportunities for inferring host epidemiological linkages. Whole genome sequencing has provided powerful insights into the dynamics of fast-evolving pathogens such as RNA viruses, but the application to slow-evolving pathogens has been met with challenges inherent to these microbes (e.g. low mutation rate, slow growth rate). In this chapter, we analyze whole genome sequences of *Leptospira* and build Bayesian time-calibrated phylogenies to explore two phenomena: (i) the source of a novel pathogen introduction to reintroduced island foxes (*Urocyon littoralis*) on Santa Rosa Island and (ii) apparent pathogen fadeout in California sea lions (*Zalophus californianus*) during a period when other evidence showed an unprecedented pause in *Leptospira* transmission. We found that sea lions were not the likely source of the *Leptospira* introduction to the island fox population, as the clade of isolates from Santa Rosa Island branches separately from the rest of the tree with a estimated median date of 1947. In the second case study, our results show that all sea lion isolates taken from the period after suspected pathogen fadeout were not descendants of the clade of isolates taken prior to fadeout, suggesting that pathogen fadeout did occur in the California sea lion population and that the post-fadeout circulation must have arisen via pathogen introduction from an external reservoir. This work demonstrates that whole genome

sequencing and analysis provide an unprecedented opportunity to resolve epidemiological relationships for slow-evolving pathogens such as *Leptospira*.

1.2 Introduction

Many pathogens are capable of infecting multiple host species, creating ample opportunity for cross-species transmission within and across ecosystems (Cleaveland et al., 2001; Taylor et al., 2001; Woolhouse, 2001). However, multi-host pathogen systems are understudied due to their complex nature and remain a major frontier in disease ecology (Buhnerkempe et al., 2015; Power & Mitchell, 2004; Viana et al., 2014). In multi-host systems, transmission patterns are difficult to characterize due to the challenge of investigating the relative contribution of each host species to the transmission dynamics as either a pathogen source or a focal host species (Viana et al., 2014). However, determining the frequency and directionality of transmission in a community of hosts is crucial for understanding the ecology of the pathogen as well as for conservation and management.

There are two dimensions of transmission within a multi-host system which drive incidence patterns in the focal host species: within-species transmission and cross-species (source to target) transmission. The frequency of transmission in each dimension can occur across a low-to-high gradient and influence the intensity and duration of an outbreak (Viana et al., 2014). Low transmission in both dimensions (cross- and within-species) creates a self-limiting outbreak, as evident for human monkeypox (Grant et al., 2020; Sklenovská & Van Ranst, 2018). When a multi-host system has low cross-species transmission, but within-species transmission can be sustained by the novel host, a single cross-species transmission event can

cause an epidemic. International health crises such as the 2014 West Africa Ebola outbreak and the 2009 H1N1 pandemic are products of singular (or few) pathogen spillover events from wildlife to humans followed by sustained human-human transmission (Bajardi et al., 2011; Bonilla-Aldana et al., 2020; Webb et al., 2015). Alternatively, when cross-species transmission is the primary driver of incidence and there is limited within-species transmission, the pathogen can appear endemic in the absence of knowledge about the route of transmission. A clear example of this phenomenon are the initial dynamics of canine distemper virus (CDV) in the wildlife of Serengeti National Park, where it is believed that CDV was repeatedly introduced into the wildlife community by unvaccinated domestic dogs outside of the park (Cleaveland et al., 2000; Viana et al., 2015). High levels of both cross-species and within-species transmission complicate accurate characterization of the relative contributions of each host species to the community and population level disease dynamics. For example, *Mycobacterium bovis* transmits readily between and among European badgers and domestic cattle, which makes it difficult to estimate the relative contributions of each host species to the observed pathogen prevalence and thus affects the ability to design effective intervention measures (Crispell et al., 2019; McDonald et al., 2018). It becomes even more difficult to distinguish between the contributions of within- and cross-species transmission when each dimension occurs at an intermediate level, transmission occurs within a community of host species, and/or the relative importance of each host species is unknown.

While identifying patterns in cross-species transmission is extremely important for understanding the risks pathogens pose for communities, transmission dynamics are particularly challenging to characterize in wildlife systems, and this remains one of the main

challenges in disease ecology (Begon et al., 1999; Buhnerkempe et al., 2015; Carslake et al., 2006; Kilpatrick et al., 2006). These challenges are amplified in studies that attempt to include multiple species. Hence, such studies are rare (Almberg et al., 2009; Kamath et al., 2016; Prager et al., 2012; Viana et al., 2015). Limited and unequal data from the different hosts can limit researchers' ability to make community-wide inferences. To enrich these datasets and maximize the possible inferences, Viana et al. (2014) suggest integrating different types of data and methodologies. For instance, epidemiological data on the timing and incidence levels of an outbreak could be paired with genomic data reporting on the relatedness of the cases. This combination of evidence would enable researchers to address questions about cross- and within-species transmission within a multi-host wildlife system.

In particular, genomic data have a unique potential to shed light on transmission links, especially with the rise of whole genome sequencing (WGS). Broadly, WGS data can provide information on temporal and spatial dynamics in pathogen-host systems, identify historic and contemporary cross- and intra-species transmission events, and detect the emergence of pathogen variants (Biek & Real, 2010). Techniques to analyze WGS data have been successfully applied in wildlife and livestock systems to investigate transmission pathways (Benton et al., 2015). Kamath et al. (2016) examined the transmission dynamics of *Brucella abortus* within the greater Yellowstone ecosystem. Through their use of a time-calibrated Bayesian phylogenetic reconstruction, they inferred that free-ranging elk act as the reservoir host for *Brucella* infections in nearby cattle, and so they were able to direct control measures away from bison, which have little role in the transmission within the host community. A similar methodology was used to investigate transmission pathways of *Mycobacterium bovis* among cattle herds and

badgers in Britain (Crispell et al., 2019), with which they were able to discern that within-species transmission occurred most frequently and that badgers transmitted more often to cattle than the reverse. These insights allowed them to guide the development of control measures that were appropriate for both species, and their analysis highlights the potential for time-calibrated Bayesian methods to be applied to bacterial genomes to uncover hidden transmission dynamics. However, both systems addressed above include livestock and hence are easier to sample to have ample data and isolates to infer transmission. This level of data is rarely available in other wildlife communities.

The time-calibrated phylogenies applied in the *Brucella* and *Mycobacterium* systems can integrate epidemiological data to reconstruct transmission histories, provide genetic inference for the chains of transmission in a system, and even identify and date cross-species transmission events due to their extremely dense sampling. However, the ability to resolve and date particular cross-species transmission events depends heavily on the samples available, the ability to culture the pathogen, and the pathogen's substitution rate. In systems with sparse sampling (i.e. most wildlife systems) of a slow-evolving, hard-to-culture pathogen, it can be quite challenging to have the power to accurately date and estimate transmission linkages. To tackle this problem, the inferences from time-calibrated phylogenies can be paired with ecological data to untangle the transmission dynamics within a wildlife system.

1.2.1 Leptospira interrogans

Leptospirosis is the most widespread zoonosis in the world and has been found in most mammals and marsupials (Adler & de la Pena Montesuma, 2010). *Leptospira interrogans* is a

slow-evolving spirochete bacterium with a low recombination rate and is one of 13 pathogenic species in the *Leptospira* genus (Levett, 2001; Thaipadungpanit et al., 2007; Vos & Didelot, 2009), the causative agents of leptospirosis. Leptospires are transmitted through direct contact with urine or tissues of an infected animal or indirectly through contaminated soil or water. After an incubation period, leptospires circulate in the blood before infecting the renal tubules, after which they are shed in the urine (Levett, 2001). Clinical signs of leptospirosis can range from asymptomatic to fatal, and infection can resolve rapidly, or become chronic, with asymptomatic chronic carriers shedding leptospires into the environment for months to years (Levett, 2001). For example, California sea lions (*Zalophus californianus*) have been shown to be chronic carriers of *L. interrogans* serovar Pomona, which has circulated in the population since the early 1980s (Lloyd-Smith et al., 2007; Prager et al., 2013).

Leptospira was historically split phenotypically into pathogenic (*L. interrogans*) and saprophytic (*L. biflexa*) bacteria with over 260 serovars determined by antibody responses (Adler & de la Pena Montesuma, 2010). Genotyping led to the reclassification of *Leptospira*, and now some serovars are found in more than one *Leptospira* species (Levett, 2001). Critically, *Leptospira* serovars are not host-specific. In particular, *L. interrogans* serovar Pomona has a very broad range of host species and is known to infect species including (but not limited to) horses, raccoons, white tailed deer, striped skunks, opossums, pigs, foxes, and sea lions (Bolt & Marshall, 1995; Lloyd-Smith et al., 2007; Timoney et al., 2011). The combination of multiple serovars, multiple hosts, complex and variable host-pathogen interactions, and diagnostic challenges in identifying the infecting serovar and strain has led to a muddled understanding of *Leptospira* ecology.

Although different serovars of *Leptospira* may have one or more primary reservoir or host species, most mammals are susceptible to infection. Studies have compared the infecting serovars of multiple hosts within a given ecosystem, but generally these studies could only draw very limited conclusions due to the cross-reactivity of different serovars during antibody testing (Bishara et al., 2002; Panaphut et al., 2002; Santos et al., 2016; Sehgal et al., 1995; Tunbridge et al., 2002). In contrast, few studies have used whole genome sequencing to examine the nature of cross-species transmission for *Leptospira*. Recent work by Allan et al. (2020) used the *Leptospira secY* gene sequence to identify infecting species and construct a phylogeny with previously sequenced isolates. Of the three human-derived PCR sequences they were able to genotype, each sequence was identified as a different *Leptospira* species. With limited knowledge of the circulating species and serovars, little could be learned about cross-species transmission in the system. Kakita et al. (2021) used whole genome sequences of *Leptospira borgpetersenii* serogroup Javanica to explore the role of cats as a maintenance host for human leptospirosis infections in Japan. They inferred that cats as well as mongoose and black rats carry and transmit related strains of *L. borgpetersenii* serogroup Javanica in the region, but were limited by small sample sizes with two or fewer isolates from three of their four host species. Although they paired their phylogenetic analysis with serological data to corroborate their genetic findings, they lacked temporal depth and enough samples per host species to infer directionality and frequency of cross-species transmission. These few studies demonstrate how powerful WGS can be in disentangling the complex, multi-host ecology of *Leptospira* in a community, but were limited by their small sample sizes and lack of supporting ecological data.

1.2.2 Study system

The coastal California ecosystem is home to diverse flora and fauna in many diverse habitats, including the California Channel Islands, an eight-island archipelago located off the coast of southern California between Point Conception and San Diego. Five of the islands (San Miguel, Santa Rosa, Santa Cruz, Anacapa, and Santa Barbara) comprise Channel Islands National Park and are managed by the National Park Service. Marine mammals such as the California sea lion (*Zalophus californianus*; CSL) and the Northern elephant seal (*Mirounga angustirostris*; ES) haul out on the shores of the islands. Only four terrestrial mammal species, island foxes (*Urocyon littoralis*), island deer mice (*Perimyscus maniculatus*), western harvest mice (*Reithrodontomys megalotis*), and island spotted skunks (*Spilogale gracilis amphiala*), are native to the islands, and no island hosts all four endemic terrestrial species.

Santa Rosa Island (SRI) hosts the endemic island fox, the island spotted skunk, and the island deer mouse (Wayne et al., 1991). Island foxes are territorial and mark their home ranges with urine (Coonan et al., 2010). Island foxes and spotted skunks exist sympatrically on the island and compete for resources with the island fox being the dominant competitor (Crooks & Van Vuren, 1995; Jones et al., 2008).

California sea lions haul out on rocks and beaches along the California coastline, when not foraging in near-shore waters, and share space with other marine mammals such as Northern elephant seals (Peterson & Bartholomew, 1967). During the late spring months, San Miguel Island serves as a rookery for CSL to give birth (Melin et al., 2000). CSL migrate seasonally along the coast of California from the Channel Islands and central California as far

north as southeast Alaska (Carretta et al., 2006). Their migration is dictated by age and sex-specific patterns. Males travel further north outside of the breeding period (May/June) with the largest males traveling the furthest (Zuerner et al., 2009).

1.3 Case Studies

1.3.1 Case study #1: Reconstructing the origin of a wildlife disease emergence event

1.3.1.1 Context

Island fox populations are naturally small and prone to fluctuations. In the 1990s, anthropogenic factors initiated a cascade of events which culminated in golden eagle predation causing severe population declines of island foxes on Santa Rosa Island (Coonan et al., 2005). Declines also occurred on nearby San Miguel and Santa Cruz Islands. Introduced feral pigs and mule deer, which had drawn the non-native golden eagles to the island, have since been removed from SRI in 1993 and 2012, respectively (Coonan et al., 2010; Knowlton et al., 2007; Roemer et al., 2001, 2002). Prior to the arrival of golden eagles, the foxes had no natural predators and, hence, no behavioral adaptations against predation. Mortality through eagle predation was high and outpaced the foxes' ability to reproduce, so the population crashed and became critically endangered (Coonan et al., 2005). By 2001 the National Park Service had taken the entire remaining SRI population of 14 foxes into captivity, and the population was temporarily extinct in the wild (Coonan et al., 2010). A multi-year captive breeding program was initiated to reintroduce foxes on the island. Reintroduction began in 2003, but the earliest releases were unsuccessful due to ongoing eagle predation and other unknown mortalities. Golden eagles were eventually relocated to the mainland. By the end of 2009, 96 captive foxes

had been released into the wild. The reintroduced population was recovering, so the captive breeding program was concluded. The fox population was delisted as an endangered species in 2016.

In fall 2010, two juvenile male foxes were found dead on SRI with evidence of *Leptospira* infection. At the time, there was no evidence that this strain was circulating on the island prior to the fox population crash, which spurred investigation into the origins of the pathogen on the island. A follow-up investigation in January 2011 obtained isolates of *L. interrogans* from island foxes and island spotted skunks. Genetic analysis by variable number tandem repeat (VNTR) typing showed that the strain was indistinguishable from the strain of *L. interrogans* serovar Pomona circulating in the California sea lion population (Zuerner & Alt, 2009; unpublished results), making sea lions a suspected source of the pathogen. Our lab recently conducted serological assays of banked sera that suggest that foxes on SRI had exposure to the same strain of *Leptospira* prior to captivity (Lloyd-Smith, 2021). Transmission was interrupted during captivity, and all captive-born foxes were seronegative upon release to the island. Yet retrospective analysis of samples collected from reintroduced foxes showed that by 2006 island-wide fox seroprevalence was already high.

Clearly, the immunologically naïve captive-born foxes were released to a landscape where *Leptospira* was circulating, but the source of the outbreak was unknown. California sea lions and Northern elephant seals were potential sources of the pathogen as they haul out on the shores of the island, and *Leptospira* is endemic in California sea lions. Island spotted skunks were another suspected source as we had detected infection in this species, and they and island foxes share the same habitat and resources.

To this point, serological data has been the most prominent indicator that the same strain was circulating in multiple host species. However, the strain typing performed on sera cannot distinguish sub-strain level differences. Thus, by using whole genome sequencing and molecular ecology techniques, we sought to independently corroborate the serological, demographic, and movement data and investigate the epidemiological linkages between foxes and other potential *Leptospira* hosts in the ecosystem.

1.3.1.2 Data

Our dataset includes forty-nine bacterial isolates from four host species: Channel Island fox, island spotted skunk, California sea lion, and Northern elephant seal, spanning the years 1988 to 2017 (Table 1.1).

Table 1.1: Frequency of *Leptospira interrogans* serovar Pomona isolates by host species and sampling year from unique individuals.

Host species	1988	2004	2006	2007	2008	2010	2011	2012	2013	2015	2017	Total by host
CSL	1	3	2	3	1	2	20	2				34
ES							3					3
Island fox							4		3	2	1	10
Spotted skunk							2					2
Total by year	1	3	2	3	1	2	29	2	3	2	1	49

1.3.1.3 Results

The time-calibrated Bayesian phylogeny was constructed with a relaxed lognormal molecular clock and HKY + Γ + I site heterogeneity model. The phylogeny indicates two clearly

distinct clades, which are broadly delineated by ecosystem and robust to our imbalanced sampling (Figure 1.1; Figure S1.1). One clade (Marine1) consists entirely of marine mammals: thirty-three CSL, sampled from 2004 to 2012, and one ES, sampled in 2011. The second clade splits into two subclades: Island and Marine2. The Island subclade contains all terrestrial host isolates and a single CSL isolate (CSL10040), whereas the Marine2 subclade contains two ES isolates from 2011 and a CSL isolate, which was sampled in 1988.

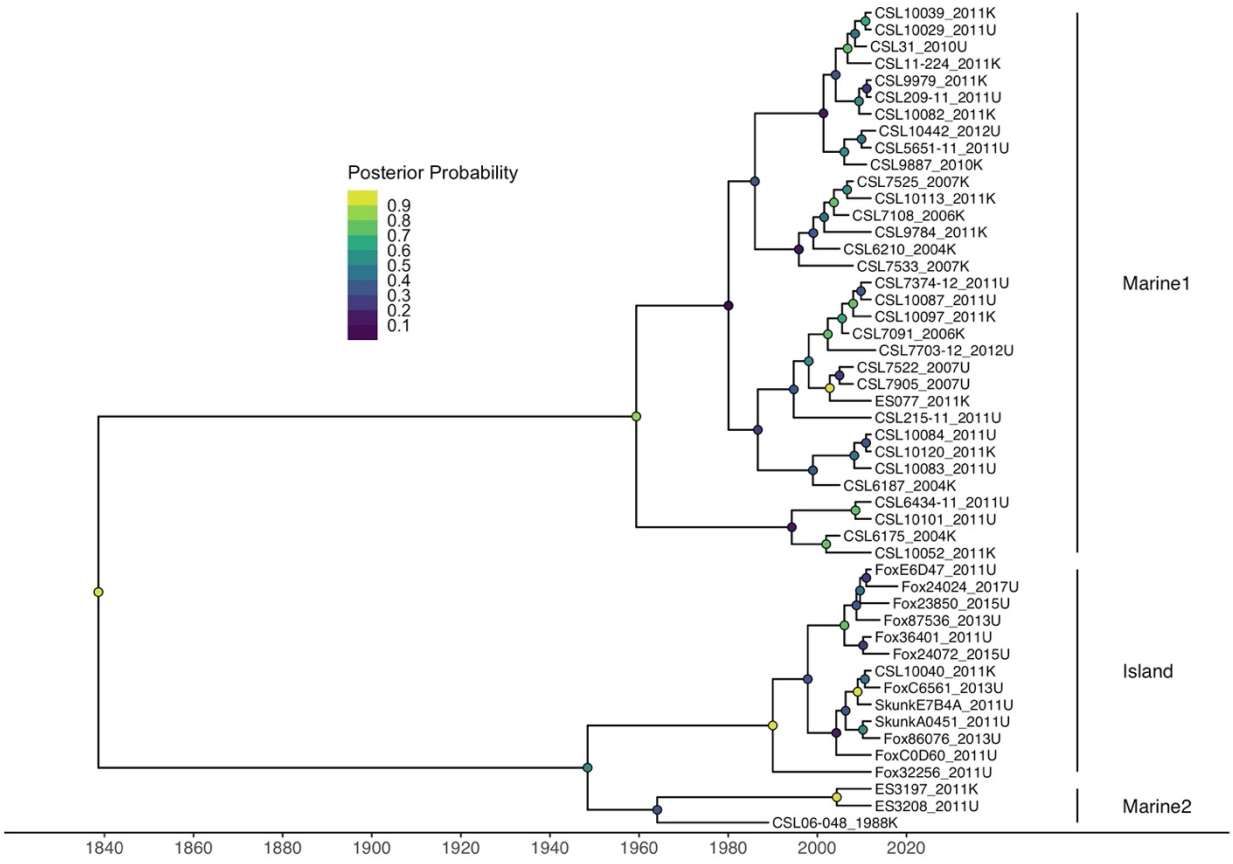

Figure 1.1: Time-calibrated maximum clade credibility tree of *Leptospira interrogans* serovar Pomona in the Channel Island ecosystem. Two major clades were identified through Bayesian phylogenetic analysis: Marine1 and Island + Marine2. Interior nodes are labelled with the posterior probability (lower values are purple, higher values are yellow). Tips are labeled with host species (CSL = California sea lions, ES = elephant seal), ID number, year, and tissue type (K = kidney, U = urine).

There are three cross-species transmission events directly evident in the phylogenetic reconstruction presented in this case study, but there is more evidence of cross-species

transmission when considering the whole tree. The Marine1 clade indicates that transmission occurs between CSL and ES, which is ecologically plausible as the two species share haul-out and rookery sites and have frequent direct contact. Additionally, these phylogenetic data are consistent with epidemiological data (e.g. serology, incidence; not published) which supports CSL as a maintenance host for *Leptospira* and ES more likely as a spillover target rather than a maintenance host population. The island clade contains two types of cross-species transmission: within the terrestrial ecosystem and between the terrestrial and marine ecosystems. The clustering in this clade suggests that island spotted skunks and island foxes had multiple cross-species transmission events prior to 2011 because the skunk isolates do not cluster separately from the foxes. Rather, they are interspersed with the fox isolates within the island clade. Additionally, there is a singular CSL (CSL10040) nestled within the island clade, indicating that cross-species transmission can occur from the terrestrial hosts on the island to the marine ecosystem, most likely from foxes to CSL.

Additional cross-species transmission events can be inferred from the deeper branching structure of the tree. Two elephant seal isolates (ES3197 and ES3208) from 2011 cluster with the 1988 CSL isolate, but we do not believe this is evidence for contemporaneous cross-species transmission, given the 23-year interval between the sampling dates. Rather, we believe that there is a separate reservoir that seeded both lineages.

The median substitution rate in these genomes was 1.16×10^{-7} substitutions per site per year, which is equivalent to 0.37 SNPs per year per genome. This rate is slower than most bacteria (range: 10^{-4}-10^{-8}) and most comparable to the genome-wide substitution rate of *Salmonella* Paratyphi A (Duchêne et al., 2016). The variation in the median clock rate was high

with the 95% highest probability density (HPD) covering 0.11 – 2.63 x 10^{-7} substitutions per site per year (Table 1.2). Using the estimated rate, we were able to estimate the date when the different clades diverged, also known as the time to most recent common ancestor (tMRCA). The estimated date for the root of the tree, when isolates from island foxes last shared a common ancestor with the larger Marine1 clade, is 1836 (HPD: 1272 - 1980). The estimated tMRCA for the Island and Marine2 clades, which represents the last date at which the terrestrial isolates shared a known ancestor with the broader marine realm, is 1947 (HPD: 1755-1988). In both instances the uncertainties on the tMRCAs are large, due to the limited number of isolates and high variation in substitution rates, but notably the intervals do not extend to the period after 2000 when the foxes entered captivity. This is consistent with the ecological data, which indicate that *Leptospira* circulated on the island prior to captivity but disappeared in the captive-born population.

Table 1.2: Posterior median estimates of genetic and temporal parameters

Parameter	Median (95% HPD)
Clock rate ($\times 10^{-7}$)	1.16 (0.11, 2.63)
Coefficient of variation$^{\perp}$	2.59 (1.60, 4.87)
tMRCA date (Root)	1836 (1272, 1980)
tMRCA date (Island + Marine2)	1947 (1755, 1988)

HPD: highest posterior density; tMRCA: time to most recent common ancestor for marine and island lineages

$^{\perp}$Measure of the variation in evolutionary rate amongst the branches

The interior nodes of the phylogeny have a wide range of posterior probabilities (PP; Figure 1.1). Crucially, the branch point at the root of the tree has a very high posterior probability (>99%), but subsequent nodes decline in posterior support. This indicates that the separation between isolates circulating in sea lions at the time the SRI outbreak began (i.e. the

Marine1 clade) and isolates involved in the SRI outbreak (the Island clade) was almost always present in the posterior set of trees, but the specific branching patterns within the clades were more variable. This further supports the conclusions from the tMRCA analyses that the source of the fox outbreak was terrestrial, not marine.

Broad clustering patterns within the tree topology suggest that multiple introductions of *L. interrogans* serovar Pomona into the California coastal ecosystem have occurred from an unknown reservoir. These repeated introductions appear to have seeded multiple chains of transmission, which circulated simultaneously in different host species. For example, the Island clade most likely had a single introduction from an external reservoir. Unfortunately, it is difficult to date when this introduction occurred due to the large posterior density on the median tMRCA. It appears likely that the sea lion isolate obtained from CSL06-048 in 1988 was seeded by an introduction independent of the one that seeded the Marine1 clade. Finally, the cluster of elephant seal isolates in 2011 signifies another potential introduction from an unknown reservoir.

1.3.2 Case study #2: Climate-driven fadeout and re-emergence of an endemic pathogen in a wildlife host

1.3.2.1 Context

Leptospira interrogans serovar Pomona has circulated in CSL since at least the mid-1980s (Greig et al., 2005; Gulland et al., 1996). The population experiences yearly, seasonal outbreaks of varying magnitude (Lloyd-Smith et al., 2007). We hypothesize that persistent circulation in the sea lion population relies on a crucial link between seasonal outbreaks during

which *Leptospira* is transmitted from chronically infected older animals (Buhnerkempe et al., 2017), to the pool of young susceptible sea lions, during contact which occurs with age- and sex-specific seasonal movements. However, serologic, epidemiological, demographic, and ecological data collected between 2010-2019 suggest that *Leptospira* disappeared entirely from the CSL population in early 2013. In separate work from our lab, we demonstrate that pathogen fadeout (i.e. disappearance of the pathogen from the sea lion population) is associated with a reduced pool of susceptible sea lions due to low survival and growth of the 2012 birth cohort and disrupted age- and sex-specific movement patterns – all likely driven by oceanographic anomalies – which broke the crucial link in the between-season transmission chain. A single *Leptospira*-infected CSL stranded on the California coast in 2016, but no other evidence of pathogen circulation was detected in the CSL population during that year. We believe that the 2016 case arose from a spillover event which failed to lead to pathogen re-emergence. In 2017, *Leptospira* reappeared in the wild and data up through 2020 suggest that it appears to have resumed its annual outbreak cycle. This study is the first recorded example of spontaneous fadeout of an endemically circulating pathogen from a wildlife population.

The evidence from field studies and CSL stranding data suggests that *Leptospira* stopped circulating entirely in the CSL population. Despite our intensive surveillance efforts, it is always possible that *L. interrogans* persisted at a low level but was undetectable by our current sampling methods. Since the pathogen reemerged in the CSL population, we have been able to collect multiple isolates from the post-fadeout period. Through whole genome sequencing, we elucidate the relationship between the pre-fadeout (prior to 2013) and post-fadeout (2017 and later) strains of the bacteria. The bacterial genomic data provide a qualitatively different type of

information, by enabling estimation of the probability that the post-fadeout bacterial lineage is an outgrowth of the pre-fadeout lineage.

1.3.2.2 Data

Our dataset includes bacterial isolates from five host species, Channel Island fox, island spotted skunk, California sea lion, Northern elephant seal, and pig (*Sus scrofa*), spanning the years 1984 to 2018 (Table 1.3). All isolates included in the first case study are also included here in addition to four CSL isolates from the post-fadeout period. We also included two isolates from neonatal swine obtained in Iowa to serve as a terrestrial mainland comparison.

Table 1.3: Frequency of *Leptospira interrogans* serovar Pomona isolates by host species and sampling year from unique individuals.

			Pre-fadeout								Fadeout	Post-fadeout		
Host species	1984	1988	2004	2006	2007	2008	2010	2011	2012	2013	2015	2017	2018	Total by host
CSL		1	3	2	3	1	2	20	2			1	3	38
ES								3						3
Island fox								4		3	2	1		10
Spotted skunk								2						2
Pig	2													2
Total by year	2	1	3	2	3	1	2	29	2	3	2	2	3	55

1.3.2.3 Results

Whole-genome sequencing of fifty-five *L. interrogans* serovar Pomona isolates enabled a phylogenetic reconstruction using 355 single nucleotide polymorphisms. The phylogeny, built with an HKY + Γ + I substitution model and uncorrelated relaxed molecular clock, indicates three broad clades (Figure 1.2). The pre-fadeout CSL, a singular ES isolate (ES077), and one post-fadeout CSL isolate (CSL14079) form Clade1. The second clade is comprised of the island isolates in one subclade and two pairs of ES (ES3208 and ES3197) and post-fadeout CSL isolates (CSL13546 and CSL13755). The two pig isolates and a single post-fadeout CSL isolate (CSL13990) form Clade3, and the 1988 CSL isolate clusters as an outgroup.

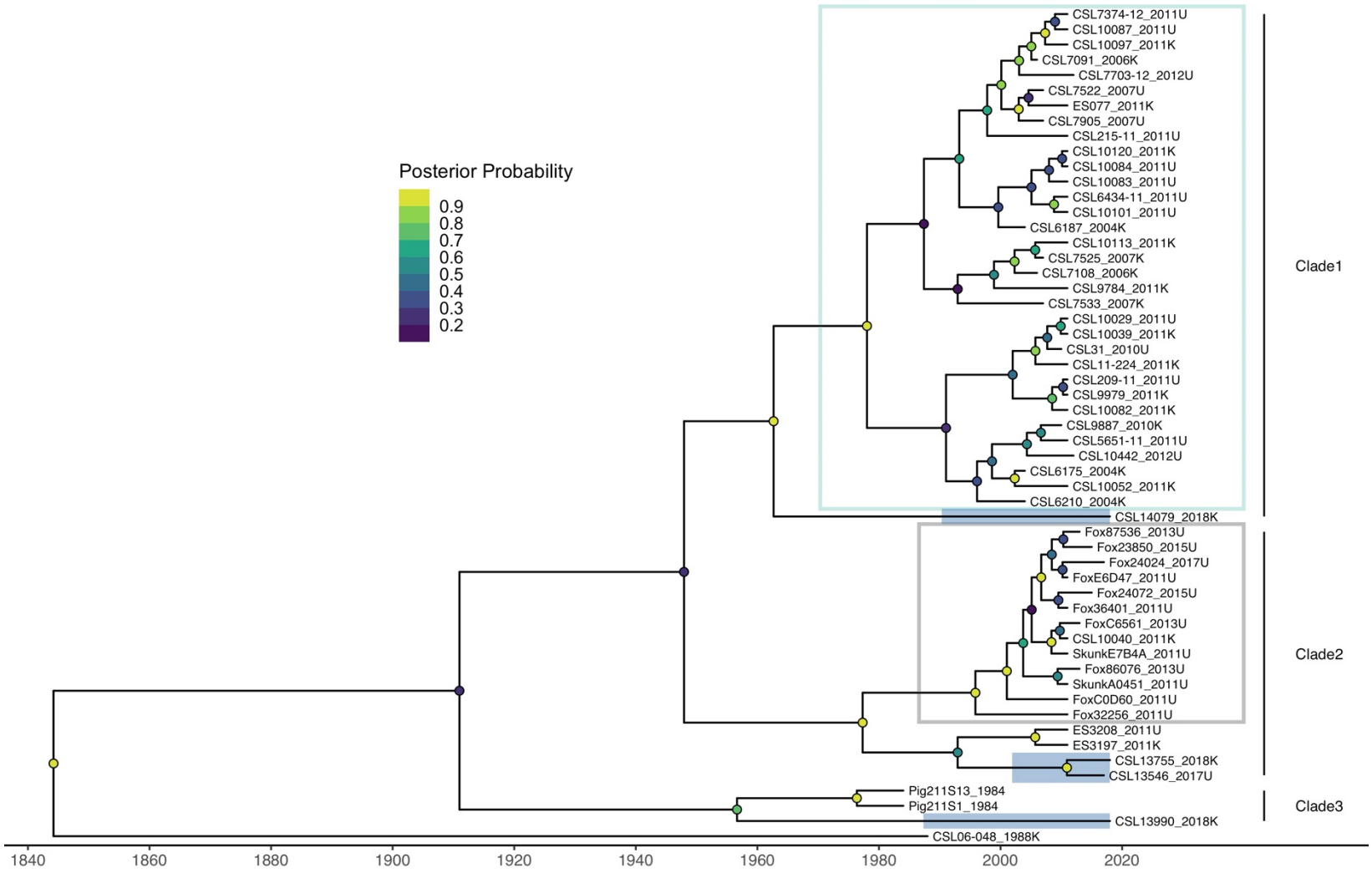

Figure 1.2: Time-calibrated maximum clade credibility tree of *Leptospira interrogans* serovar Pomona in the Channel Island ecosystem. Interior nodes are labelled with the posterior probability (lower values are purple, higher values are yellow). Tips are labeled with host species (CSL = California sea lions, ES = elephant seal), ID number, year, and tissues type (K = kidney, U = urine). CSL isolates sampled between 2000 and 2013 (pre-fadeout) are contained within the light blue box. The terrestrial island isolates are denoted by the grey box. Branches of post-fadeout CSL isolates are highlighted in dark blue.

The median substitution rate was estimated to be 1.36 x 10^{-7} substitutions per site per year, which is equivalent to 0.43 SNPs per year per genome. This evolutionary rate varied highly among the branches, with the 95% highest probability density (HPD) covering 0.04 – 2.48 x 10^{-7} substitutions per site per year (Table 1.4). The root of the tree was predicted to occur in the year 1844.

Table 1.4: Posterior median estimates of genetic and temporal parameters.

