DISSERTATION

AVIAN INFLUENZA TAKES FLIGHT: HOST MOBILITY, VIRAL PREVALENCE, AND TRANSMISSION AT LARGE SPATIAL SCALES.

Submitted by

Brooke Berger

Graduate Degree Program in Ecology

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Fall 2024

Doctoral Committee:

Advisor: Colleen Webb

Ryan Miller
Kim Pepin
Erin Gorsich
David Koons

ABSTRACT

AVIAN INFLUENZA TAKES FLIGHT: HOST MOBILITY, VIRAL PREVALENCE, AND TRANSMISSION AT LARGE SPATIAL SCALES.

Many pathogens have large geographic distributions, but we currently have little ability to predict how they may change over time. Understanding what mechanisms drive the large-scale distributions and movements of pathogens is critical to designing effective surveillance programs, disease interventions, and predictive models of disease spread. Additionally, we lack information on what spatial scale these mechanisms are most important. In this dissertation we address one of the fundamental problems in ecology, the "problem of pattern and scale", in the context of disease prevalence and spatial transmission. Are the large-scale patterns we see emergent properties of many small-scale processes, or are they a product of large-scale processes themselves? We focused on the spatial distribution, prevalence and spatial transmission of influenza A virus in its endemic host, wild waterfowl. We used a zero-inflated Bayesian CAR model to determine if local environmental persistence of the virus or regional host migration were better predictors of large-scale patterns of prevalence. We found that an unweighted host migration network better predicted high and low values of prevalence than did local drivers. To understand how these factors impacted where IAV moves in the United States (US) we investigated how local-scale transmission and regional-scale host movements influence large-scale spatial transmission and our ability to detect these transmissions. We developed a Bayesian zero-inflated binomial network model to estimate the probability of spatial transmission between watersheds pairs. We found that regional host movement was the best predictor of spatial

transmission and that Mallard ducks likely play a special role in moving the virus throughout the US. Viral movement patterns were closely associated with important waterfowl breeding and wintering habitats rather than flyways, as has been previously shown. In order to extend these analyses to other geographic areas and host species we need to construct continental scale host movement networks from movement data with differing spatial and temporal resolutions. We developed a method to simulate host movements from very few observations allowing us to match mark-recovery data to highly detailed satellite telemetry data. We used the biological information in the detailed movement data to estimate population posterior distributions of travel speed, turning angle, and direction. These quantities and an approximately Bayesian rejection scheme were used to simulate missing locations in the mark-recovery data with estimates of uncertainty. The method was validated with a telemetry dataset tracking the movement of Northern Pintail ducks, an important host of IAV. Movement networks constructed from simulated locations captured known population scale migration patterns of Pintail and exhibited similar higher order community structure. More broadly, this research contributes to our understanding of how host mobility impacts the prevalence and movement patterns of pathogens, and the spatial scale at which this mechanism is important. Our findings suggest that predicting the spread and spillover risk of IAV requires an understanding of where hosts move at the regional scale. In the future, as climate and land-use change alter the migration patterns of wild waterfowl, we can expect the distribution and movement patterns of IAV to shift as well.

ACKNOWLEDGEMENTS

First, I would like to thank my advisor, Colleen Webb for her excellent mentorship. Her clarity, honesty, compassion, and intellectual rigor are unsurpassed, and she has been a calming and generous presence in my professional life. I have been exceptionally lucky to learn from her over the past six years. I would also like to thank my committee, Ryan Miller, Kim Pepin, Erin Gorsich and Dave Koons for their thoughtful and useful feedback on this work. An additional thank you to Erin and Ryan for collaborating on these projects.

I am also grateful to Andy Ramey of the USGS Alaska Science Center for his important intellectual contributions, and for his professionalism and patience with the pace of my work. An enormous thank you to Sarah Bevins and all of the other folks who have been involved in designing data collection, swabbing birds, running samples, and curating the datasets. None of this would be possible without them.

Thank you to the brilliant Webb Lab members both past and present for their feedback, friendship, and all the fun times we've had. They are in no particular order Ben Golas, Kendra Gilbertson, HP O'Doherty, Jonathan Bertram, Sam Smith, Catherine Hertzog, Rachel Oidtman, Mark Wilbur, Christopher Brandon, and Lindsay Beck-Johnson. A very special bonus thank you to Lindsay for keeping our computers running, sharing her coding expertise, and her excellent mentorship. My deep gratitude to Clint Leach, my best beer drinking buddy who taught me so much about formulating and fitting Bayesian models. Thank you for never making me feel stupid, for giving so generously of your time, and always making me laugh.

Thank you to my family for the lifetime of love and support. It has meant the world to me. Thanks to Dustin and Ayla for making sure we had a wonderful and affordable place to live and work. To my mom and Jeff, thank you for actually being interested in what I am doing and

believing wholeheartedly that I can do whatever I put my mind to. Lastly, my deepest gratitude to my exceptional husband Bennett. You have kept me happy, sane and well fed through all of this. Your pep talks are unparalleled, and your experience, wisdom, and excellent scholarship have been an inspiration to me. I could not have done this without you.

Funding for this work was provided by USDA APHIS National Wildlife Research Center Cooperative Agreements AP21VSSP0000C048, APP22VSSP0000C009, APP22VSSP0000C063, AP23VSSP0000CO41, and USGS Cooperative Ecosystems Study Unit Agreement G21AC10170-00. The findings and conclusions in this dissertation have not been formally disseminated by the U. S. Department of Agriculture or U. S. Geological Survey and should not be construed to represent any agency determination or policy.

DEDICATION

To my mother, Angela West, who told me that math was the secret language of the universe.

Turns out she was right.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1: Introduction

The field of disease ecology has historically focused on the local scale, modeling transmission and outbreak dynamics in homogeneous populations. This is partly a consequence of the field's origins, applying models of population dynamics to parasites and pathogens. From a practical standpoint, it is also easier to collect data at the local scale and disease interventions are often focused on controlling localized outbreaks. However, the COVID-19 global pandemic and influenza A epizootics have highlighted the urgent need to investigate how we may predict and intervene in outbreaks at the continental scale. Advances in statistical inference, disease surveillance, and remote sensing technology have made data at the large scale more available than ever and has stimulated a growing sub-field of disease ecology at the landscape scale. Central to this field is understanding what drives the global pathogen distributions we see today and how we can expect them to change over time (Ostfeld et al. 2005, Meetenmeyer et al. 2012).

Many pathogens have nearly global distributions including, but not limited to, SARSCov2, HIV, influenza A virus, Batrachochytrium dendrobatids (Bd), rabies virus, and West Nile virus. Large-scale phenomena like climate change, land-use change, and globalization are predicted to intensify outbreaks of these, and other infectious diseases in humans, wildlife, and domesticated plants and animals. Several studies have already documented these changes occurring through a variety of mechanisms. Increasing temperatures have expanded the range of the Asian tiger mosquito, a known vector of Chikungunya virus, dengue, and yellow fever (Ruiz-Moreno et al. 2012). Warming temperatures have also made historically cool climates livable year-round causing some animals to abandon their annual migrations. Monarch butterflies that stopped migrating were found to have higher rates of protozoan infection than their migratory

counterparts (Altizer et al. 2013). Fragmentation of habitat and vital food resources for fruit bats in Australia has increased contact between bats, domestic animals, and humans and resulted in the spillover of Hendra virus from bats to domestic horses (Plowright et al. 2011). Human air and boat travel has been directly linked to the introductions of West Nile virus to North America, avian malaria to Hawaii, and varroasis to honeybees causing severe declines in their new host populations (Daszak et al. 2001). These, and future changes in the spatial distributions of pathogens, pose a significant risk to human health, food security, and the persistence of sensitive wildlife populations. Studying pathogens at the large scale provides critical knowledge needed to conduct disease surveillance, design interventions, and to help move towards a future of disease prediction and prevention.

Beyond these practical motivations, expanding our focus to study pathogens at large spatial scales provides us with an opportunity to address one of the fundamental problems in ecology, the 'problem of scale' (Weins 1989). The patterns we observe at one scale are often mechanistically produced at other scales of organization (Levin 1992). Are the large-scale patterns we see emergent properties of many small-scale processes (eg. competition, transmission) or are they a product of large-scale processes themselves (eg. climate, host migration)? This question is one of the primary themes of this dissertation where I examine the impact of mechanisms at both the local and regional spatial scales on the large-scale prevalence, spatial distribution, and spatial transmission of influenza A virus (IAV) in an endemic host population.

The IAV system is well suited for comparing the relative contribution of regional and local-scale factors to overall patterns on the landscape. IAV has a nearly global distribution and is endemic to many species of wild waterfowl. These hosts undergo long-distance seasonal

migrations that are hypothesized to be important drivers of viral movement and may be responsible for introducing the virus to new locations (Lam et al. 2012, Trovao et al. 2015, Yang et al. 2024). On the local scale, transmission among waterfowl is primarily environmentally mediated (van Dijik et al 2018, Ramey et al 2020). Lab studies have shown that water temperature, pH and salinity all impact the environmental persistence of the virus (Brown et al. 2007). The longer IAV is able to survive in the environment, the more opportunities there are to infect new hosts, potentially increasing local transmission and viral prevalence. Historically, most large-scale studies have not considered factors that may impact local transmission and their subsequent impact on large-scale patterns of IAV. In Chapter 2, I focus on the relative importance of local-scale environmental transmission and regional-scale host movement in predicting the large-scale spatial distribution and prevalence of the virus. In Chapter 3, I model the large-scale spatial transmission of IAV to determine whether local factors like temperature and viral prevalence, or regional-scale host movements better predict the movement of the virus throughout the United States (US).

Other studies have looked at the role of host migration in the large-scale spatial spread of IAV (Lam et al. 2012, Tian et al. 2014, Trovao et al. 2015, Yang et al. 2024). However, these studies primarily focus on characterizing how waterfowl migratory flyways have shaped viral phylogenies and the chronology of outbreaks. Flyways characterize the seasonal migration patterns of dozens of species of waterfowl and encompass millions of square kilometers. Such a large scale is less helpful for understanding where, and what species managers should sample to detect new subtypes of the virus. This focus on flyways is largely due to the scarcity of detailed data on waterfowl movement at the continental scale. Additionally, methods of data collection often differ between regions resulting in datasets with different spatio-temporal resolutions. For

example, much of the host movement data in Asia comes from highly detailed satellite telemetry and GPS tags. In North America most of these data come from long term bird banding datasets at a much coarser spatio-temporal scale. In the absence of robust frameworks to combine different data types, flyways offer one of the few ways to connect host movement to IAV introductions and epizootics across multiple countries. This mismatch in data resolution limits not just inference in IAV spread, but studies on a wide range of ecological questions including foraging and territorial behavior, habitat connectivity, and how animal migration is responding to habitat loss and climate change.

In Chapter 4 of this dissertation, I addressed this problem by developing a method to simulate host movements from very few observations. While detailed movement trackers are becoming cheaper and more accurate, they have still not reached the point where they can be deployed on a global scale. We used these scarce, but detailed data to estimate key parameters that describe waterfowl movements and use these to simulate missing locations in less expensive and more abundant mark-recovery datasets. This has the additional bonus of providing a robust framework for integrating movement datasets at two very different spatial and temporal scales and improving the resolution of the relationship between the spatial transmission of IAV and host movements. This approach is customizable and could be used for a variety of taxa to answer questions about migratory responses to climate changes, large-scale habitat connectivity, and the global spread of infectious disease. In the context of IAV the resulting simulations can be used to construct host migration networks that can elucidate the relationship between intercontinental host movements and the probability of introducing new viruses via migratory birds.

Results from all of these chapters can be used to refine viral surveillance efforts in the US. IAV frequently spills over into domestic poultry populations causing widespread mortality

and economic losses (Caron et al. 2014, Klaassen and Wille 2023). Recent epizootics have also caused deaths in a wide range of wildlife including, but not limited to, mustelids, felids, ursids, raptors, and candids (Plaza et al. 2024). The spread of IAV from birds to dairy cattle in 2024, and the presence of virions in milk, has led to significant concerns about spillover into human populations as well (Airey and Short 2024). Early detection of new subtypes of the virus and predictions of where it may spread are important for managing spillover risk to sensitive host populations in the future.

Chapter 2: Waterfowl migration drives spatio-temporal distribution of highly pathogenic avian influenza prevalence at large spatio-temporal scales.

2.1 INTRODUCTION

One of the primary goals of ecology is to understand what mechanisms generate spatio-temporal patterns at different scales, and how these scales interact (Levin 1992). In the field of disease ecology, much of the focus historically has been on drivers of local transmission and outbreaks in small-scale populations. Data are easier to collect at this scale, modeling is more straightforward when space is homogeneous, and disease interventions are often focused on halting local spread (Ostfeld et al. 2005, Meentemeyer et al. 2012). However, there has been growing interest in how transmission processes in spatially structured populations scale up to explain large-scale distributions of pathogens (Grenfell et al. 2001, Farnsworth et al. 2006, Cross et al. 2013, Nalar et al. 2014, Gorsich et al. 2022). Are these distributions the consequence of factors that influence local transmission or are they better explained by large-scale spatial phenomena like the movement of infected hosts?

Many widely distributed human and zoonotic pathogens such as SARS-CoV-2, influenza A virus, and West Nile virus infect hosts that can quickly move long distances. Incorporating migration into models of outbreak dynamics can dramatically impact predictions of pathogen spread (Altizer et al. 2011, Dougherty et al. 2018). Understanding whether local or regional-scale processes like migration better predict large-scale pathogen distributions helps us to understand how global environmental and anthropogenic changes will shift these distributions over time (Stephens et al. 2016). Forecasting changes in pathogen distributions is essential for designing

pathogen surveillance programs and disease interventions, allowing managers to prioritize sample locations and resources (Becker et al. 2019).

The avian influenza system is well suited to compare the contribution of local and regional processes to the large-scale distribution of a pathogen. Influenza A virus (IAV) has been found in wild waterfowl on nearly every continent and is hypothesized to move quickly into new locations during host migrations. It is also a pathogen of management concern because it frequently spills over into new host populations. Avian influenza viruses originally hosted by waterfowl are now endemic in a wide range of taxa including harbor seals, pigs, dogs, and horses, and are a viable source of virus that could evolve to cause the next human flu pandemic (Webster et al. 1992, Short et al. 2015). Certain subtypes of the virus (H5 and H7) regularly spillover into domestic poultry and wildlife where they can evolve into highly pathogenic strains that cause widespread mortality (Caron et al. 2014, Klaassen and Wille 2023). In 2024 the virus moved from poultry to dairy cattle and caused several human cases in farm workers (Venkatesan 2024). Spillover events have had an extremely high economic cost as well. In the last decade alone, outbreaks of waterfowl origin IAV in the United States have cost over a billion dollars to poultry producers and government agencies (Seeger et al. 2021).

To predict where these spillover events are most likely to occur in the future, we need to know how prevalent infections are on the landscape and whether local or regional phenomena drive the spatial distribution of prevalence. To address this need, we characterized the large-scale spatio-temporal patterns of IAV prevalence in US waterfowl and tested two hypotheses about the mechanisms that drive the patterns we observed. To test the *local environmental hypothesis,* we looked for patterns of prevalence at the local scale where features of the environment (e.g. water temperature, pH, salinity) likely mediate the persistence and subsequent transmission of the virus

(Garamszegi and Moller 2007, Gaidet et al. 2012, van Dijk et al. 2014). In wild waterfowl IAV transmission is primarily environmental. The virus is shed into aquatic habitats in feces and taken up by hosts while feeding. Locations with high prevalence of IAV in waterfowl may have a larger environmental reservoir of virus resulting in higher rates of local transmission, higher prevalence, and a greater risk of spillover to other species.

To test our second hypothesis, the *regional migratory hypothesis*, we looked for regional patterns of prevalence associated with host migration. IAV strains have more genetic similarities within migratory flyways suggesting the influence of large-scale host movements on the spatial distribution of some viral subtypes (Lam et al. 2012, Bahl 2013). Additionally, IAV prevalence follows seasonal patterns that appear to correspond with the life history and movement patterns of hosts. In waterfowl, prevalence of the virus increases rapidly on the summer breeding grounds prior to the southward fall migration when the numbers of young birds, which are more susceptible to infection by IAV, reach high densities (Garamszegi and Moller 2007, Farnsworth et al. 2012, Hill et al. 2012). Prevalence then decreases on the wintering grounds and generally remains low as hosts migrate north in the spring. If movement of infected birds is the main driver of spatio-temporal patterns, we expect to see prevalence spatially correlated based on where host movements occur.

While the local environment and regional migration may both play a role, we are interested in understanding their relative importance in predicting the large-scale patterns we observe. To accomplish this, we used a spatially explicit conditional autoregressive (CAR) model and model selection approach to compare the local environmental hypothesis and regional migratory hypothesis, as well as a null model in which we assumed that prevalence is not spatially structured. Within disease ecology, CAR models are most commonly used for spatial

smoothing in disease mapping, treating spatial autocorrelation as a nuisance parameter that must be accounted for to ensure realistic predictions. However, we leverage the ability of this approach to explicitly define the spatial structure we expect for each of our spatial hypotheses and determine which model best fits the observed data. This application of CAR models is an often-overlooked advantage to the approach that is rarely leveraged but highly recommended (Farnsworth et al. 2012, Ver Hoef et al. 2017). Understanding whether local environmental factors or regional host movements most impact IAV prevalence will help us predict how large-scale influenza disease risk will change in the future and allow more targeted sampling and interventions. Our results are also a step towards quantifying how processes at different scales contribute to the spread of pathogens into new hosts and territories.

2.2 METHODS

We compared a *null model* that does not include any spatial structure to three different zero-inflated spatial conditional auto-regressive (CAR) models, each representing one of our biological hypothesis about how prevalence values were correlated in space. The *local model* tested the local environmental hypothesis where prevalence values that are close in geographic space are expected to have more similar values of prevalence than those that are geographically distant. If local persistence is the main driver of spatial distribution, we expect that IAV prevalence will be positively spatially autocorrelated since factors like vegetation type, latitude, abundance of water and host habitat, and climate all tend to be more similar in areas that are closer in geographic space.

We used two models, the *unweighted regional model* and the *weighted regional model*, to test the regional migratory hypothesis. The unweighted model represents the idea that it is the migratory connections between watersheds that matter the most, not the volume of hosts that

move between them. The weighted version represents the idea that watersheds that are highly connected by high levels of host migration have more similar prevalence values than those that are weakly connected.

The null model and three spatial CAR models shared the same zero-inflated model structure (Figure 2.1). We used viral surveillance data to make inference on two key parameters; the probability that IAV occupied a given location (ρ), and the true prevalence at that location (λ). We incorporated imperfect detection and spatial structure into the prevalence portion of the model (Figure 2.1). This format allowed us to estimate prevalence at only those times and locations where viral habitat is suitable enough for the virus to be present and allowed us to determine the impact of temperature and host abundance on the probability an area was occupied by the virus (Figure 2.1, Viral Occupancy). Viral occupancy requires that hosts be present and the temperature of the environment be cool enough for environmental transmission to occur. Salinity and pH of water have also been shown to be important for persistence of IAV in water but were not included because available data show these values are generally within published tolerances. The detection part of the model allowed us to incorporate information about imperfect detection of the virus so that we could estimate the true prevalence from apparent prevalence (p). Apparent prevalence is the number of positive tests divided by the overall sample size (Figure 2.1, Detection).

We incorporated our hypotheses about the impact of spatial structure on true prevalence via a spatially structured random effect (φ) (Figure 2.1, Spatial Structure). In a CAR model, spatial information is included in an adjacency matrix that encodes the spatial structure we expect to find under each of the alternate hypotheses. This adjacency matrix is incorporated into the precision matrix ($\boldsymbol{Q^{-1}}$) of the prior distribution of the spatially structured random effect φ.

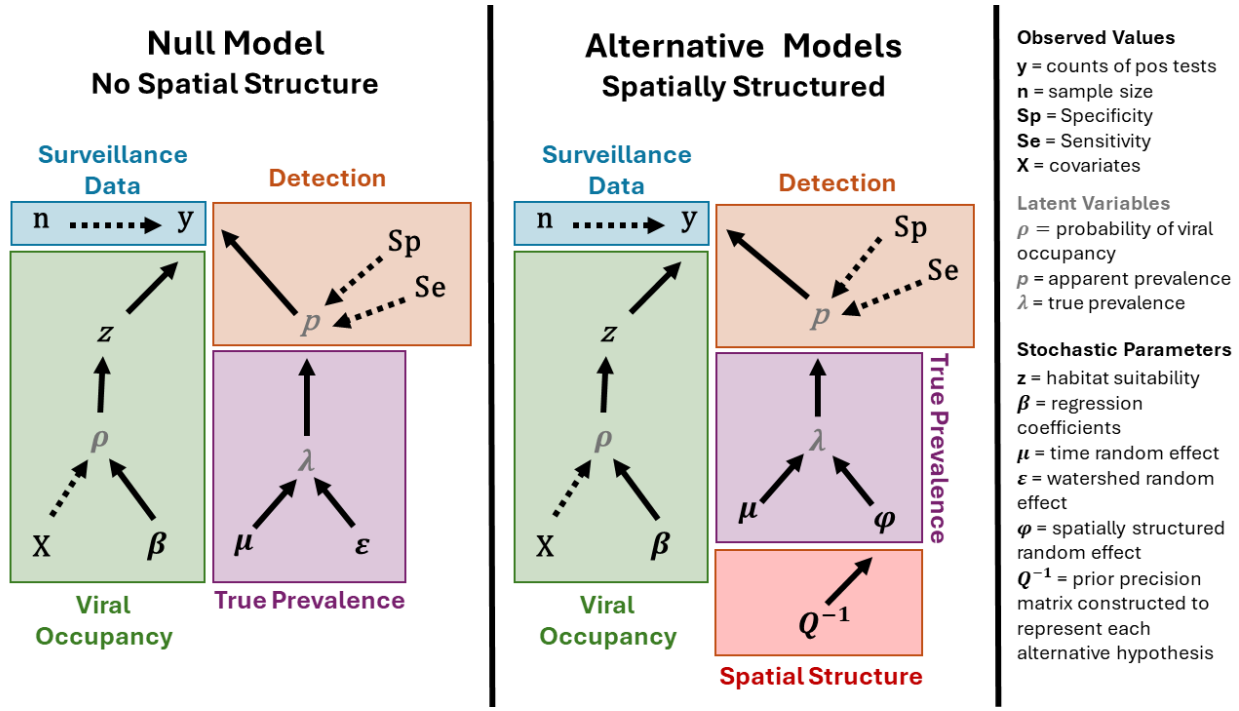

Figure 2.1: Directed acyclic graphs modified to show the basic model structure and the biological interpretation of the model components. Data and parameters directly estimated by the models are shown in black. Latent variables are shown in grey. Non-italic letters and dashed lines indicate known quantities and data.

2.2.1 Viral Data

We used viral surveillance data of potentially highly pathogenic subtypes of IAV in Mallard ducks (*Anas platyrhynchos*) within the continental United States collected by the US Department of Agriculture Animal and Health Inspection Services (USDA APHIS). Mallards were chosen as the target testing group because they have high prevalence of many subtypes of IAV, and for their abundance and widespread geographical distribution in the continental United States (Bevins et al., 2014). Samples were collected from live waterfowl and hunter harvested birds from 2007 - 2016. From 2007 – 2013 specific areas of interest for sampling priority were identified on a state level based on historic prevalence of IAV and the high numbers of waterfowl band recoveries (Farnsworth et al. 2011). From 2014 – 2016 sampling locations were prioritized with a relative ranking approach based on the presence of significant historical IAV

clusters, locations with high rates of between-flyway mixing, and temperature in accordance with the Surveillance Plan for Highly Pathogenic Avian Influenza in Wild Migratory Birds in the United States (2015). Though 43% of samples were taken in fall, this sample design still resulted in good spatio-temporal coverage of the US.

Each individual sample included an oropharyngeal and a cloacal swab that were tested for the presence of IAV and subtyped using rRT-PCR at the National Animal Health Laboratory Network. Samples that were H5 or H7 positive were sent to the National Veterinary Services Laboratories to determine pathogenicity. A detailed account of sampling and laboratory methods can be found in Bevins et al. (2014). The dataset used for model fitting was composed of 74,448 samples from across the US (15,027 positive; 59,421 negative). Data collected in 2016 was withheld for out of sample validation of the top model (4147 positive; 22,477 negative).

2.2.2 Temporal and Spatial Scale

True seasonal prevalence was estimated across five years, from 2007-2010 and 2015. Very few samples were collected in 2011-2014 resulting in poor seasonal and spatial coverage so these years were not included in model fitting. We defined biological seasons based on the life history of Mallard ducks. We chose to use these biological seasons because they align with previously observed annual shifts in viral prevalence and partially account for the effects of host age on susceptibility to IAV. Biological seasons were classified as spring migration (March - May), breeding season (Jun - Aug), fall migration (Sep – Nov), and overwintering season (Dec – Feb). Because the overwintering season spans two calendar years, we used biological years that run from March – Feb. For example, the biological year of 2012 spans March 2012 – February 2013.

Seasonal prevalences were estimated for 202 individual sub-regional watersheds in the continental US. Subregional watersheds, also referred to by their hydrologic unit code HUC4, are delineated by the USGS based on surface hydrologic features. We chose the watershed scale because it represents a discrete biological unit that is pertinent to IAV hosts and habitat as opposed to political boundaries. It is also sufficiently large to include enough samples for inference, but not so large as to ignore important spatial heterogeneity.

2.2.3 Migratory Networks

A migratory host network was constructed using publicly available bird band-recovery data for Mallards collected by the Bird Banding Lab and the Patuxent Wildlife Research Center, US Geological Survey (USGS 2017, retrieved Feb 2018). Live birds were banded and then the bands were recovered later through hunter recovery. We removed the few occurrences that were due to visual observation or live capture to improve the consistency of the dataset. We defined the migratory path of the individual as the difference in location from where the bird was banded to where the band or bird was recovered. Because individual birds did not always follow the same migratory paths every year, we only used observations that were recovered within one year or less of banding to ensure that watersheds without true migratory connections were not connected in the network. The network included all such observations from 2006-2016.

2.2.4 Viral Occupancy Covariates

Mean land surface temperature for each watershed, in each season, was calculated from monthly means of 2-meter gridded land surface temperature downloaded from NOAA Physical Sciences Laboratory from their website at https://psl/noaa.gov (Fan and van den Dool, 2008, retrieved Nov 2018). This aggregate value characterizes the general suitability of the climate that

may impact the environmental reservoir of the virus through time (Cassie 2006). In the lab, IAV can remain infective for months in waters below 17 C (Brown et al. 2009).

The total number of Mallards present in each watershed, in each season, was calculated from the sum of all observations made during the sample period in the ebird Basic Dataset (Cornell Lab of Ornithology, retrieved Jul 2022). Only observations that included a date and latitude and longitudinal coordinates were used so counts could be aligned with the surveillance data in time and space. To account for broad seasonal distributional shifts of birds as they migrate north or south, we also used the latitude of the centroid of each watershed as a covariate.

2.2.5 Model Structure and Hypotheses

We modeled counts of positive tests (y_{ijk}) for each season (i), each year (j), and in each watershed (k). We used a zero-inflated Bayesian model where counts were zero when habitat was not suitable ($z_{ijk} = 0$) and binomially distributed with sample size n_{ijk} and apparent prevalence p_{ijk} when habitat was suitable ($z_{ijk} = 1$) (Line 1, Null and Alternate Models). Habitat suitability was a Bernoulli random variable with probability of viral occupancy, ρ_{ijk}, and was influenced by latitude of the watershed, the abundance of the host, and the mean land surface temperature of the watershed. We estimated the impact of these covariates on viral occupancy in each watershed for each season, except for latitude, which was estimated for each season only. Latitude was included to determine if IAV occupancy shows broad latitudinal shifts throughout the yearly cycle.

Observed apparent prevalence, p_{ijk}, arises from the true prevalence of a watershed (λ_{ijk}). True prevalence is impacted by diagnostic test sensitivity and specificity as well as errors in sampling that result in false positives and negatives. We set sensitivity (Se) equal to 0.863 and specificity (Sp) equal to 0.984 based on values reported in the literature for domestic poultry and

expert opinion (USDA APHIS, 2017; Janice Pedersen, personal communication; Mia Torchetti, personal communication). We tried a version of the model that estimated Se and Sp but checking the model with simulated data revealed that the true values of these parameters could not be recovered. Instead, we chose to fix these values as constants.

2.2.6 Null Model

The null model had no spatial structure to which we compared the alternative models with our various spatially structured hypotheses. True prevalence (λ_{ijk}) was influenced by a non-spatial random effect (μ_{ij}), which was the average prevalence for each year and season across the entire US, and a spatially varying random effect that did not include any spatial correlation between watersheds (ε_k).

Null Model: No Spatial Structure

1) $y_{ijk} \sim \begin{cases} 0 & when\ z = 0 \\ Binomial(n_{ijk}, p_{ijk}) & when\ z = 1 \end{cases}$
2) $p_{ijk} = Se \times \lambda_{ijk} + (1 - Sp)(1 - \lambda_{ijk})$
3) $logit(\lambda_{ijk}) = \mu_{ij} + \varepsilon_k$
4) $z_{ijk} \sim Bernoulli(\rho_{ijk})$
5) $logit(\rho_{ijk}) = \beta_i \times Latitude_k + \beta_{1,ik} \times Host\ Abunance_{ijk} + \beta_{2,ik} \times Temp_{ijk} + \beta_{0,ik}$

2.2.7 Alternative Models

We used three different spatially structured CAR models to estimate the true prevalence (λ_{ijk}) of the virus under each of our alternative hypotheses. These models had the same general form as the null except that they incorporated explicit spatial dependency between watersheds with the addition of the spatially structured random variable φ_k. Besag (1974) showed that the

joint prior distribution of φ_k is a multivariate normal with mean = 0 and precision matrix **Q** (Alternative Models, Line 6).

Alternative Models: Spatial CAR Model

1) $y_{ijk} \sim \begin{cases} 0 & when\ z = 0 \\ Binomial(n_{ijk}, p_{ijk}) & when\ z = 1 \end{cases}$

2) $p_{ijk} = Se \times \lambda_{ijk} + (1 - Sp)(1 - \lambda_{ijk})$

3) $logit(\lambda_{ijk}) = \mu_{ij} + \varphi_k$

4) $z_{ijk} \sim Bernoulli(\rho_{ijk})$

5) $logit(\rho_{ijk}) = \beta_i \times Latitude_k + \beta_{1,ik} \times Host\ Abunance_{ijk} + \beta_{2,ik} \times Temp_{ijk} + \beta_{0,ik}$

Spatial Prior:

6) $\varphi_k \sim MVN(0, \boldsymbol{Q}^{-1})$, where $\boldsymbol{Q} = [(\boldsymbol{D} \times \tau)(\boldsymbol{I} - \alpha \boldsymbol{B})]$

Q was constructed from the adjacency matrix, **W**, which was populated with values that reflected the correlation structure that is expected under each of the hypotheses. **W** was an *m x m* adjacency matrix where, *m* was the number of watersheds, *{i,i}* entries were zero and the off-diagonals were the hypothesized weights. For the local model, watersheds *i* and *j* were assigned a one if they shared any portion of their boundary, and a zero otherwise.

For the regional models, we defined **W** in two different ways. We constructed migratory spatial networks from the banding data where nodes were watersheds and edges were defined in different ways to represent two variations on the regional migratory hypothesis. The first adjacency matrix was unweighted where one indicated that birds moved between each watershed pair and zero indicated no movement. The second adjacency matrix was weighted by the number of birds that moved between watersheds and were rescaled to ensure that they were bound by zero and one. Both migratory networks omitted loops, connections where the origin and

destination watershed were the same since we were only interested in between watershed movements. Both networks were undirected because the migratory network did not vary by season or year. Hosts generally fly south in fall and north in spring. A single directed network that included movement in both directions would obscure the signal of host movement when not aligned with the corresponding season. Additionally, the CAR model prior specification required that the adjacency matrix be symmetrical.

W was converted into the scaled adjacency matrix, **B**, by multiplying it by the inverse of **D**, an *m x m* diagonal matrix where *{i,i}* were the number of neighbors of watershed i and the off diagonals were zero. **B** was multiplied by an autocorrelation parameter, α, that controlled the strength of spatial correlation by weighting its influence in the prior of the precision of φ_k. To approximate an intrinsic conditional autoregressive (iCAR) model, which assumes complete spatial correlation, we set $\alpha = 0.999$. This made it possible to use a joint prior distribution in the model instead of the full conditional. Approximating the iCAR allowed us to estimate one fewer parameters, reduced computation time, and ensured that the precision matrix of the joint distribution of φ_k was positive definite.

We tried a version of the CAR models that also included a non-structured spatial random effect but found this resulted in model overfitting and the parameter was omitted. Uninformative priors were used for the remainder of model parameters following the recommendations of Gelman (1996). Prior distributions are described in the Chapter 2 Supplemental Material. We performed several checks to ensure that information about host migration was not also contained in the local model adjacency matrix and found that the local and regional **W** matrices encoded different spatial phenomena (Chapter 2 Supplemental Material).

2.2.8 Model Fitting and Validation

All analyses were conducted in R (R Core Team, 2021) and Bayesian model fitting was done using package rjags (Plummer, 2003). We used a Monte Carlo Markov Chain (MCMC) procedure of three chains for 100,000 iterations after a burn in of 100,000 iterations. Convergence was confirmed using the Gelman and Rubin's convergence diagnostic from the coda package to ensure that at least 97% of the scale reduction factors for each parameter and their confidence intervals were below 1.1 (Brooks and Gelman 1998). The deviance information criterion (DIC) was used for model selection where the model with the lowest DIC indicates the best fit (Spiegelhalter et al. 2002). Posterior distributions of prevalence and viral occupancy from the top model were each combined across years. This resulted in a posterior distribution for each season and watershed for prevalence and occupancy respectively.

Out of sample validation was used to determine how well the top model predicted the 2016 apparent prevalence data withheld from model fitting. Leave one out cross validation was not performed because CAR models have long computation times, making fitting the model to multiple subsets of the data intractable. However, this is not problematic as out of sample validation is the preferred method. Predicted apparent prevalences were calculated from the posterior distributions of true prevalence. To compare the model predictions to observed apparent prevalences in 2016 we calculated the absolute value of the residuals between model predictions and observed data. Because we are mostly interested in our model's ability to correctly predict high or low values of prevalence, we also performed a categorical validation where an apparent prevalence of 6% was selected as a threshold value based on the median value of observed apparent prevalence in the withheld dataset and values reported in the literature (Olsen et al., 2006). Watersheds with an apparent prevalence of 6% or greater were classified as having high apparent prevalence of IAV and watersheds with less than 6% as having low

apparent prevalence. We also calculated the area under the receiver operating characteristic curve (AUC) which quantifies the ability of the model to correctly estimate high or low prevalence. AUC ranges in value from 0 to 1. An AUC of 0 indicates that the model predicted high prevalence as low and low prevalence as high. An AUC of 0.5 indicates that the model could not distinguish between high and low prevalence, and a value of 1 indicates that the model correctly classified prevalence as high or low for every observation. Because only two samples were taken in the spring migration season of 2016, this season was omitted from model validation.

2.3 RESULTS

2.3.1 Model Selection and Validation

The unweighted regional model had the lowest DIC indicating that it best fit the data (Table 2.1). Thus, all parameter estimate results discussed here were from this top model. Both the unweighted and weighted regional models outperformed the local model, and all three of these spatial models were a better fit to the data than the non-spatially structured null. The unweighted regional model correctly predicted high (≥ 6%) and low (< 6%) apparent prevalence for 81% and 58% of all observations from the withheld 2016 data respectively (Figure 2.2, AUC = 0.71). Quantitatively, the model underpredicts apparent prevalence with individual watershed-level predictions not exceeding 40% prevalence compared to the maximum observed value of 90% in 2016. Particularly high observed apparent prevalences generally occurred in watersheds where fewer samples were taken.

Table 2.1: Deviance information criterion (DIC) for each model of IAV true prevalence and viral occupancy, as well as the hypothesis that each model represents. The model with the best fit has the lowest DIC.

Model	Hypothesis	DIC
Unweighted Regional Model	Regional Migration	13983
Weighted Regional Model	Regional Migration	14003
Local Model	Local Environmental	14010
Null Model	No Spatial Structure	14206

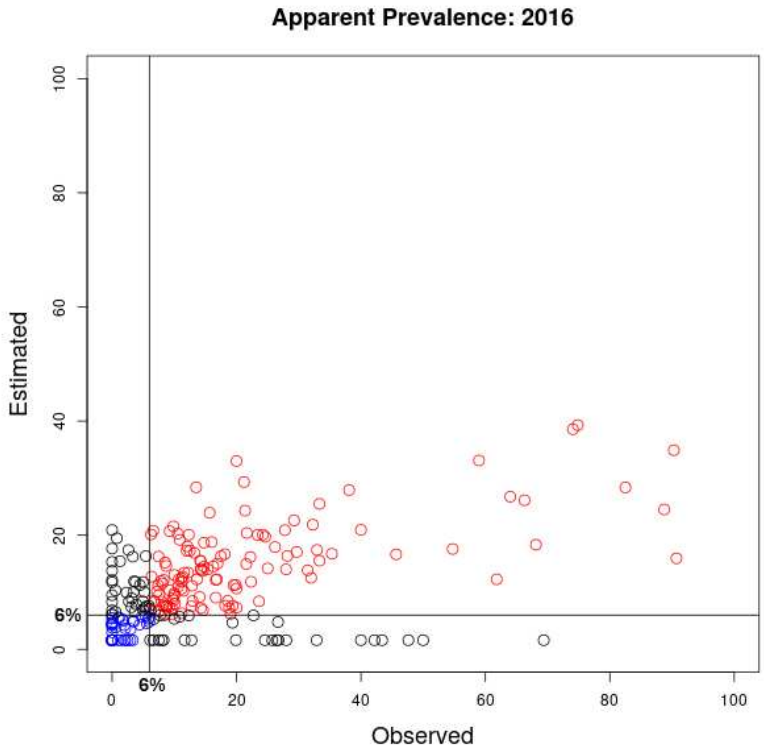

Figure 2.2: Estimated apparent prevalence of IAV from all seasons in 2016 compared to observed apparent prevalence calculated from the top model estimates of true prevalence. The black lines intersecting the x and y axis indicate the demarcation between high and low prevalence; 6%. Red circles indicate values correctly classified as high prevalence by the model and blue circles indicate values correctly classified as low prevalence by the model.

2.3.2 Spatio-temporal Patterns of True Prevalence

Most watersheds had low estimates of true prevalence in the spring migration season. The highest prevalence watersheds in this season were located in the southeastern US with a

maximum prevalence of 30% (Figure 2.3A). Uncertainty in our estimates was greatest in spring when the fewest number of samples were taken. In the breeding season, the maximum prevalence was almost double what was seen in spring and shifted to watersheds in the north (Figure 2.3B). True prevalence remained high in the northern latitudes in the fall migration season, but with notable shifts southward along the Mississippi Valley and the Intermountain West, a region bounded by the Rocky Mountains in the east and the Cascade and Sierra Mountain ranges in the west (Figure 2.3C). Maximum prevalence was lowest in the overwintering season at 25%, and the number of moderate and high prevalence watersheds diminished noticeably from the fall migration season (Figure 2.3D). The desert southwest consistently had the lowest prevalence in every season, but also the greatest uncertainty due to the lack of sampling in this region.

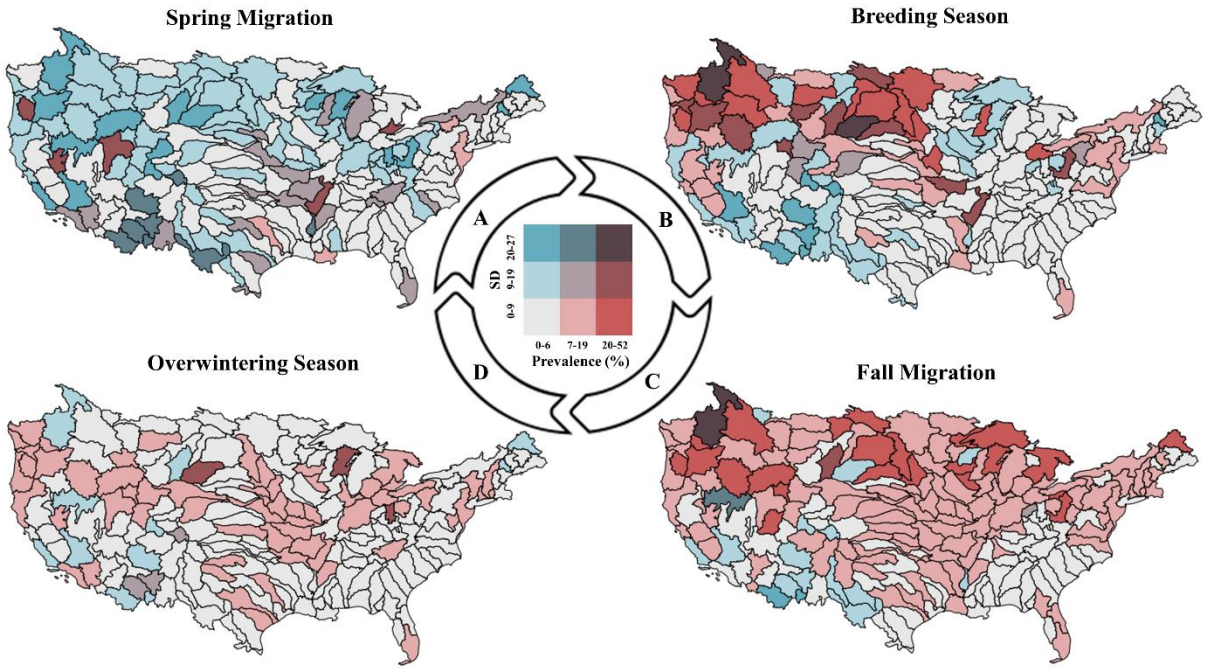

Figure 2.3: Maps showing estimated seasonal true prevalence of IAV, expressed as a percentage, and standard deviation for each watershed in the contiguous US. Prevalence estimates are the medians of posterior distributions from the top model combined across years.

Arrows show the ordering of seasons in a biological year and highlight the cyclical nature of IAV prevalence in wild waterfowl in North America.

2.3.3 Viral Occupancy

Viral occupancy probability had similar spatio-temporal patterns as true prevalence but greater uncertainty in the model estimates. Latitude had a strong positive relationship with viral occupancy during the summer breeding season (β = 2.06) and fall migration season (β = 2.15), when more virus was detected in northern latitudes, and a negative relationship in the overwintering season (β = -1.20) and spring migration season (β = -2.08) when birds migrated south to lower latitudes. The relationships of Mallard host abundance and mean land surface temperature with viral occupancy were less clear. Viral occupancy did not show a strong relationship with host abundance for almost all watersheds, with 94% of 95% equal-tailed credible intervals (CI) for the regression coefficient overlapping zero. For those few watersheds where the CI did not overlap zero, there was a positive effect of host abundance on viral occupancy. This positive effect was only observed in the fall migration and overwintering seasons on the west coast of the US and the Mississippi River Valley (Figure 2.4A). The fall migration season was the only time in which there was a clear effect of temperature on viral occupancy, though the uncertainty of these estimates was high (Figure 2.4B). There was a clear spatial pattern to how temperature impacted viral occupancy with a positive effect of temperature in the southeastern US that transitioned to an increasingly negative effect to the northwest.

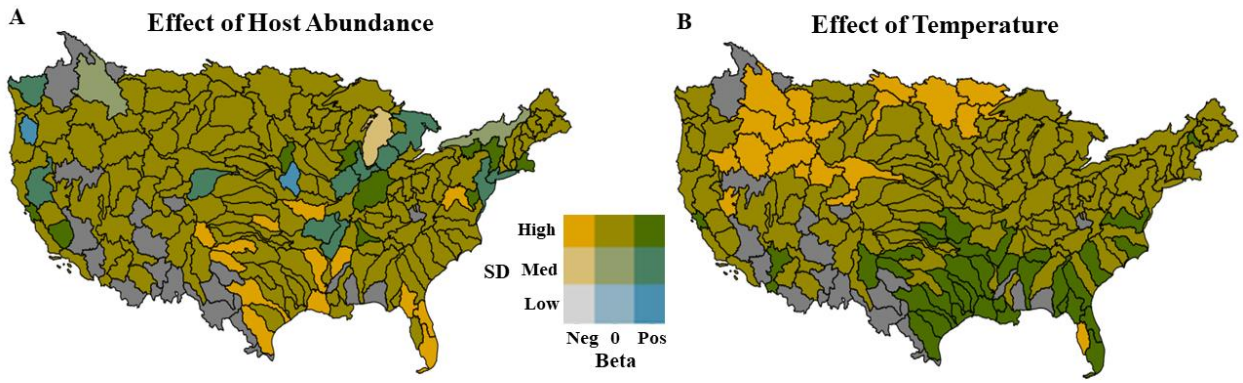

Figure 2.4: Maps showing the influence of Mallard host abundance (A) and mean land surface temperature (B) on IAV occupancy probability in each watershed in the contiguous US. Only the fall migration season is shown because this is the only season where a strong effect was observed. Beta is the fitted regression coefficient of each covariate in the top model on the log likelihood scale, categorized to show if there was a negative ($-2 \leq \beta_{fall,k} <$ -0.05), positive ($0.05 < \beta_{fall,k} \leq 2.7$), or no influence ($-0.05 \leq \beta_{fall,k} \leq 0.05$) of the covariate on occupancy. Standard deviation of these quantities was classified as low ($1.4 \leq$ SD ≤ 1.7), medium ($1.8 \leq$ SD ≤ 2.1) , or high ($2.2 \leq$ SD ≤ 2.5). Dark grey polygons were not sampled during the entire study period.