Parameter	Median (95% HPD)
Clock rate ($\times 10^{-7}$)	1.36 (0.04, 2.48)
Coefficient of variation$^{\perp}$	2.27 (1.44, 3.73)
Root height date	1844 (1515, 1966)
Monophyly of Pre-fadeout clade	0.996*
Monophyly of Pre-fadeout clade + CSL14079	0.899*

HPD: highest posterior density

$^{\perp}$Measure of the variation in evolutionary rate amongst the branches

*Mean is shown

There are four notably robust elements to the tree topology. The island isolate element (within Clade2), which contains all skunk and fox isolates in addition to a single CSL isolate (CSL10040; Figure 1.2 grey box), has high confidence across the posterior set of trees. The node defining Clade2 also has a high posterior support; this branchpoint indicates that two of the post-fadeout CSL isolates (CSL13755 and CSL13546) are highly unlikely to be found elsewhere in the tree. These isolates are more related to the 2011 ES isolates and the terrestrial island isolates than to Clade1. Likewise, the posterior probability for the node of the pre-fadeout isolates (deepest node within the blue box of Figure 1.2) is high, but the arrangement within this subclade has less posterior support, indicating the relative positions of the subclade were

much more variable across the posterior set of trees. Finally, the Clade1 juncture shows that CSL14079 is more similar to the pre-fadeout isolates than the terrestrial isolates.

Because the central hypothesis of our work is that post-fadeout isolates are not descended from pre-fadeout isolates (i.e. the post-fadeout isolates are derived from one or more separate lineages introduced from some outside reservoir), we investigated the CSL14079 isolates in more depth. To quantify how often CSL14079 was included in Clade1, we summarized two monophyly statistics (Table 1.4). The pre-fadeout clade (Figure 1.2; light blue box) was monophyletic in 99.6% of posterior trees, indicating that this topology is quite robust. When CSL14079 was included with the set of pre-fadeout isolates (to form Clade1), the monophyly was preserved in 89.9% of posterior trees, meaning that CSL14079 clusters away from the pre-fadeout clade about 10% of the time.

More broadly, this time-calibrated phylogenetic reconstruction indicates that none of the post-fadeout strain(s) are directly descended from the pre-fadeout clade. CSL14079 is most related to the Marine1 clade but does not group within the clade, as would be expected if this isolate was descended from the pre-fadeout CSL strain. The two monophyly measures show that CSL14079 is highly unlikely to cluster within the pre-fadeout clade because the monophyly of the pre-fadeout was only broken in 0.4% of posterior trees. When this evidence is combined with the lower monophyly statistic for Clade1, it supports that CSL14079 and the other post-fadeout isolates are not descendants of the pre-fadeout clade, simply that CSL14079 is more similar to the pre-fadeout strain than to the others in the tree.

The CSL13546 and CSL13755 isolates cluster with the 2011 ES pair, but these ES are an unlikely source for these post-fadeout isolates. The strain found in this ES pair was isolated

while they were in a rehabilitation facility and are believed to have acquired infection after arriving (Delaney et al., 2014). The genetic data show clearly that the source of these ES infections was not from a wild CSL, but rather from some unknown external reservoir; this suggests that these two post-fadeout CSL isolates originated from an external source as well, rather than from ES. Additional serologic evidence (unpublished) shows that ES seroprevalence plummeted between 2013 and 2017 when pathogen fadeout occurred in CSL. This evidence strongly suggests that ES are not the reservoir for *Leptospira.* It is possible (though not necessary or parsimonious) that ES acted as bridge in the re-emergence in CSL, as there were cases of leptospirosis in ES in early 2017.

Finally, CSL13990 clusters with the two pig isolates. These isolates were taken in 1984 in Iowa from neonatal swine. Clearly, the CSL infection did not originate directly from these pigs, but this clustering links this post-fadeout isolate to a strain that was once in a terrestrial host. This evidence all indicates that there is an unknown reservoir host, likely a terrestrial host on the mainland, which has seeded the ecosystem multiple times with closely related strains. At least five (and as many as eight) introductions have occurred into the ecosystem since the 1980s (Table 1.5). The 1988 CSL isolate is distinct from the pre-fadeout CSL clade, meaning at least two strains have circulated in the CSL population prior to pathogen fadeout. The island clade appears to have had its own strain introduction. The strain now circulates readily on the island with evidence of occasional spillover into CSL (CSL10040), but there is no evidence that the island terrestrial strain seeded the post-fadeout CSL outbreak. The two 2011 ES isolates are the fourth strain and are most likely a result of spillover from an unknown terrestrial host. The single leptospirosis case detected in a stranded CSL in 2016 represents another strain

introduction into the ecosystem, if our interpretation of that event as a failed re-emergence is correct. Finally, the post-fadeout CSL isolates represent at least one introduction of a new strain into the ecosystem. It is plausible that three separate introductions occurred into the CSL population during the post-fadeout period, but further analysis of these isolates and more information on potential terrestrial reservoir hosts is needed to resolve the uncertainty.

Table 1.5: Potential introductions of related strains of *Leptospira interrogans* serovar Pomona in the coastal California ecosystem.

Evidence of new strain introduction	Potential date of introduction
1988 CSL	1984-1988
Pre-fadeout CSL	Prior to 2004
Island clade	Prior to 2000
ES3208 and ES3197	2011
2016 stranded CSL	2016
Post-fadeout CSL	Probably 2017

1.4 Discussion

In these case studies, we used whole genome sequencing of *Leptospira interrogans* serovar Pomona isolates to reconstruct transmission linkages within the California coastal ecosystem. We showed that the source of the *Leptospira* outbreak in Channel Island foxes post-reintroduction was not from California sea lions. Additionally, when *Leptospira* re-emerged in the CSL population after a 4-year absence in the wild, our data provide strong corroborating evidence that the pre-emergence strain had disappeared and show that re-emergence was the result of the introduction of one or more new strains. More broadly, our data suggest that transmission occurs primarily within the terrestrial or marine ecosystems but can occasionally occur across ecosystems as well, challenging conventional notions of marine island as closed

systems and echoing past work on resource subsidies across ecosystem boundaries (Borremans et al., 2019). There is additional evidence that an unknown reservoir host (putatively a mainland terrestrial species) may be seeding independent introductions of *L. interrogans* serovar Pomona into the different host species within the California coastal ecosystem.

Our lab's recent survey of micro- and macroparasites in urban wildlife (e.g. squirrels, raccoons, rabbits, coyotes, opossum) of Southern California detected evidence of *Leptospira interrogans* exposure or infection in several species (Helman et al., unpublished). Using the microagglutination test (MAT) to detect anti-*Leptospira* antibodies, we detected evidence of exposure to serovar Pomona in coyotes, raccoons, opossums, and skunks. However, MAT can be misleading with respect to infecting serovar because it is subject to high levels of cross reactivity (Chirathaworn et al., 2014; Mummah et al., unpublished). Active *Leptospira* infections (via PCR) were detected at low levels in these four species as well. Combined, this evidence suggests that any of these mammals are prime candidates as sources of *Leptospira* transmission to marine and island communities. Another study has also shown the presence of feral swine along the California coast which have serologic evidence of infection by serovar Pomona, making them another candidate reservoir for the pathogen (Pedersen et al., 2015). In general, more research is needed to understand the contribution of coastal mainland terrestrial species to the *Leptospira* transmission dynamics of the island and marine systems.

This work was not without its challenges. As in most multi-host studies, we have uneven sampling across the four host species, which limited our ability to estimate of the frequency of cross-species transmission. The uneven sampling is in part due to difficulty in obtaining samples from hosts but also in culturing isolates. *Leptospira* is very slow-growing and difficult to culture,

which make a study of this kind very difficult to execute. Additionally, *Leptospira* has a slower substitution rate compared to other bacteria, with high variation across the branches of the tree (Duchêne et al., 2016). Slow substitution rates can result in a limited number of differences between isolates despite the temporal depth in sampling and can hinder the power with which the tree topology is formed. Uneven sampling across host species can also potentially lead to biases in inferred tree topology, including misinterpretation of the source and target species. However, in our system, we have other lines of evidence that support the inferences we have drawn (for instance, serologic data spanning the past decade indicate that ES are not a maintenance host for *Leptospira* (Prager et al., unpublished), and our long-term field surveillance shows that spotted skunks have played a steadily declining role in the SRI outbreak (Lloyd-Smith, 2021)).

In the future, greater sample sizes and temporal depth across host species would improve the estimates achieved in these case studies. Additional isolates from ES, particularly if sampled near the times when new strains were introduced to the CSL population, could reveal the potential role they play as a bridge between CSL and an unknown terrestrial host and the frequency of spillover from the unknown terrestrial host and the marine ecosystem. More post-fadeout isolates from CSL (including from later than 2018) would shed light on how many lineages are circulating after the pathogen re-introduction and may help resolve the structure of Clade2 and Clade3 in the phylogeny. Samples isolated from coastal mainland terrestrial hosts would be most critical in linking the repeated introductions of *Leptospira* in the marine and island communities to a terrestrial reservoir host. Beyond this system, obtaining broad temporal depth is crucial for accurately estimating divergence, especially for slow-evolving

pathogens. When the temporal range of isolates is too narrow, the topology can have low posterior support, as is seen within the pre-fadeout clade of CSL which has a temporal range of 7 years (2004-2011).

Within *Leptospira* more broadly, there is the potential for recombination, which could bias branch length estimation and node dating on the tree (Kakita et al., 2021; Thaipadungpanit et al., 2007; Vos & Didelot, 2009). However, the reported recombination rate for *Leptospira interrogans* is low (Thaipadungpanit et al., 2007; Vos & Didelot, 2009). For these analyses, we chose not to address recombination because our primary goal of bacterial phylogenetic reconstruction is robust to recombination (Hedge & Wilson, 2014). There is potential within these isolates and *Leptospira* more broadly to identify gene regions with high levels recombination amongst a group of related isolates, which would further our understanding of *Leptospira* genomics and the role of recombination in its evolution.

Using this unique system of *Leptospira* dynamics across a marine-terrestrial ecosystem, we have demonstrated multiple ways to infer cross-species transmission from genetic data. We illustrated clear proximate evidence of transmission when a tip is nested in a clade with isolates from other species (e.g. CSL10040 within the island clade or ES077 within the pre-fadeout CSL clade), and we showed that the deeper branching structure of the tree can indicate cross-species transmission events within and across ecosystems. However, our inferences were possible because of the extraordinary temporal depth, host breadth, and sampling coverage in our dataset. Long-term surveillance programs have the potential to capitalize on the temporal depth of their datasets and pair long-term ecological datasets with modern phylogenetic methods. Additionally, the use of historic or banked samples could aid in the application of

these methods and could also improve the temporal depth of a dataset, as it did in these case studies with the inclusion of the 1988 CSL and 1984 pig isolates. Furthermore, integrating across different long-term monitoring programs when relevant could improve the host breadth or provide vital ecological context to interpret genetic patterns. For instance, Channel Islands National Park has multiple long-term monitoring programs in place but are usually class specific (i.e. land birds, rocky intertidal zone species). Our work has shown the importance of long-term surveillance and sample collection in understanding pathogen dynamics in wildlife.

The application of this methodology and breadth of data is novel to *Leptospira* ecology, and few studies have examined the ecology of cross-species transmission for *L. interrogans*. Although other studies have applied whole genome sequencing to a limited number of samples, our studies are the first with sufficient breadth of host species and depth of temporal resolution to examine the fine structure of *Leptospira* lineages within a single serovar. Increased temporal depth and host breadth would provide still greater opportunities to learn more about the frequency and nature of cross-species transmission within the coastal California ecosystem, and for *L. interrogans* serovar Pomona more broadly. Beyond *Leptospira,* our work illustrates the range of relative transmission frequencies within and across species, and the utility of whole genome sequencing as a component of a multi-disciplinary, multi-data source study to untangle the complexity of wildlife disease systems when cross-species and within-species transmission occur with similar frequency.

1.5 Methods

1.5.1 Bacterial culturing

Sterile urine samples collected via cystocentesis were cultured by placing 1cc sterile urine into EMJH medium at 1:10 dilution (Ellinghausen & McCullough, 1965; Johnson & Harris, 1967). Culture media was inoculated with 0.1cc of the diluted urine (Adler, 2015). Cultures were stored with lids lightly open in a dark room at room temperature until they were sent to the National Animal Disease Center (NADC) in Ames, IA, where they were kept in an incubator at 30°C for at least 6 months. Cultures were either shipped to NADC or CDC in Atlanta, GA and examined periodically for growth using dark field microscopy.

1.5.2 DNA preparation and whole-genome sequencing

When sufficient genetic material was available for an isolate, DNA was extracted using Qiagen Blood Tissue Kit for whole genome sequencing (WGS). RNase was added before shipping the extractions to USDA National Animal Disease Center, where they were sequenced using Illumina MiSeq, targeting 80-fold coverage across the genome. Sequencing was conducted in three separate runs and produced 2x250 bp paired-end reads.

1.5.3 Genome assembly

Raw short reads were error-corrected and assembled using the shovill pipeline (https://github.com/tseemann/shovill). Briefly, the depth of the FASTQ files was reduced and adapters were trimmed using Trimmomatic version 0.39. Read sequencing errors were corrected and paired-end reads were pre-overlapped prior to assembly. Reads were assembled

using SPAdes (Bankevich et al., 2012). Assembled contigs were error corrected by mapping the reads back to the contigs; low quality and short contigs were removed. The resulting contigs were annotated using PROKKA (https://github.com/tseemann/prokka), requiring a minimum contig length of 200 to ensure Genbank compliance. Core and accessory genes were identified using Roary (Page et al., 2015). In turn, Roary produced a core genome multi-FASTA alignment using MAFFT v7.475. All sites were required to have a base call in the final multi-sequence alignment. The alignment used for Case Study #1 ultimately contained 3,188,030 sites, whereas the Case Study #2 alignment contained 3,157,250 sites.

1.5.4 Isolate selection

In total, seventy-two isolates were sequenced from fifty-three unique hosts (See Table S1.1). We have duplicates in two ways: some hosts have two isolates from different tissues, and some isolates were resequenced in the last sequencing run. One isolate for each unique host was selected for analysis. When a host had two isolates from different tissues (7 CSL; 1 ES), the urine-cultured isolate was selected for the sea lions while the kidney isolate was selected for the elephant seal (since all other elephant seal isolates were from kidney tissue). Nearly all of these dual isolates were in the second sequencing run, and this selection resulted in a more even across-batch selection for the phylogenetic analyses. For resequenced isolates, the isolate sequenced in the third sequencing run was selected. This resulted in 13, 13, and 12 CSL isolates being selected from each sequencing run. We did not see substantial differences between any two isolates from the same individual host. Two additional isolates were included which were

isolated from neonatal swine in Iowa; these were sequenced in a separate batch using the same library preparation and sequencing techniques. Isolate selection is summarized in upplement Table S1.1.

1.5.5 Phylogenetic reconstruction of L. interrogans over time

We reconstructed the evolutionary relationships among *L. interrogans* serovar Pomona isolates by incorporating molecular sequence and temporal data using a Bayesian Markov chain Monte Carlo (MCMC) analysis in BEAST v1.10.1 to estimate a time-calibrated phylogeny (Suchard et al., 2018). We applied a Hasegawa, Kishino, and Yano nucleotide substitution model with gamma-distributed rate heterogeneity and a proportion of invariant sites (HKY + Γ + I), and we also selected an uncorrelated relaxed molecular clock model as the clock model prior in these phylogenetic analyses (Drummond et al., 2006; Hasegawa et al., 1985). Three independent MCMC chains were run for 100 million generations, and posterior distributions were sampled every 1,000 generations. Independent chains were combined using LogCombiner and the first 10% of each independent chain was discarded as burn-in. Model parameters were assessed for convergence and sufficient effective sample sizes (>200) in Tracer v1.7.1 (Rambaut et al., 2018). The maximum clade credibility tree was identified in TreeAnnotator (Suchard et al., 2018).

1.5.6 Post hoc monophyly statistic

We constructed two monophyly statistics to measure the preservation of clades within the posterior set of trees. Monophyly statistics measure how often a set of isolates (taxon set)

is preserved as a group with no additional isolates included or any isolates from the set missing. Each posterior tree is evaluated as having the taxon set preserved (1) or not (0). The arrangement of the taxon set can change but does not affect the statistic. All isolates within the pre-fadeout set (Figure 1.2; light blue box) were included in the Pre-fadeout clade monophyly statistic. The second monophyly statistic included the same isolates as the first in addition to CSL14079 (which together makes the isolates of Clade1).

1.6 Supplement

Table S1.1: Sequenced genomes from four host species in the Channel Island Ecosystem.

Sequencing batch	Sample Name	ID	Host	Sample tissue	Year	Paired?	Resequenced?
2	CIslandFox24072_S23	CIslandFox24072	Fox	Urine	2015		
2	CIslandFox32256_S3	CIslandFox32256	Fox	Urine	2011		
2	CIslandFox36401_S2	CIslandFox36401	Fox	Urine	2011		
2	CIslandFoxC0D60_S4	CIslandFoxC0D60	Fox	Urine	2011		
2	CIslandFoxE6D47_S1	CIslandFoxE6D47	Fox	Urine	2011		
2	CIslandSkunkA0451_S5	CIslandSkunkA0451	Skunk	Urine	2011		
2	CIslandSkunkE7B4A_S6	CIslandSkunkE7B4A	Skunk	Urine	2011		
1	CSL10029_S4	CSL10029	CSL	Urine	2011		
1	CSL10039_S5	CSL10039	CSL	Kidney	2011		
1	CSL10040_S6	CSL10040	CSL	Kidney	2011		Y
3	CSL10040-Kidney_S15	CSL10040	CSL	Kidney	2011		Y
1	CSL10052_S8	CSL10052	CSL	Kidney	2011		
1	CSL10082_S9	CSL10082	CSL	Kidney	2011		Y
3	CSL10082K-1_S23	CSL10082	CSL	Kidney	2011		Y
3	CSL10082K-2_S24	CSL10082	CSL	Kidney	2011		Y
1	CSL10083_S10	CSL10083	CSL	Kidney	2011	Y	Y
2	CSL10083_S7	CSL10083	CSL	Urine	2011	Y	Y
3	CSL10083K_S22	CSL10083	CSL	Kidney	2011		Y
3	CSL10083U_S21	CSL10083	CSL	Urine	2011		Y
1	CSL10084_S11	CSL10084	CSL	Kidney	2011	Y	
2	CSL10084_S8	CSL10084	CSL	Urine	2011	Y	
1	CSL10087_S12	CSL10087	CSL	Kidney	2011	Y	
2	CSL10087_S9	CSL10087	CSL	Urine	2011	Y	
1	CSL10097_S16	CSL10097	CSL	Kidney	2011		
2	CSL10101_S10	CSL10101	CSL	Urine	2011	Y	
1	CSL10101_S13	CSL10101	CSL	Kidney	2011	Y	
1	CSL10113_S14	CSL10113	CSL	Kidney	2011		
1	CSL10120_S15	CSL10120	CSL	Kidney	2011		
3	CSL10442_S11	CSL10442	CSL		2012		
2	CSL11-224_S19	CSL11-224	CSL	Kidney	2011		
3	CSL13546_S5	CSL13546	CSL	Kidney	2017	Y	
3	CSL13546-Urine_S12	CSL13546	CSL	Urine	2017	Y	
3	CSL13755-Kidney_S13	CSL13755	CSL	Kidney	2018		
3	CSL13990-Kidney_S6	CSL13990	CSL	Kidney	2018		
3	CSL14079-Kidney_S7	CSL14079	CSL	Kidney	2018		
1	CSL6175_S1	CSL6175	CSL	Kidney	2004		
2	CSL6187_S11	CSL6187	CSL	Kidney	2004		
3	CSL6210_S17	CSL6210	CSL	Kidney	2004		Y
2	CSL6210_S18	CSL6210	CSL	Kidney	2004		Y
2	CSL7091_S12	CSL7091	CSL	Kidney	2006		
2	CSL7108_S13	CSL7108	CSL	Kidney	2006		
2	CSL7522	CSL7522	CSL	Kidney	2007	Y	
2	CSL7522	CSL7522	CSL	Urine	2007	Y	
2	CSL7525_S16	CSL7525	CSL	Kidney	2007		
2	CSL7533_S15	CSL7533	CSL	Kidney	2007		
2	CSL7905	CSL7905	CSL	Urine	2008	Y	
2	CSL7905	CSL7905	CSL	Kidney	2008	Y	
1	CSL9784_S7	CSL9784	CSL	Kidney	2011		
1	CSL9887_S2	CSL9887	CSL	Kidney	2010		Y

3	CSL9887K_S19	CSL9887	CSL	Kidney	2010		Y
1	CSL9979_S3	CSL9979	CSL	Kidney	2011		
3	E7B4A_S20	CIslandSkunkE7B4A	CIS	Urine	2011		Y
1	ES3197Kidney_S21	ES3197	NES	Kidney	2011	Y	
1	ES3197Urine_S22	ES3197	NES	Urine	2011	Y	
1	ES3208_S23	ES3208	NES	Urine	2011		
3	Fox23850_S1	CIslandFox23850	CIF	Urine	2015		
3	Fox24024_S4	CIslandFox24024	CIF	Urine	2017		
3	Fox86076_S8	CIslandFox86076	CIF	Urine	2013		
3	Fox87536_S2	CIslandFox87536	CIF	Urine	2013		
3	FoxC6561_S9	CIslandFoxC6561	CIF	Urine	2013		
2	Po-06-048_S22	Po-06-048	CSL	Kidney	1988		
3	Po06-048_S14	Po-06-048	CSL	Kidney	1988		Y
1	WCSL209-11_S18	WCSL209-11	CSL	Urine	2011		
1	WCSL215-11_S19	WCSL215-11	CSL	Urine	2011		
1	WCSL31_S17	WCSL31	CSL	Urine	2010		Y
3	WCSL31U_S18	WCSL31	CSL	Urine	2010		Y
1	WCSL5651-11_S20	WCSL5651-11	CSL	Urine	2011		
3	WCSL6434-11_S3	WCSL6434-11	CSL		2011		
2	WCSL7374-12_S24	WCSL7374-12	CSL	Urine	2011		
3	WCSL7703-12_S10	WCSL7703-12	CSL		2012		
3	WDFW2011-077_S16	WDFW2011-077	NES	Kidney	2011		Y
1	WDFW2011-077_S24	WDFW2011-077	NES	Kidney	2011		Y
4	Pig211_S1	Pig211S1	Pig	Kidney	1984		
4	Pig211_S13	Pig211S13	Pig	Kidney	1984		

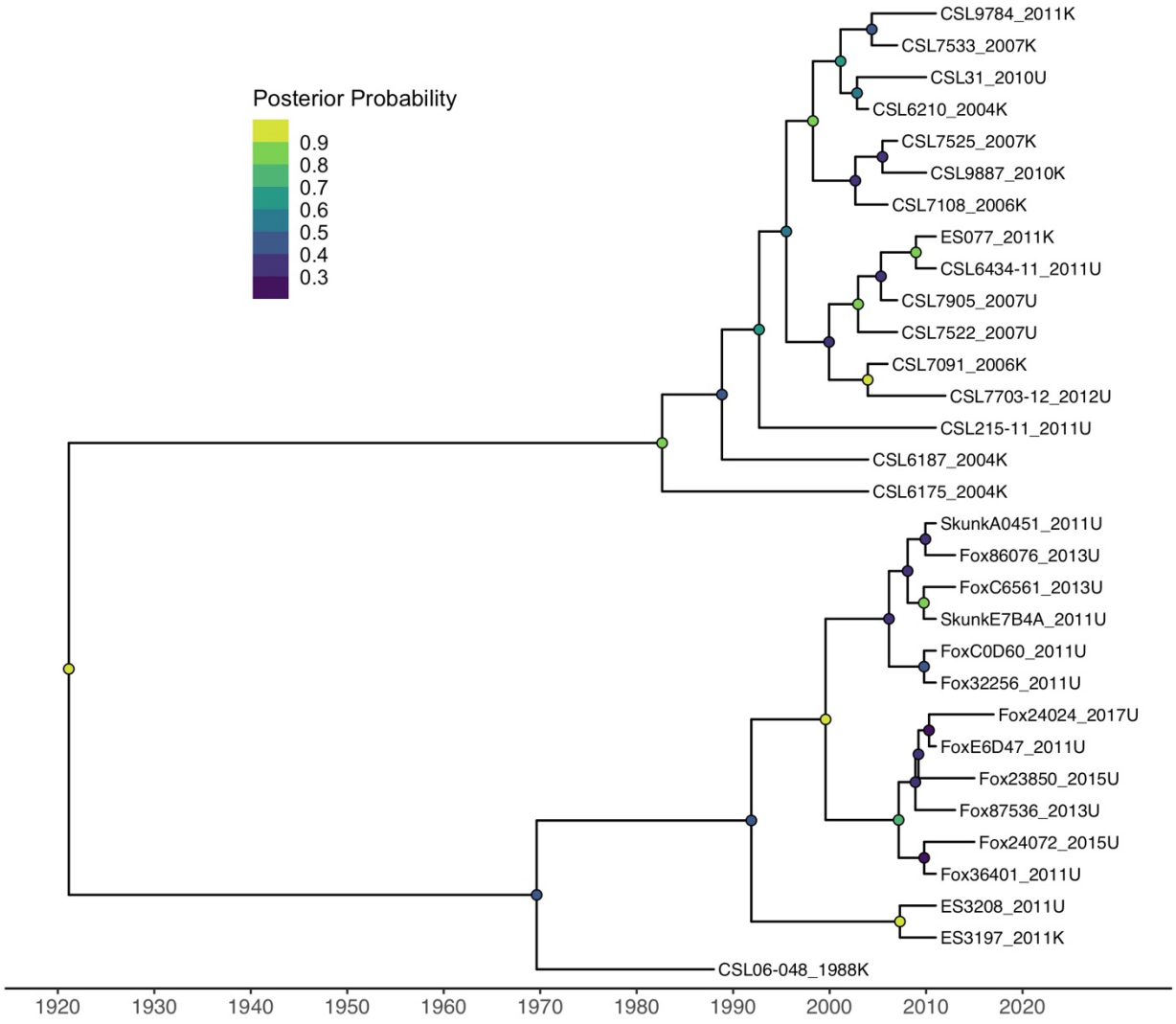

Figure S1.1: Analysis of the effect of the imbalanced sampling across host species through time. There are more 2011 CSL isolates than other host-year combinations. To analyze the effect of the inclusion of these isolates on the tree topology, we randomly subsampled the 2011 CSL isolates down to three and reconstructing the tree with all other isolates. Our analysis shows that the tree topology is robust to the exclusion of these isolates.

1.7 References

Adler, B., & de la Pena Montesuma, A. (2010). Leptospira and leptospirosis. *Veterinary Microbiology*, *140*, 287–296. https://doi.org/10.1016/j.vetmic.2009.03.012

Allan, K. J., Maze, M. J., Galloway, R. L., Rubach, M. P., Biggs, H. M., Halliday, J. E. B., Cleaveland, S., Saganda, W., Lwezaula, B. F., Kazwala, R. R., Mmbaga, B. T., Maro, V. P., & Crump, J. A. (2020). Molecular Detection and Typing of Pathogenic Leptospira in Febrile Patients and Phylogenetic Comparison with Leptospira Detected among Animals in Tanzania. *The American Journal of Tropical Medicine and Hygiene*, *103*(4), 1427–1434. https://doi.org/10.4269/ajtmh.19-0703

Almberg, E. S., Mech, L. D., Smith, D. W., Sheldon, J. W., & Crabtree, R. L. (2009). A Serological Survey of Infectious Disease in Yellowstone National Park's Canid Community. *PLoS ONE*, *4*(9), e7042. https://doi.org/10.1371/journal.pone.0007042

Bajardi, P., Poletto, C., Ramasco, J. J., Tizzoni, M., Colizza, V., & Vespignani, A. (2011). Human Mobility Networks , Travel Restrictions , and the Global Spread of 2009 H1N1 Pandemic. *PLoS ONE*, *6*(1), e16591. https://doi.org/10.1371/journal.pone.0016591

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. A., & Pevzner, P. A. (2012). SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *Journal of Computational Biology*, *19*(5), 455–477. https://doi.org/10.1089/cmb.2012.0021

Begon, M., Hazel, S. M., Baxby, D., Bown, K., Cavanagh, R., Chantrey, J., Jones, T., & Bennett, M. (1999). Transmission dynamics of a zoonotic pathogen within and between wildlife host species. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *266*(1432), 1939–1945. https://doi.org/10.1098/rspb.1999.0870

Benton, C. H., Delahay, R. J., Trewby, H., & Hodgson, D. J. (2015). What has molecular epidemiology ever done for wildlife disease research? Past contributions and future directions. *European Journal of Wildlife Research*, *61*(1), 1–16. https://doi.org/10.1007/s10344-014-0882-4

Biek, R., & Real, L. A. (2010). The landscape genetics of infectious disease emergence and spread. *Molecular Ecology*, *19*(17), 3515–3531. https://doi.org/10.1111/j.1365-294X.2010.04679.x

Bishara, J., Amitay, E., Barnea, A., Yitzhaki, S., & Pitlik, S. (2002). Epidemiological and Clinical Features of Leptospirosis in Israel. *European Journal of Clinical Microbiology and Infectious Diseases*, *21*(1), 50–52. https://doi.org/10.1007/s10096-001-0660-6

Bolt, I., & Marshall, R. B. (1995). The epidemiology of *Leptospira interrogans* serovar pomona in grower pig herds. *New Zealand Veterinary Journal*, *43*(1), 10–15. https://doi.org/10.1080/00480169.1995.35833

Bonilla-Aldana, D. K., Dhama, K., & Rodriguez-Morales, A. J. (2020). Revisiting the One Health Approach in the Context of COVID-19: A Look into the Ecology of this Emerging Disease. *Advances in Animal and Veterinary Sciences*, *8*(3). https://doi.org/10.17582/journal.aavs/2020/8.3.234.237

Borremans, B., Faust, C., Manlove, K. R., Sokolow, S. H., & Lloyd-Smith, J. O. (2019). Cross-species pathogen spillover across ecosystem boundaries: Mechanisms and theory. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *374*(1782), 20180344. https://doi.org/10.1098/rstb.2018.0344

Buhnerkempe, M. G., Prager, K. C., Strelioff, C. C., Greig, D. J., Laake, J. L., Melin, S. R., DeLong, R. L., Gulland, F. M., & Lloyd-Smith, J. O. (2017). Detecting signals of chronic shedding to explain pathogen persistence: *Leptospira interrogans* in California sea lions. *Journal of Animal Ecology*, *86*(3), 460–472. https://doi.org/10.1111/1365-2656.12656

Buhnerkempe, M. G., Roberts, M. G., Dobson, A. P., Heesterbeek, H., Hudson, P. J., & Lloyd-smith, J. O. (2015). Eight challenges in modelling disease ecology in multi-host , multi-agent systems. *Epidemics*, *10*, 26–30. https://doi.org/10.1016/j.epidem.2014.10.001

Carretta, J. V., Forney, K. A., Muto, M. M., Barlow, J., Baker, J., Hanson, B., & Lowry, M. S. (2006). U. S. Pacific marine mammal stock assessments: 2005. In *NOAA Technical Memorandum NMFS*.

Carslake, D., Bennett, M., Hazel, S., Telfer, S., & Begon, M. (2006). Inference of cowpox virus transmission rates between wild rodent host classes using space–time interaction. *Proceedings of the Royal Society B: Biological Sciences*, *273*(1588), 775–782. https://doi.org/10.1098/rspb.2005.3400

Chirathaworn, C., Inwattana, R., Poovorawan, Y., & Suwancharoen, D. (2014). Interpretation of microscopic agglutination test for leptospirosis diagnosis and seroprevalence. *Asian Pacific Journal of Tropical Biomedicine*, *4*, S162–S164. https://doi.org/10.12980/APJTB.4.2014C580

Cleaveland, S., Appel, M. G. J., Chalmers, W. S. K., Chillingworth, C., Kaare, M., & Dye, C. (2000). Serological and demographic evidence for domestic dogs as a source of canine distemper virus infection for Serengeti wildlife. *Veterinary Microbiology*, *72*(3), 217–227. https://doi.org/10.1016/S0378-1135(99)00207-2

Cleaveland, S., Laurenson, M. K., & Taylor, L. H. (2001). Diseases of humans and their domestic mammals: Pathogen characteristics, host range and the risk of emergence. *Philosophical*

Transactions of the Royal Society of London. Series B: Biological Sciences, *356*(1411), 991–999. https://doi.org/10.1098/rstb.2001.0889

Coonan, T. J., Schwemm, C. A., & Garcelon, D. K. (2010). *Decline and recovery of the Island Fox: A case study for population recovery*. Cambridge University Press.

Coonan, T. J., Schwemm, C. A., Roemer, G. W., Garcelon, D. K., & Munson, L. (2005). Decline of an Island Fox Subspecies to near Extinction. *The Southwestern Naturalist*, *50*(1), 32–41. JSTOR.