2.4 DISCUSSION

Incorporating seasonal host migration into our model allowed us to accurately predict spatial and seasonal patterns of high IAV prevalence. Model estimates of high prevalence watersheds closely matched predictions of regions suitable for the virus and areas where IAV outbreaks occurred during the study period (Belkhiria et al. 2016). Observed apparent prevalences in sparsely sampled watersheds were greater than model predictions for the validation dataset. This is a common issue with raw estimates of prevalence that are very sensitive to low sample sizes. When sample effort is not adequate to capture negative cases, apparent prevalence can be biased to larger values (Jovani and Tella 2006). Our qualitative model assessment allowed us to evaluate model performance despite the resulting inflation of the residuals for poorly sampled watersheds.

Our maps of estimated true prevalence were consistent with previously observed seasonal distributional shifts of IAV, and closely followed the distributional shifts of migratory hosts (Hill et al. 2012). The highest prevalences occurred in the breeding season in areas known to be important breeding habitat for migratory waterfowl, including the Pacific Northwest and the Prairie Pothole Region in North and South Dakota. During the fall migration season an additional cluster appeared in the northern latitudes of the Mississippi River Valley and shifted southward in winter (Figure 2.3C, 2.3D). Similar spatial patterns were observed in genetic relatedness among IAV strains (Lam et al. 2011, Fourment et al. 2017). IAV genetic lineages are generally constrained within migratory flyways, regions that characterize large-scale seasonal movements of waterfowl. Watersheds connected by host movement likely share the same subtypes and strains of IAV with similar levels of infectivity and transmissibility that in turn generate similar levels of prevalence.

Movements of wild birds are often overlooked in studies characterizing the spatio-temporal patterns of HPAI outbreaks (Gilbert and Pfeiffer 2012). In the only other study that considered the role of host movement, as well as other local factors thought to influence IAV prevalence, Gorsich et al. (2022) analyzed the same IAV surveillance dataset using statistics derived from a migration network as model covariates. They found that individual demographic and local-scale processes, not host movement, most influenced the spatial distribution of IAV. Their analysis, however, took place on a smaller spatial and temporal scale than our own, modeling monthly individual cases. Our contrasting results highlight the importance of matching the spatial and temporal scale of an analysis to the desired scale of inference. Conclusions can change dramatically with the scale considered, though each can be valid (Wiens 1989). Ours is the largest scale yet considered and the only study to use host movement data to explicitly define

spatial structure, making our conclusions appropriate for inference at the scale of the contiguous United States. Ours is also the first study to directly demonstrate a link between host movements and IAV prevalence indicating that regional-scale phenomena can shape large-scale distributions of pathogens.

Studies of other pathogens have shown that large-scale abiotic factors like climate, altitude and latitude are better predictors of the large-scale distribution and prevalence of than local biotic factors like host richness (Cohen et al. 2016). Abiotic and biotic are often used as shorthand for large and small spatial scales respectively. Host migration, however, blurs the line between these two designations. Migration is caused by large-scale seasonal climatic changes and spans distant latitudes, but ultimately results in local changes in community composition and host density. While our results are consistent with the idea that large-scale processes are most predictive of large-scale patterns, the assertion that these are due to biotic or abiotic factors is less certain.

The scale of our analysis also influences how we interpreted the impact of temperature on viral occupancy probability. While we expected that watersheds with high average temperatures would be unsuitable for the virus, spatial and seasonal patterns of occupancy were often different from what was predicted based on lab experiments of viral persistence in water. For example, despite high summer and autumn temperatures in Florida, the virus was still found in this region suggesting that within large-scale watersheds there are still suitable habitats that allow environmental transmission of the virus. At large scales, mean seasonal temperature likely does a better job predicting where hosts are, rather than the suitability of aquatic habitats for the virus. Temperature is a powerful predictor of migratory host movements because it impacts the availability of food resources and the energetic costs of thermoregulation (Si et al. 2015).

Waterfowl prefer the warmer temperatures of southern latitudes in winter and the cooler temperatures of northern latitudes in the summer. Our model estimates of viral occupancy reflected this shift in host habitat preference and is consistent with previous results showing that seasonal changes in host distributions are likely important for explaining the large-scale distribution of IAV.

Despite the clear importance of host movement in predicting viral prevalence, we saw no impact of host abundance on viral occupancy in most watersheds. However, in watersheds that hosted particularly large numbers of Mallards in the fall and winter, we did observe a positive influence. Mallards are widespread and abundant in the contiguous United States, so it may have taken particularly large numbers of these hosts to impact occupancy probability. It is also possible that the presence of other host species, acting as either a source or sink for IAV infection, may complicate the observed relationship, decreasing the impact we observed from Mallards alone.

Though regional host migratory movements were the most important driver of large-scale spatial prevalence, local viral persistence likely still plays a role. Local spatial variation in water temperature and habitat can allow the virus to transmit in areas where we predict it should not be at the regional scale. Aggregating covariates over space reduces variation and can result in underestimating the importance of spatial covariates like temperature (Cross et al. 2013). Collecting these data at sites where the virus is actually sampled may reveal a stronger impact of the environment, even when aggregated by watershed.

Unevenness in sampling effort in time and space can also limit our ability to accurately predict prevalence for all watersheds in the US, especially during the spring migration season when IAV sampling was the lowest. Low sample size is not an uncommon problem. It is

intractable to sample all regions, at all time periods, when making inference on such a large area. One benefit of a CAR model over a more traditional regression is that incorporating spatial structure allows us make prevalence predictions for regions that have never been sampled. Watersheds with estimates of high prevalence and high uncertainty could be good future targets for sampling, when combined with expert knowledge of the area. It will also be important to sample other subtypes of the virus. Existing surveillance has typically focused on potentially pathogenic subtypes of IAV but spatial patterns of prevalence can vary greatly between host species and viral strains (Hill et al. 2012). Expanding our analysis framework to other viral subtypes and additional host species will help us understand how generalizable our results are to IAV as a whole.

In the future, effective management of globally distributed pathogens will necessitate surveillance efforts and interventions that explicitly consider the spatial scale of inference. We must also understand what mechanisms drive the large-scale movement and distributions of these pathogens. We have shown that predicting high IAV prevalence across the US, and consequently identifying regions with higher spillover risk, requires an understanding of where hosts move. In the future, as hosts of pathogens with large geographic ranges change their movement patterns in response to habitat loss and climate change, we can expect regional spatio-temporal patterns of disease prevalence and risk to shift with them.

Chapter 3: Large-scale wild waterfowl movements drive spatial transmission of H5 and H7 influenza A virus in the United States.

3.1 INTRODUCTION

Though we have more data on the incidence of infectious disease than at any other time in history, large-scale spatial transmission of pathogens remains largely unpredictable. Determining what mechanisms drive the movements of pathogens across the globe is complicated by processes operating at different ecological scales including, but not limited to, environmental reservoirs of pathogen, host density and demography, and host mobility (Hesterbeek et al. 2015). Understanding the relative contribution of theses mechanisms to observed patterns of spatial transmission, and the scales at which they are important, is a crucial first step toward predicting the spread of disease and the risk of pathogen spillover into new host populations (Becker et al. 2019). Despite this, models of large-scale spatial transmission rarely include predictors at multiple spatial scales.

Recent epizootics of avian flu in domestic poultry have highlighted an urgent need to understand the spatial transmission of influenza A virus (IAV) among wild waterfowl (Russel 2016). These reservoir hosts of IAV are thought to drive much of its genetic diversification leading to spill over into other host species (Trovao et al. 2015, Yang et al. 2024). Outbreaks of H5N8, H5N5, H5N6 and H5N1 subtypes of IAV in poultry have all been linked to genetic reassortment events in wild waterfowl (Beerens et al. 2016, Hill et al. 2019, Lycett et al. 2020). Recent spillover from birds to mammals that frequently interact with humans such as racoons, dogs, cats, and dairy cattle has led to significant concern that wild birds could be the source of the next outbreak of pandemic flu in humans (Airey and Short 2024). This concern, as well as the severe economic cost of outbreaks in poultry have led to over a decade of IAV surveillance

in wild birds in the United States (US). Improving our understanding of what mechanisms drive spatial transmission among reservoir hosts, and at what spatial scale they are important, will help managers prioritize sampling locations for early detection of new potentially pathogenic subtypes of the virus.

To contribute to the broader spatial questions within disease ecology and address practical concerns of IAV surveillance, we investigated how local-scale transmission and regional-scale host movements influence the large-scale movement of wild waterfowl IAV. Locations with only a few cases of IAV have fewer opportunities for infection of hosts that subsequently move the virus to new locations. Thus, conditions that mediate local transmission may also mediate large-scale spatial transmission. Among wild waterfowl, IAV host-to-host transmission is impacted by many factors including conditions of the environment (Breban et al. 2009), host demography (van Dijk et al. 2013), and host abundance and density (Kent et al. 2022). However, local-scale predictive covariates have not yet been considered in analyses of spatial transmission. To address this gap, we evaluated how two factors related to local transmission impacted large-scale spatial transmission.

First, we considered the local temperature. In wild birds, IAV is primarily environmentally transmitted (van Dijik et al 2018, Ramey et al 2020). The virus is shed into aquatic habitats in host feces and transmitted to other hosts when they feed in contaminated water. Environmental conditions such as temperature must be suitable enough for the virus to survive and transmit to other hosts (Brown et al. 2009). Because laboratory experiments have shown that IAV persists longer in water below 17 C (Brown et al. 2009) we predicted that locations with lower temperatures would have more viral movement between them. Second, we considered the prevalence of the virus. High densities of young, immunonaive birds are

associated with higher local prevalence of IAV (van Dijk et al. 2013). Among all age groups, large aggregations of infected hosts shed more virus into the environment increasing the probability of secondary infections. This suggests that local prevalence is an indirect, but reasonable proxy reflecting the host demographic and density impacts on local transmission. We predicted that locations with higher prevalences would have more viral movement between them.

Large-scale host migration is also hypothesized to be an important driver of IAV movement patterns. Previous research on host movements has largely focused on the impact of migratory flyways on the phylogenic structure of IAV. Lam et al. (2012) found that IAV gene flow was greater within North American flyways than between them. In Asia, Tian et al. (2014) showed that within flyways, the timing of IAV outbreaks and host migrations were highly correlated. Trovao et al. (2015) linked the spatial diffusion of an H5N1 outbreak to avian flyways in Asia. Most recently, Yang et al. (2024) found that seasonal migrations were associated with highly pathogenic IAV dispersal in the northern hemisphere. All of these studies use viral genomic data, phylogeographic and phylodynamic approaches to link host migration at the flyway scale to large-scale viral movement.

These studies necessarily focus on flyways because the analyses incorporate large geographic areas and a variety of host taxa, often in regions where more detailed data on host movement is scarce. While taking a flyway-based approach can tell us a great deal about overall viral movement patterns, it lacks a detailed understanding of host movement and hence is less useful for prioritizing where viral surveillance should occur. Migratory flyways are very large-scale generalizations of seasonal waterfowl movement for dozens of bird species. Regions where flyways overlap, such as Alaska, are frequently cited as locations where new viral introductions are more likely to occur but represent enormous areas containing millions of potential hosts. To

predict viral movements at an actionable scale for intervention, such as prioritization of surveillance locations within large areas, we need to know where hosts are specifically moving and quantify how this influences viral movement. To address this need, we shifted away from the phylogeographic/flyway paradigm to see if host movement data could be used to directly predict the IAV spatial transmission network. We expected that locations connected by more host movement would have more spatial transmission between them.

Many terms have been used to describe the spatial movement of pathogens including pathogen dispersal, pathogen diffusion, regional spread, geographic spread, pathogen migration, large-scale transmission, and spatial transmission. We use the term *spatial transmission* because it clarifies that viral spatial movement occurs via at least one, but likely many, host-to-host transmission events. We investigate whether these local transmissions dictate where the virus can move, or if regional movements via an infected host most influence spatial transmission. We developed a zero-inflated Bayesian network model to quantify the role of local-scale environmental temperature, prevalence, and regional-scale host movement in shaping viral movement patterns. This model estimates the probability of spatial transmission between locations in the contiguous US and estimates the probability that we detect spatial transmissions when they occur. Understanding how these mechanisms impact spatial transmission is an important first step in producing predictive models of IAV spread. Our results also contribute to the broader understanding of how phenomena at different spatial scales impact the global spread of infectious disease.

3.2 METHODS

We constructed a spatial IAV transmission network for the contiguous United States using viral genomic data. We used this network to fit a Bayesian zero-inflated network model to

estimate the probability of spatial transmission between watershed pairs. The model estimated the effect of the number of hosts moving between locations, as well as the effect of temperature and prevalence within watersheds, on spatial transmission of the virus. The zero-inflated framework also allowed us to estimate the probability that spatial transmission was detected given that a transmission occurred. We evaluated the role of sample effort and the time between sampling events on our ability to detect spatial transmission.

3.2.1 Spatial Scale

We used sub-regional watersheds as the spatial unit for analysis. These are delineated by the US Geological Survey and are also referred to by their hydrological unit code HUC4. We chose the watershed scale because it represents a discrete biological unit that is pertinent to IAV hosts and habitat. They are sufficiently large to include enough samples for inference at large spatial scales but not so large as to ignore important spatial heterogeneity.

3.2.2 Viral Genomic Data

We used whole and partial genome sequences of potentially pathogenic (H5 and H7) influenza A viruses to make the spatial transmission network. Whole sequences contained genetic data for all eight gene segments for a given viral isolate while partial sequences contained sequences for one or more gene segments. We focused on these subtypes because they have been responsible for most IAV outbreaks in domestic animals. Samples were collected as part of a large-scale IAV surveillance effort implemented by the US Department of Agriculture Animal and Health Inspection Services (USDA APHIS). Sampling locations for surveillance were chosen based on historical incidence of IAV, wetland habitat, past presence of waterfowl, and expert opinion. A total of 58 watersheds were sampled between September 2015 and November 2017. Oropharyngeal and cloacal swabs were collected from live and hunter

harvested waterfowl of 12 different species: Blue Winged Teal, Cinnamon Teal, American Green Winged Teal, American Black Duck, Ring Necked Duck, Mallard, Northern Shoveler, Northern Pintail, Lesser Snow Goose, American Wigeon, Gadwall, and Lesser Scaup. All positive samples were subtyped using rRT-PCR at the National Animal Health Laboratory Network. All H7 and H5 viruses were then sequenced at the National Veterinary Services Laboratory. A detailed account of sampling and laboratory methods can be found in Bevins et al. (2014).

Sequence data for each gene segment were aligned using the MUSCLE algorithm with the R package 'msa' (Bodenhofer et al. 2015). The subtype of the HA and NA gene segments was known so alignment for these gene segments was performed only within subtype. For example, the genetic sequence for an HA gene segment of an H5 viral isolate was not aligned with HA gene segments from H7 viral isolates. The genetic distance between each sequence was calculated as the proportion of nucleotide substitutions between two sequences using the ape R package and an evolutionary model developed by Tamura and Nei (1993) (Paradis and Schliep 2019).

3.2.3 Spatial Transmission Network

A total of 169 viral isolates were used to construct a spatial transmission network for each gene segment. Because IAV can reassort, spatial transmission of entire viral genomes cannot be tracked over time. Instead, we used the seqTrack algorithm to construct viral genealogies for each gene segment separately. SeqTrack differs from traditional phylogenetic approaches by allowing ancestors and descendants to be present in the sample at the same time, rather than inferring hypothetical common ancestors between isolates. This approach is appropriate for quick mutating viruses like IAV and is straightforward to implement using the adegenet package in R (Jombart et al. 2011). The algorithm linked gene segment isolates by

minimizing the genetic distance between them and inferred which was ancestral based on sampling dates. We excluded any linked isolates that were sampled less than 30 days apart to ensure that descendants were not mistaken for ancestors.

All influenza A viruses share some genetic similarities. It is possible that SeqTrack linked some isolates in the sample with genetic similarities due to the fact that they are both functioning gene segments of IAV, rather than recent common ancestry. There are no published thresholds of genetic distance that can be used to distinguish between these two scenarios and any such threshold would differ by gene segment and viral subtype. Instead, we used a data driven approach to determine genetic distance thresholds for each gene segment. We calculated the number of connections in the viral genealogy for each value of genetic distance between zero and the maximum value observed. A disproportionate increase in the number of connections with only a small increase in genetic distance would indicate a natural threshold value. No such thresholds were detected so we used the value of genetic distance that allowed us to keep 99% of the connections. This eliminated all connections with unusually large genetic distances between them, relative to other connections in the genealogy.

To construct a spatial transmission network for each gene segment, isolates were assigned to watersheds based on their sample location (Figure 3.1). In these networks, the origin watershed was the location where the ancestor isolate was detected, and the destination watershed was the location where the descendant isolate was detected. Because we were interested in only large-scale spatial transmission, all connections where ancestors and descendants were detected in the same watershed were removed. We assumed that each gene segment transmitted separately. All eight spatial transmission networks, one for each gene segment, were added together into a single, directed spatial transmission network where edges

were weighted by the number of spatial transmissions that occurred. All watershed pairs without genetic links between them had a weight of zero indicating no spatial transmission.

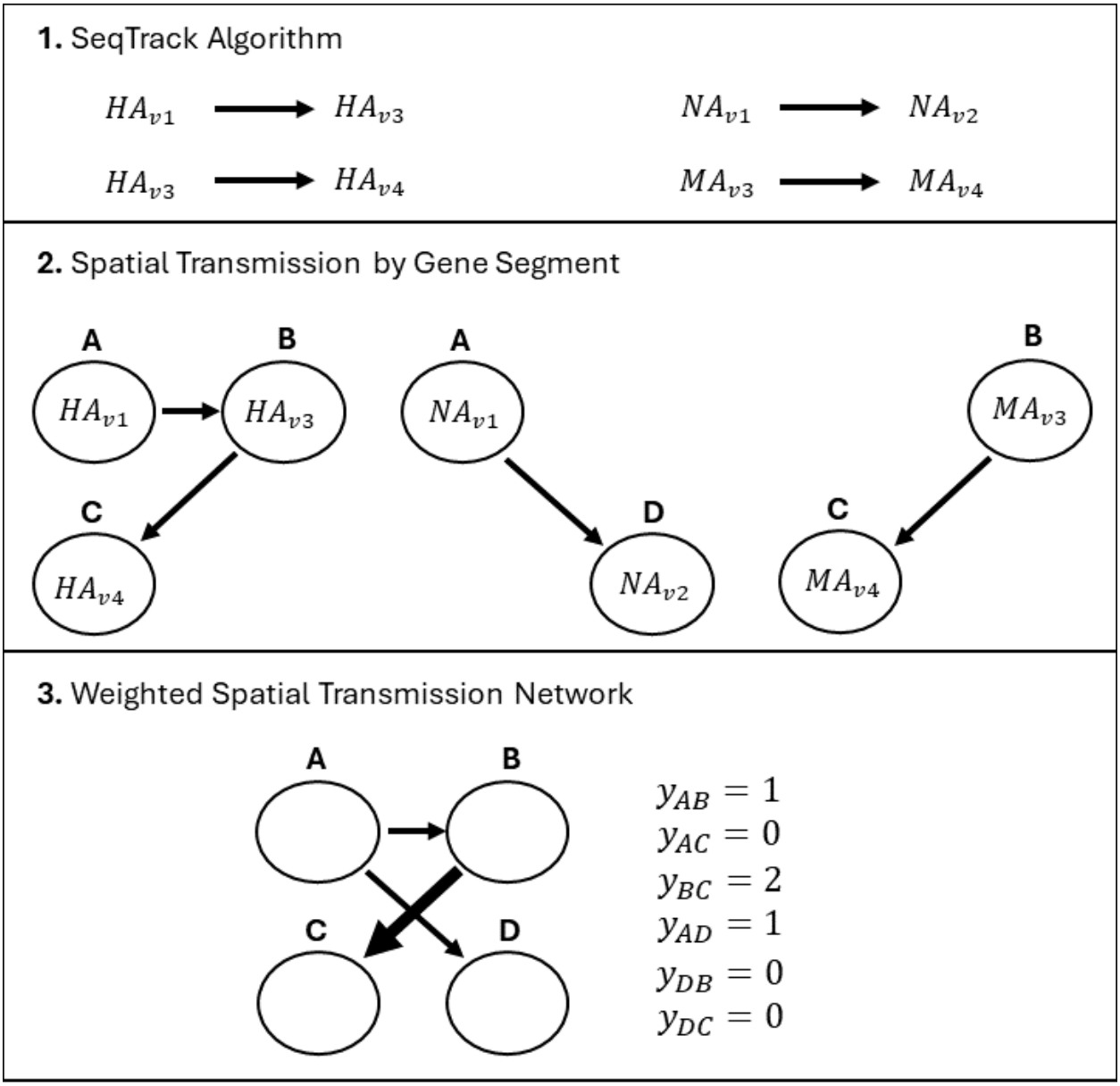

Figure 3.1: A conceptual diagram showing the process of constructing a weighted spatial transmission network for IAV from genealogies reconstructed by the SeqTrack algorithm. For simplicity, this example uses only three of the eight viral gene segments. HA denotes the hemagglutinin gene segment, NA the neuraminidase gene segment, and MA the matrix gene segment. The sub-index *v1*, *v2*, *v3*, and *v4* indicate which viral isolate the gene segments came from. In step one, the SeqTrack algorithm linked gene segments with the smallest genetic distance between them and inferred which was ancestral based on the sampling order. The

sampling order of watersheds was A, D, B, C. Ancestors are at the base of the arrows. In step 2, each ancestor and descendant were assigned to a watershed based on their sampling location. Viral isolate 1 was sampled in watershed A, isolate 2 in watershed D, isolate 3 in watershed B and isolate 4 in watershed C. This resulted in three spatial transmission networks, one for each sampled gene segment. Arrows now indicate the direction of spatial transmission. In step three, all three networks were added together to get a single weighted spatial transmission network. y denotes the number of spatial transmissions between each watershed pair. Edges y_{DA}, y_{BA}, and y_{BD} are not included in the network because spatial transmissions cannot occur backward in time. Watershed B was sampled after both D and A, and watershed D was sampled after A.