Crispell, J., Benton, C. H., Balaz, D., De Maio, N., Ahkmetova, A., Allen, A., Biek, R., Presho, E. L., Dale, J., Hewinson, G., Lycett, S. J., Nunez-Garcia, J., Skuce, R. A., Trewby, H., Wilson, D. J., Zadoks, R. N., Delahay, R. J., & Kao, R. R. (2019). Combining genomics and epidemiology to analyse bi-directional transmission of Mycobacterium bovis in a multi-host system. *ELife*, *8*, e45833. https://doi.org/10.7554/eLife.45833

Crooks, K. R., & Van Vuren, D. (1995). Resource utilization by two insular endemic mammalian carnivores, the island fox and island spotted skunk. *Oecologia*, *104*(3), 301–307. https://doi.org/10.1007/BF00328365

Delaney, M. A., Colegrove, K. M., Spraker, T. R., Zuerner, R. L., Galloway, R. L., & Gulland, F. M. D. (2014). Isolation of *Leptospira* from a Phocid: Acute Renal Failure and Mortality from Leptospirosis in Rehabilitated Northern Elephant Seals (*Mirounga angustirostris*), California, USA. *Journal of Wildlife Diseases*, *50*(3), 621–627. https://doi.org/10.7589/2013-08-195

Drummond, A. J., Ho, S. Y. W., Phillips, M. J., & Rambaut, A. (2006). Relaxed Phylogenetics and Dating with Confidence. *PLoS Biology*, *4*(5), e88. https://doi.org/10.1371/journal.pbio.0040088

Duchêne, S., Holt, K. E., Weill, F.-X., Le Hello, S., Hawkey, J., Edwards, D. J., Fourment, M., & Holmes, E. C. (2016). Genome-scale rates of evolutionary change in bacteria. *Microbial Genomics*, *2*(11), e000094. https://doi.org/10.1099/mgen.0.000094

Ellinghausen, H. C. J., & McCullough, W. G. (1965). Nutrition of Leptospira pomona and growth of 13 other serotypes: A serum-free medium employing oleic albumin complex. *American Journal of Veterinary Research*, *26*, 39–44.

Grant, R., Nguyen, L.-B. L., & Breban, R. (2020). Modelling human-to-human transmission of monkeypox. *Bulletin of the World Health Organization*, *98*(9), 638–640. https://doi.org/10.2471/BLT.19.242347

Greig, D. J., Gulland, F. M. D., & Kreuder, C. (2005). A Decade of Live California Sea Lion (*Zalophus californianus*) Strandings Along the Central California Coast: Causes and Trends, 1991-2000. *Aquatic Mammals*, *31*(1), 11–22. https://doi.org/10.1578/AM.31.1.2005.11

Gulland, F. M. D., Koski, M., Lowenstine, L. J., Colagross, A., Morgan, L., & Spraker, T. (1996). Leptospirosis in California sea lions (*Zalophus californianus*) stranded along the central California coast, 1981-1994. *Journal of Wildlife Diseases*, *32*(4), 572–580. https://doi.org/10.7589/0090-3558-32.4.572

Hasegawa, M., Kishino, H., & Yano, T. (1985). Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution*, *22*(2), 160–174. https://doi.org/10.1007/BF02101694

Hedge, J., & Wilson, D. J. (2014). Bacterial Phylogenetic Reconstruction from Whole Genomes Is Robust to Recombination but Demographic Inference Is Not. *MBio*, *5*(6), e02158-14. https://doi.org/10.1128/mBio.02158-14

Helman, S. K., Tokuyama, A. F., Mummah, R. O., Prager, K. C., Friscia, A., Byrne, H., Lynch-Alfaro, J., & Lloyd-Smith, J. O. (unpublished). *Pathogens of zoonotic and conservation concern in Los Angeles County, California*.

Johnson, R. C., & Harris, V. G. (1967). Differentiation of Pathogenic and Saprophytic Leptospires I. Growth at Low Temperatures. *Journal of Bacteriology*, *94*(1), 27–31. https://doi.org/10.1128/jb.94.1.27-31.1967

Jones, K. L., Van Vuren, D. H., & Crooks, K. R. (2008). Sudden Increase in a Rare Endemic Carnivore: Ecology of the Island Spotted Skunk. *Journal of Mammalogy*, *89*(1), 75–86. https://doi.org/10.1644/07-MAMM-A-034.1

Kakita, T., Kuba, Y., Kyan, H., Okano, S., Morita, M., & Koizumi, N. (2021). Molecular and serological epidemiology of Leptospira infection in cats in Okinawa Island, Japan. *Scientific Reports*, *11*(1), 10365. https://doi.org/10.1038/s41598-021-89872-3

Kamath, P. L., Foster, J. T., Drees, K. P., Luikart, G., Quance, C., Anderson, N. J., Clarke, P. R., Cole, E. K., Drew, M. L., Edwards, W. H., Rhyan, J. C., Treanor, J. J., Wallen, R. L., White, P. J., Robbe-Austerman, S., & Cross, P. C. (2016). Genomics reveals historic and contemporary transmission dynamics of a bacterial disease among wildlife and livestock. *Nature Communications*, *7*(1), 11448. https://doi.org/10.1038/ncomms11448

Kilpatrick, A. M., Daszak, P., Jones, M. J., Marra, P. P., & Kramer, L. D. (2006). Host heterogeneity dominates West Nile virus transmission. *Proceedings of the Royal Society B: Biological Sciences*, *273*(1599), 2327–2333. https://doi.org/10.1098/rspb.2006.3575

Knowlton, J. L., Josh Donlan, C., Roemer, G. W., Samaniego-Herrera, A., Keitt, B. S., Wood, B., Aguirre-Muñoz, A., Faulkner, K. R., & Tershy, B. R. (2007). Eradication of non-native mammals and the status of insular mammals on the California Channel Islands, USA, and Pacific Baja California Peninsula Islands, Mexico. *The Southwestern Naturalist*, *52*(4), 528–540.

Levett, P. N. (2001). Leptospirosis. *Clinical Microbiology*, *14*(2), 296–326. https://doi.org/10.1128/CMR.14.2.296

Lloyd-Smith, J. O. (2021). *Leptospirosis in endangered island foxes and California sea lions: Outbreak prediction and prevention in a changing world* (Final RC-2635). University of California, Los Angeles.

Lloyd-Smith, J. O., Greig, D. J., Hietala, S., Ghneim, G. S., Palmer, L., St Leger, J., Grenfell, B. T., & Gulland, F. M. (2007). Cyclical changes in seroprevalence of leptospirosis in California sea lions: Endemic and epidemic disease in one host species? *BMC Infectious Diseases*, *7*(1), 125. https://doi.org/10.1186/1471-2334-7-125

McDonald, J. L., Robertson, A., & Silk, M. J. (2018). Wildlife disease ecology from the individual to the population: Insights from a long-term study of a naturally infected European badger population. *Journal of Animal Ecology*, *87*(1), 101–112. https://doi.org/10.1111/1365-2656.12743

Melin, S. R., Delong, R. L., Thomason, J. R., & Vanblaricom, G. R. (2000). Attendance Patterns of California Sea Lion (zalophus Californianus) Females and Pups During the Non-Breeding Season at San Miguel Island. *Marine Mammal Science*, *16*(1), 169–185. https://doi.org/10.1111/j.1748-7692.2000.tb00911.x

Mummah, R. O., Gomez, A. C. R., Guglielmino, A., Borremans, B., Galloway, R. L., Prager, K. C., & Lloyd-Smith, J. O. (unpublished). *MAT is a minefield: A cautionary tale of a single Leptospira serovar in multiple host species*.

Page, A. J., Cummins, C. A., Hunt, M., Wong, V. K., Reuter, S., Holden, M. T. G., Fookes, M., Falush, D., Keane, J. A., & Parkhill, J. (2015). Roary: Rapid large-scale prokaryote pan genome analysis. *Bioinformatics*, *31*(22), 3691–3693. https://doi.org/10.1093/bioinformatics/btv421

Panaphut, T., Domrongkitchaiporn, S., & Thinkamrop, B. (2002). Prognostic factors of death in leptospirosis: A prospective cohort study in Khon Kaen, Thailand. *Int J Infect Dis*, *6*, 52–59.

Pedersen, K., Pabilonia, K. L., Anderson, T. D., Bevins, S. N., Hicks, C. R., Kloft, J. M., & Deliberto, T. J. (2015). Widespread detection of antibodies to *Leptospira* in feral swine in the United States. *Epidemiology and Infection*, *143*(10), 2131–2136. https://doi.org/10.1017/S0950268814003148

Peterson, R. S., & Bartholomew, G. A. (1967). *The natural history and behavior of the California sea lion*. The American Society of Mammologists.

Power, A. G., & Mitchell, C. E. (2004). Pathogen Spillover in Disease Epidemics. *The American Naturalist*, *164*(S5), S79–S89. https://doi.org/10.1086/424610

Prager, K. C., Borremans, B., Mummah, R. O., & Lloyd-Smith, J. O. (unpublished). *The role of northern elephant seals (*Mirounga angustirostris*) in Leptospira transmission dynamics in the marine ecosystem*.

Prager, K. C., Greig, D. J., Alt, D. P., Galloway, R. L., Hornsby, R. L., Palmer, L. J., Soper, J., Wu, Q., Zuerner, R. L., Gulland, F. M. D., & Lloyd-Smith, J. O. (2013). Asymptomatic and chronic carriage of Leptospira interrogans serovar Pomona in California sea lions (Zalophus californianus). *Veterinary Microbiology*, *164*(1–2), 177–183. https://doi.org/10.1016/j.vetmic.2013.01.032

Prager, K. C., Mazet, J. A. K., Dubovi, E. J., Frank, L. G., Munson, L., Wagner, A. P., & Woodroffe, R. (2012). Rabies Virus and Canine Distemper Virus in Wild and Domestic Carnivores in Northern Kenya: Are Domestic Dogs the Reservoir? *EcoHealth*, *9*(4), 483–498. https://doi.org/10.1007/s10393-013-0815-9

Rambaut, A., Drummond, A. J., Xie, D., Baele, G., & Suchard, M. A. (2018). Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. *Systematic Biology*, *67*(5), 901–904. https://doi.org/10.1093/sysbio/syy032

Roemer, G. W., Coonan, T. J., Garcelon, D. K., Bascompte, J., & Laughrin, L. (2001). Feral pigs facilitate hyperpredation by golden eagles and indirectly cause the decline of the island fox. *Animal Conservation*, *4*(4), 307–318. https://doi.org/10.1017/S1367943001001366

Roemer, G. W., Donlan, C. J., & Courchamp, F. (2002). Golden eagles, feral pigs, and insular carnivores: How exotic species turn native predators into prey. *Proceedings of the National Academy of Sciences*, *99*(2), 791–796.

Santos, R. F. dos, Silva, G. C. P. da, Assis, N. A. de, & Mathias, L. A. (2016). Aglutininas anti-Leptospira spp. Em equídeos da região sul do Brasil abatidos em matadouro-frigorífico. *Semina: Ciências Agrárias*, *37*(2), 841. https://doi.org/10.5433/1679-0359.2016v37n2p841

Sehgal, S., Murhekar, M., & Sugunan, A. (1995). Outbreak of leptospirosis with pulmonary involvement in north Andaman. *Indian J Med Res*, *102*, 9–12.

Sklenovská, N., & Van Ranst, M. (2018). Emergence of Monkeypox as the Most Important Orthopoxvirus Infection in Humans. *Frontiers in Public Health*, *6*, 241. https://doi.org/10.3389/fpubh.2018.00241

Suchard, M. A., Lemey, P., Baele, G., Ayres, D. L., Drummond, A. J., & Rambaut, A. (2018). Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evolution*, *4*(1). https://doi.org/10.1093/ve/vey016

Taylor, L. H., Latham, S. M., & woolhouse, M. E. J. (2001). Risk factors for human disease emergence. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *356*(1411), 983–989. https://doi.org/10.1098/rstb.2001.0888

Thaipadungpanit, J., Wuthiekanun, V., Chierakul, W., Smythe, L. D., Petkanchanapong, W., Limpaiboon, R., Apiwatanaporn, A., Slack, A. T., Suputtamongkol, Y., White, N. J., Feil, E. J., Day, N. P. J., & Peacock, S. J. (2007). A Dominant Clone of Leptospira interrogans Associated with an Outbreak of Human Leptospirosis in Thailand. *PLoS Neglected Tropical Diseases*, *1*(1), e56. https://doi.org/10.1371/journal.pntd.0000056

Timoney, J. F., Kalimuthusamy, N., Velineni, S., Donahue, J. M., Artiushin, S. C., & Fettinger, M. (2011). A unique genotype of *Leptospira interrogans* serovar Pomona type kennewicki is associated with equine abortion. *Veterinary Microbiology*, *150*(3–4), 349–353. https://doi.org/10.1016/j.vetmic.2011.02.049

Tunbridge, A., Dockrell, D., Channer, K., & McKendrick, M. (2002). A breathless triathlete. *The Lancet*, *359*(9301), 130. https://doi.org/10.1016/S0140-6736(02)07372-5

Viana, M., Cleaveland, S., Matthiopoulos, J., Halliday, J., Packer, C., Craft, M. E., Hampson, K., Czupryna, A., Dobson, A. P., Dubovi, E. J., Ernest, E., Fyumagwa, R., Hoare, R., Hopcraft, J. G. C., Horton, D. L., Kaare, M. T., Kanellos, T., Lankester, F., Mentzel, C., … Lembo, T. (2015). Dynamics of a morbillivirus at the domestic–wildlife interface: Canine distemper virus in domestic dogs and lions. *Proceedings of the National Academy of Sciences*, *112*(5), 1464–1469. https://doi.org/10.1073/pnas.1411623112

Viana, M., Mancy, R., Biek, R., Cleaveland, S., Cross, P. C., Lloyd-smith, J. O., & Haydon, D. T. (2014). Assembling evidence for identifying reservoirs of infection. *Trends in Ecology & Evolution*, *29*(5), 270–279. https://doi.org/10.1016/j.tree.2014.03.002

Vos, M., & Didelot, X. (2009). A comparison of homologous recombination rates in bacteria and archaea. *The ISME Journal*, *3*(2), 199–208. https://doi.org/10.1038/ismej.2008.93

Wayne, R. K., George, S. B., Gilbert, D., Collins, P. W., Kovach, S. D., Girman, D., & Lehman, N. (1991). A Morphological and Genetic Study of the Island Fox, Urocyon littoralis. *Evolution*, *45*(8), 1849–1868.

Webb, G., Browne, C., Huo, X., Seydi, O., Seydi, M., & Magal, P. (2015). A Model of the 2014 Ebola Epidemic in West Africa with Contact Tracing. *PLoS Current Outbreaks*, *1*, 1–20. https://doi.org/10.1371/currents.outbreaks.846b2a31ef37018b7d1126a9c8adf22a.Revisions

Woolhouse, M. E. J. (2001). Population Biology of Multihost Pathogens. *Science*, *292*(5519), 1109–1112. https://doi.org/10.1126/science.1059026

Zuerner, R. L., & Alt, D. P. (2009). Variable nucleotide tandem-repeat analysis revealing a unique group of Leptospira interrogans serovar Pomona isolates associated with California sea lions. *Journal of Clinical Microbiology*, *47*(4), 1202–1205. https://doi.org/10.1128/JCM.01639-08

Zuerner, R. L., Cameron, C. E., Raverty, S., Robinson, J., Colegrove, K. M., Norman, S. A., Lambourn, D., Jeffries, S., Alt, D. P., & Gulland, F. (2009). Geographical dissemination of Leptospira interrogans serovar Pomona during seasonal migration of California sea lions. *Veterinary Microbiology*, *137*(1), 105–110. https://doi.org/10.1016/j.vetmic.2008.12.017

2 Assessing infection risk factors in wildlife populations: a case study of *Leptospira* in Channel Island foxes

2.1 Abstract

Quantifying risk factors for infection can be difficult in wildlife systems due to significant challenges with the data. Two challenges in particular, interval-censored data, which can result from infrequent sampling, and time-varying covariates, are often not addressed together when estimating infection risk. In this study, we estimate risk of *Leptospira* infection in Channel Island foxes using a novel approach to address these challenges. We impute time of infection using a titer kinetics model, which allows us to apply a Cox proportional hazards model with time-varying covariates to assess risk factors. We find that low fox abundance and high 24-month cumulative precipitation provide ideal conditions for the transmission of *Leptospira.* Our innovative approach, in terms of intensive and sustained data collection and new methodologies for analysis, lays a foundation for future studies to investigate transmission risk to inform prevention and control strategies in wildlife populations.

2.2 Introduction

Transmission is the crux of disease ecology. The field of disease ecology relies on understanding pathogen transmission because it is the pivotal process that determines the population-wide effects of any pathogen (Heisey et al., 2006; Weitz et al., 2020). For instance, understanding the optimal conditions for host-to-host transmission guides the design of prevention and control measures. This was recently demonstrated in the assessment of

transmission risk of SARS-CoV-2, which prompted public health agencies across the globe to instate various public health guidelines, including social distancing, mask wearing, and vaccination (CDC, 2021; ECDC, 2021; WHO, 2020). There are countless earlier examples in human, wildlife, and livestock populations. For instance, transmission studies coupled with novel statistical methods have enabled researchers to investigate the factors that drive transmission of *Mycobacterium bovis* in cattle (Crispell et al., 2019), which guided culling and vaccination practices in Great Britain. Furthermore, studies of suburban raccoons have shown their role in the maintenance and transmission of rabies in the eastern United States and the potential issues with raccoon vaccination strategies (Reynolds et al., 2015).

As crucial as understanding transmission is, it is notoriously difficult to study because it is typically not directly observable and is influenced by many factors. Intrinsic factors such as host immunology, social behavior, population density, spatial distribution, sex, age, or immune status can affect the probability of an individual becoming infected (Anderson, 1982; Anderson & May, 1982, 1992; Kim et al., 2020; Mannelli et al., 2012; Marina et al., 2005; McCallum et al., 2009; Pourbohloul et al., 2009). Work by Viboud *et al*. (2004) explored the risk factors for influenza transmission within households and the effects of childhood vaccination against the virus. They found that preschool-age contacts were driving transmission. Their analyses show that if children were vaccinated, it would reduce secondary cases in households by as much as 41%. Their study highlights the importance of identifying the primary factors of infection risk to reduce subsequent cases. Extrinsic factors such as resource availability and environmental conditions can also alter the individual-level risk of transmission as well through impacts on host immunity and/or pathogen survival in the environment (Koelle & Pascual, 2004; Millán et

al., 2018). For instance, a study performed in China showed that Seoul hantavirus infections in Norway rats were linked to the extrinsic factor relative humidity (Li et al., 2019).

However, assessing individual-level risk is less common in studies of wildlife disease due to challenges in collecting the necessary data, and challenges in the ensuing analyses. Wildlife systems can have many extrinsic factors which affect transmission risk that are not easily measured, such as community assemblage, inter- and intra-species interactions, population sizes, and environmental, climatic, and oceanographic forces. Additionally, within wildlife systems, observation and sampling can be time- or resource-restricted and are frequently dictated by the biology of the animals of interest (Childs, 2007). These sampling schemes can create irregular and infrequent time intervals between samples, making the already difficult-to-observe process of transmission even harder to characterize. Furthermore, risk factors often vary meaningfully through time on timescales relevant to infection risk. Even in the rare instances when risk factors have been evaluated for wildlife infections, these issues arising from the sparseness, irregularity and non-stationarity of wildlife data are not typically addressed.

2.2.1 Leptospira interrogans

Pathogenic species within the genus of spirochete bacteria *Leptospira* cause the disease leptospirosis. Clinical signs can range from asymptomatic to fatal and infections can range from acute to chronic. Leptospires circulate in the blood before infecting the renal tubules, after which they are shed through urine (Levett, 2001). The pathogen transmits directly or through contact with contaminated water or soil sources, where the organism can survive and remain infectious for weeks to months (Faine et al., 1999). *Leptospira* is classified into serovars

determined by antibody responses (Adler & de la Pena Montesuma, 2010). Critically, *Leptospira* serovars are not host-specific. *L. interrogans* serovar Pomona is a multi-host pathogen with a complex ecology affected by host (intrinsic) and biotic and abiotic environmental (extrinsic) factors. This serovar has been associated with infections in foxes, skunks, and California sea lions (CSL) in the coastal California ecosystem (Dierauf et al., 1985; Gerber et al., 1993; Gulland et al., 1996; Lloyd-Smith et al., 2007; Tabel & Karstad, 1967).

Leptospira ecology has been previously shown to be associated with certain intrinsic and extrinsic conditions. Human *Leptospira* infections often rise after heavy rainfall, but it is still unknown how long *Leptospira* persist in different water and soil types (Bierque et al., 2020). A study done on Norwegian rats in an inner-city neighborhood in Vancouver, Canada suggested that the prevalence of *Leptospira* infections was associated with intrinsic factors such as the social structure and interactions between rats and indicated some level of positive density dependence exists in the transmission of this pathogen (Himsworth et al., 2013). Another study performed in Chicago neighborhoods showed that *Leptospira* infections in rats were linked to extrinsic factors such as standing water, a consequence of precipitation (Murray et al., 2020). To our knowledge no study has attempted to assess the risk factors for *Leptospira* transmission in (non-urban rat) wildlife, so there is great need and opportunity to explore the extent of the intrinsic and extrinsic factors that affect *Leptospira* transmission.

2.2.2 Channel Island foxes (Urocyon littoralis)

The California Channel Islands are an eight-island archipelago located off the coast of southern California between Point Conception and San Diego (Figure 2.1). Five of the islands

(San Miguel, Santa Rosa, Santa Cruz, Anacapa, and Santa Barbara) comprise Channel Islands National Park and are managed by the National Park Service. Only four terrestrial mammal species, namely island foxes (*Urocyon littoralis*), island deer mice (*Peromyscus maniculatus*), western harvest mice (*Reithrodontomys megalotis*), and island spotted skunks (*Spilogale gracilis amphiala*), are native to the islands, and no island hosts all four species. The focal island of this study, Santa Rosa Island, hosts the endemic Channel Island fox, the island spotted skunk, and the island deer mouse (Wayne et al., 1991).

Figure 2.1: Map of the Channel Islands off the coast of Southern California. (California Department of Forestry and Fire Protection, 2021; ESRI Inc., 2017; GEBCO Compilation Group, 2020; *NLCD 2016 Land Cover Conterminous United States*, 2019; NPS Land Resources Division, 2016)

The endemic Channel Island fox inhabits six of the islands, each forming their own subspecies (Wayne et al., 1991). Island foxes are the smallest North American canid, weighing about 1.9 kilograms (Coonan et al., 2010). Their omnivorous diet consists of small animals, plants, and insects (Roemer et al., 2001). After forming mated pairs, the foxes annually birth

litters of pups in April and May (Roemer et al., 2001). Fecundity and survival exhibit negative density dependence (Bakker et al., 2009; Coonan et al., 2010). Adult island foxes lack natural predators, so survival is generally high. However, island fox abundance is naturally small and prone to fluctuations due to environmental variation. A population viability analysis performed by Bakker et al. (2009) suggested that the fox populations were at low extinction risk except in the event of predation (by introduced predators) or disease.

Although the island fox populations have historically been small and fluctuating, anthropogenic factors initiated a cascade of effects which caused very severe population declines caused by golden eagle predation in the 1990s on San Miguel, Santa Rosa, and Santa Cruz islands (Coonan et al., 2005). In particular, only fourteen foxes (*Urocyon littoralis santarosae*) were left on Santa Rosa Island when the National Park Service took the entire population into captivity in 2000, and thus, the population was temporarily extinct in the wild (Coonan et al., 2010). A multi-year reintroduction program was initiated in 2003 to reestablish foxes on the island. By the end of 2008, 85 captive foxes had been released to the wild, and the reintroduced population was recovering, so the captive breeding program was concluded.

In fall 2010, two juvenile male foxes were found dead on Santa Rosa Island with evidence of *Leptospira* infection. At the time, there was no evidence that this pathogen had circulated on the island before, but subsequent serological assays of banked data suggested that foxes on SRI had exposure to *Leptospira interrogans* serovar Pomona prior to captivity (Lloyd-Smith, 2021). Retrospective analysis of samples collected from reintroduced foxes showed that by 2006 (during reintroduction) island-wide adult fox seroprevalence was already high (Figure 2.2). Throughout the post-reintroduction period, adult seroprevalence remained

above 55% with a subtle decline in more recent years (2013-2019). Pup seroprevalence mirrored the dynamics of adult seroprevalence during the reintroduction period but settled into low (less than 30%) seroprevalence during the post-reintroduction period. Broadly, it appears there was an initial wave of infection during the reintroduction period, and the system settled into an endemic state thereafter, with infection levels fluctuating around a steady state. Intriguingly, spotted skunks on the island showed genetic evidence of infection by the same strain of *L. interrogans* serovar Pomona (See Chapter 1).

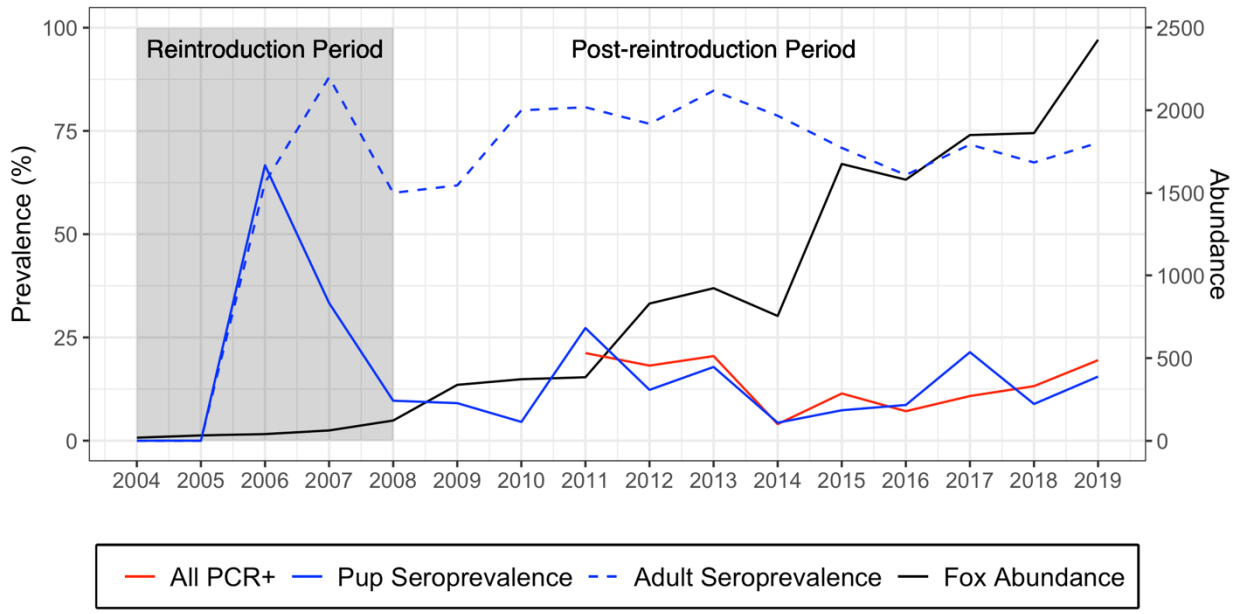

Figure 2.2: Time series of disease metrics during from 2004 to 2019. The reintroduction period occurred between 2004 and 2008 (grey shaded region). All-age prevalence of infection (indicated by PCR positivity; red), pup seroprevalence (blue solid), adult seroprevalence (blue dashed), and fox abundance (black) vary throughout the monitoring period.

2.2.3 Survival Analysis

Survival analysis is a class of statistical techniques which examine a time-to-event outcome variable (e.g., time to death, time to infection), and is conventionally used to quantify the rate of dying at a particular point in time. In our epidemiological context, we apply survival

analysis to estimate the force of infection, or infection hazard, which is fundamental to modeling and predicting transmission dynamics (Heisey et al., 2006). The hazard function, $h(t)$, is the instantaneous rate of infection and is commonly modeled using survival analysis, which evaluates the dependency of time-to-event data (e.g., time to infection) on explanatory variables. Survival analysis involves two primary quantities: the hazard function and the survival function. Assuming that the time of infection, *T*, is a continuous random variable with a probability density function *f(t)*, the hazard function is composed of the conditional probability that the time of infection will occur in the time interval $[t, t + \Delta t)$, given that infection has not occurred previously, relative to the interval width.

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t \mid T \geq t)}{\Delta t}$$

The survival function, $S(t)$, gives the probability that infection has not occurred up to time *t*, and takes the form of the complement of the cumulative distribution function:

$$S(t) = P(T \geq t), \qquad 0 < t < \infty$$

A key attribute of survival analysis is its ability to deal with censored data. Time-to-event data can be censored in four ways (uncensored, right-, left-, and interval-censored), and multiple types of censoring can occur within a single dataset. When the time of the event (e.g., infection) is known exactly, the data is uncensored. However, this type of data only occurs with continuous monitoring and is very rare in wildlife studies. If the event has not occurred by the

end of the study but may occur at a later date, the individual is right-censored (Figure 2.3A). Left-censoring is much less common than right-censoring and is defined by the occurrence of the event prior to the start of the study (Sun, 2006). Finally, interval censoring describes the situation when subjects are periodically observed, as in most wildlife trapping and sampling schemes, so the time of the event falls within an interval of time, rather than being exactly measured (Figure 2.3B, C).

There can be many factors which affect the rate of an event occurring at a particular point in time, and a Cox proportional hazards analysis examines the effect of such factors on the hazard rate (D.R. Cox, 1972). A Cox proportional hazards model is expressed by the hazard function $h(t)$, which takes the form:

$$h(t) = h_0(t) \times exp(b_1x_1 + b_2x_2 + \ldots + b_px_p)$$

where the hazard function is determined by a set of covariates $(x_1, x_2, \ldots, x_p)$. The coefficients $(b_1, b_2, \ldots, b_p)$ measure the impact of the covariate on the hazard function, and the baseline hazard, $h_0(t)$, corresponds to the hazard when all the covariates are equal to 0. The hazard ratio (HR), $exp(b_i)$, describes the relative effect of the i^{th} covariate on the hazard rate. If b_i is equal to 0 or the hazard ratio is equal to one, then there is no effect on the hazard rate. If the hazard ratio is greater than one, then the covariate is positively associated with the event occurrence and, thus, will lead to a shorter time to event.

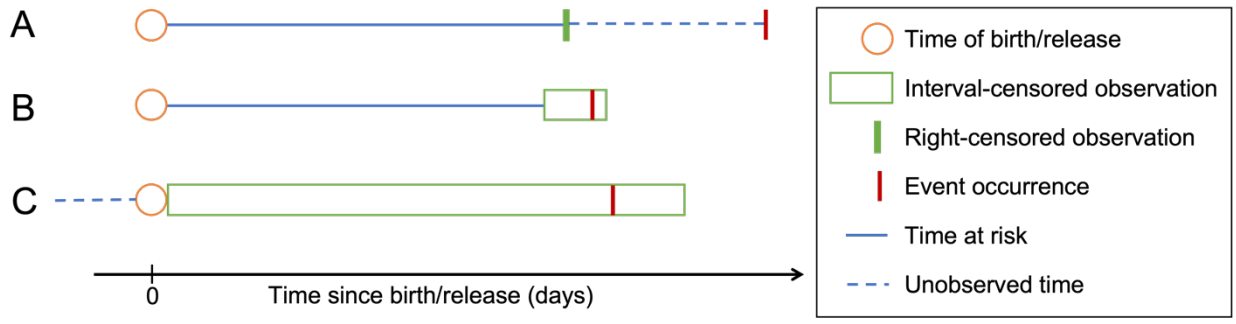

Figure 2.3: Schematic of censored data types in the context of the fox system. Solid blue lines denote survival time within the observation period, whereas the dotted blue lines indicated unobserved time. An individual enters the study (orange circles) when they are born in the wild (A, B) or are released from captivity (C) and are at risk for infection on the island. Some individuals (B, C) experience the event (red line) during the study time but are interval censored (green box). Other individuals are right censored (A) and do not experience the event during the time of the study but may or may not experience the event thereafter. The bottom panel also illustrates a long interval, where an individual was negative when released from captivity but was positive upon its first capture.

Previous studies have successfully applied survival analysis, in particular Cox proportional hazards regression (CPH), to understand factors affecting infection risk (Bui et al., 2018; Hagan et al., 2004; Satayathum et al., 2006). Human epidemiology studies have frequently applied CPH to investigate intrinsic risk factors of infection, but CPH has been less frequently applied to evaluate extrinsic risk factors, which often vary through time. Similar studies in wildlife systems using CPH are limited, but Almberg et al. (2012) used CPH regression to model the pack-level risk of mange infection in wolves. They found that areas of highest resource quality, which are also the most likely to host a high density of wolves, were associated with the highest infection risk to wolves. However, this study did not address the added complications of interval-censored data or time-varying covariates, which are generally inherent in wildlife data and sampling schemes.

In wildlife systems, subjects are sampled in infrequent trapping or other immobilization events. This inherently leads to interval censoring because the time of infection falls between

the last negative test and first positive test. Because the fieldwork required to obtain biometrics on infection status can be heavily seasonal due to logistical or biological constraints, the intervals between sequential samples from individual animals can be irregular or extend over long time periods (e.g., up to a year or more). Studies that explore the impacts of interval censoring show that long intervals can introduce significant bias (Vandormael et al., 2018). Therefore, it is common to impute a time of infection to make the data right-censored only, which allows the use of more standard survival analysis techniques.

Data imputation provides a way to cope with the technical challenge of interval-censoring, giving rise to augmented datasets with (estimated) precise event times, which enable the use of standard statistical techniques. However, Vandormael et al. (2018, 2020) have shown that imputing the infection time with the midpoint or endpoint of a censored interval can systematically bias estimates of the hazard function. Another common method is to impute the time of infection by random sampling from a uniform distribution. Although this is a better approach because it avoids systematic bias, it assumes that the hazard is constant across the interval and any potentially informative information about the timing of infection is discarded (Vandormael et al., 2018).

Recent methodological advances have harnessed the power of dependably varying biomarkers to infer the time of infection (Borremans et al., 2016, In preparation; Pepin et al., 2019; Wilber et al., 2020). Antibody responses to a pathogen can occur in a predictable way and can be described using a titer kinetics model, which estimates the peak antibody titer and antibody decay rate. Using longitudinal serological data, the antibody kinetics can be used to back-infer the time of infection from an individual's antibody titer. In this way, longitudinal

serological data can greatly improve estimation of incident infections (Borremans et al., 2016). Seminal work in this field by Borremans et al. (2016) used lab-infected mice to characterize an antibody decay function and demonstrated model improvement with the incorporation of additional intrinsic factors such as differing types of antibodies or individual age. However, this study utilized data from frequently-sampled individuals, which is generally not possible in field systems (Borremans et al., 2015). In recent developments, Wilber et al. (2020) used quantitative serology in an extended survival analysis when inferring seasonal infection risk for influenza A virus in swine. Their model accounted for interval censoring and explored intrinsic factors, host demography and antibody boosting caused by re-exposure to the virus, but their inquiry did not include extrinsic factors.

Inclusion of extrinsic factors in survival analysis can be challenging because extrinsic factors of interest frequently vary over time (e.g., precipitation). Intrinsic factors such as host population density or demographic rates can be time-varying as well. Time-varying factors are covariates whose effect on infection risk is constant but whose values change over the time an individual is at risk. For example, if rainfall drives the risk of infection but varies through time, then changes in rainfall must be considered over the period that an individual is at risk of infection. The inclusion of time-varying covariates in infection risk analyses requires special formulation of the time-at-risk and requires right-censored-only data. Thus, few studies address both interval-censored data and time-varying covariates, even in analyses of human data, and no studies have done this in wildlife. In this study, we aim to identify key factors that affect incident infections of *Leptospira* in Channel Island foxes while addressing key data issues that arise in wildlife systems, namely large censoring intervals and time-varying covariates. The

insights gained through this study will contribute to our understanding of the basic ecology of the system and can be used to develop management strategies and project possible changes under future climate scenarios. Furthermore, the methods developed in this study are generalizable to other wildlife systems and provide a new set of tools that will enable disease ecologists and epidemiologists to better assess infection hazard in their study systems.

2.3 Methods

2.3.1 Study period and Study cohort

From 2000 to 2009, the National Park Service (NPS) monitored the health of the island foxes brought into the captive breeding program and their captive-born offspring. All captive individuals had serum samples taken 1-4 months prior to reintroduction into the wild. During reintroduction (2003 - 2009) the NPS continued to monitor the wild fox population and conduct health evaluations on individuals captured through target trapping. Sampling coverage was low during the first couple of years for logistical reasons, but annual evaluations began in 2006. From 2009 to 2019, following the conclusion of the captive breeding program, serum and urine were collected via target trapping and a large-scale structured sampling program with 18 trapping grids run annually from reintroduced and wild-born foxes.

The cohort analyzed in this study consisted of 1226 foxes. To be included in this dataset, individuals were required to have a known release date from captivity (with a negative microagglutination test (MAT) result in their last captive test) or a known year of birth in the wild. All individuals that only ever tested seronegative during the study period (2004-2019) were included in the dataset. We defined infection of a fox to be when a fox born in the wild or

a known seronegative fox tested MAT positive for *L. interrogans* serovar Pomona or serovar Autumnalis. Serovar Autumnalis has been shown to be cross-reactive with serovar Pomona and generally has higher titer magnitudes in foxes than serovar Pomona, despite Pomona being the infecting serovar (Mummah et al., unpublished). No individuals in this study are classified as left-censored (i.e. no individual experienced the event prior to the study start time).

2.3.2 Choice of timescale

For survival analysis, it is crucial to choose an appropriate timescale (Fieberg & DelGiudice, 2009). The timescale is defined as the unit of time over which you are estimating risk (e.g. days, weeks, years) from the time where risk accrual begins (e.g. at birth). The choice of timescale governs the interpretation of survival times and allows a clear understanding of how age and temporal factors are accounted for within the models. In our study, we define our timescale as "days since birth/release". Foxes kept in captivity begin their time-at-risk when they were released to island (Figure 2.3C). However, the majority of the foxes in our dataset were wild-born post-reintroduction, and, therefore, begin their time-at-risk at birth, assumed to be the midpoint (April 1) of the island fox birthing season (Coonan et al., 2010). These choices affect our ability to estimate age effects in the models. Because much of our dataset are foxes which are wild-born and whose time-at-risk begin at birth, the time-at-risk in the wild and the individual's age are conflated and cannot be separated. This kind of direct correlation with time-at-risk is intractable in CPH, and, thus, we cannot estimate the effect of age on infection risk with our chosen timescale.

2.3.3 Data Imputation

Rather than using a standard technique to impute the time of infection for each individual in an uninformed manner, we generated biologically-informed time of infection estimates using the titer kinetics model described in Borremans et al. (Borremans et al., In preparation). This model uses longitudinal and quantitative serology data to estimate the peak antibody titer and antibody titer decay rate for an individual within the bounds of its infection interval (defined by an individual's last negative test and first positive test; Figure 2.3 - green boxes). Given one or more antibody titer values, the model can estimate a posterior probability on the time of infection for every day within the bounded interval. The level of precision varies by individual with some posteriors being more informative than others (Figure S2.1). Using these probabilities as weights, we construct an ensemble of augmented datasets by bootstrap sampling the times of infection for all foxes in an informed way; for each dataset, we assume that the event date is known exactly, which enables the incorporation of time-varying covariates more easily.

2.3.4 Counting process formulation

To incorporate time-varying covariates into a proportional hazards model, the data must be transformed into a counting process formulation (Fieberg & DelGiudice, 2009). An individual's time-at-risk (Figure 2.3; solid blue lines) is subdivided into smaller intervals over which the time-varying covariates can be assigned. This formulation assures that survival time is accurately accrued and that the covariates can be updated through time, with appropriate

temporal resolution (e.g. monthly or annually). In this study, we divided every individual's time at-risk into monthly intervals (Table 2.1).

Table 2.1: An example of a counting process formulation for a single individual's risk time.

Full time at-risk (days)	Full time at-risk (by date)	Subdivided time at-risk (days)	Subdivided time at-risk (by date)
[0, 100)	[04-01-2005, 7-10-2005)	[0, 30)	[04-01-2005, 05-01-2005)
		[30, 61)	[05-01-2005, 06-01-2005)
		[61, 91)	[06-01-2005, 07-01-2005)
		[91, 100)	[07-01-2005, 07-10-2005)

2.3.5 Candidate risk factors for infection with Leptospira

Based on our knowledge of the host-pathogen system and the ecology of Santa Rosa Island, we investigated potential risk factors for individual foxes to become infected with *Leptospira.* We formed three groups of covariates: individual, biotic environmental, and abiotic environmental (Table 2.2). Sex, the only individual-level variation we accounted for, was recorded upon birth within captivity or first capture of the fox on the island.

Table 2.2: Potential factors affecting *Leptospira* infection risk in 1226 foxes between 2004 and 2019. Descriptive statistics are given for the unscaled covariates. The covariates were normalized for the analysis.

Covariate description	Mean	SD
Individual level		
Sex (Male)	n=647 (52.8%)	
Abiotic environmental		
1-month cumulative precipitation (precip1)	0.90	1.46
12-month cumulative precipitation (precip12)	10.70	4.53
24-month cumulative precipitation (precip24)	20.64	5.67
Monthly average temperature	56.68	4.78
Monthly average relative humidity	73.11	11.95
Biotic environmental		
Fox abundance	812.56	763.40
Skunk abundance	16.69	14.68
Pup seroprevalence	0.15	0.16

The biotic environmental category includes factors that would conventionally be considered intrinsic to the host-pathogen system (fox abundance, skunk abundance, pup seroprevalence). We also included skunk abundance in this category because island spotted skunks have been shown to carry the same strain of *Leptospira* as island foxes (Chapter 1). Fox abundance was estimated annually by the National Park Service using live trapping and VHF radio telemetry data from collared individuals. Between 2004 and 2009, fox captures were primarily obtained through target trapping. In 2009, NPS biologists switched to a grid trapping scheme. Throughout this entire period, the skunk population was monitored via by-catch in fox traps. Skunk abundance was estimated using the number of skunk captures per trap-night. Estimation of skunk abundance is fundamentally uncertain because the trapping strategy and effort varied across the study time, skunks were not always individually marked in the early years of the study, and traps were frequently saturated by foxes, particularly in the later years of our study.

The abiotic environmental category contains extrinsic factors precipitation, temperature, and relative humidity. Daily measurements for these variables were obtained for Santa Rosa Island from 2003-2020 from the Western Regional Climate Center (WRCC) in Reno, Nevada. Monthly averages were calculated for temperature and relative humidity. For precipitation, we created three covariates that captured cumulative total precipitation over the past 1, 12, and 24 months. The *precip1* variable was intended to capture the immediate effect of rainfall, which could affect *Leptospira* transmission via puddles, wet vegetation, or changes in animal behavior. The *precip12* and *precip24* capture the past one and two rainy seasons in the highly seasonal Mediterranean climate on Santa Rosa Island, which affect the quantity of

surface water in streams and pools and have bottom-up effects across the island ecosystem (Power & Rudolph, 2018). These longer-term variables would also capture the impact of extended drought periods.

Because we know that a primary driver of infection risk during an outbreak is the prevalence of infectious individuals in the island fox population, we also included yearly pup seroprevalence as an index of the island-wide force of infection each year (Anderson & May, 1992). Although PCR positivity in urine samples would be a more direct measure of active shedding island-wide, we lack PCR data in the early phase of the outbreak and reintroduction (2004-2010). However, between 2011 and 2019, pup seroprevalence and PCR positivity are strongly correlated (Figure 2.2), so we used pup seroprevalence as a proxy for active infections on the island. To maintain our focus on non-disease factors while controlling for the influence of infection prevalence, in our final analyses we first determined the best model that excludes pup seroprevalence. We then tested whether the model's conclusions were robust to the addition of pup seroprevalence as a proxy for outbreak context.

2.3.6 Imputation and Hazard modeling

We performed 10,000 bootstrap runs in which time of infection for each individual was imputed using weights proportional to the posterior probabilities generated by the titer kinetic model. The full dataset was subdivided into a counting process formulation by month, and time-varying covariates were assigned to each interval. All covariates were standardized to have a mean of 0 and a standard deviation of 1. Then, we fit a series of univariate Cox

proportional hazards models to assess the effect of single covariates on the hazard rate using the *survival* package in R version 4.0.5 (R Core Team, 2021; Therneau, 2021).

We formulated a multivariate model with factors from all three categories. Because factors within a category tended to be correlated and we wanted to avoid multicollinearity (Figure S2.2), we chose a single factor from each one. In addition to the lone individual factor (sex), the best-performing covariate from each environmental category was selected to form a multivariate model. We compared different multivariate models using Aikake information criterion (AIC) scores to balance parsimony with goodness of fit (Burnham & Anderson, 2002). We evaluated our final multivariate model with and without pup seroprevalence included, to evaluate the robustness of the core findings, and the potential benefit of including disease data in such a model when available.

2.4 Results

2.4.1 Univariate models

Clear patterns arose within the three categories of factors (individual, biotic environmental, and abiotic environmental) in the univariate analysis (Table 2.3; Figure 2.4). Twelve- and twenty-four-month cumulative precipitation appeared to increase the risk of infection by 6% and 12%, respectively, with every increase in one standard deviation in the covariates (4.53 and 5.67 inches, respectively). One-month precipitation, monthly average temperature, and monthly average relative humidity showed weak trends but were not significant predictors of risk. All biotic environmental variables had significant relationships with the risk of infection in univariate analyses. Contrary to the standard expectations of density

dependence in pathogen transmission, greater fox abundance was associated with a decrease in the risk of infection as the population increases. Skunk abundance had the opposite effect. As expected, pup seroprevalence, a proxy for force of infection, had the largest effect size of any covariate. For a 16% (1 standard deviation) increase in pup seroprevalence, the individual risk of infection increased by 28%.

Table 2.3: Estimated hazard ratio (HR) and center 95% from univariate Cox proportional hazards model with imputed time of infection estimates of potential factors affecting *Leptospira* infection risk in 1226 foxes between 2004 and 2019. The variation represented by the interval is generated from the 10,000 bootstrap runs.

	HR (95% Interval)
Individual level	
Sex (Male)	1.04 (1.00, 1.09)
Abiotic environmental	
1-month cumulative precipitation (precip1)	1.04 (0.94, 1.13)
12-month cumulative precipitation (precip12)	1.06 (1.01, 1.11)
24-month cumulative precipitation (precip24)	1.12 (1.07, 1.18)
Monthly average temperature	0.97 (0.89, 1.07)
Monthly average relative humidity	0.96 (0.90, 1.04)
Biotic environmental	
Fox abundance	0.93 (0.90, 0.96)
Skunk abundance	1.11 (1.06, 1.16)
Pup seroprevalence	1.28 (1.21, 1.36)

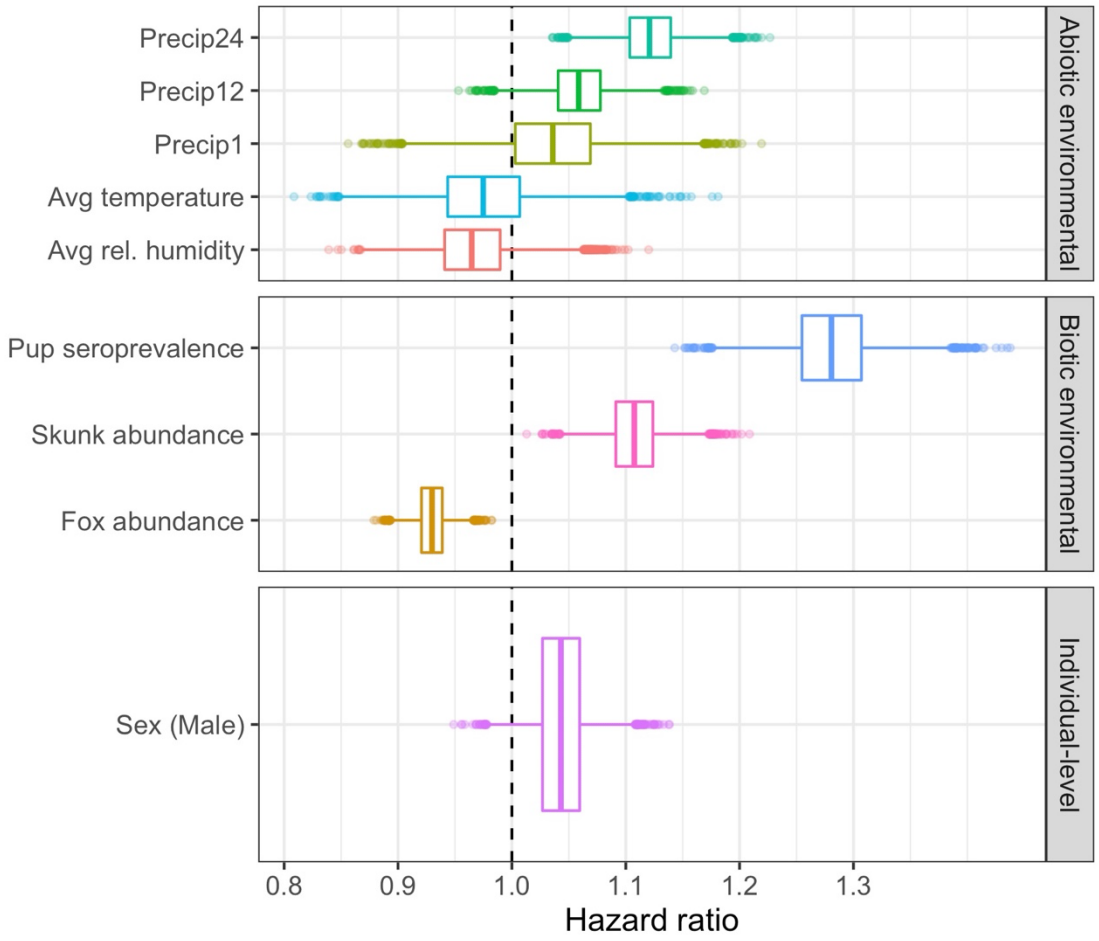

Figure 2.4: Distribution of hazard ratios estimated by univariate Cox proportional hazards model with imputed time of infection estimates. The variation represented by the boxplots is generated from the 10,000 bootstrap runs. The edges of the boxes denote the interquartile range and the middle line of the box represents the median value.

2.4.2 Multivariate model

To build our base multivariate model, we included the best-performing covariate from each category. In addition to the single individual factor (sex), we chose to include 24-month cumulative precipitation (precip24) as it had the greatest effect on infection risk in the univariate models in the abiotic environmental category. We chose to include fox abundance over skunk abundance for the biotic environmental category. These two covariates are strongly negatively correlated (r = -0.81, Figure S2.2), and because foxes were the focal host species in

this study and there were major sources of bias in the skunk abundance data, the fox abundance covariate was retained for analysis in the multivariate model. We additionally evaluated the model with and without the inclusion of pup seroprevalence as a proxy for active infection. Thus, our multivariate model included sex, fox abundance, and 24-month cumulative precipitation (precip24). We also tested the effect of including monthly average temperature or relative humidity as well, but neither covariate contributed significantly to the model fit, as assessed by minor changes to AIC scores (Table S2.1).

The effects observed in the univariate models were all preserved in the multivariate model, with greater effect sizes and reduced uncertainties (Figure 2.5A). When the other covariates in the model were held constant, 24-month cumulative precipitation exhibited a stronger effect of increasing infection risk (median HR = 1.15), and fox abundance exhibited a stronger protective effect (median HR = 0.90). When pup seroprevalence was included in the model, these effects were qualitatively robust but diminished in magnitude (Table 2.4; Figure 2.5B). Pup seroprevalence had a median hazard ratio of 1.21 after controlling for sex, precipitation, and fox abundance, which means that the risk increases by 21% with every one standard deviation increase in pup seroprevalence.

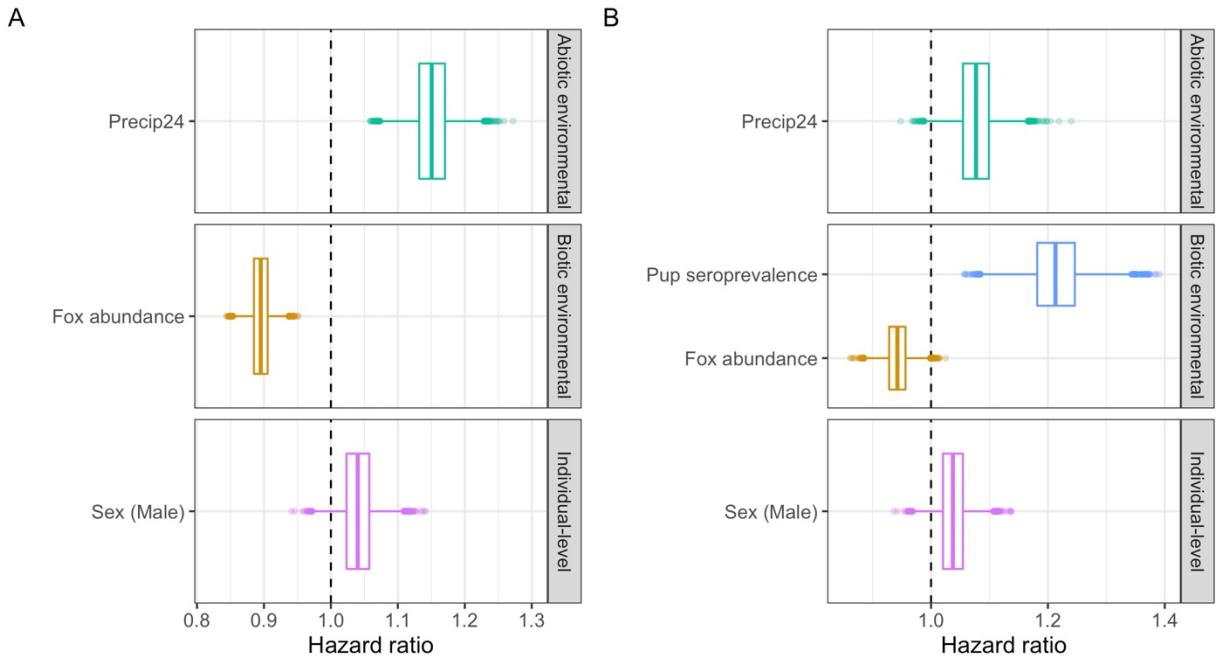

Figure 2.5: Distribution of hazard ratios estimated by multivariate Cox proportional hazards model with imputed time of infection estimates. Panel A shows the final multivariate model, which includes sex, yearly fox abundance, and 24-month cumulative precipitation. Panel B shows the same model with the inclusion of yearly pup seroprevalence.

Table 2.4: Estimates of median hazard ratios (HR) of multivariable Cox proportional hazards model with imputed times of infection.

	Without pup SP	With pup SP
	HR (95% Interval)	**HR (95% Interval)**
Individual level		
Sex (Male)	1.04 (0.99, 1.09)	1.04 (0.99, 1.09)
Abiotic environmental		
24-month cumulative precipitation (precip24)	1.15 (1.10, 1.21)	1.08 (1.01, 1.15)
Biotic environmental		
Fox abundance	0.90 (0.87, 0.93)	0.94 (0.90, 0.98)
Pup seroprevalence		1.21 (1.12, 1.31)

2.4.3 Time-to-infection curves

To put these model findings into concrete terms to assist population managers, we compared median time-to-infection (the time corresponding to a 0.5 probability of becoming infected) for males and females at varying levels of fox abundance and 24-month cumulative precipitation (Figure 2.6). The low, medium, and high levels of fox abundance and 24-month cumulative precipitation were defined by the 10th, 50th, and 90th percentiles of each covariate. Regardless of stratification level, most median time-to-infection estimates are greater than one year (Table 2.5). Females had 2- to 4-week longer times than those of males across all levels of abundance and rainfall. The shortest median times predicted by our best-fit model (males: 360 days; females: 364 days) occurred when foxes are at low abundance with high precipitation.

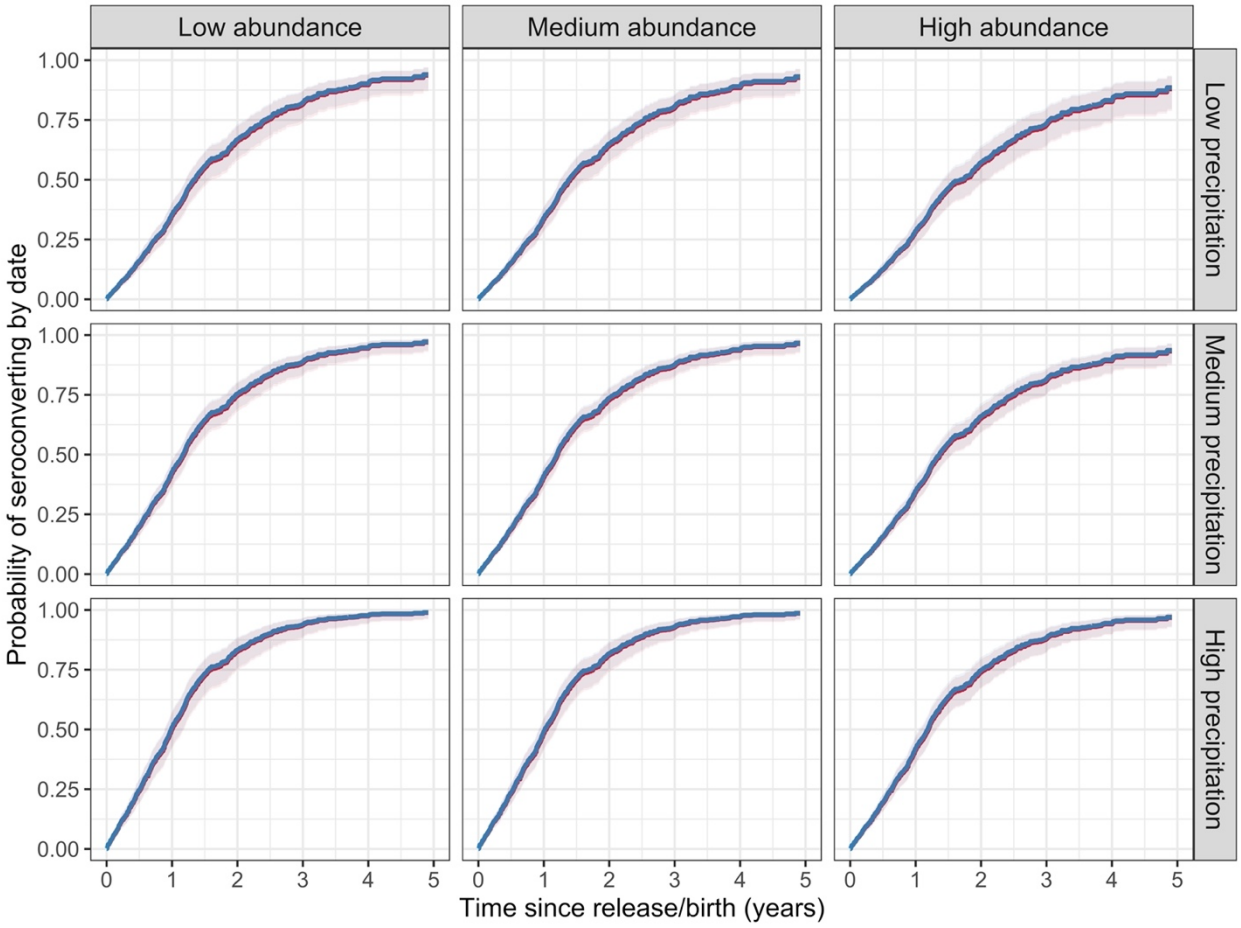

Figure 2.6: Time-to-infection curves for low, medium, and high levels of fox abundance and 24-month cumulative precipitation when stratified by sex. Low, medium, and high levels of abundance and precipitation are defined by the 10th, 50th, and 90th percentiles of each covariate (Figure S2.3). Within each subplot, the curves are stratified by sex. The male lines (blue) occlude the female lines (red), but in all plots, the female line is slightly lower.

Table 2.5: Median survival times (days) given sex, 24-month cumulative precipitation, and fox abundance. The median survival time is the time corresponding to a 50% probability of becoming infected.

	Low abundance	Medium abundance	High abundance	
Female	497	512	640	**Low precipitation**
Male	486	504	617	
Female	433	441	503	**Medium precipitation**
Male	426	437	497	
Female	364	377	437	**High precipitation**
Male	360	371	432	

We also compared the model's predictions for the probability of infection during the different periods of time-at-risk on the island under different conditions. After the first year at-risk, male foxes had, at most, a 51% chance of becoming infected (Table 2.6). If the first year of time-at-risk occurred during less ideal conditions for infection (i.e. high fox abundance and low cumulative precipitation), the probability of infection was as low as 28%. After 2 years at-risk, the highest probability that male foxes were infected was 83%, which occurs under low fox abundance and high precipitation. The probability of infection was greater than 56% under all abundance and precipitation conditions for both sexes after 2 years at-risk.

Table 2.6: Probability of becoming infected when stratified by sex, fox abundance, and 24-month precipitation after 1 and 2 years at risk.

		Low abundance	Medium abundance	High abundance	
1-year	**Female**	0.35	0.33	0.28	**Low precipitation**
2-year		0.66	0.64	0.56	
1-year	**Male**	0.35	0.34	0.29	
2-year		0.67	0.65	0.57	
1-year	**Female**	0.42	0.40	0.34	**Medium precipitation**
2-year		0.74	0.73	0.65	
1-year	**Male**	0.43	0.41	0.35	
2-year		0.75	0.74	0.66	
1-year	**Female**	0.50	0.48	0.41	**High precipitation**
2-year		0.82	0.81	0.74	
1-year	**Male**	0.51	0.49	0.42	
2-year		0.83	0.82	0.75	

2.5 Discussion

Transmission is the center of disease ecology and crucial to understand. However, this process is difficult to study as transmission is typically not directly observable and is influenced by many factors. Assessing risk in a wildlife species increases the difficulty and is rare due to challenges in collecting the necessary data, and in analyzing the typically sparse data that can

be obtained. In this study, we assessed the key factors that affect individual infection risk in a wildlife species, while introducing new approaches to handle the inherent data challenges of large censoring intervals and time-varying covariates. By incorporating quantitative serology-based estimates of individual foxes' times of infection, we were able to estimate the key factors that influenced *Leptospira* infection risk in Channel Island foxes.

In a comparison of cumulative rainfall over different periods, we found that 24-month cumulative rainfall had a greater effect on infection risk than 12-month cumulative rainfall, and one-month rainfall did not have a discernable influence. This indicates that multi-annual trends in precipitation and, notably, the previous two rainy seasons are more important in driving infection risks than current climate conditions. Thus, under a changing climate, the long-term effects on the island's hydrology will be more important in driving new infections than short-term fluctuations. We postulate that this effect may be governed by changes in the distribution of standing water on the island. The underlying mechanisms are unknown, but we hypothesize that effects act on a range of timescales. For example, a recent review of *Leptospira* in water and soil environments proposes that rainfall re-suspends leptospires in the soil, making them more accessible to new hosts and potentially increasing risk of exposure (Bierque et al., 2020). Given the timescales of the survival of *Leptospira* in water and soil environments (days to months), it does not seem likely that this is the dominant effect driving the 24-month pattern. On the contrary, persistent changes in surface water that could provide hospitable environments for leptospires that are shed in the months (or years) after the rainfall could enable further transmission on a longer timescale. Further research is needed on the persistence of *Leptospira* in the environment to resolve the mechanisms of the precipitation

effect. Additional information about the island hydrology and the existence and persistence of transient pools could also improve the evaluation of risk for cumulative rainfall. Some work has been done to investigate the spatial and temporal variation of pooled water on the island and its response to variable precipitation, but the resolution of the data is much coarser than what is required to test the importance of the hydrology at this scale (Power & Rudolph, 2018).

We also found that greater fox abundance was associated with reduced risk of *Leptospira* exposure. This finding differs significantly from the traditional assumption in disease ecology that a higher density of hosts leads to higher contact rates and, in turn, higher transmission. It is crucial to note that fox abundance increased steadily throughout the study period, so this effect could be confounded with other temporal trends. Of particular concern is confounding between the increasing fox abundance and the multi-annual dynamics of the *Leptospira* outbreak since the risk of infection is expected to decrease as *Leptospira* transitioned from epidemic to endemic circulation on the island (Figure 2.2). However, when pup seroprevalence was added as a proxy for overall prevalence of active infections on the island, the effect remained in the model, though with two-fold diminished strength. We believe a dominant contributor of this negative density dependence of risk may arise from the stabilization of island fox social structure as the reintroduced population became established on the island. As foxes were reintroduced to the island landscape (which was devoid of established fox territories since the wild population had been extinct for several years), biologists observed many long-range movements after release, which could have been movements in search of suitable habitat to establish a home range (Coonan et al., 2010). Telemetry data (unpublished) showed that movements were more frequent and longer-range in the years immediately after

reintroduction, then they gradually diminished in frequency and distance over the ensuing decade as home ranges were established. As the population grew, foxes were less likely to move outside their home ranges and disperse across the island, potentially due to stronger interference competition (Lloyd-Smith, 2021). This reduced overall mixing in the fox population, which could explain the decreased risk of new infections despite the higher population density.

The model predicts the shortest median times-to-infection at low fox abundance and high 24-month cumulative precipitation. These conditions were met during the early phase of fox reintroduction in 2005 and 2006. In the winter months of 2005, there were fewer than 40 foxes on the island, and the water table was at a peak with more than 30 inches of cumulative precipitation in the previous two years (more than 1.5 standard deviations above the average 24-month cumulative precipitation) (Figure S2.4). According to our model, these combined conditions were highly conducive to *Leptospira* transmission and may have driven the initial outbreak of *Leptospira* in the population. Precipitation patterns may also have driven the subtle decrease in adult seroprevalence seen in recent years (2013-2019; Figure 2.2), as recent drought conditions may have led to suboptimal circumstances for *Leptospira* transmission (Figure S2.4).

As expected, including a measure of the level of infection on the island through pup seroprevalence significantly improved model fit and moderated the effects of fox abundance and 24-month cumulative seroprevalence. It makes sense that that having more disease present on the island would increase the risk of infection and decrease the expected time to infection. Assuming there are unobserved factors that increase risk to individuals, they will give rise to elevated risk via their direct effect on the individual and via their population-scale effect

(i.e., because the risk to other individuals is high, population prevalence will be high, and, therefore, force of infection will also be high). In a sense, these factors will be double counted in a model that does not control for population prevalence or force of infection. By including pup seroprevalence in our model, we have controlled for both routes of influence: individual-level and population-level. However, in most systems and on other islands, this information would be unavailable at the beginning of an outbreak and would require intense monitoring to have accurate estimates throughout the study period. Thus, we excluded it in our primary multivariate model to assess infection risk when disease-related data were unavailable.

This work could be broadened to a spatially explicit estimation of risk on the island to better understand what local, sub-island conditions may have influenced risk of infection. Few environmental covariates are uniformly distributed across the island. For example, due to island topography, the presence and duration of seasonal pooled water can vary across the landscape, thus shaping the risk of infection (Power & Rudolph, 2018). Many of these spatially varying covariates (e.g., local fox density, local skunk density, local standing water, local infection prevalence) are challenging and time-consuming to collect and add an additional burden in the analysis. However, if these data were available, they would increase model precision and enable more targeted control.