3.2.4 Host Movement Network

We constructed a weighted, directed host movement network from bird banding data on Mallard ducks (*Anas platyrhynchos*). We chose to focus on Mallards because they are widespread and thought to have a disproportionate impact on the spatio-temporal transmission dynamics of IAV in the United States (Hinshaw et al. 1986, Alfonzo et al. 1995, Ip et al. 2008). These data, collected by the USGS Bird Banding Laboratory, are the best available information on continental scale bird movement in the western hemisphere. We used 73,695 banding and recovery records for Mallards collected between 2006 and 2017 to approximate overall movement patterns in the continental US, rather than directly matching movement data to the years genetic samples were taken since the samples were from multiple species. We only used individuals that were marked and recovered within a six-month period to eliminate round trip movement that would skew movement patterns. The edges of the host movement network were weighted by the number of Mallards moving between watersheds. Only between watershed movements were included in the network. We expected that watersheds connected by more Mallard movement would have a greater number of spatial transmissions between them.

3.2.5 Local-Scale Covariates

Temperature in each watershed was approximated using monthly means of 2-meter gridded land surface temperature downloaded from NOAA Physical Sciences Laboratory from

their website at https://psl/noaa.gov (Fan and van den Dool, 2008, retrieved 5 April 2019). Monthly data were averaged for each watershed in the period between September 2015 and November 2017. In lab experiments, the persistence of the virus in water was best in temperatures below 17 C (Brown et al. 2009). We predicted that cooler watersheds would have more spatial transmission between them.

Viral prevalence was calculated as the proportion of positive H7 and H5 subtype samples in each watershed over the study period. These data are also from the IAV surveillance dataset described in the viral genomic data section. Because not every positive H5 and H7 sample was sequenced, a greater number of samples were used to calculate prevalence than were used to construct the spatial transmission network. We expected that watersheds with higher prevalence would have more spatial transmission between them.

3.2.6 Detection Covariates

Sample effort was the total number of samples taken in each watershed over the entire study period. We expected that watersheds with more sample effort would have a greater probability of detecting a spatial transmission. Sample delay was the number of days between when an origin and destination watershed were sampled. In cases where watersheds were sampled more than once the earliest sample date at the origin watershed and the latest sampling date at the destination watershed were used. We expected that the greater the amount of time between sampling events, the less likely a spatial transmission would be detected. We predicted that the amount of time between sampling events would also impact the probability of a transmission occurring. The longer the sample delay, the more opportunities for spatial transmission to occur. For this reason, sample delay was also included as a spatial transmission covariate.

3.2.7 Network Model and Fitting

The number of observed spatial transmissions (y_{ij}) between each origin watershed (i) and destination watershed (j) was determined by whether spatial transmission could occur and our ability to detect it (Figure 3.2). We used a zero-inflated Bayesian model where no spatial transmissions were observed when spatial transmission was not probable ($z_{ij} = 0$). When spatial transmission did occur ($z_{ij} = 1$), the number of spatial transmissions was binomially distributed with sample size at the origin watershed (n_i) and detection probability (φ_{ij}) (Line 1, Spatial Transmission Model). Detection probability was a function of the sample effort at the origin ($Effort_i$) and destination ($Effort_j$) watersheds as well as the amount of time between sampling the two watersheds ($Delay_{ij}$) (Line 2, Spatial Transmission Model). Because a given watershed can act as an origin in multiple watershed pairings, as well as a destination, we also included a random effect for origin watershed ($\alpha_{i,\varphi}$) and destination watershed ($\alpha_{j,\varphi}$). The probability that a spatial transmission occurred between two watersheds (p_{ij}) was a function of viral prevalence at the origin ($Prev_i$) and destination ($Prev_j$) watershed, the number of birds moving between the watersheds ($Move_{ij}$), the temperature at the origin ($Temp_i$) and destination watershed ($Temp_j$), and the sample delay ($Delay_{ij}$). A random effect for the origin ($\alpha_{i,p}$) and destination ($\alpha_{j,p}$) watersheds was also included. Specification of the prior distributions of the parameters can be found in Chapter 3 Supplemental Material.

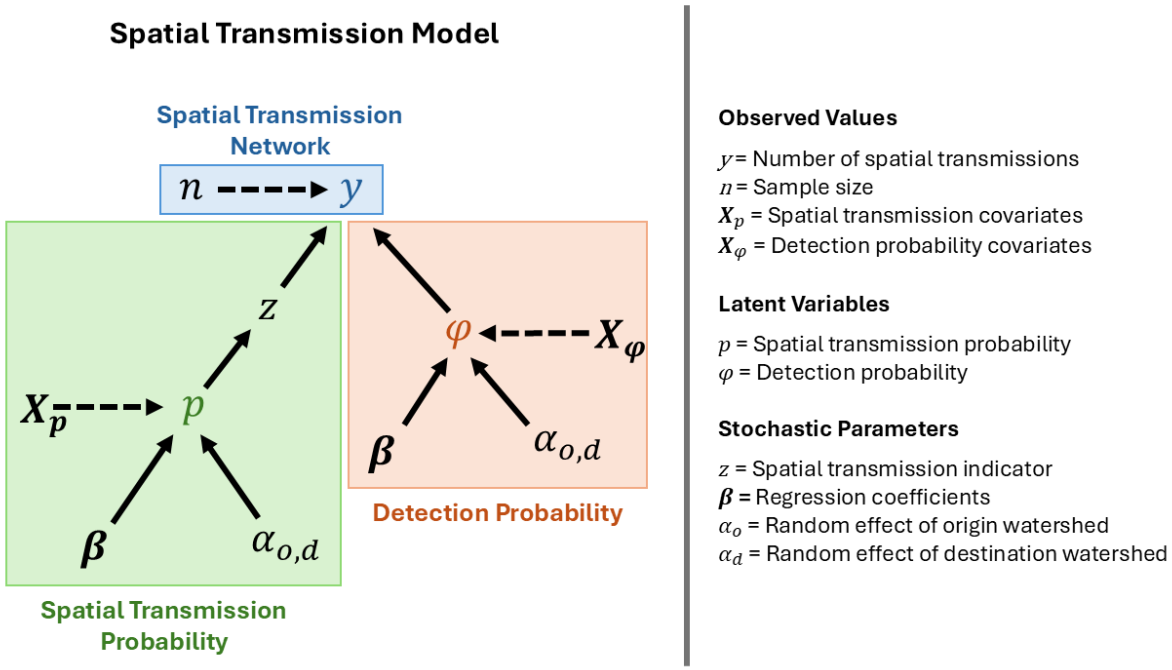

Figure 3.2: A directed acyclic graph modified to show the zero-inflated network model structure and the biological interpretation of the model components. Dashed lines indicate observed data and solid lines indicate parameters.

Spatial Transmission Model

1) $y_{ij} \sim \begin{cases} 0 & when\ z_{ij} = 0 \\ Binomial(n_i, \varphi_{ij}) & when\ z_{ij} = 1 \end{cases}$

2) $logit(\varphi_{ij}) = \beta_{X,\varphi}(Effort_i + Effort_j + Delay_{ij} + 1) + \alpha_{i,\varphi} + \alpha_{j,\varphi}$

3) $z_{ij} \sim Bernoulli(p_{ij})$

4) $logit(p_{ij}) = \beta_{X,\varphi}(Prev_i + Prev_j \times Move_{ij} + Temp_i \times Temp_j + Delay_{ij} + 1) + \alpha_{i,p} + \alpha_{j,p}$

All analyses were conducted in R (R Core Team, 2021) and Bayesian model fitting was done using package rjags (Plummer, 2003). We used a Markov Chain Monte Carlo (MCMC) procedure of three chains for 10,000 iterations after a burn in of 10,000 iterations. Chain

convergence was confirmed when the R-hat convergence diagnostic was equal to one for every estimated parameter and by visual inspection of trace plots. To ensure that the model could recover known parameters we tested the model with simulated data.

Posterior predictive checks were used to validate model estimates. We simulated data with the fitted model and calculated the Bayesian test statistic for the mean, standard deviation, maximum number of spatial transmissions, and the number of no transmissions in the distribution of the simulated dataset. Bayesian p-values should be within the range of 0.1 - 0.9 and values closer to 0.5 are ideal.

3.3 RESULTS

3.3.1 Model Validation

The distribution of data simulated from the fitted model was close to the observed data, accurately capturing the true mean ($p = 0.5$) and number of no-transmission watershed pairs ($p = 0.51$). The maximum number of spatial transmissions ($p = 0.15$) in the simulated dataset was lower than the observed data resulting in a smaller standard deviation ($p = 0.13$) than was seen in the observed data. These values indicate that the model was a reasonable fit to the observed data. A plot of the simulated data compared to the observed data (Figure S3.1) can be found in the Chapter 3 Supplemental Material.

3.3.2 Spatial Transmission Network

A majority of the connections in the SeqTrack genealogies occurred between isolates that came from hosts of different species (65%). After spatially aggregating the SeqTrack connections to make the spatial network, 20% of spatial transmissions occurred within the same watershed. After removing these connections and those with a sample delay of less than 30 days, 381 watershed pairs remained with at least one spatial transmission between them. There were

another possible 1799 watershed pairings where no spatial transmission was observed. The largest intersample time was two years and the most spatial transmissions that occurred between any watershed pair in this time was 11.

3.3.3 Network Model

The local-scale covariates did not appear to influence the probability of spatial transmission (Figure 3.3). There was no clear influence of temperature or prevalence at either the origin or destination watersheds. Sample delay did have a positive impact, however. Spatial transmission was more likely between watersheds that were sampled farther apart in time. The host movement network had a strong positive influence on the probability of spatial transmission of AIV. Watersheds with more Mallards moving between them had a higher probability of spatial transmission than those with fewer hosts moving between them. A majority of viral movements occurred between flyways (Figure 3.4). Spatial transmissions between the Central and Mississippi Flyways were highly probable and the most probable spatial transmissions within the Pacific Flyway were all tightly concentrated on the west coast. The Atlantic Flyway had fewer spatial transmissions overall and showed very little within flyway viral movement.

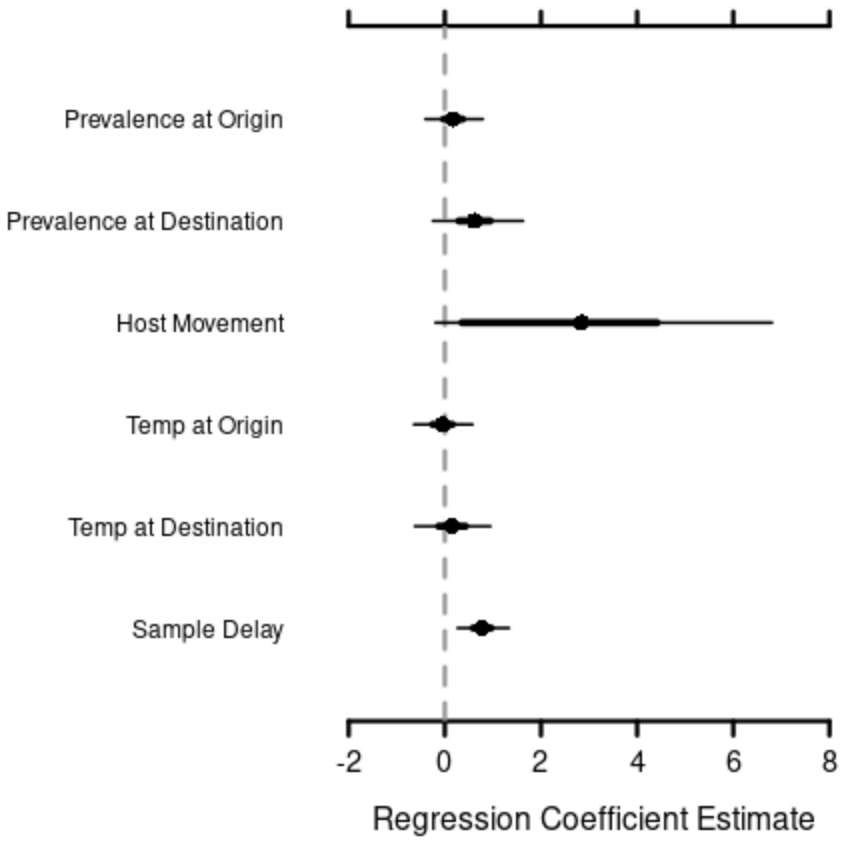

Figure 3.3: Model estimates of the influence of each covariate on the probability of spatial transmission of IAV. The regression coefficient estimate is on the log-odds scale. The point is the median of the posterior distribution, the thick line is the 50% Bayesian credible interval, and the thin line is the 95% Bayesian credible interval.

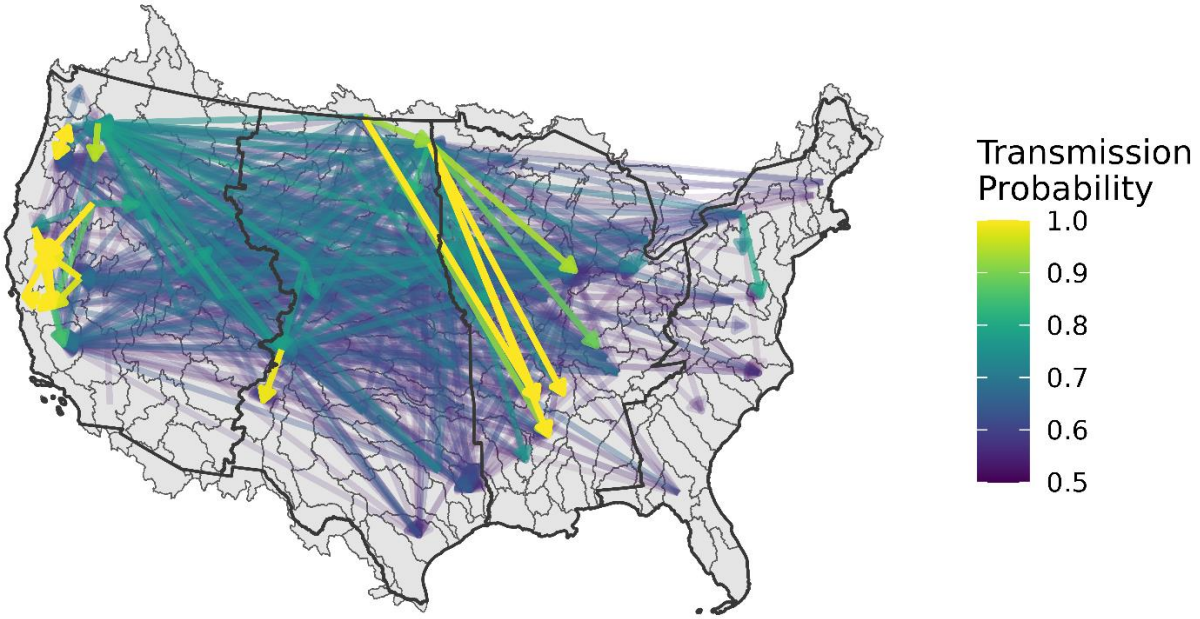

Figure 3.4: Map showing model estimates of the most probable ($p \geq 0.5$) spatial transmissions of IAV. Arrows indicate the direction of spatial transmission. The gray polygons are subregional watersheds, and the larger, black polygons show the four administrative flyways in the contiguous United States. From left to right they are the Pacific Flyway, Central Flyway, Mississippi Flyway, and Atlantic Flyway.

3.3.4 Detection Probability

Detecting a spatial transmission when it occurred was quite uncommon. The estimated detection probability for a given watershed pair was 4% on average. Surprisingly, sample effort at both the origin and destination watersheds had no clear impact on our ability to detect spatial transmissions (Figure 3.5). Sample delay, however, did have a negative influence on detection probability. We were more likely to detect spatial transmission when watersheds were sampled closer together in time (Chapter 3 Supplemental Material, Figure S3.2).

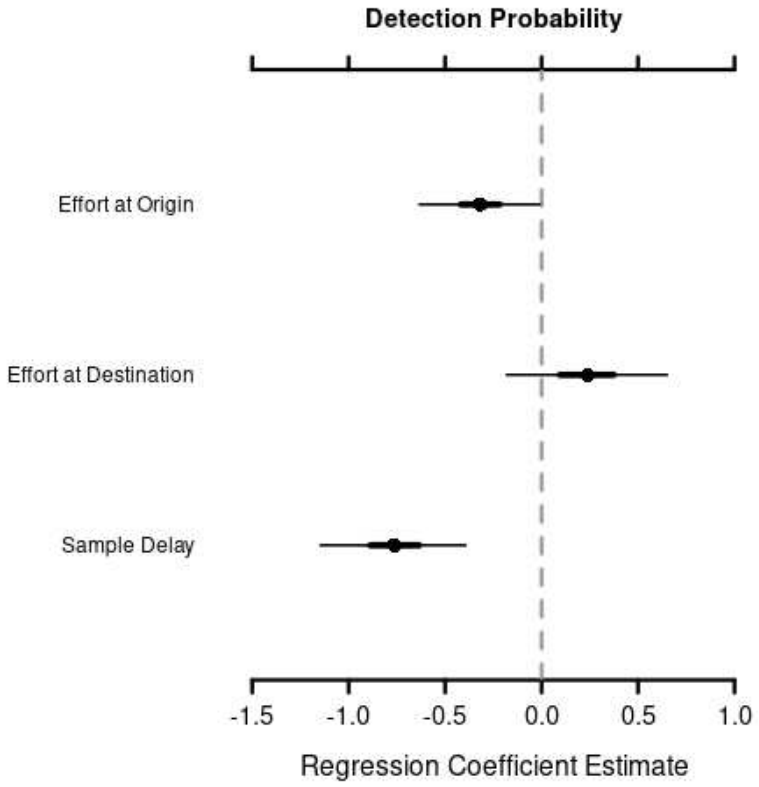

Figure 3.5: Model estimates of the influence of each covariate on the probability of detecting spatial transmission of IAV. The regression coefficient estimate is on the log-odds scale. The point is the median of the posterior distribution, the thick line is the 50% Bayesian credible interval, and the thin line is the 95% Bayesian credible interval.

3.4 DISCUSSION

The network model was a good fit to the observed data though it underestimated the maximum number of transmissions between watersheds. This is unsurprising given that a majority of watershed pairs had no observed transmissions between them and likely pulled model estimates lower. While the model did not capture the overall magnitude it did simulate more transmission events for watersheds with the highest number of observed spatial transmissions. This indicates that our estimated regression coefficients for the impact of local temperature, prevalence, and regional host movements are useful in predicting where we see more, or less spatial transmission occurring.

Mallards may have a special role in moving potentially pathogenic subtypes of IAV. Large-scale movement patterns of Mallards were predictive of spatial transmission of the virus even though a majority of transmissions were between different host species. It is possible that the Mallard movement network had good explanatory power because it was simply a good approximation of overall waterfowl movement in the US. However, Mallards are abundant throughout the US and often shed more virus in their feces than other host species (Costa et al. 2011). They may have a disproportionate impact on the spatial distribution of IAV by increasing the environmental reservoir of the virus. Mallards are also widely distributed and may be good spatial mixers, ensuring that the virus circulates throughout the country leading to more infections in other species.

Even though long-distance Mallard movement was an important predictor of spatial transmission, we saw little evidence of migratory flyways influencing where the virus was likely to move. This finding contradicts much of what has been found with phylogeographic analyses of IAV movement patterns. There are several possible factors that may explain why our results show different spatial patterns of viral movement. First, we considered a large variety of viral subtypes. A majority of phylogeographic studies restricted their analyses to H5N1 IAV because this subtype has been the cause of several recent outbreaks (Tian et al. 2015, Trovao et al. 2015, Yang et al. 2024). Restricting within subtype results in a pool of more closely related viruses and geographic restriction within flyway may be more likely in this case. Second, we used all eight viral gene segments to make our spatial transmission network rather than focusing on only the HA and NA gene segments as is more common. Internal gene segments of the virus also have an important role in spatial transmission patterns. The nonstructural protein (NS) and matrix protein (MA) gene segments have been shown to greatly impact the transmissibility and virulence of

IAV (Blaurock et al. 2021, Leyson et al. 2021). This can influence which hosts are infected and thus where the virus can move. Both Hill et al. (2016) and Bahl et al. (2013) used genetic data from all eight viral gene segments to reconstruct phylogenetic networks of IAV and found significant evidence for between flyway viral movement. In the US, Hill et. al (2016) also found increased geographically longitudinal movement of virus when there was more reassortant and interspecies transmissions. This finding suggests that incorporating all viral gene segments may improve detection of between-flyway spatial transmission because more reassortant transmission events are included.

It is important to note that we may not have seen an impact of flyways on spatial transmission because we did not explicitly look for it. This was a deliberate choice since using flyways as spatial units to explain the movements of IAV predicates on north/south migratory movements of hosts, ignoring other large-scale movements that may impact viral movement. Additionally, flyways are not highly accurate representations of host movement in North America. Current administrative flyways are based on bird banding data from the 1930s and are only roughly consistent with modern migratory communities. A network analysis of modern Mallard movement data found that administrative flyways in the US were significantly different from actual migratory communities and that these communities differed greatly by species (Buhnerkempe et al. 2016). These discrepancies between flyway designations and actual movement patterns could obfuscate the more detailed viral spatial transmission we are interested in.

Our results highlight that regional host movements between important breeding and wintering habitats, regardless of flyway, likely play an important role in spatial transmission. The most probable spatial transmissions were predicted to occur between important breeding and

wintering habitats for migratory birds and along major river valleys (Figure 3.6). Wetlands in the Prairie Pothole Region in the US and Canada host millions of breeding waterfowl each year and serve as an important migration corridor (Smith 1995). Spatial transmission from this region to the lower Mississippi Alluvial Valley was highly probable. This valley is one of the most important overwintering habitats for Mallards and other waterfowl in the US (Reinecke et al. 1992). In the western US, spatial transmissions were closely associated with the Middle Rio Grande Valley, the Sacramento River Valley, and the Columbia River all of which host important wintering and migratory habitats for waterfowl.

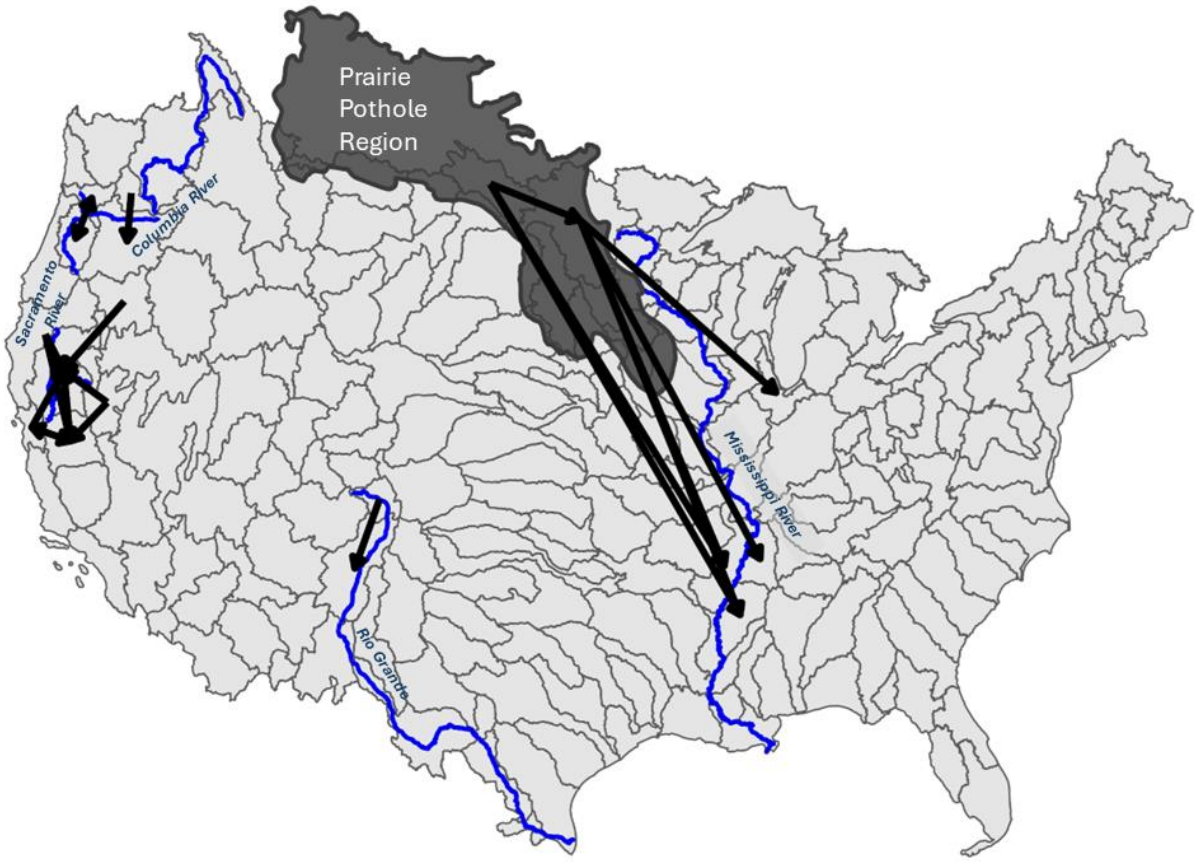

Figure 3.6: Map showing the most probable ($p > 0.9$) spatial transmissions of IAV in the contiguous US and their spatial proximity to important regions for waterfowl breeding, migration and overwintering. Arrows indicate the direction of spatial transmission. Grey polygons are watersheds.

Phylogeographic studies of IAV gene flow have also called attention to locations where there was strong evidence of between flyway movement, all corresponding to important waterfowl breeding habitats. In North America, Lam et al. (2011) singled out the Prairie Pothole Region as an important location for between flyway mixing of IAV. In Asia, Trovao et al. (2015) highlighted several wetland breeding habitats that served as important links for between flyway movement of the virus. This highlights the importance of sampling hosts at these locations to ensure a clear picture of viral movement around the globe, and to help us understand how these movements may change over time. Land use change and shifting climactic conditions in the US have significantly altered waterfowl's use of important habitats and have introduced greater interannual variation in host migration patterns and densities (Beatty et al. 2016). Future alteration of waterfowl habitats could have a dramatic impact on host movement patterns and consequently spatial transmission of IAV (Rashford et al. 2015).

Local habitats are clearly an important factor driving spatial transmission of IAV. However, in our model, local-scale temperature that influences environmental persistence within these habitats showed no clear impact on spatial transmission. Nor did local-scale prevalence, an indirect metric of local transmission. This highlights several technical challenges in connecting local conditions to larger scale spatial transmission. First, water temperature at individual sites within watersheds can vary greatly. In the future, sampling water temperature at sites where hosts are actually sampled will be important to provide a more nuanced picture of the impacts of the environment on viral persistence and perhaps show a stronger link to spatial transmission. Second, the ability of a virus to locally persist, transmit between hosts, and then move to a distant location depends on conditions in all the watersheds that are visited between when the isolates were sampled, not just the origin and destination watersheds. Constructing spatial

transmission networks within a shorter time period could help estimate more direct spatial transmissions between locations, and may improve our ability to connect viral movement to environmental conditions that impact viral persistence. However, truncating our dataset to only include watersheds that were sampled within three months of each other would eliminate 84% of observed transmissions events. Additional data would be needed to conduct this analysis at a smaller temporal scale. Lastly, different viral subtypes likely differ in their ability to persist in certain environmental conditions. Differing responses to local temperature or host density could obfuscate any clear relationship in a pooled analyses like ours. Determining if there are differences would require measuring water temperature, pH, salinity and host density at the locations where hosts were sampled.

Despite these technical challenges, there also is evidence that the inability of local-scale covariates to predict large-scale IAV spatial transmission is indicative of a larger trend in disease ecology and beyond. In one of the few multi-scale studies of pathogen distributions, large-scale phenomena like climate better explained the global spatial distributions of chytrid fungus, west Nile virus, and Lyme disease than did local factors like host density and richness (Cohen et al. 2016). Large-scale studies of species richness have found that the spatial extent and grain of analyses greatly impact what is predictive of spatial patterns of richness (Rahbek 2004). One proposed explanation is that a given variable is most predictive at whatever spatial scale it most differs (McGill. 2010, Cohen et al. 2016). This idea can be logically extended to predicting large-scale spatial transmission of IAV. Local host density and water conditions can vary greatly between local habitats. When these predictors are aggregated over larger areas like watersheds, they appear more homogeneous, reducing their ability to predict large-scale spatial transmission. In contrast, host movements are highly variable between watersheds allowing us to discern their

impact on large-scale spatial transmission of IAV. This does not mean that local environmental conditions and host-to-host transmission are unimportant at large scales, but rather that they are less predictive than large-scale host movement patterns.

Models of spatial transmission that incorporate information about host movement help epidemiologists and wildlife managers design spatially structured interventions like vaccination, travel restrictions, or culling to halt the spread of disease (Riley 2007, Keeling et al. 2001). While spatial models have linked human mobility data to the spread of SARS CoV-2 (Guan et al. 2021), measles outbreaks (Kraemer et al. 2019), and annual cases of human influenza (Lemay et al. 2014) less work has been done linking wildlife movement data to spatial transmission. To our knowledge, ours is the first study to directly link host movement data and spatial transmission of IAV. Our results show that detailed knowledge of host movements, and of Mallards in particular, is important for understanding where to prioritize surveillance for potentially highly pathogenic subtypes of IAV at an actionable scale.

Our model estimates of detection probability also provide valuable information about where to prioritize surveillance and the overall magnitude of spatial transmission. On average, sampling over three years detected an estimated 4% of the most probable spatial transmission events. This indicates that a watershed pair with ten observed connections actually had 250 spatial transmissions occur between them. Interestingly, there was no indication that increasing sample effort within any of the watersheds would improve our ability to detect these spatial transmissions. This suggests that where sampling occurs may matter more than the total number of samples taken. Our model estimates of spatial transmission probability provide useful information on where sampling will most likely detect spatial transmissions. Additionally, these provide valuable hypotheses of where new subtypes of the virus may move after an introduction

into North America. Moving forward, incorporating connections from Canada and Alaska will provide a more complete picture of important spatial transmission routes to locations with high risk of spillover into at-risk populations. Additionally, incorporating host movement data for other waterfowl species could help us understand their relative contribution to spatial transmission and help prioritize surveillance to particular taxa within the areas we have already highlighted as important for surveillance.

Chapter 4: A simulation method to estimate population-scale movement from multi-sensor animal tracking datasets.

4.1 INTRODUCTION

The field of movement ecology seeks to understand the causes and mechanisms of organismal movement as well as its consequences, from the scale of the individual all the way up to entire ecosystems (Nathan et al. 2008). Animal movement data are collected using a diverse array of techniques that often do not match one another in spatial or temporal resolution. These data are often expensive and difficult to obtain, making it more difficult to characterize movement patterns over large areas and long time periods. Addressing the spatio-temporal mismatch between different datasets, and combining them in a rigorous way, will allow us to expand our understanding of animal movement to larger spatio-temporal scales. Such understanding is critical because animal movement data have been used to better understand a diverse range of ecological questions including habitat preference (Aarts et al. 2008), foraging and territorial behavior (Bennison et al. 2017, Giuggioli and Kenkre, 2014), migratory responses to climate change (Davidson et al. 2020), habitat connectivity (Bastille-Rousseau and Wittemyer, 2020), and the spread of infectious disease (Dougherty et al. 2018).

Recent advances in this field have been driven by improvements in tracking technology and computation power. Because many animals can travel long distances very quickly, it has been exceedingly difficult until recently to obtain detailed tracking information on migratory populations. The launch of the Argos satellites in the 1970s and the widespread use of GPS technology starting in the 1990's has made detailed movement data increasingly available for a wide range of terrestrial and aquatic migratory taxa. GPS and satellite telemetry can track

animals with a high degree of accuracy at a fine temporal and spatial resolution resulting in many observations per individual. While costs have decreased over time, these methods are still expensive and are generally only deployed on relatively few individuals. Other methods of tracking like radio and acoustic tags allow researchers to mark more individuals using relatively inexpensive equipment but require a large investment in hands-on monitoring to get many observations through time. Mark-resight, mark-recapture, and mark-recovery methods that use numeric tags are the most cost effective and allow researchers to mark substantially more individuals. However, these methods usually result in only two locations per individual, the location where they were tagged and the location they were recaptured or died. Ultimately, the ideal method for tracking long distance migrations would be accurate, cost effective, require low monitoring effort, and result in many observations per individual over a large geographic area. Figure 4.1 summarizes some of the strengths and weaknesses of each of these methods relative to this 'ideal'. Additional factors to consider when deciding which tracking method is best are the habitat of the animal being tracked, their anatomy, and their body size relative to the tag.

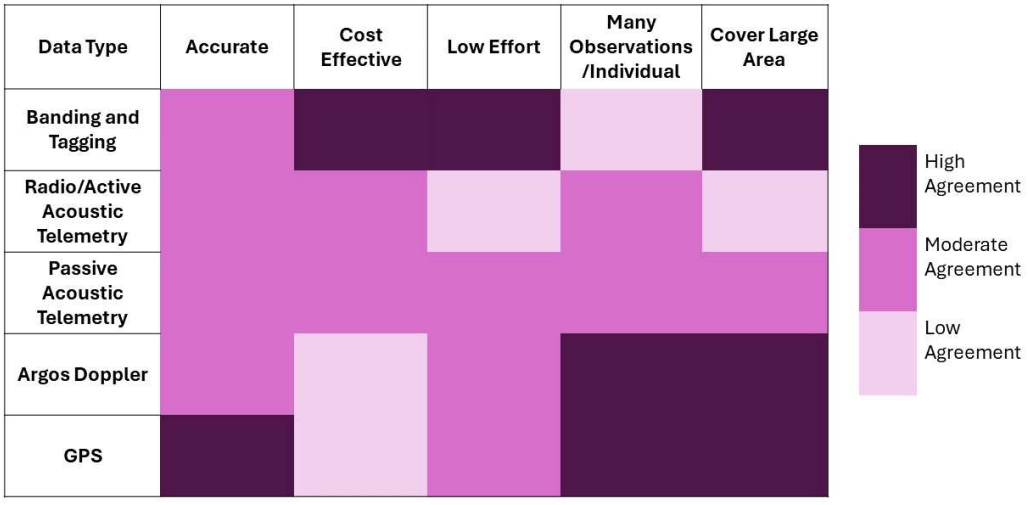

Figure 4.1: Important characteristics of each major animal tracking method. Color indicates the degree to which each animal tracking method agrees with the characteristic of an ideal tracking method that would be accurate, cost-effective, low effort and result in many observations/individual while covering a large geographic area.

While tracking technology continues to become smaller, cheaper, and more accurate, there is currently no method that satisfies the ideal. Additionally, a newly developed tag cannot help researchers deal with the spatio-temporal mismatch of data that has already been collected. These data are exceedingly valuable and allow researchers to understand how animal movements have changed over time. Animal tracking databases, such as Movebank, contain billions of locations collected using a wide range of methods, and cite multi-sensor datasets as a particular challenge to work with (https://www.movebank.org/cms/movebank-content/about-movebank). Banding/tagging data, Argos, and GPS data all possess some attributes of the ideal tracking method and have complementary strengths. Bird bands, for example, are inexpensive to use and low effort to track since recoveries are primarily reported by the public. Argos and GPS tags are accurate and crucially provide detailed information about how individuals move. A combination of these two datasets would combine their strengths as well, getting us closer to the hypothetical ideal. However, no framework currently exists that combines these highly detailed data sources with data that have only two locations per individual.

Statistical methods have been developed to integrate different movement datasets. However, their focus is on combining data that were collected at similar detailed temporal scales but differed in the time intervals between location fixes as well as sources and size of measurement error. Buderman et al. (2015) used B-splines to connect locations recorded at irregular intervals from Argos satellite and radio telemetry. However, this method requires more than two locations to infer movement paths and so cannot be applied to mark-recovery data. Continuous-time correlated random walk models of animal movement have only been tested on highly detailed and accurate GPS data and are also not designed for mark-recovery methods (Johnson et al. 2008, Breed et al. 2012). Brownian bridge models could be used to infer spatial

movement from just two locations but lack a straightforward way of incorporating knowledge of host movement behavior into the model, resulting in greater simulation error (Horne et al. 2007).

We developed a statistical method that leverages the abundance and affordability of mark-recovery data and the detail and accuracy of satellite telemetry and GPS data to simulate population scale migratory movement. This method extracts biological information about animal movement from detailed Argos/GPS data and uses an approximate Bayesian computation (ABC) rejection scheme to simulate locations for banding data to fill in the gap between when an individual was marked and when they were recovered. The resulting simulated locations match the spatio-temporal resolution of the Argos/GPS data, allowing researchers to combine mark-recovery data with the more detailed data types. ABC methods have been used to deal with missing observations in movement data, and to address discrepancies between the scale of movement processes and observations but have not been implemented to match the spatio-temporal resolution of two different observation methods (McDermott et al. 2017, Ruiz-Suarez et al. 2020). Our ABC method is customizable to the taxa and spatio-temporal resolution of interest and provides estimates of simulation error. Additionally, this method does not require any knowledge of an animal's habitat, allowing us to glean useful information about movement even when habitat data are unavailable.

We tested our ABC method on movement data of migrating Northern Pintail ducks (*Anas acuta*). This species is widely distributed across the northern hemisphere and undergoes yearly long-distance seasonal migrations from overwintering habitats in the south to breeding habitats in the north. Pintails are a reservoir host of influenza A virus (IAV), and their seasonal migrations may introduce new subtypes of the virus from Asia to North America (Ramey et al. 2010). Understanding Pintail migratory movement will help researchers understand where

specifically the virus may be introduced and where it might move after an introduction. Pintails have been marked with numeric bands for nearly 100 years in the United States (US), and band-recoveries by hunters are fairly common due to the long history of hunting in the US. In contrast, Pintail movement data in Asia come primarily from Argos satellite telemetry and GPS tracking, providing detailed movement data, but for much fewer individuals. We use the detailed information contained in the telemetry/GPS dataset to simulate movement between the marking and recovery locations in the banding dataset. Simulated locations and satellite/GPS data can be used to construct a population scale movement network for Pintails that can help predict the movement and prevalence of this important virus of concern. While we used our ABC method to address Pintails and IAV, this framework can be used to combine movement data of differing resolutions for a wide range of species and ecological questions.

4.2 METHODS

4.2.1 Overall Framework

We used a functional movement model informed by detailed location data to calculate three quantities, derived from movement trajectories that described an animal's movement in discrete-time and space. Speed described how quickly an individual moved between locations, relative turning angle described how their trajectory changed at each location, and direction described whether individuals made a left or right turn at each location. We used this biological information to simulate locations between marking and recovery locations for individuals that only had two location observations. We repeated these simulations many times for each individual and used an approximate Bayesian computation (ABC) rejection algorithm to select only the simulated paths that ended close to the known recovery location, resulting in a distribution of locations at each time-step. We used the means of these distributions to estimate

the missing movement path of an individual between when they were marked and when they were recovered and used the variance of the distributions to approximate uncertainty (Figure 4.2).

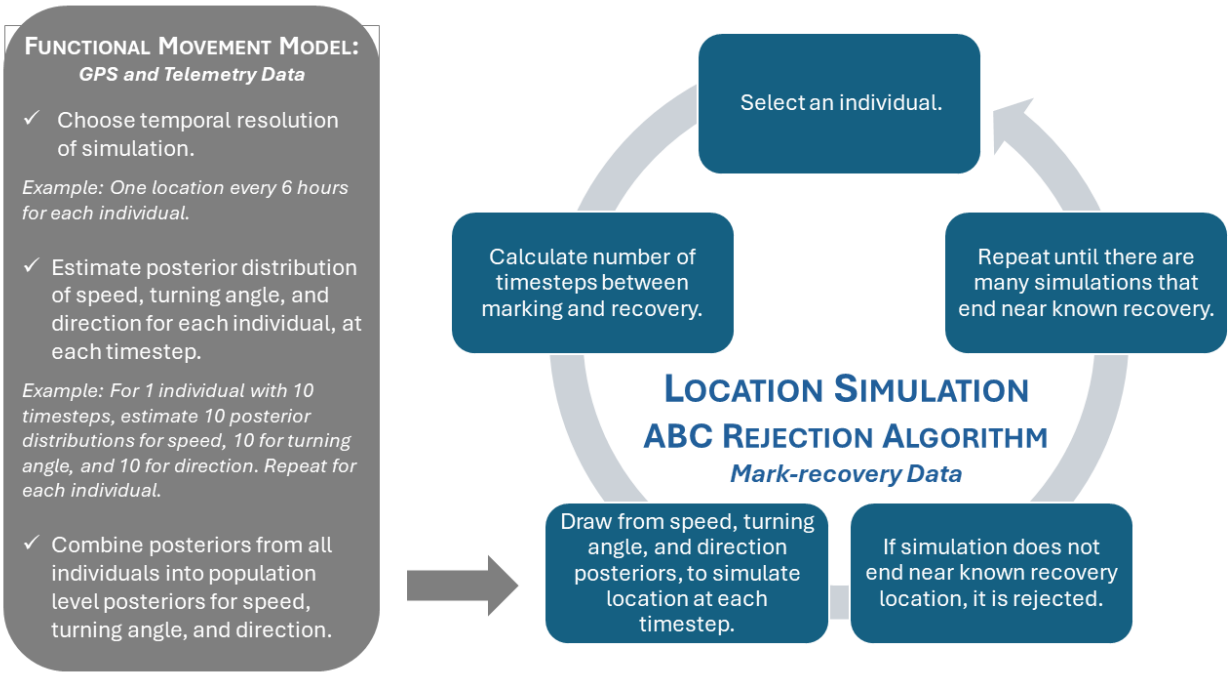

Figure 4.2: Conceptual diagram outlining the process of location simulation. The grey box shows how biological quantities that describe animal movement are estimated from GPS and Argos telemetry data. The blue boxes show the main steps in location simulation from mark-recovery data. The grey arrow shows the step at which information from detailed movement data is incorporated into the simulation.

To test the accuracy of the method, we withheld known location data for some individuals from the functional movement model, used the derived quantities to simulate locations for these individuals, and applied the ABC rejection algorithm. We visually compared individual movement paths to simulated paths. Because our method uses population scale biological information to simulate population scale movement patterns, we also compared the properties of population scale movement networks constructed from observed and simulated

locations. We first describe our simulation validation methods in general terms and then provide specific details on how it was applied to location data for migratory Pintail ducks.

4.2.2 Estimate Derived Quantities

We extracted biological information about how individuals moved from detailed movement data with a previously developed functional movement model (FMM) (Buderman et al. 2016). The model fits smooth continuous movement paths to these data to simulate spatial locations forward in discrete time. For each individual, we used these simulated locations to estimate posterior distributions of speed between each time-step, relative turning angle, and direction at each time-step. The choice of time interval should balance capturing the movement of interest while minimizing the total number of time-steps in order to decrease simulation time for the mark-recovery data. We joined the individual, time-step level posteriors into population scale posterior distributions for speed and relative turning angle. To estimate the population posterior distribution of direction we calculated the proportion (probability) of left turns for all individuals at every time-step. We drew from these three distributions to simulate missing locations for individuals that only had two location observations in total. Locations were simulated at the same time-step interval as speed and relative turning angle.

4.2.3 Simulation Approach

The simulation started at the known location where individual *k* was marked. Relative turning angle was necessarily defined relative to the trajectory of the individual in the previous time-step. Therefore, we could not draw from the relative turning angle posterior distribution for the first simulation of location as there was no previous point to inform the relative nature of turning angle at the marking location. Instead, we used the 'angle' function from the rearrr package in R to calculate the angle ($\theta_{t=0}$) between the marking location, which is centered at the

origin of a Cartesian coordinate system, and the recovery location. We used the known end point to initialize a simulation, rather than taking a random draw from the unit circle, to slow the rate of error propagation. Speed was considered a distance because the intervals between each time-step were the same and the denominator of distance/time could be ignored. Both speed (*r*) and relative turning angle (φ) are polar coordinates that we converted to the Cartesian coordinates of longitude and latitude at each time-step. To simulate the first location at time *t=1*, we drew from the distribution for speed and calculated a new longitude ($long_t$) and latitude (lat_t) using equations 1 and 2 respectively.