We implemented an improved approach for imputing time of infection data by using a quantitative serology model. However, our knowledge of time of infection gained through our quantitative serology model is limited by the slow decay rate of antibodies against *L. interrogans* serovar Pomona and serovar Autumnalis and considerable individual variation in peak antibody titer magnitude (Borremans et al., In preparation). Some posterior distribution

estimates of the time of infection are nearly flat, so they are functionally as imprecise as randomly sampling from a uniform distribution across the infection interval (Figure S2.1). Our time of infection estimates included only one type of biomarker, the level of serum antibodies, as it was the only reliable biomarker available. Estimates in other systems could be improved by including multiple biomarkers (e.g. PCR status, serum chemistry) which vary across timescales and have different decay rates (Borremans et al., 2016). For example, for *Leptospira* infections in California sea lions, serum chemistry indicates active infection and varies on the scale of weeks (Prager et al., 2020). This multiple-biomarker approach was recently applied to model the antibody kinetics of IgM and IgG during early SARS-CoV-2 infection since these antibodies reach peak levels at different times during an infection (Borremans et al., 2020).

Understanding transmission and infection risk in a wildlife system is difficult due to the challenges in collecting the necessary data that is at an appropriate resolution and accounts for the multitude of possible contributing factors. Most wildlife systems have limited data at a fine-scale resolution reminiscent of human epidemiological studies, but even with infrequent sampling and the complication of time-varying factors, we gained insights into the individual risk factors for infection. Our innovative approach, in terms of intensive and sustained data collection and new methodologies for analysis, lays the groundwork for future wildlife studies to investigate transmission risk to inform prevention and control strategies.

2.6 Supplement

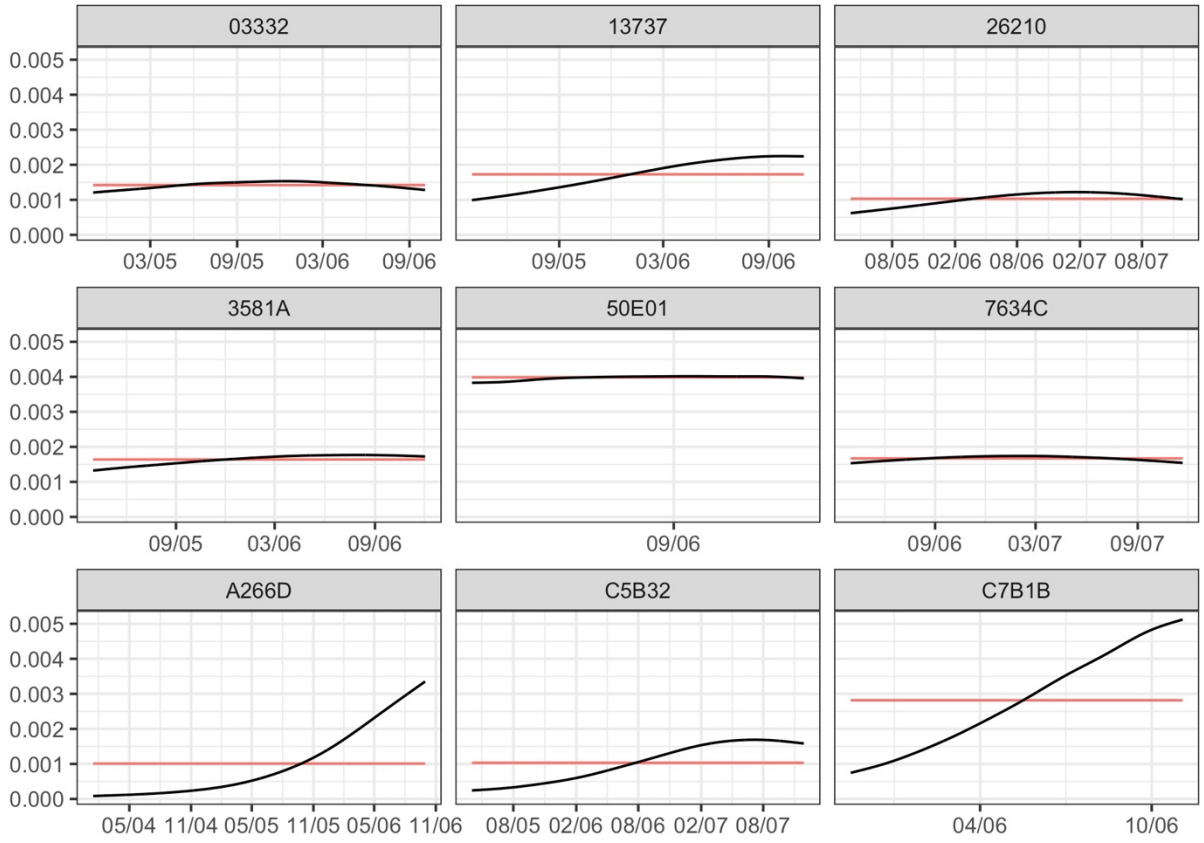

Figure S2.1: A selected set of time of infection posterior probabilities generated via the quantitative serology model. Each panel represents a single fox and are denoted by their ID numbers. The black lines show the posterior probabilities for time of infection between an individual's last negative test and first positive test. The red lines represent a uniform distribution for each fox.

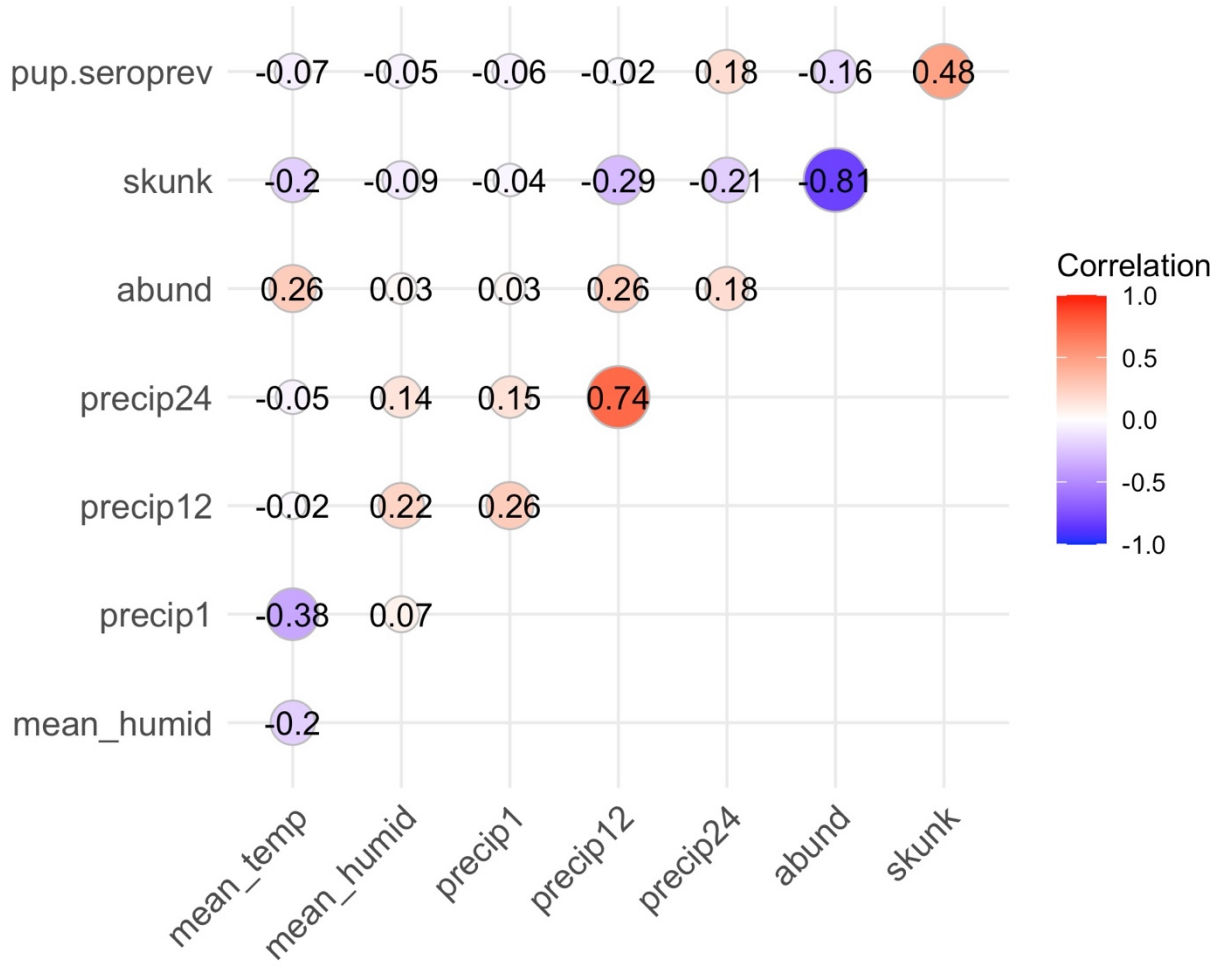

Figure S2.2: Correlation matrix among potential factors affecting risk of *Leptospira* infection.

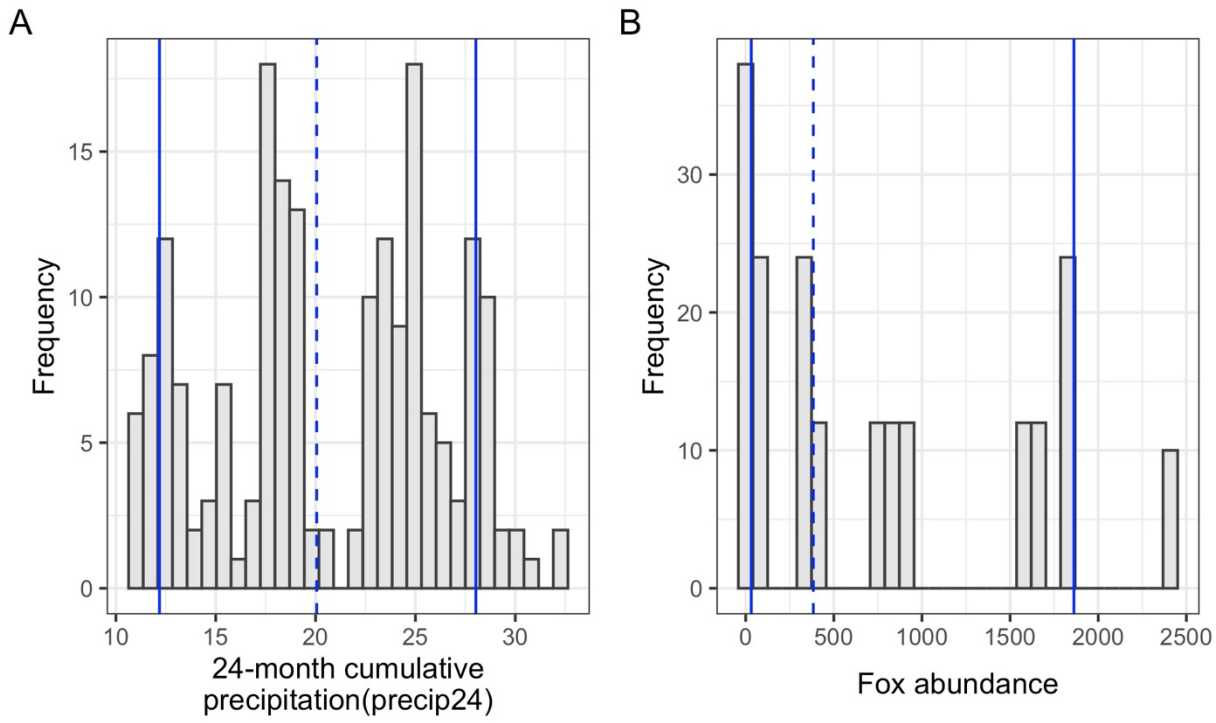

Figure S2.3: Distributions of 24-month cumulative precipitation and fox abundance. The 10th, 50th (dashed), and 90th percentiles define the low, medium, and high levels of each covariate for the time of infection comparisons.

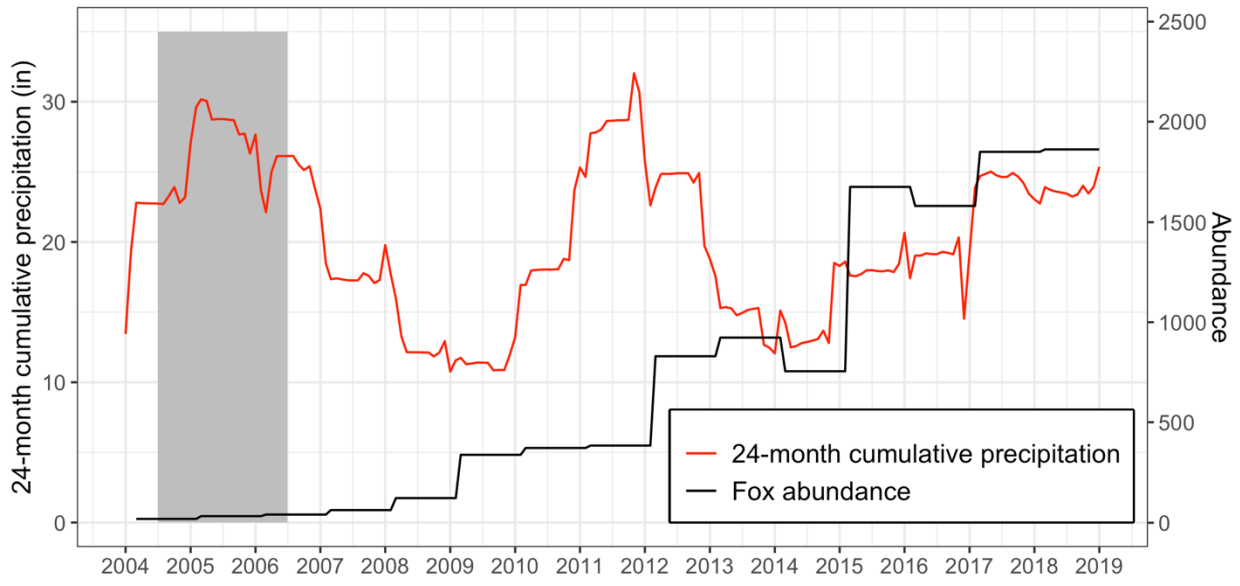

Figure S2.4: Fifteen-year time series of risk factors of *Leptospira*. Twenty-four-month cumulative precipitation is shown in red. The black line illustrates the annual fox population abundance through time. The grey box indicates the period during reintroduction where the risk factors were optimized for transmission of *Leptospira.*

Table S2.1: Median Akaike information criterion (AIC) for multivariate Cox proportional hazards models across 10,000 bootstrap runs. Lower AIC values indicate models that achieve superior balance of parsimony and goodness of fit, with AIC differences of 10 or more providing a strong basis for model selection.

Model	Median AIC
Sex	8199.456
Sex + Fox abundance	8199.456
Sex + Fox abundance + Precip24	8187.731
Sex + Fox abundance + Precip24 + Mean temperature	8185.594
Sex + Fox abundance + Precip24 + Mean rel. humidity	8185.902
Sex + Fox abundance + Precip24 + Pup seroprevalence	8177.134

2.7 References

Adler, B., & de la Pena Montesuma, A. (2010). Leptospira and leptospirosis. *Veterinary Microbiology*, *140*, 287–296. https://doi.org/10.1016/j.vetmic.2009.03.012

Almberg, E. S., Cross, P. C., Dobson, A. P., Smith, D. W., & Hudson, P. J. (2012). Parasite invasion following host reintroduction: A case study of Yellowstone's wolves. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1604), 2840–2851. https://doi.org/10.1098/rstb.2011.0369

Anderson, R. M. (1982). Transmission Dynamics and Control of Infectious Disease Agents. In R. M. Anderson & R. M. May (Eds.), *Population Biology of Infectious Diseases* (pp. 149–176). Springer. https://doi.org/10.1007/978-3-642-68635-1_9

Anderson, R. M., & May, R. M. (1982). Directly transmitted infectious diseases: Control by vaccination. *Science*, *215*(4536), 1053–1060. https://doi.org/10.1126/science.7063839

Anderson, R. M., & May, R. M. (1992). *Infectious diseases of humans: Dynamics and control*. Oxford university press.

Bakker, V. J., Doak, D. F., Roemer, G. W., Garcelon, D. K., Coonan, T. J., Morrison, S. A., Lynch, C., Ralls, K., & Shaw, R. (2009). Incorporating ecological drivers and uncertainty into a demographic population viability analysis for the island fox. *Ecological Monographs*, *79*(1), 77–108. https://doi.org/10.1890/07-0817.1

Bierque, E., Thibeaux, R., Girault, D., Soupé-Gilbert, M.-E., & Goarant, C. (2020). A systematic review of Leptospira in water and soil environments. *PLOS ONE*, *15*(1), e0227055. https://doi.org/10.1371/journal.pone.0227055

Borremans, B., Gamble, A., Prager, K., Helman, S. K., McClain, A. M., Cox, C., Savage, V., & Lloyd-Smith, J. O. (2020). Quantifying antibody kinetics and RNA detection during early-phase SARS-CoV-2 infection by time since symptom onset. *ELife*, *9*, e60122. https://doi.org/10.7554/eLife.60122

Borremans, B., Hens, N., Beutels, P., Leirs, H., & Reijniers, J. (2016). Estimating Time of Infection Using Prior Serological and Individual Information Can Greatly Improve Incidence Estimation of Human and Wildlife Infections. *PLOS Computational Biology*, *12*(5), e1004882. https://doi.org/10.1371/journal.pcbi.1004882

Borremans, B., Mummah, R. O., Guglielmino, A., Galloway, R. L., Hens, N., Prager, K. C., & Lloyd-Smith, J. O. (In preparation). *Estimating time of infection strictly from field data: Modeling antibody decay in Channel Island foxes (*Urocyon littoralis*) infected with* Leptospira interrogans.

Borremans, B., Vossen, R., Becker-Ziaja, B., Gryseels, S., Hughes, N., Van Gestel, M., Van Houtte, N., Günther, S., & Leirs, H. (2015). Shedding dynamics of Morogoro virus, an African arenavirus closely related to Lassa virus, in its natural reservoir host Mastomys natalensis. *Scientific Reports*, *5*(1), 10445. https://doi.org/10.1038/srep10445

Bui, U. T., Finlayson, K., & Edwards, H. (2018). Risk factors for infection in patients with chronic leg ulcers: A survival analysis. *International Journal of Clinical Practice*, *72*(12), e13263. https://doi.org/10.1111/ijcp.13263

Burnham, K. P., & Anderson, D. R. (2002). *Model Selection and Inference: A Practical Information-Theoretic Approach* (2nd ed.). Springer-Verlag. http://dx.doi.org/10.1007/b97636

California Department of Forestry and Fire Protection. (2021). *California County Boundaries*. U.S. Bureau of Reclamation, California Department of Conservation, California Department of Fish and Game, California Department of Forestry and Fire Protection, National Oceanic and Atmospheric Administration. https://gis.data.ca.gov/datasets/CALFIRE-Forestry::california-county-boundaries/about

CDC. (2021, July 26). *COVID-19 and Your Health*. Centers for Disease Control and Prevention. https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/prevention.html

Childs, J. E. (Ed.). (2007). *Wildlife and emerging zoonotic diseases: The biology, circumstances and consequences of cross-species transmission*. Springer.

Coonan, T. J., Schwemm, C. A., & Garcelon, D. K. (2010). *Decline and recovery of the Island Fox: A case study for population recovery*. Cambridge University Press.

Coonan, T. J., Schwemm, C. A., Roemer, G. W., Garcelon, D. K., & Munson, L. (2005). Decline of an Island Fox Subspecies to near Extinction. *The Southwestern Naturalist*, *50*(1), 32–41. JSTOR.

Crispell, J., Benton, C. H., Balaz, D., De Maio, N., Ahkmetova, A., Allen, A., Biek, R., Presho, E. L., Dale, J., Hewinson, G., Lycett, S. J., Nunez-Garcia, J., Skuce, R. A., Trewby, H., Wilson, D. J., Zadoks, R. N., Delahay, R. J., & Kao, R. R. (2019). Combining genomics and epidemiology to analyse bi-directional transmission of Mycobacterium bovis in a multi-host system. *ELife*, *8*, e45833. https://doi.org/10.7554/eLife.45833

Dierauf, L. A., Vandenbroek, D. J., Roletto, J., Koski, M., Amaya, L., & Gage, L. J. (1985). An epizootic of leptospirosis in California sea lions. *Journal of the American Veterinary Medical Association*, *187*(11), 1145–1148.

D.R. Cox. (1972). Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society: Series B*, *84*, 187–220.

ECDC. (2021, February 4). *Questions and answers on COVID-19: Prevention*. European Centre for Disease Prevention and Control. https://www.ecdc.europa.eu/en/covid-19/questions-answers/questions-answers-prevention

ESRI Inc. (2017). *ArcGIS Desktop* (10.6.0.8321) [Computer software]. ESRI Inc. https://desktop.arcgis.com/en/

Faine, S., Adler, B., Bolin, C., & Perolat, P. (1999). *Leptospira and leptospirosis* (2nd ed.). MediSci. https://research.monash.edu/en/publications/leptospira-and-leptospirosis

Fieberg, J., & DelGiudice, G. D. (2009). What time is it? Choice of time origin and scale in extended proportional hazards models. *Ecology*, *90*(6), 1687–1697. https://doi.org/10.1890/08-0724.1

GEBCO Compilation Group. (2020). *GEBCO 2020 Grid*. United Nations Educational, Scientific, and Cultural Organization (UNESCO) - International Hydrographic Organization (IHO) and Intergovernmental Oceanographic Commission (IOC). doi:10.5285/a29c5465-b138-234d-e053-6c86abc040b9

Gerber, J. A., Roletto, J., Morgan, L. E., Smith, D. M., & Gage, L. J. (1993). Findings in pinnipeds stranded along the central and northern California coast, 1984-1990. *Journal of Wildlife Diseases*, *29*(3), 423–433.

Gulland, F. M. D., Koski, M., Lowenstine, L. J., Colagross, A., Morgan, L., & Spraker, T. (1996). Leptospirosis in California sea lions (*Zalophus californianus*) stranded along the central California coast, 1981-1994. *Journal of Wildlife Diseases*, *32*(4), 572–580. https://doi.org/10.7589/0090-3558-32.4.572

Hagan, H., Thiede, H., & Des Jarlais, D. C. (2004). Hepatitis C Virus Infection Among Injection Drug Users: Survival Analysis of Time to Seroconversion. *Epidemiology*, *15*(5), 543–549. https://doi.org/10.1097/01.ede.0000135170.54913.9d

Heisey, D. M., Joly, D. O., & Messier, F. (2006). The fitting of general force-of-infection models to wildlife disease prevalence data. *Ecology*, *87*(9), 2356–2365. https://doi.org/10.1890/0012-9658(2006)87[2356:tfogfm]2.0.co;2

Himsworth, C. G., Bidulka, J., Parsons, K. L., Feng, A. Y. T., Tang, P., Jardine, C. M., Kerr, T., Mak, S., Robinson, J., & Patrick, D. M. (2013). Ecology of Leptospira interrogans in Norway Rats (Rattus norvegicus) in an Inner-City Neighborhood of Vancouver, Canada. *PLoS Neglected Tropical Diseases*, *7*(6), e2270. https://doi.org/10.1371/journal.pntd.0002270

Kim, S., Ko, Y., Kim, Y.-J., & Jung, E. (2020). The impact of social distancing and public behavior changes on COVID-19 transmission dynamics in the Republic of Korea. *PLOS ONE*, *15*(9), e0238684. https://doi.org/10.1371/journal.pone.0238684

Koelle, K., & Pascual, M. (2004). Disentangling Extrinsic from Intrinsic Factors in Disease Dynamics: A Nonlinear Time Series Approach with an Application to Cholera. *The American Naturalist*, *163*(6), 901–913. https://doi.org/10.1086/420798

Levett, P. N. (2001). Leptospirosis. *Clinical Microbiology*, *14*(2), 296–326. https://doi.org/10.1128/CMR.14.2.296

Li, Y., Cazelles, B., Yang, G., Laine, M., Huang, Z. X. Y., Cai, J., Tan, H., Stenseth, N. Chr., & Tian, H. (2019). Intrinsic and extrinsic drivers of transmission dynamics of hemorrhagic fever with renal syndrome caused by Seoul hantavirus. *PLOS Neglected Tropical Diseases*, *13*(9), e0007757. https://doi.org/10.1371/journal.pntd.0007757

Lloyd-Smith, J. O. (2021). *Leptospirosis in endangered island foxes and California sea lions: Outbreak prediction and prevention in a changing world* (Final RC-2635). University of California, Los Angeles.

Lloyd-Smith, J. O., Greig, D. J., Hietala, S., Ghneim, G. S., Palmer, L., St Leger, J., Grenfell, B. T., & Gulland, F. M. (2007). Cyclical changes in seroprevalence of leptospirosis in California sea lions: Endemic and epidemic disease in one host species? *BMC Infectious Diseases*, *7*(1), 125. https://doi.org/10.1186/1471-2334-7-125

Mannelli, A., Bertolotti, L., Gern, L., & Gray, J. (2012). Ecology of *Borrelia burgdorferi sensu lato* in Europe: Transmission dynamics in multi-host systems, influence of molecular processes and effects of climate change. *FEMS Microbiology Reviews*, *36*(4), 837–861. https://doi.org/10.1111/j.1574-6976.2011.00312.x

Marina, C. F., Fernández-Salas, I., Ibarra, J. E., Arredondo-Jiménez, J. I., Valle, J., & Williams, T. (2005). Transmission dynamics of an iridescent virus in an experimental mosquito population: The role of host density. *Ecological Entomology*, *30*(4), 376–382. https://doi.org/10.1111/j.0307-6946.2005.00711.x

McCallum, H., Jones, M., Hawkins, C., Hamede, R., Lachish, S., Sinn, D. L., Beeton, N., & Lazenby, B. (2009). Transmission dynamics of Tasmanian devil facial tumor disease may lead to disease-induced extinction. *Ecology*, *90*(12), 3379–3392. https://doi.org/10.1890/08-1763.1

Millán, J., Cevidanes, A., Chirife, A. D., Candela, M. G., & León-Vizcaíno, L. (2018). Risk factors of *Leptospira* infection in Mediterranean periurban micromammals. *Zoonoses and Public Health*, *65*(1), e79–e85. https://doi.org/10.1111/zph.12411

Mummah, R. O., Gomez, A. C. R., Guglielmino, A., Borremans, B., Galloway, R. L., Prager, K. C., & Lloyd-Smith, J. O. (unpublished). *MAT is a minefield: A cautionary tale of a single Leptospira serovar in multiple host species*.

Murray, M. H., Fidino, M., Fyffe, R., Byers, K. A., Pettengill, J. B., Sondgeroth, K. S., Killion, H., Magle, S. B., Rios, M. J., Ortinau, N., & Santymire, R. M. (2020). City sanitation and

socioeconomics predict rat zoonotic infection across diverse neighbourhoods. *Zoonoses and Public Health*, *67*(6), 673–683. https://doi.org/10.1111/zph.12748

NLCD 2016 Land Cover Conterminous United States. (2019). U.S. Geological Survey. https://doi.org/10.5066/P937PN4Z

NPS Land Resources Division. (2016). *Administrative Boundaries of National Park System Units 3/31/2016—National Geospatial Data Asset (NGDA) NPS National Parks Dataset*. National Park Service. http://science.nature.nps.gov/nrdata

Pepin, K. M., Pedersen, K., Wan, X.-F., Cunningham, F. L., Webb, C. T., & Wilber, M. Q. (2019). Individual-Level Antibody Dynamics Reveal Potential Drivers of Influenza A Seasonality in Wild Pig Populations. *Integrative and Comparative Biology*, *59*(5), 1231–1242. https://doi.org/10.1093/icb/icz118

Pourbohloul, B., Ahued, A., Davoudi, B., Meza, R., Meyers, L. A., Skowronski, D. M., Villaseñor, I., Galván, F., Cravioto, P., Earn, D. J. D., Dushoff, J., Fisman, D., Edmunds, W. J., Hupert, N., Scarpino, S. V., Trujillo, J., Lutzow, M., Morales, J., Contreras, A., … Brunham, R. C. (2009). Initial human transmission dynamics of the pandemic (H1N1) 2009 virus in North America. *Influenza and Other Respiratory Viruses*, *3*(5), 215–222. https://doi.org/10.1111/j.1750-2659.2009.00100.x

Power, P., & Rudolph, R. (2018). Quantifying Surface Water on Santa Rosa Island, California, Following a Major Five-Year Drought. *Western North American Naturalist*, *78*(4), 530. https://doi.org/10.3398/064.078.0402

Prager, K., Buhnerkempe, M. G., Greig, D. J., Orr, A. J., Jensen, E. D., Gomez, F., Galloway, R. L., Wu, Q., Gulland, F. M., & Lloyd-Smith, J. O. (2020). Linking Longitudinal and Cross-sectional Biomarker Data to Understand Host-Pathogen Dynamics: Leptospira in California Sea Lions (Zalophus californianus) as a Case Study. *PLoS Neglected Tropical Diseases*.

R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Reynolds, J. J. H., Hirsch, B. T., Gehrt, S. D., & Craft, M. E. (2015). Raccoon contact networks predict seasonal susceptibility to rabies outbreaks and limitations of vaccination. *Journal of Animal Ecology*, *84*(6), 1720–1731. https://doi.org/10.1111/1365-2656.12422

Roemer, G. W., Smith, D. A., Garcelon, D. K., & Wayne, R. K. (2001). The behavioural ecology of the island fox (Urocyon littoralis). *Journal of Zoology*, *255*(2001), 1–14. https://doi.org/10.1017/S0952836901001066

Satayathum, S. A., Muchiri, E. M., Ouma, J. H., Whalen, C. C., & King, C. H. (2006). Factors affecting infection or reinfection with *Schistosoma haematobium* in coastal Kenya:

Survival analysis during a nine-year, school-based treatment program. *The American Journal of Tropical Medicine and Hygiene*, *75*(1), 83–92.

Sun, J. (2006). *The statistical analysis of interval-censored failure time data*. Springer.

Tabel, H., & Karstad, L. (1967). The renal carrier state of experimental Leptospira pomona infections in skunks (Mephitis mephitis). *American Journal of Epidemiology*, *85*(1), 9–16.

Therneau, T. (2021). *A Package for Survival Analysis in R* (R package version 3.2-11) [Computer software]. https://CRAN.R-project.org/package=survival

Vandormael, A., Dobra, A., Bärnighausen, T., de Oliveira, T., & Tanser, F. (2018). Incidence rate estimation, periodic testing and the limitations of the mid-point imputation approach. *International Journal of Epidemiology*, *47*(1), 236–245. https://doi.org/10.1093/ije/dyx134

Vandormael, A., Tanser, F., Cuadros, D., & Dobra, A. (2020). Estimating trends in the incidence rate with interval censored data and time-dependent covariates. *Statistical Methods in Medical Research*, *29*(1), 272–281. https://doi.org/10.1177/0962280219829892

Viboud, C., Boëlle, P.-Y., Cauchemez, S., Lavenu, A., Valleron, A.-J., Flahault, A., & Carrat, F. (2004). Risk factors of influenza transmission in households. *International Congress Series*, *1263*, 291–294. https://doi.org/10.1016/j.ics.2004.01.013

Wayne, R. K., George, S. B., Gilbert, D., Collins, P. W., Kovach, S. D., Girman, D., & Lehman, N. (1991). A Morphological and Genetic Study of the Island Fox, Urocyon littoralis. *Evolution*, *45*(8), 1849–1868.

Weitz, J. S., Beckett, S. J., Coenen, A. R., Demory, D., Dominguez-Mirazo, M., Dushoff, J., Leung, C.-Y., Li, G., Măgălie, A., Park, S. W., Rodriguez-Gonzalez, R., Shivam, S., & Zhao, C. Y. (2020). Modeling shield immunity to reduce COVID-19 epidemic spread. *Nature Medicine*, *26*(6), 849–854. https://doi.org/10.1038/s41591-020-0895-3

WHO. (2020, December 1). *Mask use in the context of COVID-19*. World Health Organization. https://www.who.int/publications-detail-redirect/advice-on-the-use-of-masks-in-the-community-during-home-care-and-in-healthcare-settings-in-the-context-of-the-novel-coronavirus-(2019-ncov)-outbreak

Wilber, M. Q., Webb, C. T., Cunningham, F. L., Pedersen, K., Wan, X., & Pepin, K. M. (2020). Inferring seasonal infection risk at population and regional scales from serology samples. *Ecology*, *101*(1). https://doi.org/10.1002/ecy.2882

3 Foxes on the move: Novel methods in spatial movement reconstruction

3.1 Abstract

The technology available to study animal movement is becoming ever more advanced, leading to data with increasingly fine-scale resolution. However, collecting this data is a resource-limited endeavor, and many long-term studies generate datasets, which have coarser spatial and temporal resolution, during the course of other types of research and monitoring. These data create an opportunity to develop an approach which capitalizes on ample lower-quality data because these studies often have untapped information that is conventionally not used in estimating movement. In this study, by introducing an innovative methodology to integrate a novel spatial data type, obtained through expert interpretation of field notes, we demonstrate a method to construct wildlife movement trajectories from location data of varying resolution, using Santa Rosa Island foxes as a case study. Our approach lays the groundwork to reap the full benefit of rich, long-term monitoring datasets, which could provide vital insights into a species' movement ecology and better inform conservation and management.