Initializing A Simulation

1. $long_{t=1} = long_{t=0} + r_{t=0} \times \cos(\theta_{t=0})$
2. $lat_{t=1} = lat_{t=0} + r_{t=0} \times \sin(\theta_{t=0})$

For simulated locations *t > 1*, relative turning angle must be converted to the Cartesian scale, and we must choose if the individual turns left or right. Hence, the turning angle, (φ_t), was modified by two variables: z_t and θ_{t-1} as shown in equations 3 and 4.

All Other Time-steps

3. $long_{t+1} = long_{t-1} + r_t \times \cos(\varphi_t z_{long,t} + \theta_{t-1})$
4. $lat_{t+1} = lat_{t-1} + r_t \times \sin(\varphi_t z_{lat,t} + \theta_{t-1})$

z_t dictated whether the individual turned left (z_t = *1*) or right (z_t = *-1*) and was drawn at each time-step with probability of left turn p_{left} and probability of right turn $p_{right} = 1 - p_{left}$. A separate z_t was drawn for latitude ($z_{lat,t}$) and longitude ($z_{long,t}$). Relative turning angle was drawn from the posterior distribution for φ_t, and the angle between the location at *t-2* and the

location at *t-1* on the Cartesian scale (θ_{t-1}) adjusted φ_t to the Cartesian coordinate system. A new speed, relative turning angle, $z_{lat,t}$ and $z_{long,t}$ were drawn for each time-step. Figure 4.3 summarizes the entire simulation process for one individual, over three simulated time-steps of a single simulation iteration.

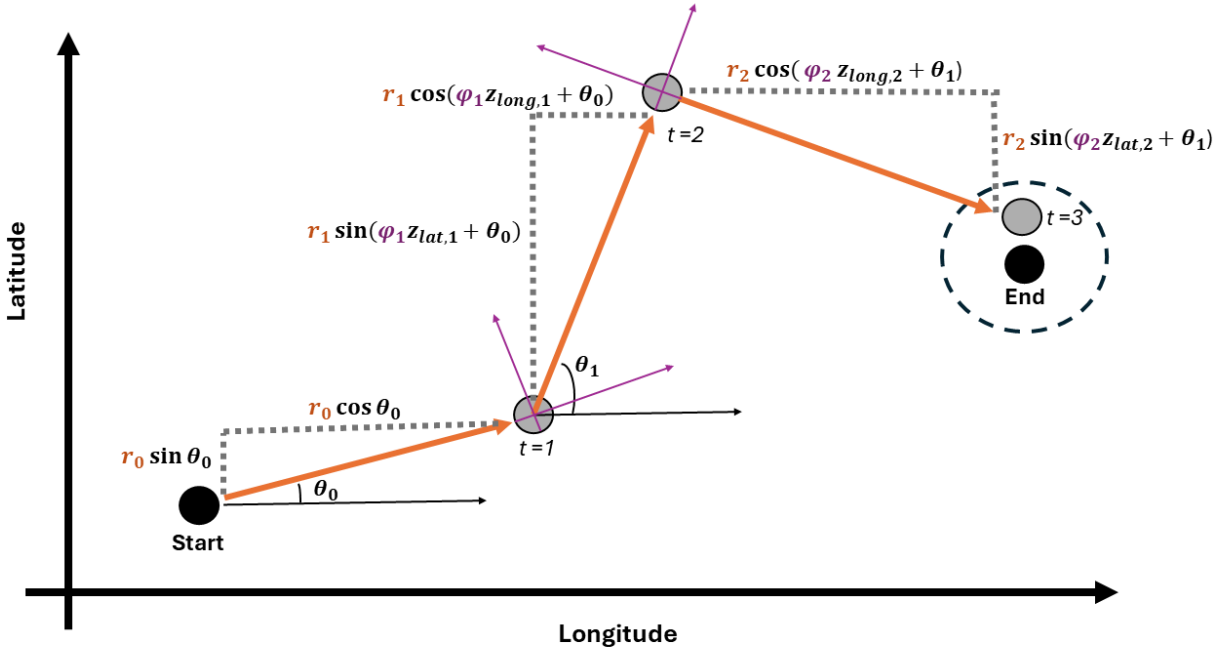

Figure 4.3: A single, successful simulation of three time-steps that ended within a buffer distance of the known endpoint. Black circles indicate known starting and ending locations. Grey circles are simulated locations. The buffer distance is shown as a black dashed circle. Grey dashed lines show the change in latitude and longitude calculated from adjacent equations. Orange lines are the distance an individual traveled and were drawn from the posterior for speed, r_t. At time *t = 0*, the trajectory of travel was the angle from the starting location to the known ending location ($\theta_{t=0}$). For all other time-steps, travel trajectory was determined by a draw from the posterior for relative turning angle (φ_t), shown in purple. The purple x and y-axes at each simulated time-step show the Cartesian scale rotated relative to the previous trajectory. The angle between the locations at *t-2* and *t-1* (θ_{t-1}), shown as a black arc, was added to φ_t to transform relative turning angle to the Cartesian scale, shown as a horizontal black arrow. $z_{lat,t}$ and $z_{long,t}$ determine whether the individual turns left or right.

4.2.4 ABC Rejection Algorithm

Once locations were simulated for all time-steps, we checked if the coordinates at time-step $t = n$ were within a set distance, or *buffer distance*, of the known ending location. If the final location was within this zone, the simulation was saved; if not, the simulation was rejected. This process was repeated for m number of simulations. Saved "successful" simulations approximated posterior distributions of latitude and longitude at each time-step for each individual. We used the mean of these distributions as the point estimate for location. We chose the mean, rather than the median, to allow a greater influence of the tails of the distribution. This allowed the simulated travel paths to curve more rather than forming a relatively straight trajectory between the known start and end locations. Additional considerations for choosing how to project the data, dealing with observation error, choosing the number of simulations, and choosing a buffer distance are covered in detail in the Chapter 4 Supplemental Material.

4.2.5 Method Validation

We used a subset of the individuals with detailed tracking histories to validate the simulation method. These individuals' tracking data were not used to fit the FMM or to calculate the population posterior distributions of speed, turning angle, or direction. Simulated locations were visually compared to the true observed locations to determine how well individual movement paths matched. We used spatial networks to quantify how well the simulation method reconstructed population scale movement patterns in the observed data. Nodes in the networks were defined as locations and edges were weighted by the number of individuals moving between them. Movement data are often encoded as spatial networks and network science provides a suite of analytical tools for comparing their structures (Jacoby et al. 2012, Jacoby and Freeman 2016, Bastille-Rousseau et al. 2018). To characterize movement patterns, we looked at

which locations were connected in the network (node and edge membership), the strength of those connections (edge weight), and which locations were grouped into communities (community membership). We used edge density, reciprocity, diameter, transitivity and the number of strong components to evaluate the higher order structure of the spatial network. A conceptual diagram showing how movement networks were made (Figure S4.1) can be found in the Chapter 4 Supplemental Material.

To understand how simulation error impacted network structure we estimated 95% Bayesian credible intervals for each of the network statistics. We randomly sampled one of a hundred quantiles from the posteriors for latitude and longitude at each time-step, for each individual. The resulting locations were converted to a spatial network, and the properties of this network calculated. This process was repeated 1000 times to form a distribution for each of the network properties mentioned above. We report the lower 0.025 quantile and the upper 0.975 quantile of these network property distributions to approximate the error of the point estimates of the network properties. A conceptual diagram (Figure S4.2) and mathematical justification explaining this process can be found in the Chapter 4 Supplemental Material.

We used one-step forward simulation and out-of-sample validation to check the performance of our method. One-step forward simulation was used to determine if simulations could recover observed locations without error propagation. Simulating without error propagation tells us if the method fundamentally works or not. To eliminate error propagation, we used the known location at each time-step to simulate a location one step forward in time. This allowed us to see how much error was introduced in a single time-step. These simulated locations were visually compared to the observed locations to determine how well the simulation performed.

The one-step-forward simulations were also used to quantify the importance of using the detailed movement data to estimate biologically informed quantities for speed, turning angle, and direction with the FMM. The simulated locations were made into a directed spatial movement network called the *biological network*. We compared the biological network to a *uniform network* that represented an appropriate, comparative null hypothesis and was constructed from locations that were simulated using derived quantities drawn from uniform distributions. This ensured that every value for speed, relative turning angle, and direction was equally probable. These distributions had the same minimum and maximum value as the biological distributions but otherwise contained no biological information about how individuals actually move. We compared the properties of these two networks to the *observed network* to evaluate how well they replicated the true population scale network structure. The observed network was constructed from the actual observed locations of the individuals withheld for validation.

Out-of-sample validation was used to determine how well the simulation method approximated the real locations with error propagation. To simulate locations with error propagation we used only the start and end locations of the detailed tracking data for the individual withheld for validation. This ensures the locations for each individual resemble mark-recovery data. We called the resulting network the *mark-recovery network*. We compared network statistics between the mark-recovery network and the observed network to determine how well the simulation reproduced the known population scale movement patterns in the observed data.

We also wanted to understand if our simulation method resulted in more realistic movement networks than simply connecting the marking and recovery locations with a straight line. To do this we compared the mark-recovery network to a *straight-line network*, which a null

hypothesis for this comparison. The straight-line network was constructed by connecting the observed start and ending locations of each validation individual with a straight line and including all nodes that intersected these lines. Connections for each individual were added together to form a single weighted directed network. Network statistics for the mark-recovery and straight-line networks were compared to determine which showed more realistic population scale movement patterns.

3.2.6 Case Study: Northern Pintail Ducks

Dataset

We used detailed movement data of Northern Pintail ducks marked with GPS and Argos tags to develop and test the full method. These data provided high resolution spatial and temporal information about long distance migratory movements. GPS and Argos satellite telemetry locations were gathered for 390 birds between January 2000 and December 2009 and covered up to one year of tracking for some individuals. The time intervals between fixes were different for most birds due to differing equipment and censorship of poor-quality locations in the Argos dataset. Birds were tagged in Hong Kong, Japan, and the United States (Table 4.1). We randomly chose 90% of individuals (n = 351) to fit the functional movement model and withheld 10% of individuals (n = 39) for out-of-sample validation of the overall method. Of the 36 birds withheld for model checking 18 were marked in Japan, two were marked in Hong Kong, one was marked in Texas, and 15 were marked in California.

GPS locations were assumed to have zero error as is standard and none of these data points were censored. Argos data accuracy was classified into location classes, each with its own estimated error radius (See Douglas et al. 2012). All Z class observations, which indicate invalid locations, were removed from the dataset. Additional low accuracy locations were removed

using the Douglas Argos-filter (Douglas et al. 2012). Because we wished to simulate large-scale migratory movement, we removed all individuals that were tracked for fewer than five days since these observations generally reflected short distance local movement. Most individuals were tracked for fewer than six months. Because all bird locations were recorded in the northern hemisphere on either side of the Pacific Ocean, we used an Albers equal area conic projection with a central meridian at -162 degrees longitude to center our projection on the Pacific Ocean.

Table 4.1: Summary of detailed movement data used to fit the functional movement model and validate the overall method.

Marking Location	# Individuals	Tag Type	Citation
Butte County, California, USA	160	Argos	Miller et al. 2005
Miyagi, Hokkaido and Yamagata Prefectures, Japan	139	Argos	Hupp et al. 2019
Hong Kong, China	31	GPS (6), Argos (23)	Sullivan et al. 2018
Texas and New Mexico, USA	60	Argos	Haukos et al. 2006

Estimate Derived Quantities

We used the R package ctmcmove to fit the continuous time functional movement model to the location data for a given individual at 24-hour intervals (Hanks 2018). This time increment ensures that we captured important long-distance movements while minimizing the number of time-steps that were simulated to reduce computation time. The estimated movement paths were used to calculate posterior distributions for speed (km/24 hrs) and relative turning angle (radians) at each time-step for each individual following Buderman et al. (2015). Posterior distributions for all time-steps and all individuals were joined into a single population scale posterior distribution for each quantity. The population posterior distribution of direction was determined

using the trajr package to calculate the proportion of left turns in the posterior of all locations estimated by the FMM (McLean and Skowron Volponi, 2018).

Simulations

We ran 100,000 simulations for each withheld individual, which balanced computation time with achieving a reasonable number of successful simulations per bird. We saved all simulations that ended within 128 km of the known recovery locations of each individual. This is the average size of watersheds later used to construct the movement networks used for validation. We chose watersheds because they are biologically relevant to waterfowl, can be defined uniformly across many countries with different political boundaries, and provide sufficient movement detail while being at a larger scale than potential observation error.

ABC Rejection Algorithm and Method Validation

To check how the overall method performed on these data, we estimated locations for the 36 withheld birds using the population posteriors of speed, turning angle, and direction from the 324 individuals used to fit the FMM. The simulated movement trajectories were then filtered using the ABC rejection algorithm. We visually compared movement paths and compared network statistics for all movement networks. We used the mean latitude and longitude of the resulting location posterior distributions to construct the biological network. Nodes of the network were watersheds and the edges connecting them were the number of birds that move between them. We also used the withheld birds to construct the uniform, straight-line and observed networks and calculated network statistics as well as their 95% credible intervals. We used a community detection algorithm developed by Buhnerkempe et al. (2016) to compare migratory community structure.

4.3 RESULTS

4.3.1 One-step Forward Simulation without Error Propagation

The method was able to recover the true locations when simulated without error propagation. One-step ahead simulated locations and true observed locations were a close visual match when using speed, turning angle and direction from the functional movement model. Figure 4.4 compares the one-step ahead simulations to the observed movement locations for four Pintail ducks. Two individuals were randomly chosen from each continent to demonstrate the difference in simulating locations with biologically derived quantities (Figure 4.4A) and the uniform derived quantities (Figure 4.4B). The simulated biological network closely resembled the observed network, and 95% credible intervals for each statistic contained the true network statistic in all cases except the number of edges and reciprocity, which were overestimated (Table 4.2). The biological network captured 96% of the true nodes in the observed network. Metrics of community structure, including transitivity and the number of strong components, closely matched the structure of the observed network.

Drawing speed, turning angle and direction from uniform distributions resulted in significantly more error in the simulated locations (Figure 4.4B). Only 54% of the watersheds in the uniform network were also in the observed network. Credible intervals for network statistics were much wider than the biological network and did not contain the true network values for reciprocity, transitivity, or number of strong components (Table 4.2). With the exception of diameter, the medians of the biological network statistics were all closer to the real values.

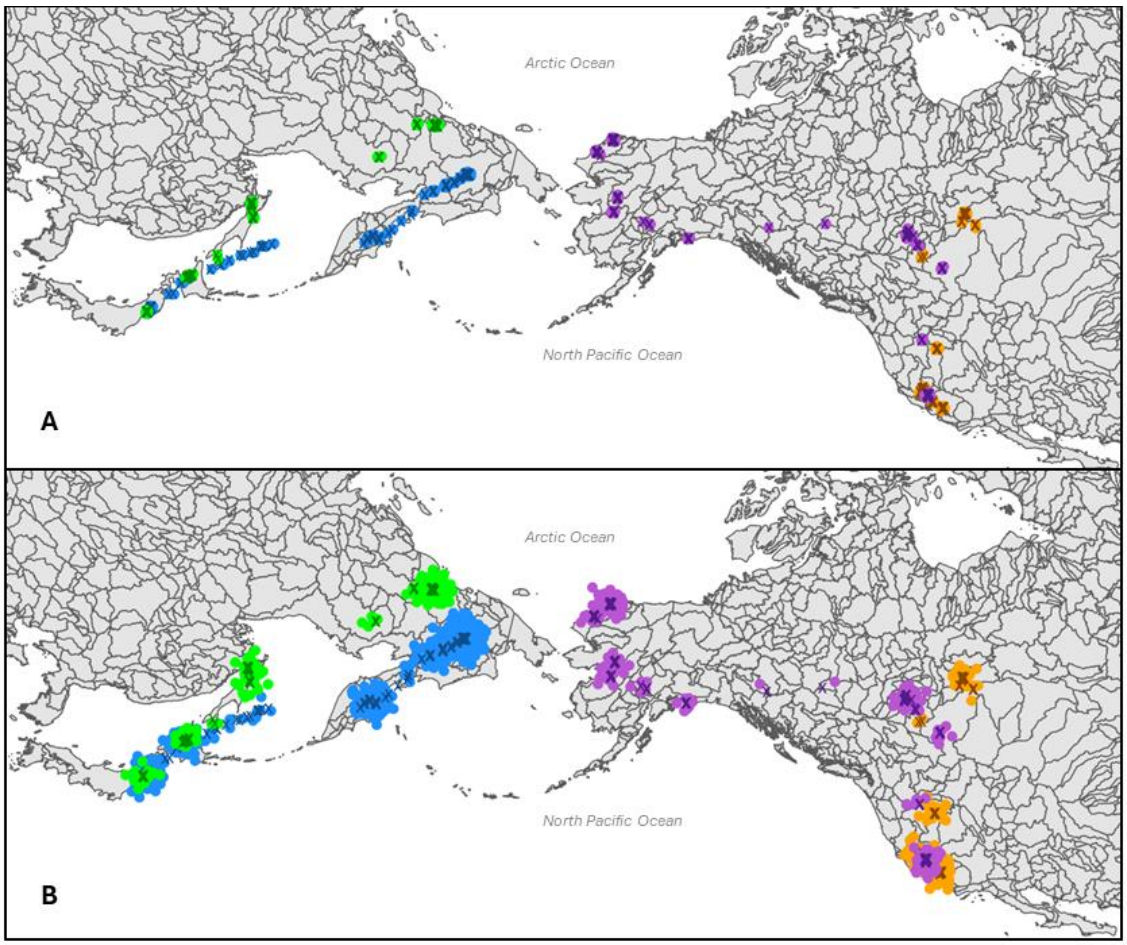

Figure 4.4: Simulations without error propagation for four Northern Pintail ducks, each shown in a different color. Dots show the simulated locations and x's show the observed locations. Panel A shows mean locations simulated from biologically derived quantities from the functional movement model. Panel B shows locations simulated from a uniform distribution with the same lower and upper bounds as the biologically derived distribution.

Table 4.2: Network statistics highlighting the difference in network structure when locations are simulated with biologically informed derived quantities. The biological and uniform networks were simulated without error propagation. Statistics for these networks are the median of the posterior distributions. 95% credible intervals are in parentheses and are bolded if they contain the observed network statistic. The observed network was constructed from the true locations of all 36 validation birds.

Network	# Watersheds	# Edges	Edge Density	Reciprocity	Diameter	Transitivity	# Strong Components
Observed	139	257	0.0134	0.4	42	0.23	39
Biological	148 **(138 - 238)**	291 (260 - 837)	0.0134 **(0.013-0.016)**	0.53 (0.45 - 0.66)	113 **(40 - 108)**	0.24 **(0.20 - 0.37)**	43 **(21 - 46)**
Uniform	240 **(64 - 353)**	949 **(175 - 1645)**	0.0165 **(0.012-0.045)**	0.7 (0.57 - 0.75)	105 **(30 - 314)**	0.46 (0.34 - 0.48)	18 (13 - 26)

4.3.2 Out of Sample Validation with Error Propagation

Each individual bird had a different number of successful simulations, and we were able to obtain posterior distributions for 36 of the 39 individuals withheld for simulation validation. Half of these individuals were marked in Japan and the other half were marked in California, USA. Those individuals with no successful simulations had the most time-steps to simulate due to their long tracking histories. There was a strong negative relationship between the number of time-steps and the number of successful simulations ($Rho = -0.73$, $p < 0.005$).

The simulation method with error propagation produced more realistic population scale migration paths than the straight-line network. The mark-recapture network included 69% of the watersheds in the observed network and shared five out of their top ten most heavily weighted connections. For all reported network statistics, except the number of watersheds and number of edges, the mark-recapture network values were closer to the observed network. Notably, the mark-recapture network had a similar number of reciprocated ties between watersheds as the observed network while the straight-line network had none. However, the simulation

overestimated the number of watersheds and the number of connections between them and underestimated the median in-degree and out-degree of the observed network.

The community structure of the mark-recapture network was similar to the observed network. Mark-recapture network transitivity and the number of strong components were closer in value to the observed network than the straight-line network was, and the 95% credible intervals for these statistics all included the observed network statistic (Table 4.3). The observed and mark-recapture networks had the same number of migratory communities, except one additional Asian community in the mark-recapture network. Mark-recapture network communities were more contiguous than the observed network which showed unvisited locations between some watersheds that were grouped into the same migratory communities.

Table 4.3: Network statistics showing whether the simulated mark-recapture network or the straight-line network better reproduced the observed population scale movement patterns. Statistics for the mark-recovery network are the median of the posterior distribution. 95% credible intervals are in parentheses and are bolded if they contain the observed network statistic. Credible intervals cannot be calculated for the straight-line network. The observed network was made with locations from the 32 validation birds that had successful simulations.

Network	# Watersheds	# Edges	Edge Density	Reciprocity	Diameter	Transitivity	# Strong Components
Observed	101	186	0.018	0.41	30	0.25	40
Mark-recapture	153 (128-258)	282 (252-619)	0.012 (0.008-0.015)	0.48 (0.46-0.65)	37 **(28-54)**	0.22 **(0.21-0.29)**	45 **(28-51)**
Straight-line	136	188	0.01	0	31	0.13	136

4.3.3 Northern Pintail Migration Patterns

In east Asia, the island of Sakhalin in Russia appears to be an important hub for Pintail ducks dispersing northward (Figure 4.5). This hub structure was maintained in the mark-recapture network but not in the straight-line network. In North America, a majority of the validation birds migrated north to Alaska, though a few traveled to Hudson Bay instead. On their

way north, most North American Pintail moved inland, away from the coast, before continuing north to Alaska. Simulated locations stayed closer to the west coast than what was actually observed in the validation dataset (Figure 4.5). However, costal migration paths were common in the dataset used to estimate the derived quantity posterior distributions and represent realistic migration paths for this species.