3.2 Introduction

3.2.1 Background

Movement is a defining characteristic of an animal's ecology because it underlies many biological phenomena of an animal's life, including searching for food, finding a mate, avoiding predators, and socializing (Kays et al., 2015; Nathan et al., 2008). By understanding an animal's movement across a landscape, scientists can learn how they interact with their ecosystem

across timescales. For example, movement associated with foraging or predator avoidance varies from day-to-day, but juvenile dispersal and habitat selection occur over longer periods of time. These patterns in animal movement can be used to estimate population density, resource use, home ranges, and dispersal rates, all of which can inform conservation and management decisions. For example, GPS telemetry data has enabled researchers to identify predictable feeding areas in Adélie penguins (*Pygoscelis adeliae*), informing the creation of the South Orkney Islands Marine Protected Area (Trathan et al., 2018).

Movement rates and patterns also influence the speed and spatial spread of infectious diseases within a host population (Gaidet et al., 2010; Reynolds et al., 2015; VanderWaal et al., 2018; Vicente et al., 2007). For directly transmissible diseases, it is well-established that contacts between susceptible and infectious hosts drive transmission. In Yellowstone National Park, outbreaks of sarcoptic mange between grey wolf packs were driven by the proximity to and range overlap with the nearest infected pack (Almberg et al., 2012). Other work has shown that seasonal pulses in gastrointestinal nematode infections in domestic sheep and saiga (*Saiga tatarica tatarica*) coincide with annual saiga migration (Morgan et al., 2007). Analyzing movement data can thus provide unique insights in terms of individual and population level ecology while simultaneously addressing questions about major biological phenomena like disease outbreaks.

The range of technologies used to study animal movements has exploded in the last 50 years (Wilmers et al., 2015). Non-invasive tracking technology can be used to study animal movement rates, home ranges, and migrations for individuals via camera traps or large groups by satellite imagery (Fidino et al., 2021; Kucera & Barrett, 1993; Nagy et al., 2017; Rowcliffe et

al., 2016; Young et al., 2019). Alternatively, animals can be captured and uniquely tagged externally (e.g. bands, ear tags, toe clipping, paint, or dye) or internally (e.g. chemical markers, passive integrated transponder (PIT) tags) (Silvy et al., 2012). These identifying markings allow researchers to obtain location data at the time of capture and recapture, through camera traps or automatic PIT tag readers, and they enable long-term monitoring of populations. While they provide valuable information about movement behaviors, they are unable to actively monitor an animal's space use.

Very high frequency (VHF) radio telemetry pioneered active tracking technology, in which animals are equipped with radio transmitters and tracked using specialized antennas, allowing for prolonged monitoring at the individual level. VHF telemetry can provide a more general location or direction from which the collar's signal was detected, or it can specify a location point obtained through more complex telemetry known as triangulation. Telemetry triangulation is accomplished by taking at least 2 bearings based on the direction of the strongest signal heard and using the intersection of those bearings to estimate the position of the radio transmitter (British Columbia et al., 1998). However, technologies such as VHF still requires frequency trips into the field to collect detailed location data, which limits the number of obtainable observations. In recent years, with advances in global positioning system (GPS) and battery technology, autonomous sensors that are able to collect fine scale movement data have gained popularity (Craighead, 1982; Tomkiewicz et al., 2010). These advancements facilitate the collection of highly accurate animal location information at a fine temporal scale. Meanwhile, the devices themselves have gotten ever-smaller while maintaining functionality, increasing the number of small or wide-ranging species that can be tracked (Kays et al. 2015).

With the evolution of movement sensor technology, the age of large spatiotemporal datasets has followed.

With the advancements in tracking technology, statistical approaches have evolved in parallel to analyze the generated datasets (Kie et al., 2010; Tomkiewicz et al., 2010; Walter et al., 2015). The first method to estimate animal locations was developed in 1947 using minimum convex polygons (Mohr, 1947), but this method has been heavily criticized for its sensitivity to sampling duration (Powell, 2000; Swihart & Slade, 1985), sampling strategy (Börger et al., 2006), and serial autocorrelation (Laver & Kelly, 2008; Swihart & Slade, 1997). Kernel-based methods, which use the density and associated error of observations in a location, followed and were some of the first methods to enable parametric estimation of space use, home ranges, and utilization distributions (Benhamou & Cornélis, 2010; Bullard, 1999; Mohr, 1947; Worton, 1989). Work by Getz et al. (2007; 2004) expanded these kernel-based methods into a nonparametric framework using local nearest-neighbor convex hulls (LoCoH), which more readily account for hard landscape boundaries common in real-world systems. They additionally expanded the approach to account for measurements through time, allowing the estimation of trajectories, or the path of an animal's movements. Methods such as Brownian bridge modeling or continuous-time discrete-space (CTDS) models have been developed in recent years to account for the availability of high-resolution data (Hanks et al., 2015; Kranstauber et al., 2012).

Despite the many advances in animal tracking technology and analyses of the resulting datasets, an important problem in movement analyses has been left unexplored (Cagnacci et al., 2010). The types and time intervals of data collected by the advancing sensor technology vary significantly across species, technology used, and purpose of study. When research studies

are focused specifically on animal movements, more precise and frequent observations are collected. For example, Gurarie et al. (2009) used 763 GPS fixes of an individual northern fur seal over 38 days to uncover more about their behavioral ecology and develop novel methods for identifying behaviors. When location data are collected more opportunistically or without the sole purpose of studying animal movements (e.g., for survival analysis or demographic studies), the data often have lower temporal and spatial resolution but are still suitable to answer the questions of interest (Mereu et al., 2015; White & Shenk, 2001; Winterstein et al., 2001). Furthermore, limited resources and logistical constraints can reduce the frequency of location data collection. When location data are less precise and less frequent, it is non-trivial to determine whether the data can be sufficiently utilized to estimate the location or movement trajectories of individual animals. Indeed, these datasets are often overlooked in formal movement analyses because they are not well suited for use in existing methodological approaches. Furthermore, there could be unappreciated opportunities to augment datasets with partial or imprecise information that is often overlooked or deemed unsuitable for spatial analysis. For instance, field notes often contain a wealth of unused information on the locations of animals that only an expert on the system or field site would be able to interpret. By harnessing expert knowledge on the system and landscape characteristics, such notes could be converted into discrete areas where the animal was located (i.e. not specific point locations or circular buffers, but “polygon” areas defined by landscape features), and such data could be useful in reconstructing animal movement patterns when higher-resolution data are sparse.

The integration of these unconventional “polygons” with traditional but temporally infrequent location data presents a new frontier in estimating movement trajectories. Most

methods used to estimate animal movement trajectories or home ranges mentioned previously rely on precise, frequent location data in the form of specific coordinates collected via GPS or telemetry triangulation. There are currently no methods to accommodate imprecise "polygon" data or a mix of the two data types (GPS and "polygon"). Further challenges arise in estimating animal movement trajectories from such integrated data while simultaneously addressing irregular or sparse observations.

Recent work by Buderman et al. (2016, 2017) addressed a component of the challenge of analyzing telemetry datasets collected for purposes other than fine-scale movement analyses, by applying a functional data analysis approach to estimating movement and behavior of reintroduced Canada lynx (*Lynx canadensis*) in Colorado. Functional data analysis uses functions to interpolate the locations between observation times, smoothing through the imprecise location data to estimate and visualize the broad migrations of the animal. Buderman et al. (2016, 2017) presented two types of spatial data with differing error structures: Argos and GPS. Argos data have a range of seven error categories, whereas GPS data falls into the most precise Argos category. To incorporate these data types, they parametrically modelled the error structures of each data type and incorporated them into a Markov Chain Monte Carlo (MCMC) spline fitting algorithm, which additionally allowed them to derive movement behaviors from their estimated trajectories. Buderman et al. (2016) had less frequent data than most movement studies with one observation every few days (as opposed to hourly observations). Our study will adapt their method to individuals with even less temporal density. Overall, their method cannot address "polygon" data or the irregular error structure associated with these

data but a similar application of functional data analysis will address the challenge of temporal infrequency.

Herein, we propose a new way of estimating animal movement trajectories from irregularly-shaped polygon location data arising from VHF telemetry, alongside more spatially precise location data, by combining functional data analysis with a spatial resampling algorithm. We use location data from the reintroduced Channel Island fox population on Santa Rosa Island, California, which were collected at great effort and expense for purposes other than movement analyses but provide a unique opportunity to develop and test these novel methods. Furthermore, this method enables a disease outbreak reconstruction using data from a long-term study of island foxes during reintroduction.

3.2.2 Island fox reintroduction on Santa Rosa Island, California Channel Islands

The Channel Island fox (*Urocyon littoralis*) is the smallest north American canid and is endemic to six of the eight Channel Islands off the coast of Southern California (Moore & Collins, 1995). They are diurnal and omnivorous, mainly subsisting on invertebrates, fruits, and rodents. They lack natural predators and, therefore, have high survival and a docile nature (Coonan et al., 2010). Island foxes form monogamous pairs and establish territories that they inhabit year-round, with little overlap with territories of adjacent pairs (Roemer et al., 2001). Females typically give birth between April and May to litters of two to three kits on average. The offspring usually disperse in the fall to find a mate and establish their own ranges.

In the 1990s, the Channel Island fox population on Santa Rosa Island (SRI) experienced severe population declines due to a cascade of anthropogenic factors, and they became critically endangered (Coonan et al., 2005). In 2000, the National Park Service (NPS) initiated a captive breeding program, bringing the surviving 14 foxes into captivity which left the population temporarily extinct in the wild (Coonan et al., 2010). Reintroduction efforts began in 2003 and, at the conclusion of the breeding program in 2009, 96 captive foxes had been released into the wild. The reintroduced population continued to grow, and the fox population was delisted as an endangered species in 2016.

3.2.2.1 Post-reintroduction outbreak of Leptospira

In fall 2010, shortly after the conclusion of the captive breeding program, two juvenile male foxes were found dead on SRI with evidence of leptospirosis, a kidney disease caused by the bacteria *Leptospira* (See Chapters 1 and 2). There was no evidence at the time that this strain of *Leptospira* had previously circulated on the island, which spurred investigation into the origins of the pathogen on the island. Retrospective analysis of banked serum samples revealed widespread seropositivity against *Leptospira* since 2006 at least (Figure 3.1). With limited serological data prior to 2006, this phenomenon prompted questions about whether we could reconstruct the outbreak prior to 2006 and potentially learn about the locations of the earliest cases.

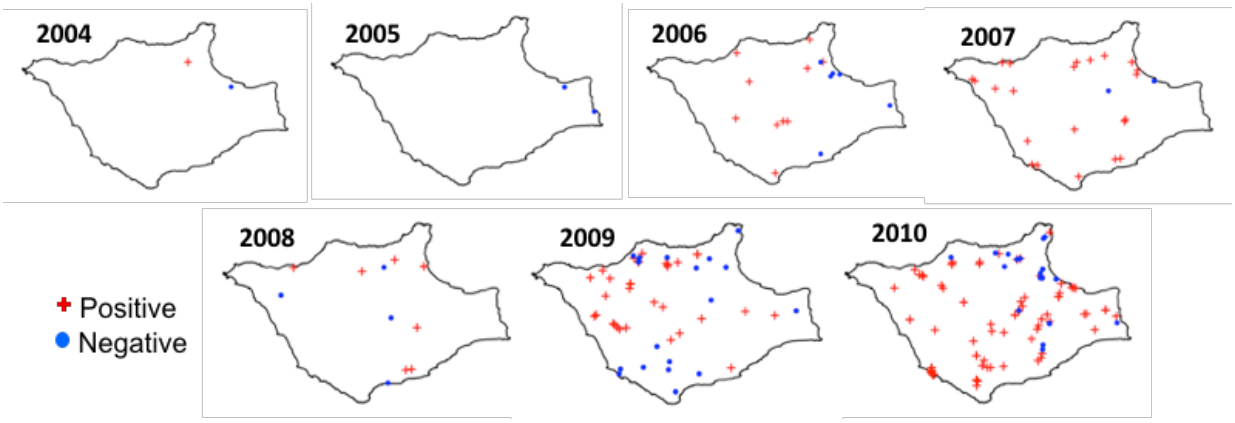

Figure 3.1: Fox serostatus against *Leptospira* from 2004 to 2010. Seropositivity is shown by red crosses, whereas seronegativity is illustrated with blue dots. Beginning in 2006, there was widespread exposure on the island, which continues through 2010. Note: The serological reactivity profile of the one positive fox in 2004 does not look like that of the outbreak-infected animals. We suspect that it is a false positive or a non-pathogen strain.

3.2.2.2 Origin reconstruction

Reconstructing the spatiotemporal origin of an outbreak requires two primary pieces of information: individual location data and temporal information on serostatus. Recent work by Borremans et al. (In preparation) presents a model which utilizes changing antibody titers through time to estimate the time an individual was infected (See Chapter 2). The output from this model provides the necessary temporal component to estimating the spatiotemporal outbreak origin. In the present work, we demonstrate the utility of our movement reconstruction algorithm by pairing it with the temporal infection data to gain new insights into when and where the outbreak originated.

3.3 Methods

3.3.1 Spatial data collection

3.3.1.1 Fox trapping

The NPS has conducted annual fox trapping since 2004 to monitor the reintroduced population on SRI. From 2004 to 2008 while the wild population was still small, target trapping was conducted on an as-needed basis in locations known to be occupied by foxes. In 2009, the NPS began a structured 18-ladder grid trapping program on SRI to estimate the annual fox population size and density (Figure S3.1). Each of the ladder grids was run for 6 consecutive nights in July or August each year. Additional target trapping efforts were conducted from July to January to install radio collars, for targeted sample collection, and to administer vaccines not completed on the grids.

3.3.1.2 Fox telemetry

From 2003 to 2006, all reintroduced and wild-born foxes were collared, PIT-tagged, and tracked with very high frequency (VHF) radio telemetry collars (Table S3.1). Locations were collected weekly via VHF, although the resolution on these locations could be quite poor at times since the primary purpose of these weekly fixes was to detect any mortalities. Additional locations were obtained through opportunistic detections. From 2007 onward, only a subset of the wild population (40-50 individuals) was collared at any time due to the growing population size (Figure S2.4). After the conclusion of the captive breeding program in 2009, telemetry efforts shifted from weekly to biweekly observations, which coincided with the shift to the grid trapping system.

Precise GPS coordinates were obtained for collared foxes at reintroduction release, trap capture, carcass collection, or visual confirmation. Location data with GPS-like precision were also generated via triangulation telemetry and estimated using the program Locate II (Nams, 1990). For the subsequent analyses, triangulation telemetry and precise GPS coordinates will both be considered "GPS data."

The less precise telemetry data were generated via single measurements with directional "Yagi" antennas or nondirectional "Omni" antennas, which were both used for routine telemetry detections. During telemetry surveillance when a collar signal was heard but the fox was not seen or triangulated, the fox PIT tag, time, and a verbal description of the general area of the signal were recorded. If the received radio signal was very strong or came from a well-defined landscape feature such as a canyon, the name of the feature was recorded. If the signal was heard with the directional antenna, the location at which the signal was detected and the direction from which it came were recorded. However, when the nondirectional antenna was used, only the location at which the signal was detected was recorded. These methods resulted in a written list of general locations, landscape features, and directions for individual foxes, rather than a set of GPS coordinates.

3.3.2 Digitizing and filtering field data

To convert the written telemetry locations into a format compatible with geospatial analysis, the recorded descriptions were translated by Angela Guglielmino, who worked as a fox biologist on SRI for more than 10 years. Each unique place description (n=3,812) was converted to a spatial polygon, mapped using the 'add polygon' tool in Google Earth Pro, and saved as a

.kml file (Figure 1). The bounds of each polygon were informed by the island topography (e.g. canyons, coastline, ridges, roads), making the shapes highly irregular. For subsequent analyses, each polygon shapefile (.kml) was imported and processed in R version 4.1.0 using the *rgdal* package (Bivand et al., 2021; R Core Team, 2021).

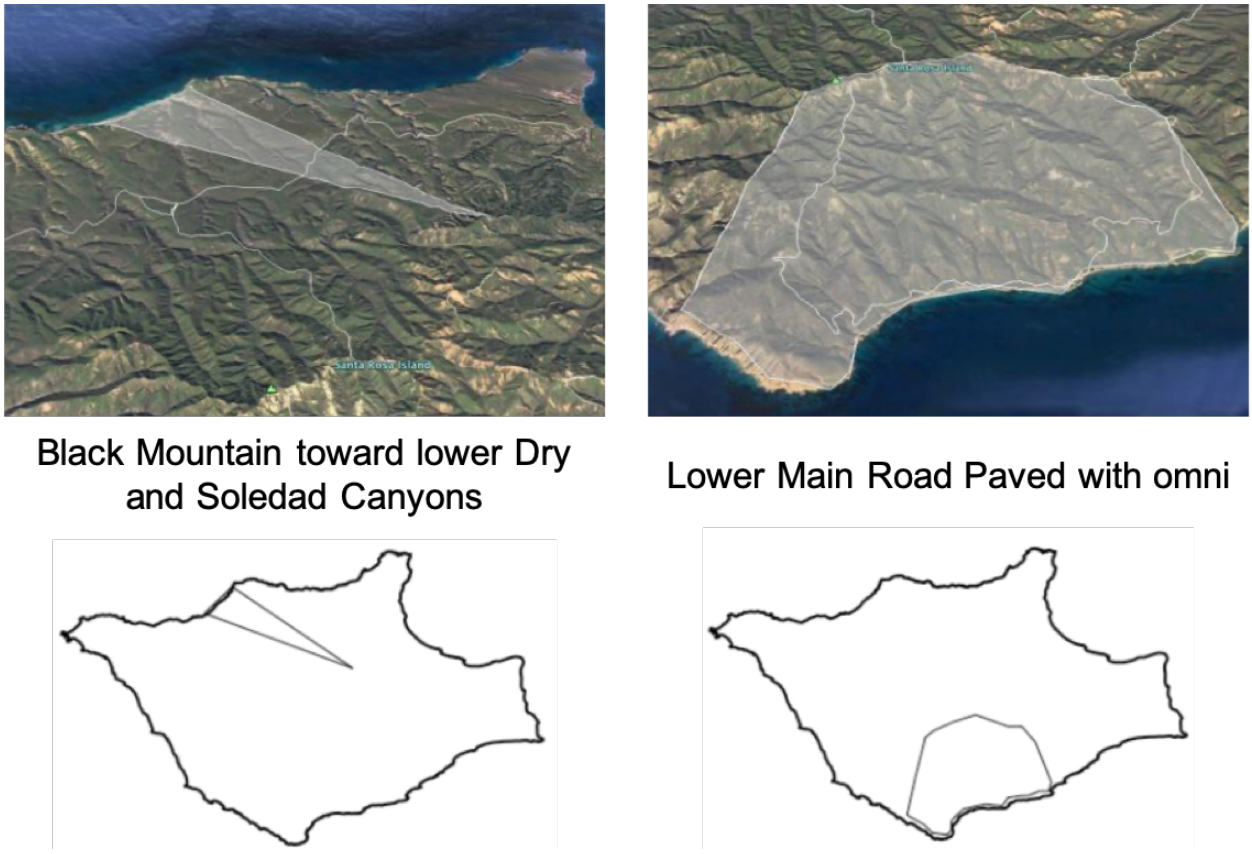

Figure 3.2: Spatial polygons converted from telemetry data. Verbal descriptions of telemetry locations were mapped in Google Earth Pro and exported to R as a .kml file for analysis.

3.3.2.1 Polygon filtering

Among the approximately 3,800 unique polygon locations in our dataset, the area of fox location polygons varies drastically from less than 1km^2 to the entire area of the island (214km^2; Figure 3.3A). To incorporate these polygons into a movement analysis, they must be at a

resolution that yields useful information about the location of the fox, while not adding too much noise. For example, polygons that are the same size as the island add no additional information on location, and only serve as confirmation that the fox was still alive. As such, we excluded polygons with areas more than 2km^2 (Figure 3.3B), which is reported as the size of the fox home range (Coonan et al., 2010). This reduced the polygon dataset to 848 unique polygons.

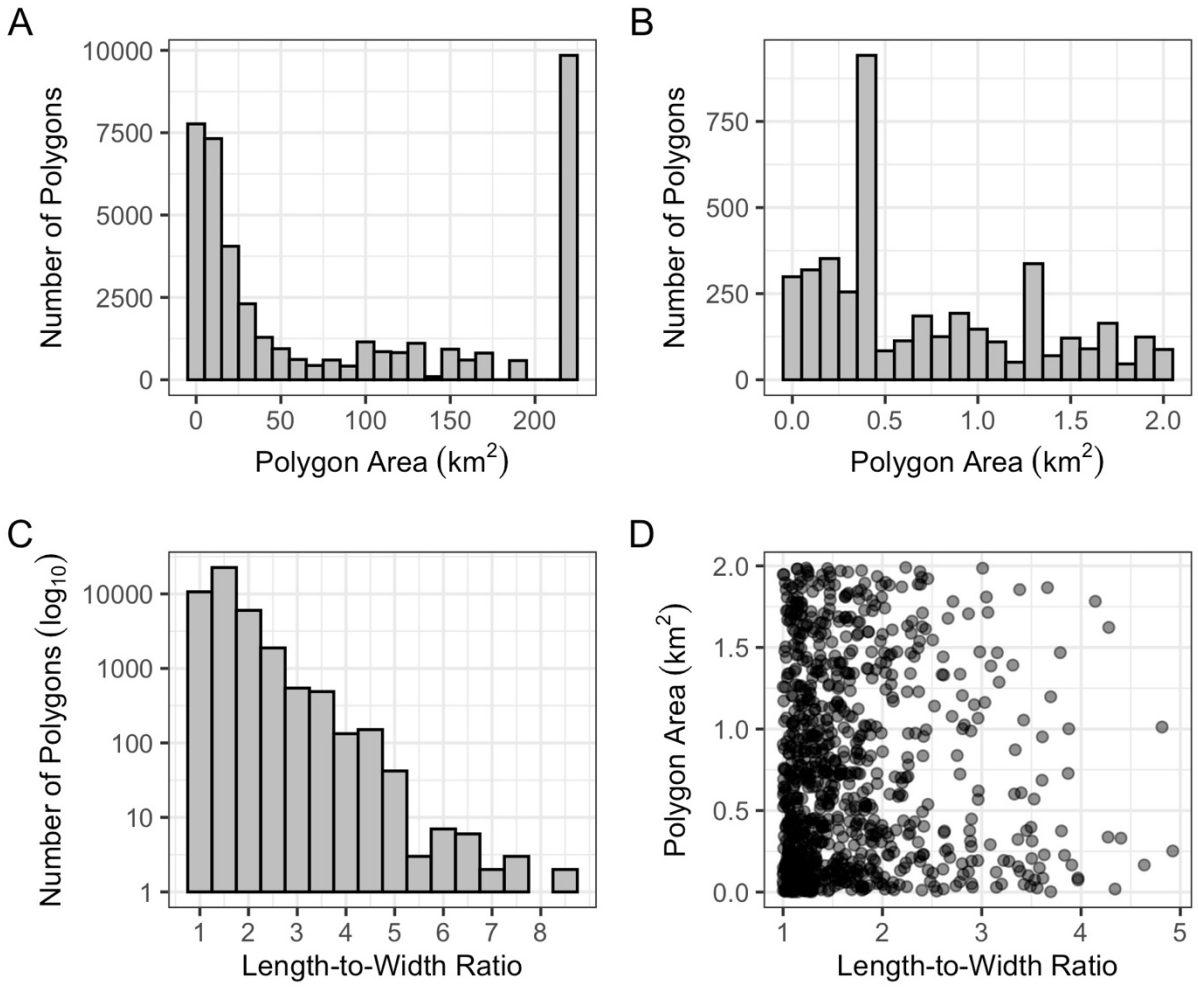

Figure 3.3: Summary distributions of the non-unique polygons included in the spatial dataset. The distribution of non-unique polygon areas (n=42,635) within the entire dataset varies widely (A). Thus, only polygons with areas less than 2km^2 (n=4,209) were included (B). The length-to-width ratio was also used to filter the polygons (C), with ratios less than 5 included. These criteria resulted in the joint distribution of unique combinations of polygon area and LWR shown in Panel D.

The length-to-width ratio (LWR) was also used to filter out imprecise polygons (Figure 3.3C). The maximum LWR across all polygons was 8.28 and results in a long, thin polygon. To filter the dataset to more elliptical polygons, which would mirror the typical error structure of location data, and exclude only the most extreme shapes, we included polygons with a LWR less than 5.

3.3.3 Dataset filtering

To investigate the fundamental challenge of studying movement from a mixed spatial resolution dataset, we focused on data in the early period of fox reintroduction between November 2003 and March 2007. Both of these data types were collected intensively on a high proportion of individuals, making this dataset ideal for the application and testing of the method. The dataset contains 61 individual foxes which have location data prior to March 2007. Any individual time series of locations which contained a gap of more than 365 days was split into two time series, unified by the individual PIT tag, to exclude the gap from trajectory estimation. Furthermore, an individual fox (or segment of the time series if it was split) was required to have more than 10 observations.

3.3.4 Resampling algorithm to integrate spatial data types

Because our dataset contained both precise GPS or triangulated locations (GPS data) and location data without precise coordinates (polygons), we needed to integrate the data into a single type (i.e., point data) that was compatible with appropriate statistical approaches while accounting for the spatial error of both. To do this, we created bootstrapped datasets by

iteratively resampling point locations from within each observed location on its observation date. For a GPS data point, the location was sampled from a circular 1km^2 buffer (clipped to the island coastline) around the original GPS location; this is greater than the spatial uncertainty of the instrument but is intended to represent the spatial scale of fox movements for foraging and other purposes. For a polygon, a location was uniformly randomly sampled from within the area of the polygon, which allows the full extent of the area to be incorporated into the movement estimation. Each iteration of this bootstrap (for every individual fox) leads to a series of point locations for each date the fox was located.

3.3.5 Estimate movement through time

The resampling algorithm yielded an ensemble of datasets with discrete point locations at irregular and varying time intervals. We built on the functional data analysis (FDA) approach applied in Buderman et al. (2016) but adapted it to the characteristics of our data. In comparison to their study, our system is wholly different in that our polygon data are irregular in shape and cannot be parametrically described, but a similar application of an FDA framework addresses the challenges of temporal infrequency in our dataset while integrating both spatial data types.

3.3.5.1 Functional data analysis

Functional data analysis uses smoothing functions called splines to interpolate data. There are many varieties of splines. Buderman et al. (2016) use a type of spline called a B-spline, which forms a linear combination of basis functions. However, this type of spline cannot

be readily applied to our data due to the irregularity of our sampling. The function cannot be fit over large gaps between observations. Instead, we employed a more flexible approach using a cubic smoothing spline. This type of smoothing spline allows the user to define the dimension of the basis (in our case, d = 2) and the smoothing parameter. The smoothing parameter can alternatively be fit by various procedures (e.g. generalized cross-validation, maximum likelihood, Aikake information criterion).

3.3.5.2 Spline-based spatial interpolation

We fit a smoothing spline to each dimension of the location data (easting and northing) as a function of time. For each date within the interval of time an individual was located on the island, a location is estimated by a fitted cubic regression spline, where the smoothing parameter was fixed at 0.1. For each fox, 100 resampled datasets were created and fitted, giving rise to an ensemble of fitted splines. The fits were then summarized by taking the median location in each coordinate direction (easting and northing) for each date in the interval of time an individual was located on the island, with a 95% percentile envelope of uncertainty. All data processing and analysis was performed in R version 4.1.0 (R Core Team, 2021). The fitting of the model to each resampled dataset was performed using the R package *npreg*.

3.3.5.3 Effects of data quantity

To investigate the effects of sampling frequency, we used a model fox with data collected twice per week on average (fox 53313, n = 187 location data points; Figure 3.7) and subsampled its full dataset to target four mean frequency levels: once per week (n = 94), every

other week (n = 44), once per month (n = 22), and every other month (n = 10) (Figure 3.11). The spline-based interpolation was applied to each subsampled dataset for comparison.

3.3.6 Reconstructing the Leptospira outbreak origin

To identify the spatiotemporal origin of the outbreak, we focused on foxes with location data in the early reintroduction period prior to November 2006 that also had anti-*Leptospira* antibody titers from a positive microagglutination test (MAT) during this period. These antibody titers were used to estimate posterior distributions on each fox's time of infection (TOI) using the titer kinetics model developed by Borremans et al. (In preparation; see Chapter 2). Twenty-five foxes have TOI estimates and location data available between 2004 and 2006 (Figure 3.4).

Seven individuals lacked location data at the beginning of their TOI interval (e.g. Fox C5B32; Figure 3.4 orange lines), but all seven were born in the wild in the year of their first observation. Because they were pups and were unlikely to have dispersed to a new home range at their first observation in late summer/early fall, we assume that the location of their first observation is near the location of their birth, and we augmented their location dataset between their birth (assumed to be 1 April) and their first observation by resampling their first location at a rate of one observation per week.

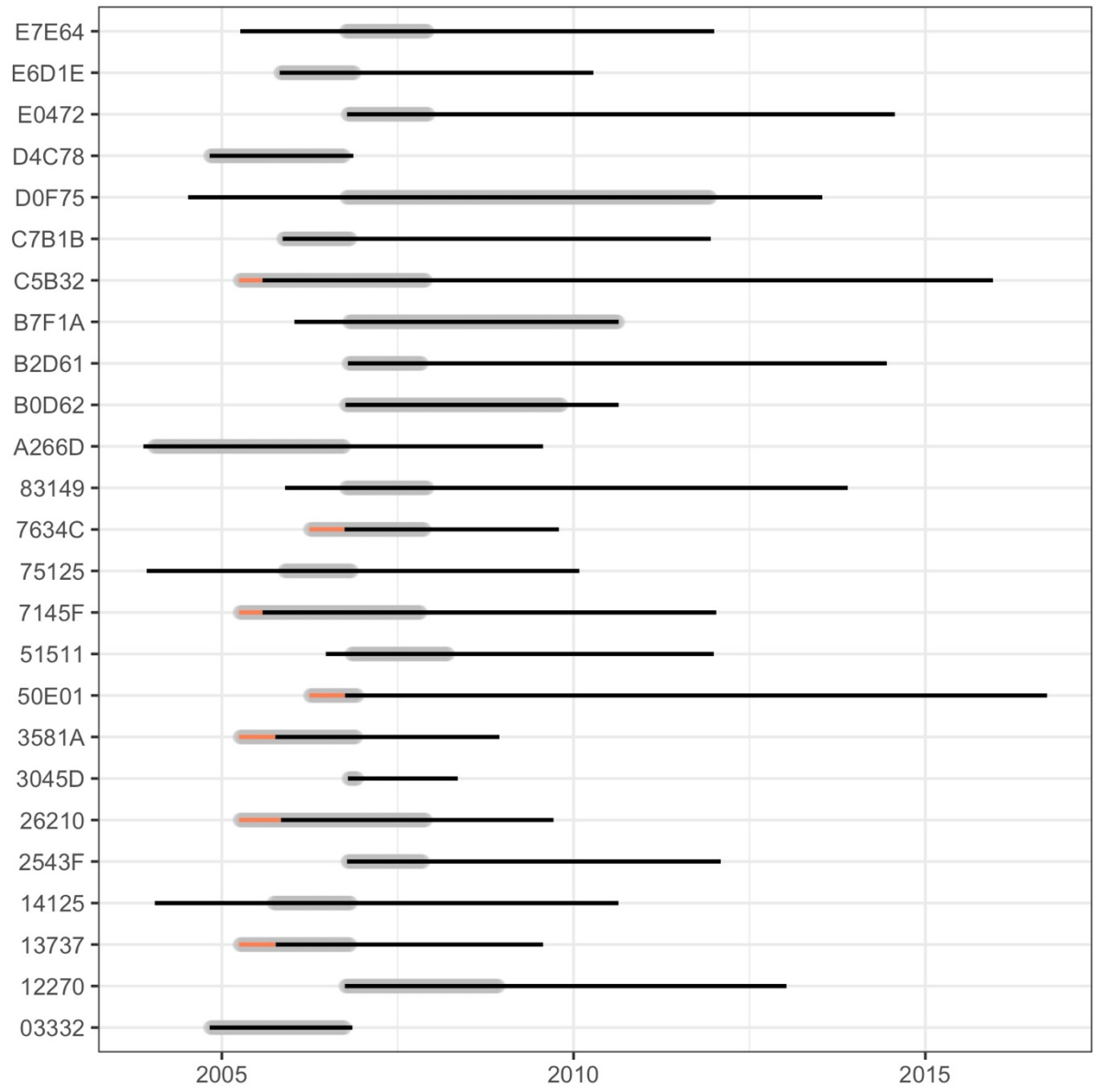

Figure 3.4: Foxes with time of infection estimates and location data between 2004 and 2006. Time of infection intervals (grey bars) were estimated using the titer kinetic model. Date ranges of location data for each fox are shown by the black lines. Intervals shown by the orange lines were imputed, assuming that the pup's first location was representative of the area in which it was born and lived for the first six months of its life.

3.3.6.1 Visualizing the outbreak reconstruction

For the twenty-five individuals, we estimated a location for every day within their observation interval (Figure 3.4) using the interpolation algorithm. Each location estimate was

paired with their daily probability of infection. For each individual, the cumulative probability of infection was calculated by day. To visualize the population-level risk of infection across the island, we created a grid of hexagons across the island map. For a given hex in a given quarter, we plotted the maximum cumulative probability of infection across all individuals that were located in that hex during that time. For example, an individual with an 80% probability of infection which spent time on four hexes during a quarter will be shown in the 80% color if there were no other foxes with a higher probability of infection during that time. After an individual is confirmed infected, it remains at the highest probability of infection for the remainder of its location data.