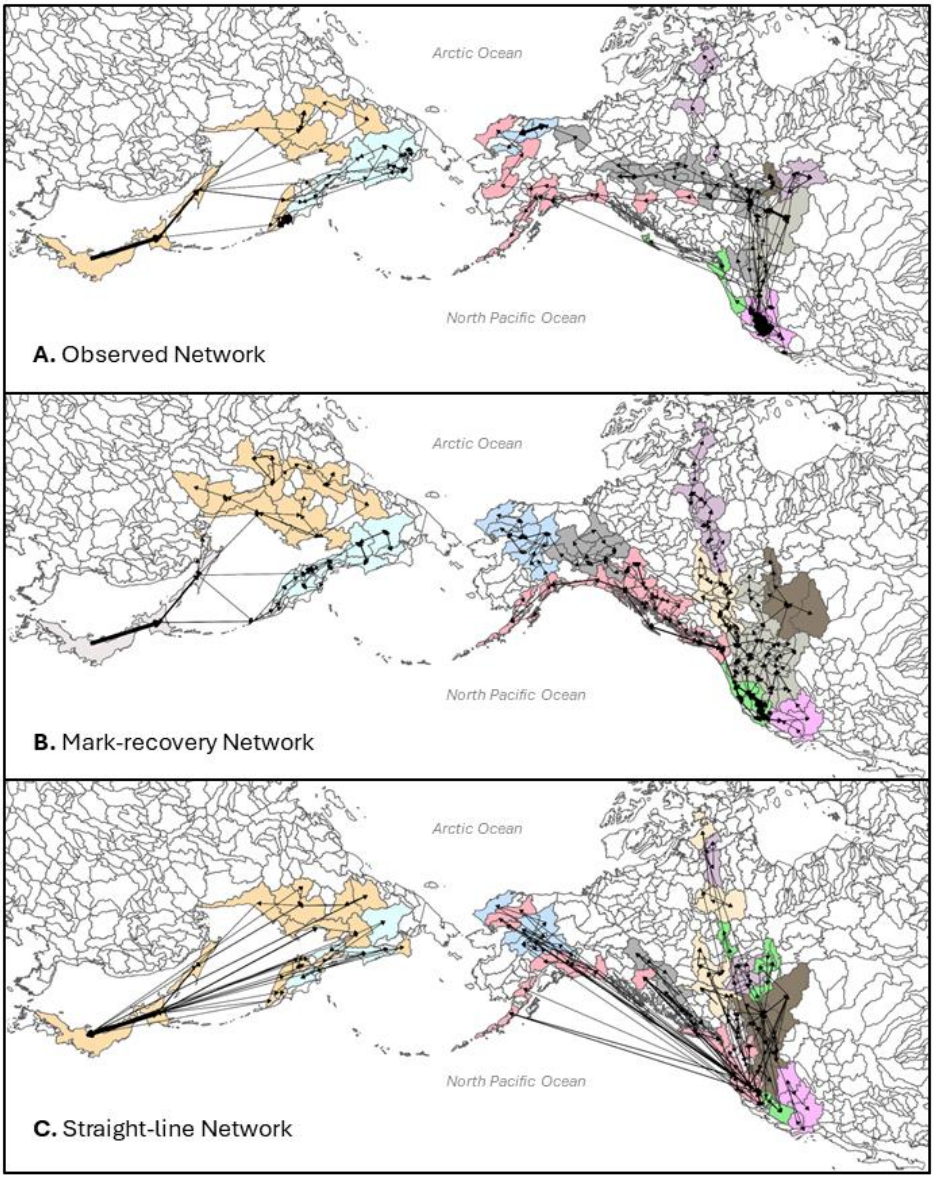

Figure 4.5: Maps showing the observed network (A), the mark-recovery network (B) and the straight-line network (C). The mark-recovery network was simulated with error propagation.

Arrows indicate the direction of travel and line thickness indicates the number of Northern Pintail ducks moving between the watersheds. Watersheds that are the same color were grouped into a migratory community with a community detection algorithm. Community color is consistent across all three maps.

4.4 DISCUSSION

Simulating locations without error propagation confirmed that the method fundamentally works. The resulting network captured almost all the watersheds that were actually visited by migrating individuals indicating a high degree of accuracy for a single time-step of simulation. Incorporating biological information about three characteristics of Pintail movement patterns greatly improved the accuracy and precision of the simulations. Statistics measuring the structure of the biological network were closer to what was actually observed and had far less uncertainty than the uniform network. When the uniform network was simulated with error propagation, the significant error introduced by ignoring the biological information resulted in mostly unusable simulations.

The biologically informed simulation and ABC rejection algorithm approach simulated realistic movement paths, with error propagation, for individuals with only two known locations. The method particularly excelled at recovering overall migratory community structure. However, the simulated network included several watersheds that were not visited by Pintail in the observed network resulting in more contiguous communities. This may be due to two, non-mutually exclusive factors. First, all simulated locations are subject to error, and this error increased the number of adjacent watersheds that were included in the network. Second, the derived quantities of speed, relative turning angle and direction were all drawn from population scale posterior distributions. The simulated networks may better represent the movement of the Pintail duck population at large, rather than the movement of these 36 specific individuals.

Indeed, the simulated movement networks are consistent with known migration patterns of Pintail ducks at large. Pintails that winter in Japan generally migrate north to Sakhalin Island in Russia, an important hub for northward migration to other regions of Russia. From there they disperse to the Kamchatka Peninsula or the Kolyma region of northern Russia, with some traveling far to the northeast to the Chukotka Peninsula (Hupp et al. 2011). With only the known starting and ending locations of 18 individuals marked in Japan, we were able to reconstruct these movement patterns including the hub structure of Sakhalin Island. In North America, Pintails marked in California generally use three different migration paths (Miller et al. 2005 and 2010). Some individuals first travel to the Prairie Pothole Region (PPR) in southern Alberta and then continue to Alaska or Hudson Bay. Others follow the western coastline, while others fly over the ocean directly to Alaska. Again, with only 18 individuals we were able to accurately simulate migratory paths to the PPR, Hudson Bay and Alaska that reflect these known migration strategies.

There is strong interest in whether large-scale movements of Pintail ducks, and other migratory waterfowl, are responsible for the spatial transmission of Influenza A virus (Lam et al. 2012, Russel et al. 2016). The lack of detailed movement data at the continental scale has been cited as a roadblock to understanding the role of migratory host movement in the spatial spread of the virus (Hupp et al. 2011, Trovao et al. 2015, Yang et al. 2024). Our simulation method can be used to address this issue, providing detailed and accurate locations with estimates of error to help predict large-scale movements and introductions of the virus.

An additional benefit of our simulation method is that it performed well without incorporating any information about the environment the animals move through. Migratory waterfowl require a large number of calories during migration and flight paths are thought to be

chosen based on weather and land use factors that impact foraging conditions (Si et al. 2015). However, these types of data are not always available, particularly in the case of historical movement datasets when land use and even climate could have been drastically different. Additionally, the influence of the environment on animal movement is often scale dependent and requires knowledge of which data are most predictive of movement at the desired scale of inference (Avgar et al. 2012). While factors that influence large-scale waterfowl migration have been fairly well studied, information on the influence of the environment on movement patterns of other organisms is generally scarce. Our method provides a useful alternative to infer movement patterns from mark-recapture data when environmental data are unavailable. For those species for which we have a detailed understanding about how environmental conditions influence movement, our method could be extended to incorporate this information to reduce error propagation.

Overall, our method is well suited to highly directed, large-scale movement like that of migratory birds. The method is not appropriate for simulating round-trip paths and may struggle to reproduce highly curved or sinuous trajectories. Using the mean or median of the posterior distributions of location tends to smooth and straighten trajectories, while more extreme changes in locations are found in the tails of these distributions. This makes the scale of inference a particularly important choice before implementing this method. Animal paths generally appear random or highly sinuous when viewed at close scale, but zooming out to a larger spatial and temporal scale often reveals directed movement patterns. The chosen spatial and temporal scale of simulations should reflect this consideration for the animal of interest. Additionally, this method is best for recovering population scale movement patterns rather than reconstructing movement paths of particular individuals. Studies using movement data to learn about the

foraging and territorial behavior of animals (Fuller and Harrison 2010), habitat preferences (Maciel and Lutscher 2013), and disease outbreaks (Poletto et al. 2012, VanderWaal and Ezenwa 2016) are often interested in how individual heterogeneity impact these phenomena. Locations simulated with our method are likely not appropriate for addressing these local-scale questions. Additionally, our method is restricted to simulating movement in two dimensions limiting its use for species that move in three dimensions like aquatic and arboreal animals.

Computation power and time are an important consideration for any simulation-based method. We were able to complete 100,000 simulations for each of the 36 validation individuals on a desktop computer in three days when running in parallel on 18 cores. Implementing our method on a large number of individuals in a short amount of time is best done in parallel on a high-performance computing cluster. For those that already have access to high performance computing our method is more cost effective, and requires less effort, than marking additional individuals with GPS tags with an average cost of $2000 USD each. It is important to note that simulation time is greatly dependent on the number of time-steps that are being simulated. Choosing larger increments of time will greatly speed up the simulation process, increase the number of successful simulations, and improve estimates of location error. The estimation of simulation error provided by ABC provides important context for how movement patterns may vary and how this impacts analyses that use the simulated data.

Our method can be used to increase the spatio-temporal resolution of existing mark-recapture data to match more detailed tracking data. This provides a robust framework to integrate multi-sensor datasets. Implementing our method on lower cost mark-recovery data allows researchers a cost-effective way to extract more value and information from both detailed tracking data and mark-recapture data. Our method is easy to implement across large geographic

areas and provides detailed and realistic location information. Until GPS tracking technology becomes affordable enough to deploy on thousands of individuals at a time, our simulation method represents a compelling alternative with many of the characteristics of the 'ideal' tracking method.

Chapter 5: Conclusion

In this dissertation I have examined the relative contribution of local and regional drivers of the spatial distribution, prevalence, and spatial transmission of Influenza A virus in endemic hosts (IAV). In Chapter 2 I showed that regional host migration better predicted the large-scale seasonal patterns of prevalence within the United States. While there was some indication that temperature may play a role in shaping spatial patterns of prevalence it is likely that at the watershed scale this factor better explains the spatial distribution of hosts rather than the environmental persistence of the virus. In Chapter 3, I am the first to directly link host movement data to spatial transmission of IAV and show that large-scale Mallard movement was an important predictor of where the virus moves within the US. Interestingly, Mallard movement was closely associated with spatial transmission between many species of waterfowl further supporting the idea that Mallards likely have a special role in the spatial dissemination of potentially highly pathogenic IAV in North America.

Results from both of these chapters show that understanding where hosts move will be an important component for prioritizing surveillance efforts for IAV as well as predicting future invasions of the virus. As waterfowl migration patterns shift with changing temperatures and altered agricultural land use, we can expect that both the prevalence of the virus, and where it moves, to shift as well. To understand if these findings are generalizable to other host species, and low pathogenic IAV strains, we need to expand both these analyses to other taxa and locations. This highlights the importance of Chapter 4 where I developed a host movement simulation method that addressed two pressing needs. First, the need for a rigorous framework to integrate movement datasets collected at different spatio-temporal scales. Second, the need to

extract as much information as possible from the movement data that are available. Movement networks constructed from simulated locations will allow us to extend the analyses in Chapters 2 and 3 to larger geographic areas and other host species.

Linking local conditions to larger-scale IAV distribution and movement is also complicated by the lack of environmental data from viral sampling locations. There are many unanswered questions about how long the virus survives under natural and heterogeneous conditions, how different subtypes of the virus response to these conditions, and how other organisms in aquatic environments mediate transmission. In this dissertation I have had to rely on remotely sensed temperature and omit other potentially important measures of environmental suitability such as pH and salinity. Collecting data on water chemistry (e.g. temperature, salinity, pH) at the locations where viral sampling occurs, as well as environmental sampling for the virus, will help us to better understand its ability to persist in natural environments. These data may also reveal a stronger influence on large-scale prevalence and spatial transmission.

The results in both Chapters 2 and 3 showed strong support for regional phenomena driving large-scale spatial patterns, but little influence of local factors in predicting either prevalence or spatial transmission of IAV at the large scale. Multiscale analyses of disease systems like ours are rare but are critically important when comparing mechanisms across scales (Garira 2020). A study at the scale of sampling sites may find that water temperature is hugely important for predicting local prevalence, while a study at a larger scale may show no effect at all. Neither mechanism is unimportant but one is more predictive at the scale of interest. This scale dependence is a well-known challenge in ecology (Levin 1992) but is particularly important when studying ecological phenomena at large spatial scales. The larger the scale, the

more possible mechanisms there are to shape the patterns we see, from individual host immunity to global climate.

Collecting data at large spatial scales is logistically challenging, particularly in wildlife populations (See Stallknecht 2007, Mysterud et al. 2023). However, these challenges should not dissuade us from tackling important large-scale questions about how pathogen distributions are changing. Landscape and macroecological scale studies of disease systems will provide crucial information for designing wildlife surveillance programs to help us anticipate zoonotic spillover events (Stephens 2016, Delgado et al. 2023). Characterizing large-scale host and vector movement patterns, and understanding how they are changing, will help us model the global spread of pathogens and disease. However, in a recent survey of disease ecology practitioners both movement ecology and landscape ecology were among the least popular areas of research today (Brandell et al. 2021). It is time for this to change. Movement ecology offers quantitative tools to get the most out of existing animal movement data and exciting opportunities to quantify the role of host mobility in both the local and global spread of disease. Looking forward, there should be a greater emphasis on incorporating the skills, tools, and perspectives from these disciplines into our study of infectious disease.

Literature Cited

Aarts, G., M. MacKenzie, B. McConnell, M. Fedak, and J. Matthiopoulos. 2008. Estimating space-use and habitat preference from wildlife telemetry data. Ecography 31: 140-160.

Airey, M., and K. R. Short. 2024. High pathogenicity avian influenza in Australia and beyond: could avian influenza cause the next human pandemic? Microbiology Australia. https://doi.org/10.1071/MA24040.

Alfonso, C. P., B. S. Cowan, and H. van Campen. 1995. Influenza A viruses isolated from waterfowl in two wildlife management areas of Pennsylvania. Journal of Wildlife Diseases, 31: 179–185.

Altizer S., R. Bartel, and B. A. Han. 2011. Animal migration and infectious disease risk. Science 331:296-302. doi: 10.1126/science.1194694. PMID: 21252339.

Altizer, S., R. S. Ostfeld, P. T. J. Johnson, S. kutz, and C. D. Harvell. 2013. Climate change and infectious diseases: From evidence to a predictive framework. Science 341: 514-519.

Avgar, T., A. Mosser, G. S. Brown, and J. M. Fryxell. 2012. Environmental and individual drivers of animal movement patterns across a wide geographical gradient. Journal of Animal Ecology 82: 96-106. https://doi.org/10.1111/j.1365-2656.2012.02035.x

Bahl, J., S. Krauss. D. Kuhnert, M. Fourment, G. Raven, S. P. Pryor, L. J. Niles, A. Danner, D. Walker, I. H. Mendenhall, Y. C. F. Su, V. G. Dugan, R. A. Halpin, T. B. Stockwell, R, J. Webby, D. E. Wentworth, A. J. Drummond, G. J. D. Smith, and R. G. Webster. 2013. Influenza A virus migration and persistence in North American wild birds. PLoS Pathogens 9: https://doi.org/10.1371/journal.ppat.1003570.

Bastille-Rousseaum G., and G. Wittemyer. 2020. Characterizing the landscape of movement to identify critical habitat and corridors. Conservation Biology 35: 346-359. https://doi.org/10.1111/cobi.13519

Bastille-Rousseaum G., I. Douglas-Hamilton, S. Blake, J. M. Northrup, and G. Wittemyer. 2018. Applying network theory to animal movements to identify properties of landscape space use. Ecological Applications 28: 854-864. https://doi.org/10.1002/eap.1697.

Beatty, W. S., D. C. Kesler, E. B. Webb, L. W. Naylor, A. H. Raedeke, D. D. Humburg, J. M. Coluccy, and G. J. Soulliere. 2016. How will predicted land-use change affect waterfowl spring stopover ecology? Inferences from an individual-based model. Journal of Applied Ecology 54: 926-934. https://doi.org/10.1111/1365-2664.12788.

Becker, D. J., A. D. Washburne, C. L. Faust, E. A. Mordecai, and R. K. Plowright. 2019. The problem of scale in the prediction and management of pathogen spillover. Phil. Trans. R. Soc. B 374: 20190224 http://doi.org/10.1098/rstb.2019.0224

Beerens, N., R. Heutink, S. A. Bergervoet, F. Harders, A. Bossers, and G. Koch. 2016. Multiple reassorted viruses as cause of highly pathogenic avian influenza A(H5N8) virus epidemic, the Netherlands, 2016. 2017. Emerging Infectious Diseases 23: 1974-1981. https://doi.org/10.3201/eid2312.171062

Belkhiria, J., M. A. Alkhamis, and B. Martiez-Lopez. 2016. Application of species distribution modeling for avian influenza surveillance in the United States considering the North America migratory flyways. Scientific Reports 6: https://doi.org/10.1038/srep33161

Bennison, A., S. Bearhop, T. W. Bodey, S. C. Votier, W. J. Grecian, E. D. Wakefield, K. C. Hamer, and M. Jessopp. 2017. Search and foraging behaviors from movement data: A

comparison of methods. Ecology and Evolution 8: 13-24. https://doi.org/10.1002/ece3.3593.

Besag, J. 1974. Spatial interaction and the statistical analysis of lattice systems. Journal of the Royal Statistical Society Series B (Methodological): 192-236.

Bevins, S. N., K. Pedersen, M. W. Lutman, J. A. Baroch, B. S. Schmit, D. Kohler, T. Gidlewski, D. L. Nolte, S. R. Swafford, and T. J. DeLiberto. 2014. Large-scale avian influenza surveillance in wild birds throughout the United States. PLoS ONE: 9, e104360. https://doi.org/10.1371/journal.pone.0104360

Blaurock, C., A. Breithaupt, D. Scheibner, O. Bagato, A. Karger, T. C. Mettenleiter, and E. M. Abdelwhab. 2021. Preferential selection and contribution of non-structural protein 1 (NS1) to the efficient transmission of panzootic avian influenza H5N8 virus clades 2.3.4.4A and B in chickens and ducks. Journal of Virology 95: https://doi.org/10.1128/jvi.00445-21.

Bodenhofer U., E. Bonatesta, C. Horejš-Kainrath, and S. Hochreiter. 2015 Msa: An R package for multiple sequence alignment. Bioinformatics 31: 3997–3999.

Brandell, E. E., D. J. Becker, L. Sampson, and K. M. Forbes. 2021. Demography, education, and research trends in the interdisciplinary field of disease ecology. Ecology and Evolution 11: 17581-17592. https://doi.org/10.1002/ece3.8466.

Breban, R., J. M. Drake, D. E. Stallknecht, and P. Rohani. 2009. The role of environmental transmission in recurrent avian influenza epidemics. PLoS Computational Biology 5: https://doi.org/10.1371/journal.pcbi.1000346.

Breed, G.A., D. P. Costa, I. D. Jonsen, P. W. Robinson and J. Mills-Flemming. (2012) State-space methods for more completely capturing behavioral dynamics from animal tracks. Ecological Modelling 235: 49–58.

Brooks, S. P. and A. Gelman. 1998. General methods for monitoring convergence of iterative simulations. Journal of Computational and Graphical Statistics 7: 434-455.

Brown, J. D., D. E. Swayne, R. J. Cooper, R. E. Burns, and D. E. Stallknecht. 2007. Persistence of H5 and H7 avian influenza viruses in water. Avian Diseases 51: 285-289. https://doi.org/10.1637/7636-042806R.1

Brown, J. D., G. Goekjian, R. Poulson, S. Valeika, and D. E. Stallknecht. 2009. Avian influenza virus in water: Infectivity is dependent on pH, salinity and temperature. Veterinary Microbiology 136: 20-26. doi:10.1016/j.vetmic.2008.10.027.

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

Buhnerkempe, M. G., C. T. Webb, A. A. Merton, J. E. Buhnerkempe, G. H. Givens, R. S. Miller, and J. A. Hoeting. Identification of migratory bird flyways in North America using community detection on biological networks. Ecological Applications 26: 740-751 https://doi.org/10.1890/15-0934.

Caron, A., V. Grosbois, E. Etter, N. Gaidet, and M. de Garine-Wichatitsky. 2014. Bridge hosts for avian influenza viruses at the wildlife/domestic interface: An eco-epidemiological framework implemented in southern Africa. Preventative Veterinary Medicine 117: 590-600.

Cassie, D. 2006. The thermal regime of rivers: a review. Freshwater Biology 51: 1389-1406.

Cohen, J. M., D. J. Civitello, A. J. Brace, E. M. Feichtinger, C. N. Ortega, J. C. Richardson, E. L. Sauer, X. Liu, and J. R. Rohr. 2016. Spatial scale modulates the strength of ecological processes driving disease distributions. Proc Natl Acad Sci U S A 113: E3359-64. doi: 10.1073/pnas.1521657113. Epub 2016 May 31. PMID: 27247398; PMCID: PMC4914148.

Costa, T. P., J. D. Brown, E. W. Howerth, and D. E. Stallknecht. 2011. Variation in viral shedding patterns between different wild bird species infected experimentally with low-pathogenicity avian influenza viruses that originated from wild birds. Avian Pathology 40: 119-124. https://doi.org/10.1080/03079457.2010.540002

Cross, P. L., D. Caillaud, and D. M. Heisey. 2013. Underestimating the effects of spatial heterogeneity due to individual movement and spatial scale: infectious disease as an example. Landscape Ecology 28: 247-257. doi :10.1007/s10980-012-9830-4

Daszak, P., A. A. Cunningham, and A. D. Hyatt. 2001. Anthropogenic environmental change and the emergence of infectious diseases in wildlife. Acta Tropica 78: 103-116. https://doi.org/10.1016/S0001-706X(00)00179-0.

Davidson, S., G. Bohrer, E. Gurarie, S. Lapoint, P. J. Mahoney, N. T. Boelman, J. U. H. Eitel, L. R. Prugh, Lee A. Vierling et al. 2020. Ecological insights from three decades of animal movement tracking across a changing Arctic. Science 370: 712-715. doi: 10.1126/science.abb7080

Delgado, M., N. Ferrari, A. Fanelli, S. Muset, L. Thompson, J. M. Sleeman, C. L. White, D. Walsh, C. Wannous, and P. Tizzani. 2023. Wildlife health surveillance: gaps, needs and opportunities. Rev. Sci. Tech. 42: 161-172. doi: 10.20506/rst.42.3359.

Dougherty, E. R., D. P. Seidel, C. J. Carlson, O. Spiegel, and W. M. Getz. 2018. Going through the motions: incorporating movement analyses into disease research. Ecology Letters 21: 588-604. https://doi.org/10.1111/ele.12917.

Douglas, D. C., R. Weinzierl, S. C. Davidson, R. Kays, M. Wikelski, and G. Bohrer. 2012. Moderating Argos location errors in animal tracking data. Methods in Ecology and Evolution 3: 999-1007. doi: 10.1111/j.2041-210X.2012.00245.x.