3.4 Results

3.4.1 Integrated spatial dataset

Our dataset contained 61 foxes with 1,774 GPS locations from March 2004 to May 2016 (Table S3.2). The inclusion of the polygon data added an additional 2,719 observations to the dataset and 476 unique polygons (Figure 3.6). The proportion of the polygon data of all available data for any individual ranged from 0% to 88% (Figure 3.5A), and five individuals had no polygon data available during this period. The median interval between observations of an individual's time series was 14 days and the median number of observations per month was 2 (Figure 3.5B, C). Without the inclusion of the polygon data, thirteen individuals would have been excluded by the ten-observation threshold.

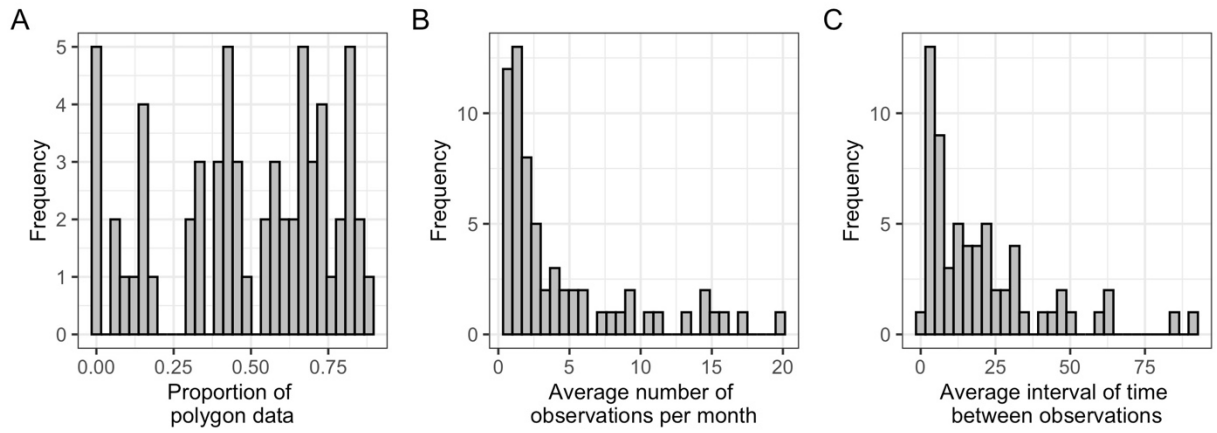

Figure 3.5: Histograms of metrics of the spatial dataset with the inclusion of the polygon data across 61 individuals.

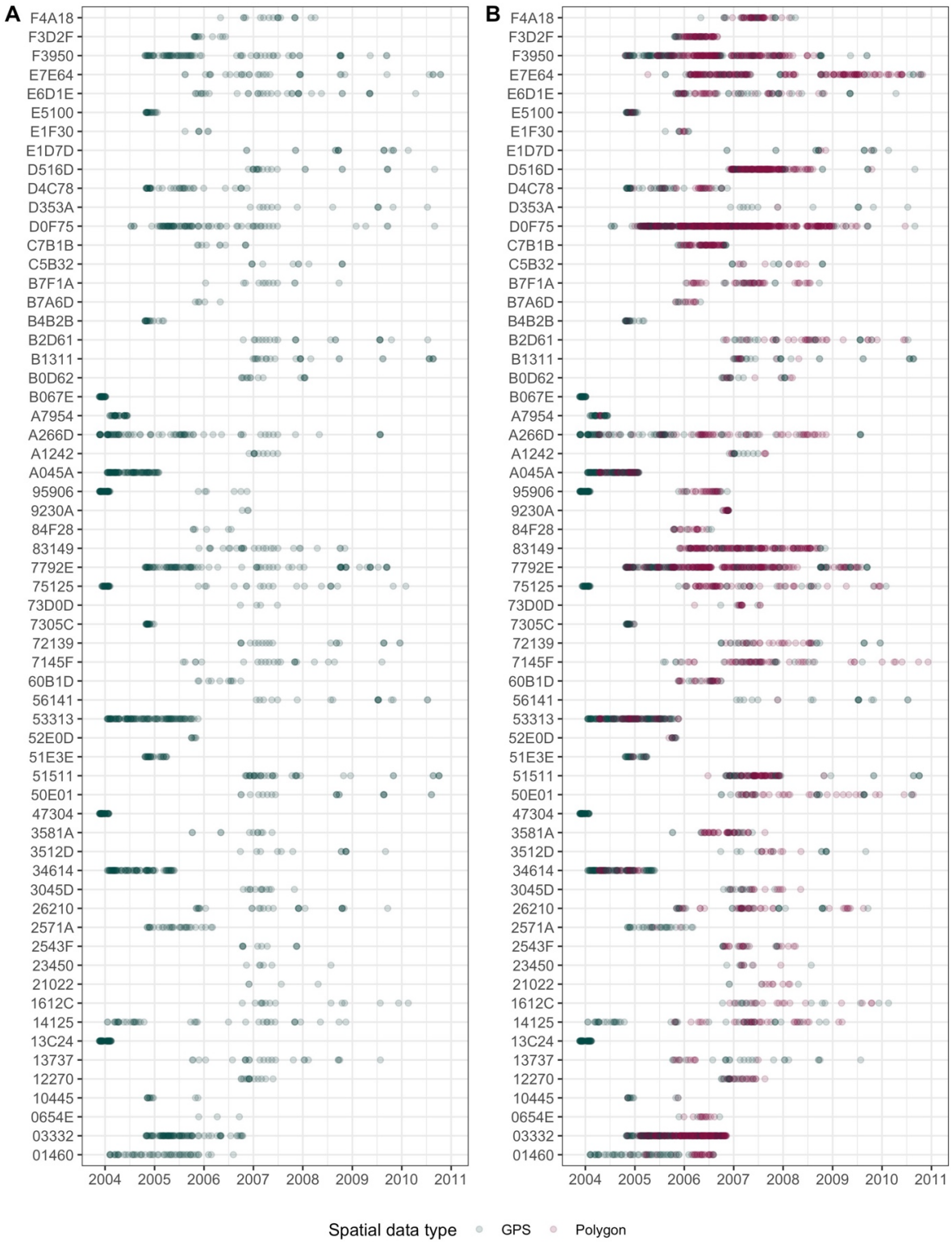

Figure 3.6: Temporal distribution of filtered location data for 61 individual foxes with observations prior to March 2007. Panel A shows the distribution of GPS points (teal) for individual foxes beginning in late 2003 and truncated at the end of 2010. With the inclusion of the polygon data filtered to less than 2km^2 (B), the number of observations included in our dataset in total more than doubles.

3.4.2 Spline-based movement trajectory estimation

With the spatial dataset, including both GPS and polygon data, we estimated movement trajectories by fitting smoothing splines independently to both easting (Figure 3.7A) and northing (Figure 3.7B) coordinate directions of resampled locations. When the spline-fit coordinates were then paired based on observation date, a two-dimensional trajectory was recreated on the island (Figure 3.7C), where the line represents the median trajectory of the individual through time. The 95% envelope represents the bounds of the center 95% of trajectories (Figure 3.7; dashed lines). The bounds of the 95% envelope are well-within a reasonable range (~1 km^2) for daily movements, but many apparent data points fall outside these bounds. The polygon data points are resampled in the algorithm using the full polygon area but are visualized using the coordinates of the centroid, so they may not fall within the bounds of the envelope. The spline fitting smooths through the locations and balances the residuals which results in the median and center 95% envelope to appear as an average, running between the datapoints.

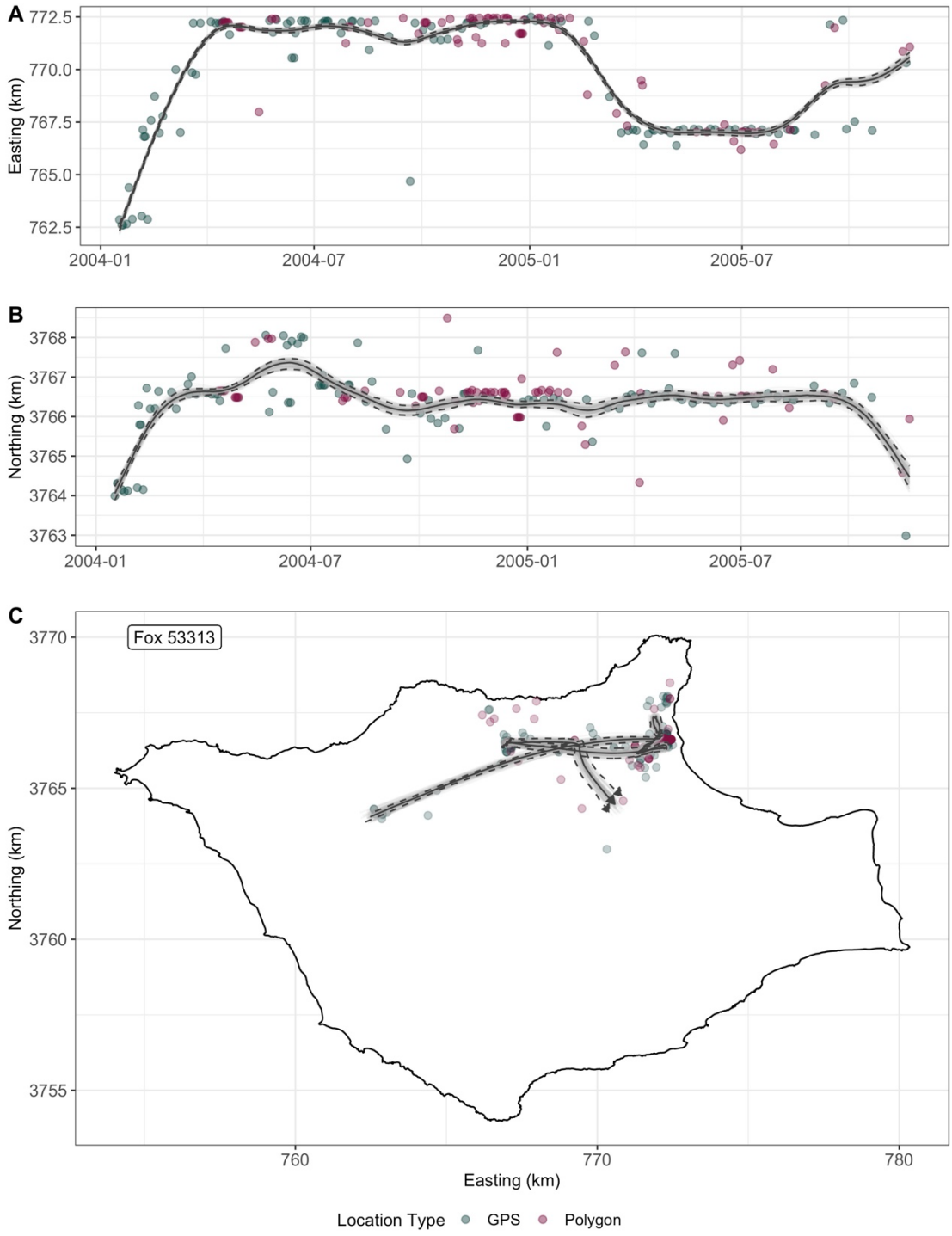

Figure 3.7: Movement trajectories fit via smoothing splines for fox 53313. Smoothing splines were fit to the easting and northing coordinates separately (left column). Each plot shows the GPS (teal) and polygon (purple) locations on the date they were recorded. Polygon data points represent the coordinates of the polygon centroids, but the fitting was done with resampled locations. Individual spline fits are shown by the light grey lines. The solid black line denotes the median value across the 100 bootstrap fits, and the center 95% interval is illustrated by the dashed black lines.

3.4.2.1 Effects of different spatial data types

To explore the impact of our new method to include polygon data in our movement trajectory estimates, we examined three examples that illustrate where incorporating polygon data may add meaningful information. These three cases are illustrative of two dimensions: the proportion of polygon data to GPS data and temporal resolution. For the first, individual foxes may have few or many polygon-based observations and the number of polygons compared to GPS points is likely to influence the utility of these data. The second dimension, temporal resolution, represents how often a fox is being observed through time. Higher temporal resolution means the fox was observed more often, which can affect the quality of movement trajectory estimation. We expected the inclusion of polygon data to matter most when the proportion of polygons was high or when the temporal frequency was low.

Table 3.1: Metrics on focal individuals of movement reconstruction.

PIT tag	Date interval	Time interval (months)	Max interval between obs. (days)	No. of GPS data points	No. of polygon data points (% of total)	Average obs. per month
A7954	02/05/04 - 06/14/04	4	16	30	6 (17%)	9.00
F3D2F	10/20/05 - 09/03/06	11	22	12	55 (82%)	6.09
73D0D	03/18/06 - 07/16/07	16	193	4	11 (73%)	0.94

The inclusion of polygon data can increase the temporal frequency in a more significant way when a fox has few GPS points, as demonstrated by fox F3D2F (Figure 3.9). Over an eleven-month span, this individual only has twelve GPS locations, which are highly concentrated in the

first two months (Figure 3.9 column 2). If we only used GPS data, this fox would have low temporal frequency and clearly unrepresentative location estimates, as seen in Figure 3.9B. By including polygon data, the data available increases four-fold, and the inclusion stabilizes the estimate of the trajectory in the second half of this individual's observation interval (Figure 3.9A). The inclusion of polygon data in this case refines the estimates to a smaller area of the island and fills out the temporal resolution of the data. Conversely, including the GPS data extends the temporal range of the data and shows a period when the individual was in a different region of the island.

In the last two examples, we demonstrated the utility of polygon data in enhancing the temporal frequency and range of an individual's dataset. However, when there are large gaps (relative to the full interval of observation time), the smoothing algorithm is forced to fit a curve to the gap with no information on where the fox is located. This situation can lead to a break down of the method and artefacts from the spline fitting as shown by the northward U-shaped curve in Figure 3.10C. Even in the best case scenario when both data types are combined, the remaining gap in the time series leads to a quadratic curve over the period and the trajectory of the fox is projected into the ocean. Across all three divisions of data type for fox 73D0D, the temporal irregularity of the data drives large variation in the fits. The incorporation of polygon data improves the estimate, but because the polygon points are temporally proximate to the GPS data points, they have a limited impact in improving the estimate. This illustrates the general importance of temporal regularity in the observations, independent of spatial resolution.

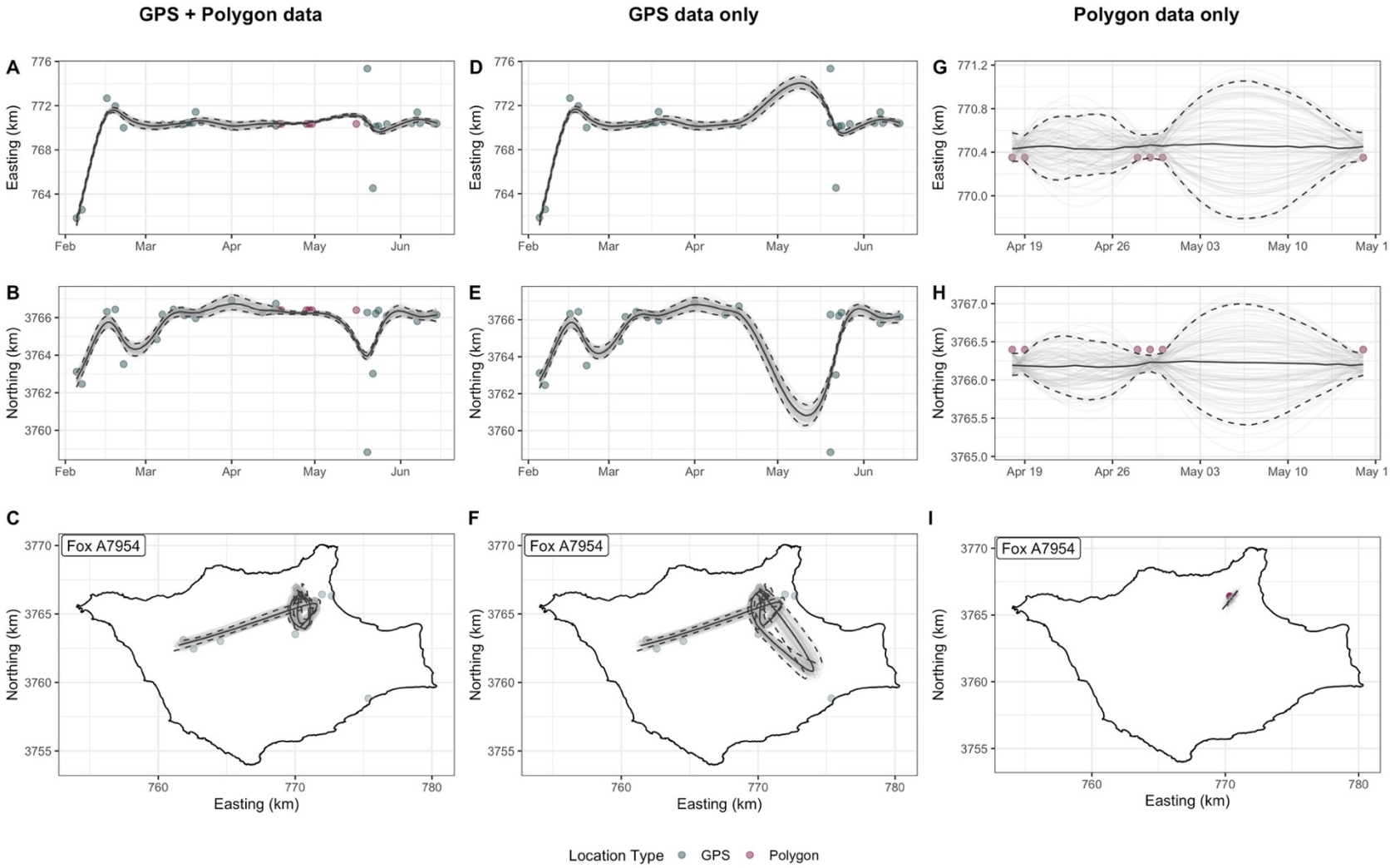

Figure 3.8: Fitted movement trajectories of fox A7954 by data type. The integration of polygon data to a GPS data rich dataset can improve the temporal regularity of observations. The first column (A, B, C) shows the trajectory estimate using all available data. The second (D, E, F) and third (G, H, I) columns show the estimated trajectories with GPS and polygon data, respectively. The first row (A, D, G) shows the fit of the easting coordinate. The second row (B, E, H) shows the spline fits for the northing coordinate, and the bottom row (C, F, I) shows the trajectory on the map of Santa Rosa Island.

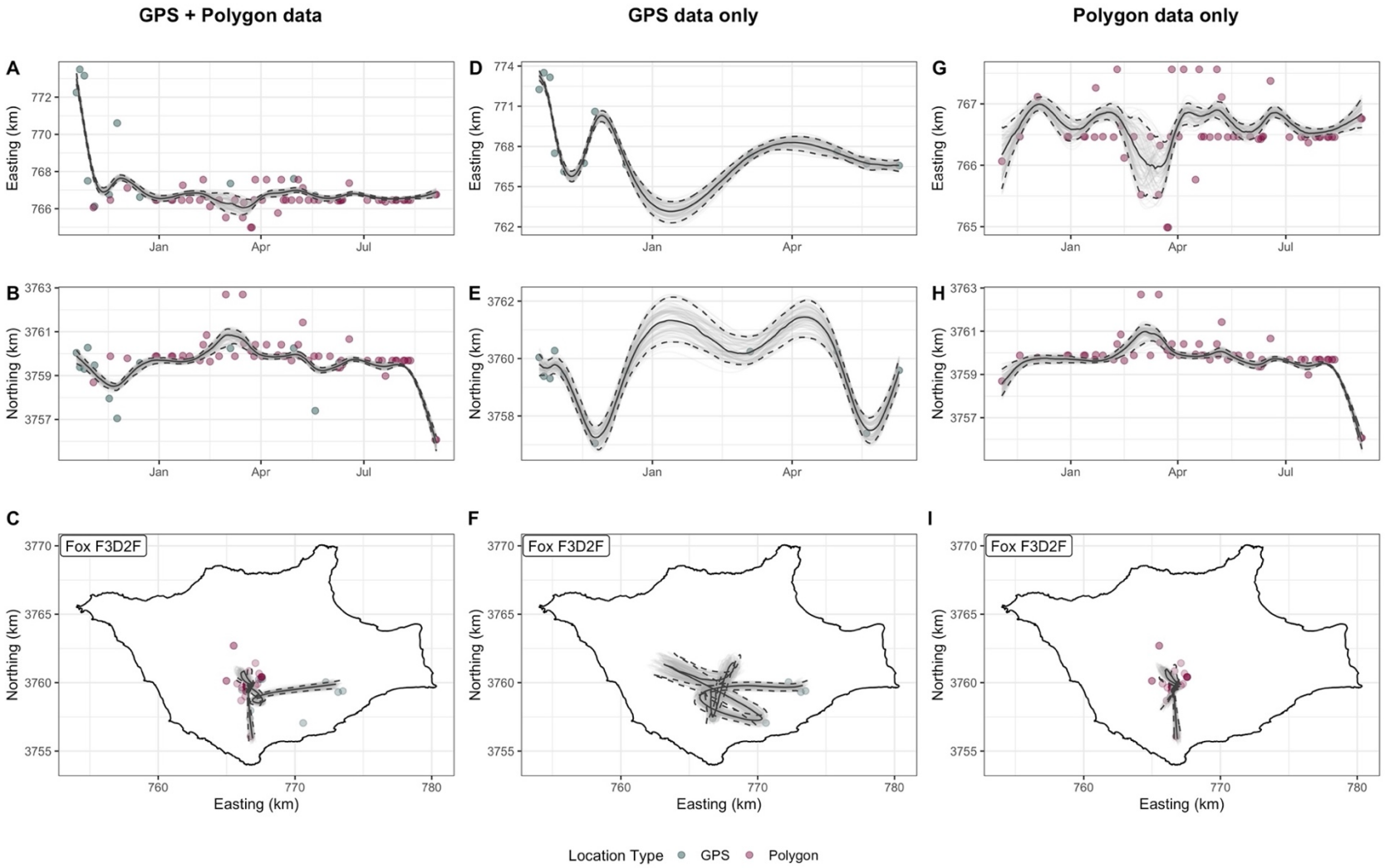

Figure 3.9: Fitted movement trajectories of fox F3D2F by data type. The integration of polygon to a GPS data poor dataset has the greatest impact in narrowing the trajectory estimates. The first column (A, B, C) shows the trajectory estimate using all available data. The second (D, E, F) and third (G, H, I) columns show the estimated trajectories with GPS and polygon data, respectively. The first row (A, D, G) shows the fit of the easting coordinate. The second row (B, E, H) shows the spline fits for the northing coordinate, and the bottom row (C, F, I) shows the trajectory on the map of Santa Rosa Island.

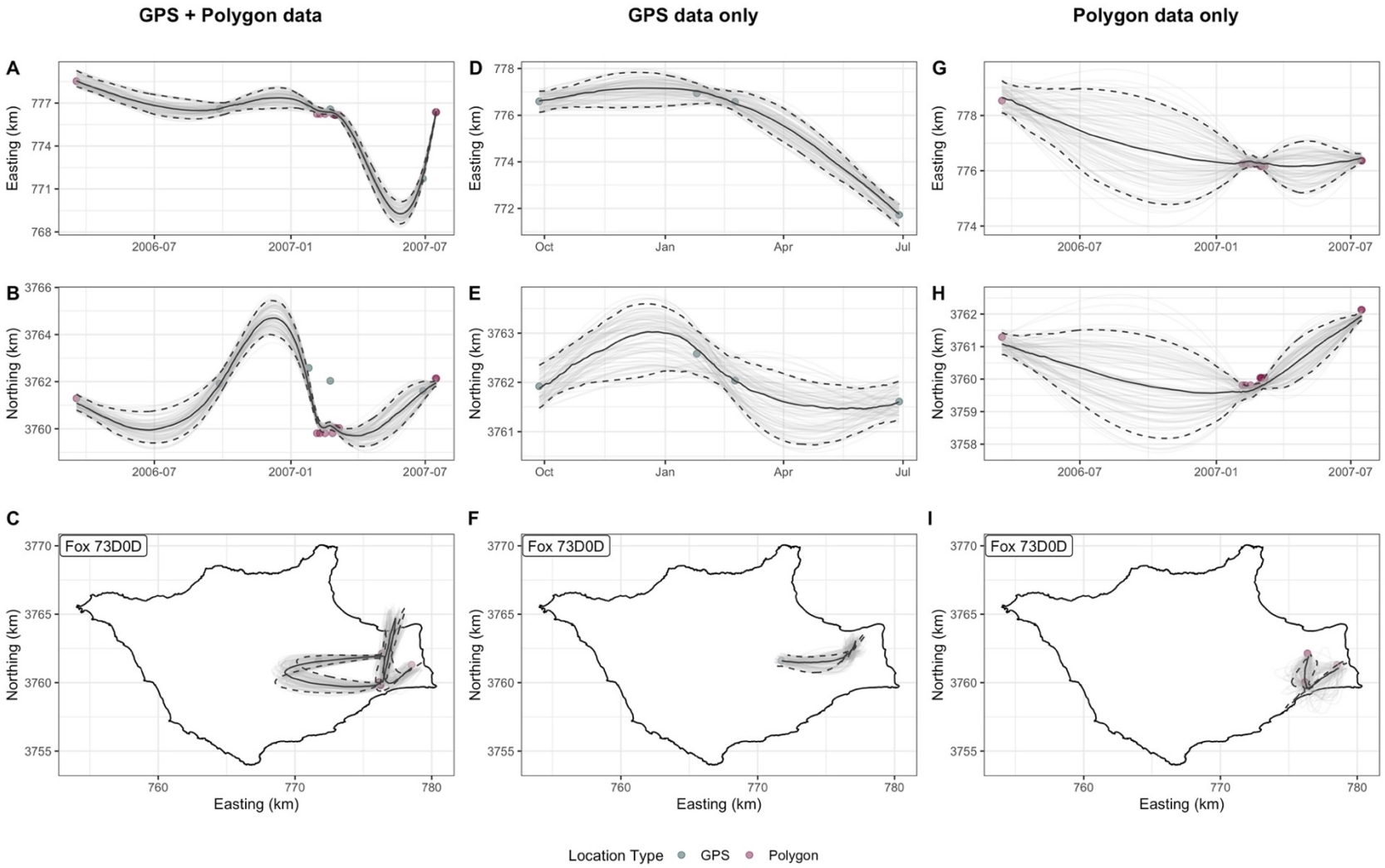

Figure 3.10: Fitted movement trajectories of fox 73D0D by data type. The inclusion of polygon data does not always improve estimates if the data are too concentrated during one period of time. The first column (A, B, C) shows the trajectory estimate using all available data. The second (D, E, F) and third (G, H, I) columns show the estimated trajectories with GPS and polygon data, respectively. The first row (A, D, G) shows the fit of the easting coordinate. The second row (B, E, H) shows the spline fits for the northing coordinate, and the bottom row (C, F, I) shows the trajectory on the map of Santa Rosa Island.

3.4.2.2 Effects of spatial and temporal frequency

To further investigate the effects of data quantity, we used a model fox (fox 53313) and subsampled its full dataset to target four mean frequency levels: once per week, every other week, once per month, and every other month (Figure 3.11). In general, as the observations become less frequent, the level of variation in the trajectories increases. The trajectories for observations once or twice (Figure 3.11; black line) a week look quite similar. Observations taken every other week or once per month) show more variation in the set of trajectories due to the smoothing spline's estimation over more gaps with less data. However, these less frequent observations still identify the same region as the location of the fox. When the data is sampled down to every other month (n = 10 observations), the model fit yields limited information on the trajectory of the fox but may be more useful in estimating its home range.

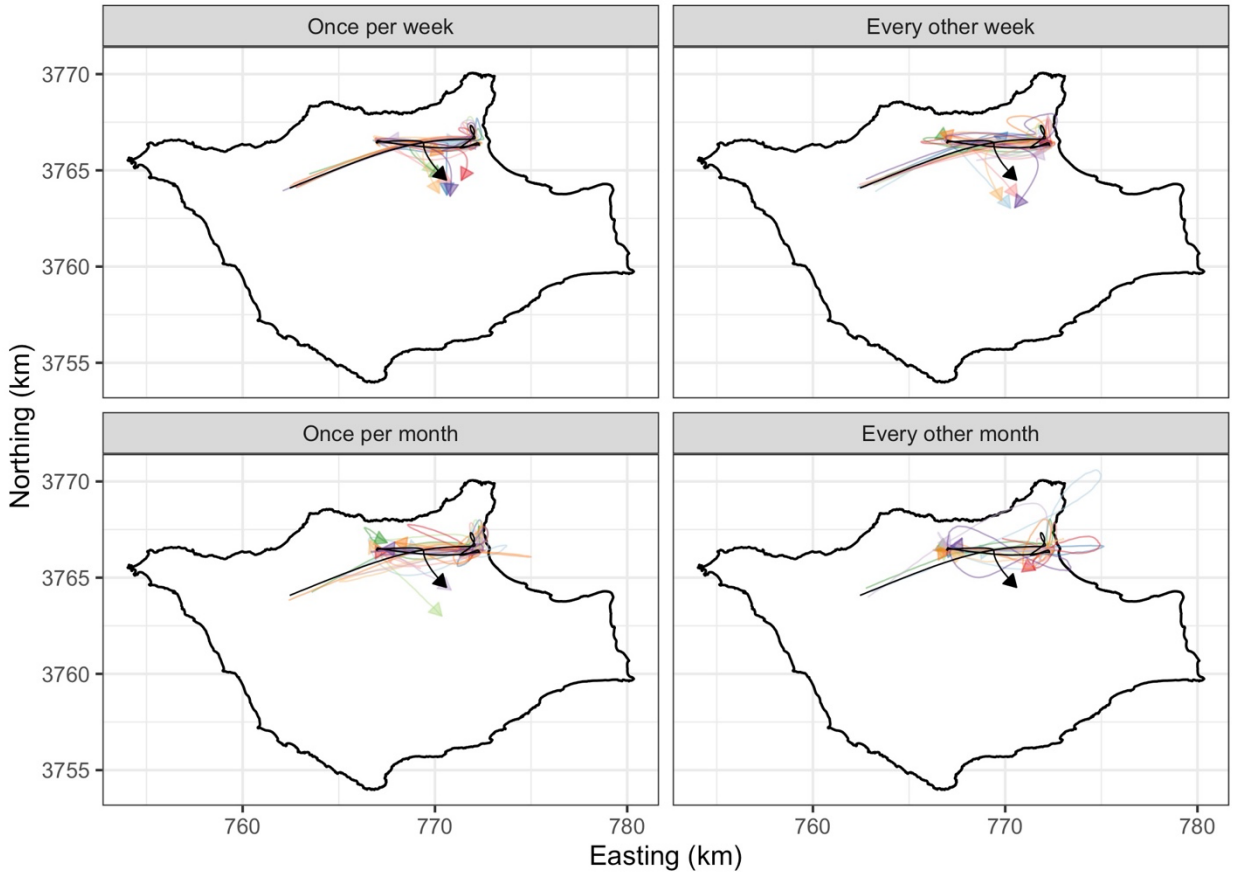

Figure 3.11: Estimated movement trajectories for down-sampled location datasets of fox 53313. The full location dataset for fox 53313 was sampled to result in four observation frequencies: once per week, every other week, once per month, and every other month (as shown by the four panels). Each down-sampling was performed 10 times for each observation frequency; the resulting median trajectories from the full algorithm are shown by different colored lines in each panel. The median trajectory for the full dataset is shown in black.

3.4.3 Extending the spatial reconstruction to the population-level: Estimating the spatiotemporal origin of a Leptospira outbreak

The pattern in the 2006 serosurvey, as shown in Figure 3.1 and echoed in Q3 and Q4 of 2006 (Figure 3.12) indicated that the *Leptospira* outbreak was island-wide by the end of 2006. Without this reconstruction, this was the extent of our knowledge about where and when the outbreak began. By developing and applying this spatial method and combining it with the time of infection estimates (via the titer kinetics model by Borremans et al. (In preparation)), we can

conclude with fair confidence that the earliest cases occurred on the north shore of the island, most likely in mid-to-late 2005. This is a great advancement in understanding when and where the outbreak started, which required the movement reconstructions to achieve.

By intersecting the cumulative probabilities that an individual was infected during a given quarter at a specific area of the island, we can visualize where and when the outbreak may have started (Figure 3.12). By quarter 2 (April - June) of 2005, there were two areas of the island where a fox had at least a 25% chance of having been infected by then. The probability grows in these two areas over the remaining quarters of 2005. By the end of 2005, *Leptospira* infections were more likely than not to have occurred in at least three areas of the island. Throughout 2006, there were multiple zones with cumulative probabilities greater than 30% with the probability growing stronger through the year, approaching quarter 3 when the island-wide serosurvey revealed widespread exposure across SRI (Figure 3.1).

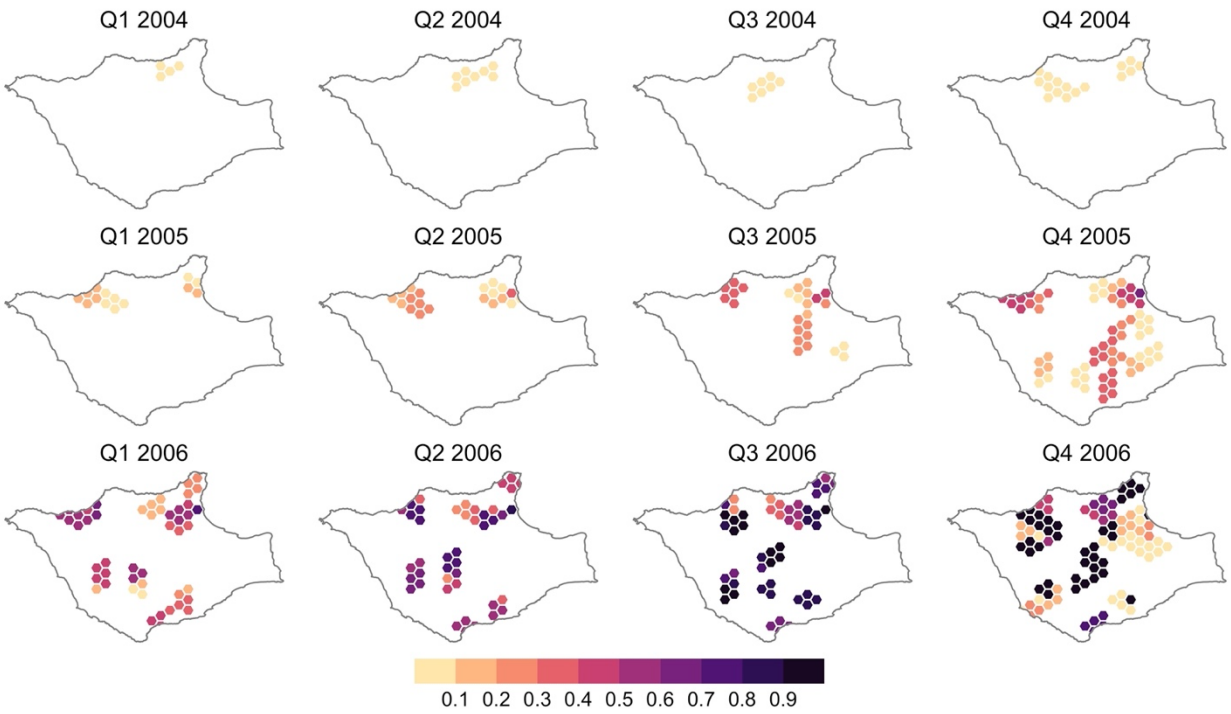

Figure 3.12: Spatiotemporal probability of infection for twenty-five foxes between 2004 and 2006. Colored grid cells represent the presence of foxes with a non-zero probability of having been infected by that time. The color scale represents the cumulative probability that an individual had been infected by the end of the quarter in question. The darker the color becomes, the higher the probability that an infected fox was present in that cell at that time. Once an individual reaches a probability of 1, they remain there.

3.5 Discussion

Using location data from Channel Island foxes, we have developed novel ways of utilizing field data and integrating differing spatial data types to estimate animal movement through time. We have shown that incorporating imprecise polygon data, derived from field notes and expert interpretation, can enhance spatial datasets and improve estimation of movement trajectories. The inclusion of polygon data nearly doubled the number of locations we included in the dataset across the 61 individuals and reduced the variation in the fits of the trajectories, especially for individuals that have few precise location estimates. This method has broad

applications to long-term monitoring datasets that were not collected to explicitly study movement.

Our data demonstrates the interacting roles of spatial and temporal resolution in enabling movement reconstruction. The quality of the reconstruction depends first on the precision of the data points (as shown by Fox A7954 in Figure 3.8). The time series for this fox consists mostly of GPS points, and the resulting median trajectory shows little variation across the bootstrapped datasets. Next the quality of the reconstruction depends on the quantity of data. When an individual's time series consists primarily of polygon data (i.e. Fox F3D2F; Figure 3.9), the algorithm results in a similarly confident median trajectory. The incorporation of the polygon data in this instance also has the benefit of filling the temporal gaps in the time series. Data quantity and temporal regularity are related, but data quantity does not imply temporal regularity as is seen with Fox 73D0D (Figure 3.10). When the incorporation of polygon data increases the amount of data available but concentrates the data in time, the algorithm gives problematic results (as seen by the trajectory projected into the ocean (Figure 3.10). In sum, the quality of the movement reconstruction depends heavily on the spatial and temporal resolution of the data.

When observations are collected at lower temporal frequencies, they can still be useful for home range estimation. Sufficient temporal frequency of the data is dependent on the species under study. In our system, island foxes form stable, territorial home ranges and do not frequently complete long-range dispersals; an individual can be predicted to stay in a restricted area. If the foxes were a well-mixed population with large, frequent movements across the island, estimation of their movements would require more frequent observations. In systems

where animals move more broadly, the presence of temporal gaps will cause issues in the movement estimates, and such a system will have higher data demands.

Our algorithm utilized smoothing splines to interpolate location data through space and time. We chose smoothing splines because they offer a flexible approach by tuning a single smoothing parameter either manually or through a generalized cross-validated procedure. For our analysis, we fixed the smoothing parameter (at 0.1 on a scale of 0 to 1) so the curvature was consistent across individuals. If we had chosen a larger smoothing parameter, the trajectories would have become less curved and more like connected lines between the observation points. Depending on the nature of the system and dataset, more or less curvature may be needed, but further study is required to uncover the extent to which the temporal and spatial resolution of the dataset influences the choice of smoothing parameter. Additional choice of the order of the spline is required. We chose a cubic spline as it is commonly used for spline fitting, but a higher order spline would add more flexibility to the fitting and may be necessary for other datasets.

The choice of spline regimes can also impact the results and interpretation of the trajectories. An alternative choice of B-spline framework would require the data to be more regular through time and is much more sensitive to temporal gaps in the data. Thus, the algorithm would fail for individuals with insufficient temporal regularity or data quantity. The smoothing spline we chose converges for all individuals but can give biologically unreasonable results with insufficient data. Additional information on the extent of the influence of data quantity and resolution on the choice of spline is required to improve the biological plausibility of some of our current estimates.

In outbreak origin reconstruction, the temporal patterns suggest that the origin of the *Leptospira* outbreak most likely occurred in the second half of 2005. The individuals driving the high probabilities of infection in two areas of the island in late 2005 had little to no interaction with each other, so we believe these spatial patterns are most consistent with multiple introductions of *Leptospira* from a reservoir within the island. Further insights could be gained if this reconstruction were paired with additional genetic or ecological data such as the data presented in Chapter 1. As skunks are the other terrestrial island host known to be infected with the same species and serovar of *Leptospira interrogans*, it is most likely that skunks reintroduced the bacteria to the fox population during reintroduction.

Tracing a wildlife outbreak back to its origin is difficult, even in human systems with a plethora of data, as has been witnessed in the current COVID pandemic. Doing as much in a wildlife system is entirely unprecedented. However, this analysis has several caveats. In the early years, nearly 90% of the population was sampled, but there is a chance that we missed an early positive case. Additionally, there are three foxes which are considered positive by immunohistochemistry and lack MAT data but were not included in the reconstruction. Together this indicates that there could be positive foxes on the landscape that we are not capturing in our reconstruction, which could influence the estimation and interpretation of the spatiotemporal origin. Furthermore, the outbreak origin reconstruction indicates that the outbreak began along the north shore of the island (Figure 3.10), which is known to have haul out sites of California sea lions, another carrier of the bacteria. Without the additional genetic or ecological data, which implicate skunks, an introduction from sea lions could be plausible.

We have demonstrated the benefit of including polygon data in trajectory estimation, but translation of the recorded location descriptions into digital polygons was onerous and time-consuming. In future studies, polygon creation could be improved by having pre-defined areas on the island that the field notes map onto, rather than translating each phrase into its own precise polygon. These pre-defined areas could be delineated by topographical features much like our polygons were. However, some level of spatial resolution could be lost by this method. There is also scope to apply artificial intelligence methods, whether trained by an expert or untrained and working from landscape features, to identify polygons.

There is great opportunity for data imputation methods to be built-in to modern spatial statistical techniques. An extension of this paper would be to integrate the resampling of the polygons (to account for their shape/error) into the Bayesian framework developed by Buderman et al. (2016). They assumed the spatial dataset was fixed and, instead, described the error distributions of their observations parametrically. To integrate the polygon data into their framework, the spatial data could be resampled for the MCMC estimation of the spline. Our resampling framework could be more broadly incorporated into other statistical techniques and provides an avenue to integrate polygon-type location data into other spatial methods as long as the data has sufficient temporal frequency for the analysis.

In our study, the island boundary serves as an impediment to movement estimation. The island coastline is a hard ecosystem boundary for the foxes, but such boundaries have been rarely addressed in estimation techniques (e.g. LoCoH and kernel density estimation; Benhamou & Cornélis, 2010; Getz et al., 2007). In practice, studies decide to clip the home ranges or trajectories at the boundary which could bias the estimates because it does not

redistribute the density appropriately. We did not explicitly incorporate a way to address the island boundary, but the polygon resampling approach we developed restricts observations to the terrestrial surface. However, the spline fitting can fit curves, which extend beyond the island boundary (as it did in Figure 3.10C). A more restrictive spline fitting or explicit approach to reject any splines which project movement into the ocean could address this shortcoming. In general, boundaries need to be more regularly accounted for in movement estimation analyses and spatial methods developments.

Rich datasets are often collected, arising from long-term monitoring programs with regular trapping or collared sentinel populations. However, these datasets have not been usable for movement analyses because there were no methods to incorporate the irregular data that can occur. The addition of these novel data to spatial datasets enriches movement estimation in studies that were not originally designed to study movement. Our work opens a new opportunity to use these unconventional data and to integrate them with other movement data at any level of spatial or temporal resolution.

3.6 Supplement

Table S3.1: Total number of actively collared foxes in each season. Season is defined by the fox-year from March to February. On average, 40-50 collars were active at any time during the season.

Season (Mar – Feb)	Number of collared foxes
2003	12
2004	23
2005	28
2006	55
2007	63
2008	92
2009	72
2010	63

Figure S3.1: Map of 18 ladder grids on Santa Rosa Island. Beginning in 2009, grids were run for six consecutive nights to mark and recapture Channel Island foxes.

Table S3.2: Tabulation of individual location data for the 61 foxes included in the dataset.

PIT tag	Date of first observation	Date of last observation	Time between first and last observations (months)	Total number of observations	Number of GPS data points	Number of polygon data points (% of total)	Average observations per month	Maximum interval between two observations (days)
01460	02/05/04	08/06/06	30	84	55	29 (35%)	2.80	72
03332	10/30/04	11/10/06	25	324	84	240 (74%)	12.96	27
0654E	11/23/05	09/19/06	10	15	3	12 (80%)	1.50	69
10445	11/08/04	11/17/05	12	13	11	2 (15%)	1.08	310
12270	10/02/06	08/20/07	10	33	14	19 (58%)	3.30	67
13737	10/07/05	07/27/09	45	33	23	10 (30%)	0.73	300
13C24	11/20/03	02/16/04	3	44	44	0 (0%)	14.67	16
14125	01/19/04	03/12/09	62	77	40	37 (48%)	1.24	344
1612C	10/11/06	02/18/10	40	32	14	18 (56%)	0.80	140
21022	11/29/06	04/23/08	17	12	4	8 (67%)	0.71	238
23450	11/11/06	07/27/08	20	11	6	5 (45%)	0.55	225
2543F	10/13/06	04/01/08	18	29	10	19 (66%)	1.61	194
2571A	11/08/04	03/05/06	16	33	30	3 (9%)	2.06	42
26210	11/04/05	09/19/09	46	62	25	37 (60%)	1.35	272
3045D	10/18/06	05/10/08	19	26	12	14 (54%)	1.37	132
34614	01/19/04	05/27/05	16	124	83	41 (33%)	7.75	25
3512D	09/27/06	09/03/09	36	22	13	9 (41%)	0.61	291
3581A	10/06/05	08/20/07	22	46	11	35 (76%)	2.09	210
47304	11/21/03	01/30/04	2	39	39	0 (0%)	19.50	10
50E01	10/02/06	08/20/10	46	53	19	34 (64%)	1.15	181
51511	06/25/06	12/30/11	66	104	39	65 (63%)	1.58	320

51E3E	10/22/04	04/01/05	6	31	27	4 (13%)	5.17	37
52E0D	09/10/05	11/03/05	2	11	6	5 (45%)	5.50	21
53313	01/17/04	11/21/05	22	187	111	76 (41%)	8.50	26
56141	01/22/07	07/02/11	54	20	19	1 (5%)	0.37	353
60B1D	11/18/05	09/30/06	10	31	10	21 (68%)	3.10	50
7145F	07/31/05	01/12/12	78	76	22	54 (71%)	0.97	267
72139	09/30/06	12/20/09	39	44	19	25 (57%)	1.13	326
7305C	10/29/04	12/28/04	2	21	18	3 (14%)	10.50	15
73D0D	03/18/06	07/16/07	16	15	4	11 (73%)	0.94	193
75125	12/08/03	02/08/04	2	29	29	0 (0%)	14.50	7
75125	11/20/05	01/31/10	50	70	19	51 (73%)	1.40	651
7792E	10/22/04	09/14/09	59	291	92	199 (68%)	4.93	169
83149	11/25/05	11/07/08	36	161	27	134 (83%)	4.47	39
84F28	10/13/05	07/21/06	9	23	6	17 (74%)	2.56	37
9230A	10/11/06	11/24/06	1	17	3	14 (82%)	17.00	14
95906	11/21/03	02/08/04	3	46	46	0 (0%)	15.33	6
95906	11/20/05	11/15/06	12	32	6	26 (81%)	2.67	651
A045A	01/17/04	02/02/05	13	147	89	58 (39%)	11.31	15
A1242	11/30/06	08/26/07	9	17	10	7 (41%)	1.89	46
A266D	11/21/03	07/27/09	68	142	81	61 (43%)	2.09	251
A7954	02/05/04	06/14/04	4	36	30	6 (17%)	9.00	16
B067E	11/20/03	01/06/04	2	32	32	0 (0%)	16.00	8
B0D62	10/05/06	03/10/08	17	24	14	10 (42%)	1.41	192
B1311	12/29/06	10/04/12	70	46	32	14 (30%)	0.66	345
B2D61	10/18/06	07/12/10	45	53	21	32 (60%)	1.18	195
B4B2B	10/22/04	03/09/05	5	21	17	4 (19%)	4.20	38
B7A6D	10/29/05	05/01/06	7	15	5	10 (67%)	2.14	35
B7F1A	01/13/06	09/25/08	32	55	12	43 (78%)	1.72	144

C5B32	12/21/06	10/19/08	22	19	11	8 (42%)	0.86	493
C7B1B	11/13/05	11/05/06	12	70	10	60 (86%)	5.83	27
D0F75	07/10/04	09/01/10	74	552	76	476 (86%)	7.46	274
D353A	12/09/06	07/04/11	55	19	18	1 (5%)	0.35	354
D4C78	10/30/04	11/15/06	25	64	37	27 (42%)	2.56	85
D516D	11/25/06	09/01/10	46	177	21	156 (88%)	3.85	332
E1D7D	11/12/06	02/18/10	39	20	17	3 (15%)	0.51	360
E1F30	08/14/05	02/03/06	6	11	6	5 (45%)	1.83	101
E5100	10/29/04	01/21/05	3	27	18	9 (33%)	9.00	11
E6D1E	10/29/05	04/14/10	54	75	34	41 (55%)	1.39	338
E7E64	04/07/05	01/01/12	81	185	33	152 (82%)	2.28	234
F3950	10/22/04	09/12/09	59	242	78	164 (68%)	4.10	214
F3D2F	10/20/05	09/03/06	11	67	12	55 (82%)	6.09	22
F4A18	05/03/06	04/01/08	23	56	17	39 (70%)	2.43	171

3.7 References

Almberg, E. S., Cross, P. C., Dobson, A. P., Smith, D. W., & Hudson, P. J. (2012). Parasite invasion following host reintroduction: A case study of Yellowstone's wolves. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1604), 2840–2851. https://doi.org/10.1098/rstb.2011.0369

Benhamou, S., & Cornélis, D. (2010). Incorporating Movement Behavior and Barriers to Improve Kernel Home Range Space Use Estimates. *The Journal of Wildlife Management*, *74*(6), 1353–1360. https://doi.org/10.1111/j.1937-2817.2010.tb01257.x

Bivand, R., Keitt, T., Rowlingson, B., Pebesma, E., Sumner, M., Hijmans, R., Baston, D., Rouault, E., Warmerdam, F., Ooms, J., & Rundel, C. (2021). *rgdal: Bindings for the "Geospatial" Data Abstraction Library* (1.5-23) [Computer software]. https://CRAN.R-project.org/package=rgdal

Börger, L., Franconi, N., De Michele, G., Gantz, A., Meschi, F., Manica, A., Lovari, S., & Coulson, T. (2006). Effects of sampling regime on the mean and variance of home range size estimates. *Journal of Animal Ecology*, *75*(6), 1393–1405. https://doi.org/10.1111/j.1365-2656.2006.01164.x

Borremans, B., Mummah, R. O., Guglielmino, A., Galloway, R. L., Hens, N., Prager, K. C., & Lloyd-Smith, J. O. (In preparation). *Estimating time of infection strictly from field data: Modeling antibody decay in Channel Island foxes (*Urocyon littoralis*) infected with* Leptospira interrogans.

British Columbia, Ministry of Environment, L., and Parks, Resources Inventory Branch, Resources Inventory Committee (Canada), & Terrestrial Ecosystems Task Force. (1998). *Wildlife radio-telemetry*. Ministry of Environment, Lands, and Parks, Resources Inventory Branch for the Terrestrial Ecosystems Task Force, Resources Inventory Committee.

Buderman, F. E., Hooten, M. B., Ivan, J. S., & Shenk, T. M. (2016). A functional model for characterizing long-distance movement behaviour. *Methods in Ecology and Evolution*, *7*(3), 264–273. https://doi.org/10.1111/2041-210X.12465

Bullard, F. (1999). *Estimating the home range of an animal: A Brownian bridge approach. Johns Hopkins University*. Master thesis.

Cagnacci, F., Boitani, L., Powell, R. A., & Boyce, M. S. (2010). Animal ecology meets GPS-based radiotelemetry: A perfect storm of opportunities and challenges. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1550), 2157–2162. https://doi.org/10.1098/rstb.2010.0107

Coonan, T. J., Schwemm, C. A., & Garcelon, D. K. (2010). *Decline and recovery of the Island Fox: A case study for population recovery*. Cambridge University Press.

Coonan, T. J., Schwemm, C. A., Roemer, G. W., Garcelon, D. K., & Munson, L. (2005). Decline of an Island Fox Subspecies to near Extinction. *The Southwestern Naturalist*, *50*(1), 32–41. JSTOR.

Craighead, F. C. (1982). *Track of the grizzly*. Random House.

Fidino, M., Gallo, T., Lehrer, E. W., Murray, M. H., Kay, C. A. M., Sander, H. A., MacDougall, B., Salsbury, C. M., Ryan, T. J., Angstmann, J. L., Amy Belaire, J., Dugelby, B., Schell, C. J., Stankowich, T., Amaya, M., Drake, D., Hursh, S. H., Ahlers, A. A., Williamson, J., … Magle, S. B. (2021). Landscape-scale differences among cities alter common species' responses to urbanization. *Ecological Applications*, *31*(2). https://doi.org/10.1002/eap.2253

Gaidet, N., Cappelle, J., Takekawa, J. Y., Prosser, D. J., Iverson, S. A., Douglas, D. C., Perry, W. M., Mundkur, T., & Newman, S. H. (2010). Potential spread of highly pathogenic avian influenza H5N1 by wildfowl: Dispersal ranges and rates determined from large-scale satellite telemetry. *Journal of Applied Ecology*, *47*(5), 1147–1157. https://doi.org/10.1111/j.1365-2664.2010.01845.x

Getz, W. M., Fortmann-Roe, S., Cross, P. C., Lyons, A. J., Ryan, S. J., & Wilmers, C. C. (2007). LoCoH: Nonparameteric Kernel Methods for Constructing Home Ranges and Utilization Distributions. *PLoS ONE*, *2*(2), e207. https://doi.org/10.1371/journal.pone.0000207

Getz, W. M., & Wilmers, C. C. (2004). A local nearest-neighbor convex-hull construction of home ranges and utilization distributions. *Ecography*, *27*(4), 489–505. https://doi.org/10.1111/j.0906-7590.2004.03835.x

Gurarie, E., Andrews, R. D., & Laidre, K. L. (2009). A novel method for identifying behavioural changes in animal movement data. *Ecology Letters*, *12*(5), 395–408. https://doi.org/10.1111/j.1461-0248.2009.01293.x

Hanks, E. M., Hooten, M. B., & Alldredge, M. W. (2015). Continuous-time discrete-space models for animal movement. *The Annals of Applied Statistics*, *9*(1). https://doi.org/10.1214/14-AOAS803

Kays, R., Crofoot, M. C., Jetz, W., & Wikelski, M. (2015). Terrestrial animal tracking as an eye on life and planet. *Science*, *348*(6240), aaa2478–aaa2478. https://doi.org/10.1126/science.aaa2478

Kie, J. G., Matthiopoulos, J., Fieberg, J., Powell, R. A., Cagnacci, F., Mitchell, M. S., Gaillard, J.-M., & Moorcroft, P. R. (2010). The home-range concept: Are traditional estimators still relevant with modern telemetry technology? *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1550), 2221–2231. https://doi.org/10.1098/rstb.2010.0093

Kranstauber, B., Kays, R., LaPoint, S. D., Wikelski, M., & Safi, K. (2012). A dynamic Brownian bridge movement model to estimate utilization distributions for heterogeneous animal movement. *Journal of Animal Ecology*, *81*(4), 738–746. https://doi.org/10.1111/j.1365-2656.2012.01955.x

Kucera, T. E., & Barrett, R. H. (1993). In My Experience: The Trailmaster® Camera System for Detecting Wildlife. *Wildlife Society Bulletin (1973-2006)*, *21*(4), 505–508.

Laver, P. N., & Kelly, M. J. (2008). A Critical Review of Home Range Studies. *The Journal of Wildlife Management*, *72*(1), 290–298.

Mereu, M., Agus, B., Addis, P., Cabiddu, S., Cau, A., Follesa, M. C., & Cuccu, D. (2015). Movement estimation of Octopus vulgaris Cuvier, 1797 from mark recapture experiment. *Journal of Experimental Marine Biology and Ecology*, *470*, 64–69. https://doi.org/10.1016/j.jembe.2015.05.007

Mohr, C. O. (1947). Table of Equivalent Populations of North American Small Mammals. *The American Midland Naturalist*, *37*(1), 223–249. https://doi.org/10.2307/2421652

Moore, C. M., & Collins, P. W. (1995). Urocyon littoralis. *Mammalian Species*, *489*, 1. https://doi.org/10.2307/3504160

Morgan, E. R., Medley, G. F., Torgerson, P. R., Shaikenov, B. S., & Milner-Gulland, E. J. (2007). Parasite transmission in a migratory multiple host system. *Ecological Modelling*, *200*(3–4), 511–520. https://doi.org/10.1016/j.ecolmodel.2006.09.002

Nagy, C., Weckel, M., Monzón, J., Duncan, N., & Rosenthal, M. R. (2017). Initial colonization of Long Island, New York by the eastern coyote, *Canis latrans* (Carnivora, Canidae), including first record of breeding. *Check List*, *13*(6), 901–907. https://doi.org/10.15560/13.6.901

Nams, V. O. (1990). *Locate II* [Pacer Comput. Software].

Nathan, R., Getz, W. M., Revilla, E., Holyoak, M., Kadmon, R., Saltz, D., & Smouse, P. E. (2008). A movement ecology paradigm for unifying organismal movement research. *Proceedings of the National Academy of Sciences*, *105*(49), 19052–19059. https://doi.org/10.1073/pnas.0800375105

Powell, R. A. (2000). Animal home ranges and territories and home range estimators. *Research Techniques in Animal Ecology: Controversies and Consequences*, *442*, 65–110.

R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Reynolds, J. J. H., Hirsch, B. T., Gehrt, S. D., & Craft, M. E. (2015). Raccoon contact networks predict seasonal susceptibility to rabies outbreaks and limitations of vaccination. *Journal of Animal Ecology*, *84*(6), 1720–1731. https://doi.org/10.1111/1365-2656.12422

Roemer, G. W., Smith, D. A., Garcelon, D. K., & Wayne, R. K. (2001). The behavioural ecology of the island fox (*Urocyon littoralis*). *Journal of Zoology*, *255*(1), 1–14.

Rowcliffe, J. M., Jansen, P. A., Kays, R., Kranstauber, B., & Carbone, C. (2016). Wildlife speed cameras: Measuring animal travel speed and day range using camera traps. *Remote Sensing in Ecology and Conservation*, *2*(2), 84–94. https://doi.org/10.1002/rse2.17

Silvy, N., Lopez, R., & Peterson, M. (2012). *Techniques for marking wildlife* (Vol. 1, pp. 230–257).

Swihart, R. K., & Slade, N. A. (1985). Testing For Independence of Observations in Animal Movements. *Ecology*, *66*(4), 1176–1184. https://doi.org/10.2307/1939170

Swihart, R. K., & Slade, N. A. (1997). On Testing for Independence of Animal Movements. *Journal of Agricultural, Biological, and Environmental Statistics*, *2*(1), 48. https://doi.org/10.2307/1400640

Tomkiewicz, S. M., Fuller, M. R., Kie, J. G., & Bates, K. K. (2010). Global positioning system and associated technologies in animal behaviour and ecological research. *Philosophical Transactions of the Royal Society B: Biological Sciences*. https://doi.org/10.1098/rstb.2010.0090

Trathan, P. N., Warwick-Evans, V., Hinke, J. T., Young, E. F., Murphy, E. J., Carneiro, A. P. B., Dias, M. P., Kovacs, K. M., Lowther, A. D., Godø, O. R., Kokubun, N., Kim, J. H., Takahashi, A., & Santos, M. (2018). Managing fishery development in sensitive ecosystems: Identifying penguin habitat use to direct management in Antarctica. *Ecosphere*, *9*(8), e02392. https://doi.org/10.1002/ecs2.2392

VanderWaal, K., Perez, A., Torremorrell, M., Morrison, R. M., & Craft, M. (2018). Role of animal movement and indirect contact among farms in transmission of porcine epidemic diarrhea virus. *Epidemics*, *24*, 67–75. https://doi.org/10.1016/j.epidem.2018.04.001

Vicente, J., Delahay, R. J., Walker, N. J., & Cheeseman, C. L. (2007). Social organization and movement influence the incidence of bovine tuberculosis in an undisturbed high-density badger Meles meles population. *Journal of Animal Ecology*, *76*(2), 348–360. https://doi.org/10.1111/j.1365-2656.2006.01199.x

Walter, W. D., Onorato, D. P., & Fischer, J. W. (2015). Is there a single best estimator? Selection of home range estimators using area-under-the-curve. *Movement Ecology*, *3*(1), 10. https://doi.org/10.1186/s40462-015-0039-4

White, G. C., & Shenk, T. M. (2001). Population Estimation with Radio-Marked Animals. In *Radio Tracking and Animal Populations* (pp. 329–350). Elsevier. https://doi.org/10.1016/B978-012497781-5/50014-1

Wilmers, C. C., Nickel, B., Bryce, C. M., Smith, J. A., Wheat, R. E., & Yovovich, V. (2015). The golden age of bio-logging: How animal-borne sensors are advancing the frontiers of ecology. *Ecology*, *96*(7), 1741–1753. https://doi.org/10.1890/14-1401.1

Winterstein, S. R., Pollock, K. H., & Bunck, C. M. (2001). Analysis of Survival Data from Radiotelemetry Studies. In *Radio Tracking and Animal Populations* (pp. 351–380). Elsevier. https://doi.org/10.1016/B978-012497781-5/50015-3

Worton, B. J. (1989). Kernel Methods for Estimating the Utilization Distribution in Home-Range Studies. *Ecology*, *70*(1), 164–168. https://doi.org/10.2307/1938423

Young, J. K., Golla, J. M., Broman, D., Blankenship, T., & Heilbrun, R. (2019). Estimating density of an elusive carnivore in urban areas: Use of spatially explicit capture-recapture models for city-dwelling bobcats. *Urban Ecosystems*, *22*(3), 507–512. https://doi.org/10.1007/s11252-019-0834-6

4 Conclusions for foxes in the Channel Islands

My work in the coastal California ecosystem is a component of a broader project, which seeks to characterize *Leptospira* ecology. There are many elements of this broader project that were not explicitly included in this dissertation, namely important ecological and biomarker data of multiple host species. My dissertation provides new insights and important corroboration for the broad conclusions of this multi-faceted project. This work has revealed a new understanding of the history of cross-species transmission in this system and provides a qualitatively different set of insights and new methods to examine the origin of the *Leptospira* outbreak. Furthermore, I present the first formal analysis of risk factors for *Leptospira* in a wildlife species.

4.1 Identifying the source and spatiotemporal origin of the *Leptospira* outbreak

A primary objective of my work was to identify the source of the pathogen introduction that led to the leptospirosis outbreak in the reintroduced fox population on Santa Rosa Island (SRI). Our initial hypothesis was that *Leptospira* spilled from California sea lions (CSL) into the fox population during the reintroduction period. The analysis in Chapter 1 is inconsistent with this hypothesis and shows that the introduction was not likely from CSL. The phylogenetic reconstruction shows a nested clade structure, with deep separation between the SRI isolates from foxes and skunks and a clade of CSL isolates which were circulating at the time of the SRI outbreak. We estimated this divergence to have occurred decades to centuries ago, well before the fox population collapse in the 1990s and recent *Leptospira* outbreak. Other analyses of serological data indicate that the pathogen was circulating on SRI during the 1980s and 1990s,

at least, and that skunks were the likely reservoir of the bacteria while the foxes were in captivity. The clustering of the fox and skunk isolates in a separate clade in the phylogeny supports this evidence and illustrates that the strains circulating in the fox and skunk populations are most similar.

The phylogenetic analysis also indicates that cross-species transmission of *Leptospira* occurs within and across ecosystems. There is evidence of transmission from fox-to-skunk, sea lion-to-elephant seal, and fox-to-sea lion. The deeper clade structure of the phylogenies also indicates past transmission events across ecosystems. We hypothesize that there is an undiscovered reservoir of *Leptospira*, which has seeded multiple lineages in the coastal California ecosystem.

Another component of my work reconstructed the spatiotemporal origin of the outbreak with unprecedented resolution. Despite a limited number of serological samples during this critical period of the outbreak, we were able to estimate the timing and location of the first cases by constructing the movement trajectories of the foxes present during this early period and intersecting their movements with time of infection estimates generated by a serum antibody kinetics model. I estimate that the outbreak occurred in mid-to-late 2005 on the northern shore of the island. These findings tentatively support multiple introductions of *Leptospira* into the reintroduced fox population, and show no evident link to CSL haulout sites, further supporting island spotted skunks as the source of the outbreak.

Broadly, it is clear that California sea lions were not the proximate source of the recent outbreak on Santa Rosa Island. A terrestrial host species on the island, most likely island spotted skunks, introduced the bacteria to the establishing wild fox population. The foxes were

reintroduced onto a landscape where *Leptospira* was already circulating and by mid-to-late 2005 the outbreak was most likely initiated.

4.2 Implications for other island fox populations

There remains a risk of *Leptospira* introduction to other Channel Island fox populations. I have shown that higher cumulative rainfall over a two-year period and smaller fox abundance increases the risk of infection. The effect of fox abundance may be more about the particular conditions of the reintroduced population than about low population abundance generally. These conditions were met during 2005 when the *Leptospira* outbreak on SRI was estimated to originate. The previous two years had high levels of rainfall, and the fox population was fewer than 40 individuals. More broadly, these findings have implications on the introduction of the pathogen to fox populations on other Channel Islands. The risk of introduction would be greatest after an extended rain period with a destabilized population.

There is also a question of where an introduction to another island would originate. Although our evidence suggests that CSL were not the source of the outbreak on SRI, the possibility remains that CSL can transmit and introduce the pathogen to another island. Our evidence also suggests there exists an unidentified pathogen reservoir in the coastal California ecosystem. Identification of this unknown reservoir would shed light on potential sources of risk to other fox populations. Other research done at UCLA shows that raccoons in Los Angeles County have high seroprevalence against the same serovar of *Leptospira* involved in this system. Domestic dogs and pigs can be infected with *Leptospira* and could serve as potential terrestrial sources of the pathogen.

Understanding the transmission routes of *Leptospira* on Santa Rosa Island would guide future pathogen mitigation and elucidate risks to other species (e.g., humans, spotted skunks, other marine mammals). SRI is part of Channel Islands National Park and open to the public. *Leptospira*, as a potentially fatal infection, poses a risk to visitors (both human and other wildlife) on the island. We have shown that *Leptospira* transmission is influenced by long-term patterns in rainfall, potentially through the maintenance of environmental reservoirs for the pathogen in standing water or moist soil. By identifying potential sources of the pathogen through environmental DNA or PCR, we could understand the distribution of risk across the landscape.

This interdisciplinary body of research demonstrates the importance of long-term monitoring programs. Although wildlife data are often imperfect, maximal insights can be gained by creatively applying existing statistical techniques and inventing novel methods. By developing approaches to integrate multiple data streams, we can resolve challenges of analyzing complex wildlife disease data to understand transmission in multi-host wildlife systems and provide evidence-based recommendations to mitigate threats to at-risk species and to humans.