Douglas, D. C., R. Weinzierl, S. C. Davidson, R. Kays, M. Wikelski, and G. Bohrer. 2012. Moderating Argos location errors in animal tracking data. Methods in Ecology and Evolution 3: 999-1007. https://doi.org/10.1111/j.2041-210X.2012.00245.x.

eBird Basic Dataset. Version: EBD_relJul-2022. Cornell Lab of Ornithology, Ithaca, New York. Jul 2022.

Fan, Y., and H. van den Dool. 2008. A global monthly land surface air temperature analysis for 1948-present, J. Geophysical Research 113, D01103, doi:10.1029/2007JD008470.

Farnsworth, M. L., J. A. Hoeting, N. T. Hobbs, and M. W. Miller. 2006. Linking chronic wasting disease to mule deer movement scales: a hierarchical Bayesian approach. Ecological Applications 16: 1026-1036.

Farnsworth, M. L., Kendall, W. L., P. F. Doherty, R. S. Miller, G. C. White, J. D. Nichols, K. P. Burnham, and A. B. Franklin. 2011. Targeted surveillance for highly pathogenic avian influenza in migratory waterfowl across the conterminous United States, Chapter 12. Pandemic influenza viruses: Science, surveillance and public health. The Pennsylvania Academy of Science, Easton, PA.

Farnsworth, M. L., Miller, R. S., Pedersen, K., Lutman, M. W., Swafford, S. R., Riggs, P. D., and Webb, C. T. 2012. Environmental and demographic determinants of avian influenza

viruses in waterfowl across the contiguous United States. PLoS ONE 7: e32729.https://doi.org/10.1371/journal.pone.00327

Fourment, M., A. E. Darling, and E. C. Holmes. 2017. The impact of migratory flyways on the spread of avian influenza virus in North America. BMC Evolutionary Biology 17 https://doi.org/10.1186/s12862-017-0965-4

Fuller, A. K., and D. J. Harrison. 2010. Movement paths reveal scale-dependent habitat decisions by Canada lynx. Journal of Mammalogy 91: 1269-1279. https://doi.org/10.1644/10-MAMM-A-005.1

Gaidet, N., A. Caron, J. Cappelle, G. S. Cumming, G. Balança, S. Hammoumi, G. Cattoli, C. Abolnik, R. Servan de Almeida, P. Gil, S. R. Fereidouni, V. Grosbois, A. Tran, J. Mundava, B. Fofana, A. B. Ould El Mamy, M. Ndlovu, J. Y. Mondain-Monvalr, P. Triplet, W. Hagemeijer, W. B. Karesh, S. H. Newman, and T. Dodman. 2012. Understanding the ecological drivers of avian influenza virus infection in wildfowl: a continental-scale study across Africa. Proc. R. Soc. B 279: 1131-1141.

Garamszegi, L. Z. and A. P. Moller. 2007. Prevalence of avian influenza and host ecology. Proc. R. Soc. B 274: 2003-2012.

Garira, W. 2020. The research and development of process for multiscale models of infectious disease systems. PLoS Comput Biol. 16: 10.1371/journal.pcbi.1007734

Gelman, A., A. Jakulin, M. Grazia Pittau, and Y. Su. 2008. A weakly informative default prior distribution for logistic and other regression models. The Annals of Applied Statistics 2: 1360-1383.

Gilbert M., and D. U. Pfeiffer. 2012. Risk factor modeling of the spatio-temporal patterns of highly pathogenic avian influenza (HPAIV) H5N1: A review: Spatial and Spatio-temporal Epidemiology 3: 173-183.

Giuggioli, L, and V. M. Kenkre. 2014. Consequences of animal interactions on their dynamics: emergence of home ranges and territoriality.

Gorsich, E. E., C. T. Webb, A. A. Merton, J. A. Hoeting, R. S. Miller, M. L. Farnsworth, S. R. Swafford, T. J. DeLiberto, K. Pedersen, A. B. Franklin, R. G. McLean, K. R. Wilson, and P. F. Doherty Jr. 2021. Continental-scale dynamics of avian influenza in U.S. waterfowl are driven by demography, migration, and temperature. Ecological Applications 31: doi: 10.1002/eap.2245.

Grenfell, B. T., O. N. Bjornstad, and J. Kappey. 2001. Travelling waves and spatial hierarchies in measles epidemics. Nature 141: 716-723.

Guan, G., Y. Dery, M. Yechezkel, I. Ben-Gal, D. Yamin, and M. L. Brandeu. 2021. Early detection of COVID-19 outbreaks using human mobility data. PLOS One 16: https://doi.org/10.1371/journal.pone.0253865.

Hanks, E. 2018. ctmcmove: Modeling animal movement with continuous-time discrete-space Markov chains. R package version 1.2.9. https://CRAN.R-project.org/package=ctmcmove.

Haukos, D. A., M. R. Miller, D. L. Orthemeyer, J. Y. Takekawa, J. P. Fleskes, M. L. Casazza, W. M. Perry, J. A. Moon. 2006. Spring migration of Northern Pintails from Texas and New Mexico, USA. Waterbirds 29: 127-136. https://doi.org/10.1675/1524-4695.

Hesterbeek, H., R. M. Anderson, V. Andreasen, S. Bansal, D. De Angelis, C. Dye, K. T. D. Eames, W. J. Edmunds, S. D. W. Frost, S. Funk, T. D. Hollingsworth, V. Isham, P.

Klepac, J. Lessler, J. O. Lloyd-Smith, C. J. E. Metcalf, D. Mollinson, L. Pellis, J. R. C. Pulliam, M. G. Roberts, C. Viboud. 2015. Modeling infectious disease dynamics in the complex landscape of global health. Science 347: https://doi.org/10.1126/science.aaa4339.

Hill, N. J., E. J. Ma, B. W. Meixell, M. S. Lindberg, W. M. Boyce, and J. A. Runstadler. 2016. Transmission of influenza reflects seasonality of wild birds across the annual cycle. Ecology Letters 19: 823-1006. doi: 10.1111/ele.12629.

Hill, N. J., Takekawa, J. Y., Cardona, C. J., Meixell, B. W., Ackerman, J. T., Runstadler, J. A., and Boyce, W. M. 2012. Cross-seasonal patterns of avian influenza virus in breeding and wintering migratory birds: a flyway perspective. Vector-Borne and Zoonotic Diseases 12: 243–253. https://doi.org/10.1089/vbz.2010.024

Hill, S. C., R. Hansen, S. Watson, V. Coward, C. Russell, J. Cooper, S. Essen, H. Everest, K. V. Parag, S. Fiddaman, S. Reid, N. Lewis, S. M. Brookes, A. L. Smith, B. Sheldon, C. M. Perrins, I. H. Brown, and O. G. Pybus. 2019. Comparative micro-epidemiology of pathogenic avian influenza virus outbreaks in a wild bird population. Philosophical Transactions of the Royal Society B 374: https://doi.org/10.1098/rstb.2018.0259.

Hinshaw. V. S., R. G. Webster, and B. Turner. 1980. The perpetuation of orthomyxoviruses and paramyxoviruses in Canadian waterfowl. Canadian Journal of Microbiology, 26, 622–629.

Horne, J. S., E. O. Garton, S. M. Krone, and J. S. Lewis. 2007. Analyzing animal movements using Brownian bridges. Ecology 88: 2354 – 2363.

Hupp, J. W., N. Yamaguchi, P. L. Flint, J. M. Pearce, K. Tokita, T. Shimada, A. Ramey, S. Kharitonov, and H. Higuchi. 2011. Variation in spring migration routes and breeding

distribution of Northern Pintails *Anas acuta* that winter in Japan. Journal of Avian Biology 42: 289-300. https://doi.org/10.1111/j.1600-048X.2011.05320.x.

Hupp, J.W., T. L. Tibbitts, and D. C. Douglas. 2019. Tracking data for Northern Pintails (Anas acuta) (ver 1.0, September 2019): U.S. Geological Survey data release, https://doi.org/10.5066/P90A6HW3

Ip H. S., P. L. Flint, J. C. Franson, R. J. Dusek, D. V. Derksen, R. E. Gill, C. R. Ely, J. M. Pearce, R. B. Lanctot, S. M. Matsuoka, D. B. Irons, J. B. Fischer, R. M. Oates, M. R. Peterson, T. F. Fondell, D. A. Rocque, J. C. Pederson and T. C. Rothe. 2008. Prevalence of influenza A viruses in wild migratory birds in Alaska: patterns of variation in detection at a crossroads of intercontinental flyways. Virology Journal 5: 71.

Jacoby, D. M. P. and R. Freeman. 2016. Emerging network-based tools in movement ecology. Trends in Ecology and Evolution 31: 301-314.

Jacoby, D. M. P., E. J. Brooks, D. P. Croft, and D. W. Sims. 2012. Developing a deeper understanding of animal movements and spatial dynamics through novel application of network analyses. Methods in Ecology and Evolution 3: 574-583. https://doi.org/10.1111/j.2041-210X.2012.00187.x

Johnson, D. S., J. M. London, M. A. Lea, and J. W. Durban. 2008. Continuous-time correlated random walk model for animal telemetry data. Ecology 89: 1208-1215.

Jombart, T., R. M. Eggo, P. J. Dodd, and F. Balloux. 2010. Reconstructing disease outbreaks from genetic data: a graph approach. Heredity 106: 383-390.

Jovani, R. and J. Tella. 2006. Parasite prevalence and sample size: misconceptions and solutions. Trends in Parasitology 22: 214-218.

Keeling, M. J., M E. Woolhouse, D. J. Shaw, L. Matthews, M. Chase-Topping, D. T. Haydon, S. J. Cornell, J. Kappey, J. Wilesamith, and B. T. Grenfell. 2001. Dynamics of the 2001 UK foot and mouth epidemic: stochastic dispersal in a heterogeneous landscape. Science 294: 813-817. 10.1126/science.1065973.

Kent, C. M, A. M. Ramey, J. T. Ackerman, J. Bahl, S. N. Bevins, A. S. Bowman, W. M. Boyce, C. J. Cardona, M. L. Casazza, T. D. Cline, S. E. De La Cruz, J. S. Hall, N. J. Hill, H S. Ip, S. Krauss, J. M. Mullinax, J. M. Nolting, M. Plancarte, R. L. Poulson, J. A. Runstadler, R. D. Slemons, D. E. Stallknecht, J. D. Sullivan, J. Y. Takekawaa, and D. J. Prosser. Spatiotemporal changes in influenza A virus prevalence among wild waterfowl inhabiting the continental United States throughout the annual cycle. Scientific Reports 12: 13083. https://doi.org/10.1038/s41598-022-17396-5

Klaassen, M. and M. Wille. 2023. The plight and role of wild birds in the current bird flu panzootic. Nat Ecol Evol 7, 1541–1542. doi:10.1038/s41559-023-02182-x.

Kraemer, M. U. G., N. Golding, D. Bisanzio, S. Bhatt, D. M. Pigott, S. E. Ray, O. J. Brady, J. S. Brownstein, N. R. Faria, D. A. T. Cummings, O. G. Pybus, D. L. Smith, A. J. Tatem, S. I. Hay, and R. C. Reiner Jr. 2019. Utilizing general human movement models to predict the spread of emerging infectious diseases in resource poor settings. Scientific Reports 9: https://doi.org/10.1038/s41598-019-41192-3.

Lam, T. T., H. S. Ip, E. Ghedin, D. E. Wentworth, R. A. Halpin, T. B. Stockwell, D. J. Spiro, R. J. Dusek, J. B. Bortner, J. Hoskins, B. D. Bales, D. R. Yparraguirre, and E. C. Holmes. 2011. Migratory flyway and geographical distance are barriers to the gene flow of influenza virus among North American birds. Ecology Letters 15: 24-33.

Lemey, P., A. Rambaut, T. Bedford, N. Faria, F. Bielejec, G. Baele, C. A. Russell, D. J. Smith, O. G. Pybus, D. Brockman, M. A. Suchard. 2014. Unifying viral genetics and human transportation data to predict the global transmission dynamics of human influenza H3N2. PLOS Pathogens 10: https://doi.org/10.1371/journal.ppat.1003932

Levin, S. A. 1992. The problem of pattern and scale in ecology: The Robert H. MacArthur Award Lecture. Ecology 73: 1943-1967.

Leyson, C. M., S. Youk, H. L. Ferreira, D. L. Suarez, and M. Pantin-Jackwood. 2021. Multiple gene segments are associated with enhanced virulence of clade 2.3.4.4 H5N8 highly pathogenic avian influenzas virus in Mallards. Journal of Virology 95: https://doi.org/10.1128/jvi.00955-21.

Lycett, S. J., A. Pohlmann, C. Staubach, V. Caliendo, M. Woolhouse., M. Beer, T. Kuiken, and Global Consortium for H5N5 and Related Influenza Viruses. 2020. Genesis and spread of multiple reassortants during the 2016/2017 H5 avian influenza epidemic in Eurasia. PNAS 117: ttps://doi.org/10.1073/pnas.2001813117

Maciel, G. A., and F. Lutscher. 2013. How individual movement response to habitat edges affects population persistence and spatial spread. The American Naturalist 182: 42-45. https://doi.org/10.1086/670661.

McDermott, P. L., C. K. Wikle, and J. Millspaugh. 2017. Hierarchical nonlinear spatio-temporal agent-based models for collective animal movement. Journal of Agriculture, Biological, and Environmental Statistics 22: 2094-312.

McGill, B. J. 2010. Ecology. Matters of scale. Science 328: 575-576. doi: 10.1126/science.1188528.

McLean, D. J., and Skowron Volponi, M. A. 2018. trajr: An R package for characterization of animal trajectories. Ethology 124: 440-448. https://doi.org/10.1111/eth.12739.

Meentemeyer, R. K., S. E. Haas, and T. Václavík. 2012. Landscape epidemiology of emerging infectious diseases in natural and human-altered ecosystems. Annual Review of Phytopathology 50: 379-402.

Miller, M. R., J. Y. Takekawa, D. S. Battaglia, R. T. Golightly, and W. M. Perry. 2010. Spring migration and summer destinations of Northern Pintails from the coast of Southern California. The Southwestern Naturalist 55: 501-509.

Miller, M. R., J. Y. Takekawa, J. P. Fleskes, D. L. Orthmeyer, M. L. Casazza, and W. M. Perry. 2005. Spring migration of Northern Pintails from California's Central Valley wintering area tracked with satellite telemetry: routes, timing, and destinations. Canadian Journal of Zoology 83: 1314-1332.

Mysterud, A., H. Viljugrein, P. Hopp, R. Anderson, H. Bakka, S. L. Benestad, K. Madslein, T. Moldal, G. R. Rauset, O. Strans, L. Tran, T. Vikoren, J. Vage, and C. M. Rolandsen. 2023. Challenges and opportunities using hunters to monitor chronic wasting disease among wild reindeer in the digital era. Ecological Solutions and Evidence 4: https://doi.org/10.1002/2688-8319.12203.

Nallar, R., Z. Papp, T. Epp, F. A. Leighton, S. R. Swafford, T. J. DeLiberto, R. J. Dusek, H. S. Ip, J. Hall, Y. Berhane, S. E. J. Gibbs, and C. Soos. 2015. Demographic and spatiotemporal patterns of avian influenza infection at the continental scale, and in relation to annual life cycle of migratory host. PLoS ONE 10: e0130662. Doi: 10.1371/journal.pone.0130662.

Nathan, R., W. M. Getz, E. Revilla, M. Holyoak, R. Kadmon, D. Saltz, and P. E. Smouse. 2008. A movement ecology paradigm for unifying organismal movement research. PNAS 105: 19052-19059.

Olsen, B., V. J. Munster, A. Wallensten, J. Waldenstrom, A. D. M. E. Osterhaus, and R. A. M. Fouchier. 2006. Global patterns of influenza A virus in wild birds. Science 312: 384-388.

Ostfeld, R. S., G. E. Glass, and F. Keesing. 2005. Spatial epidemiology: an emerging (or re-emerging) discipline. Trends in Ecology and Evolution 20: 328-336.

Paradis E. and K. Schliep. 2019. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics, 35, 526-528. doi:10.1093/bioinformatics/bty633.

Patterson, T. A., B. J. McConnell, M. A. Fedak, M. V. Bravington, and M. A. Hindell. 2010. Using GPS data to evaluate the accuracy of state-space methods for correction of Argos satellite telemetry error.

Plaza, P. I., V. Gamarra-Toledo, J. Rodriguez Engui, and S. A. Lambertucci. 2024. Recent changes in patterns of mammal infection with highly pathogenic avian influenza A(H5N1) virus worldwide. Emerging Infectious Diseases 30: ISSN: 1080-6059

Plowright, R. K., P. Foley, H. E. Field, A. P. Dobson, J. E. Foley, P. Eby, and P. Daszak. 2011. Urban habituation, ecological connectivity and epidemic dampening: the emergence of Hendra virus from flying foxes (Pteropus spp.). Proc. R. Soc. B. 278: 3703–3712. https://doi.org/10.1098/rspb.2011.0522.

Plummer, M. 2023. rjags: Bayesian Graphical Models using MCMC. R package version 4-14. URL https://CRAN.R-project.org/package=rjags.

Poletto, C., M. Tizzoni, and V. Colizza. 2012. Heterogeneous length of stay of hosts' movement and spatial epidemic spread. Sci Rep 2: https://doi.org/10.1038/srep00476

R Core Team. 2021. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rahbek, C. 2004. The role of spatial scale and the perception of large-scale species-richness patterns. Ecology Letters 8: 224-239. https://doi.org/10.1111/j.1461-0248.2004.00701.x

Ramey A. M., A. B. Reeves, J. Z. Drexler, J. T. Ackerman, S. De La Cruz, A.S. Lang, C. Leyson, P. Link, D. J. Prosser, G. J. Robertson, J. Wight, S. Youk, E. Spackman, M. Pantin-Jackwood, R. L. Poulson, and D. E. Stallknecht. 2020. Influenza A viruses remain infectious for more than seven months in northern wetlands of North America. Proc Biol Sci 287:20201680.

Ramey, A. M., J. M. Flint, H. S. Ip, D. V. Derksen, J. C. Franson, M. J. Petrula, B. D. Scotton, K. M. Sowl, M. L. Wege, and K. A. Trust. 2010. Intercontinental reassortment and genomic variation of low pathogenic avian influenza viruses isolated from Northern Pintails (Anas acuta) in Alaska: examining the evidence through space and time. Virology 401: 179-189. https://doi.org/10.1016/j.virol.2010.02.006

Rashford, B. S., R. M. Adams, J. Wu, R. A. Voldseth, G. R. Gutenspergen, B. Werner, and W. C. Johnson. 2015. Impacts of climate change on land-use and wetland productivity in the Prairie Pothole Region of North America. Regional Environmental Change 16: 515-526.

Reinecke, K. J., C. W. Shaiffer, and D. Delniki. 1992. Band reporting rates of Mallards in the Mississippi Alluvial Valley. The Journal of Wildlife Management 56: 526-531. https://doi.org/10.2307/3808868.

Riley, S. 2007. Large-scale spatial-transmission models of infectious disease. Science 316: 1298-1301.

Ruiz-Moreno, D., I. Sanchez Vargas, K. E. Olson, and L. C. Harrington. 2012. Modeling dynamic introduction of chikungunya virus in the United States. PLoS Negl Trop Dis 6: e1918. https://doi.org/10.1371/journal.pntd.0001918.

Ruiz-Suarez S., V. Leos-Barajas, I. Alvarez-Castro, and J. M. Morales. 2020. Using approximate Bayesian inference for a "steps and turns" continuous-time random walk observed at regular time intervals. PeerJ 8: e8452.

Russell, C. A. 2016. Sick birds don't fly…or do they? Science 354: 174-175.

Seeger, R. M., A. D., Hagerman, K. K. Johnson, D. L. Pendell, and T. L. Marsh. 2021. When poultry take a sick leave: Response costs for the 2014-2015 highly pathogenic avian influenza epidemic in the USA. Food Policy 102: https://doi.org/10.1016/j.foodpol.2021.102068.

Short, K. R., M. Richard, J. H. Verhagen, D. van Riel, E. J. Schrauwen, J. M. van den Brand, B. Mänz, R. Bodewes, and S. Herfst. 2015. One health, multiple challenges: The inter-species transmission of influenza A virus. One Health 1: 1-13. doi:10.1016/j.onehlt.2015.03.001. PMID: 26309905; PMCID: PMC4542011.

Si, Y., Q. Xin, H. H. T. Prins, W. F. de Boer, and P. Gong. 2015. Improving the quantification of waterfowl migration with remote sensing and bird tracking. Science Bulletin 60: 1984-1993. DOI 10.1007/s11434-015-0930-9.

Smith, G. W. 1995. A critical review of the aerial and ground surveys of breeding waterfowl in North America. U.S. Department of Interior, National Biological Service Biological Science Report 5.

Spiegelhalter, D. J., N. G. Best, B. R. Carlin, and A. van der Linde. 2002. Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society Series B, Statistical Methodology 64: 583-616.

Stephens, P. R., S. Altizer, K. F. Smith, A. A. Aguirre, J. H. Brown, S. A. Budischak, J. E. Byers, T. A. Dallas, T. J. Davies, J. M. Drake, V. O. Ezenwa, M. J. Farrell, J. L. Gittleman, B. A. Han, S. Huang, R. A. Hutchinson, P. Johnsons, C. L. Nunn, D. Onstad, A. Park, G. M. Vazquez-Prokopec, J. P. Schmidt, and R. Poulin. 2016. The macroecology of infectious diseases: a new perspective on global-scale drivers of pathogen distributions and impacts. Ecology Letters 19: 1159-1171. https://doi-org.ezproxy2.library.colostate.edu/10.1111/ele.12644.

Sullivan, J. D., J. Y. Takekawa, K. A. Spragens, S. H. Newman, X. Xiao, P. J. Leader, B. Smith, and D. J. Prosser. 2018. Waterfowl spring migratory behavior and avian influenza transmission risk in the changing landscape of the East-Australasian flyway. Frontiers in Ecology and Evolution 6: https://doi.org/10.3389/fevo.2018.00206/.

Tamura, K. and M. Nei. 1992. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in Humans and Chimpanzees. Molecular Biology and Evolution 9: 512-526.

Trovao, N. S., M. A. Suchard, G. Baele, M. Gilbert, and P. Lemey. 2015. Bayesian inference reveals host-specific contributions to the epidemic expansion of influenza A H5N1. Molecular Biology and Evolution 32: 3264-3275

USDA. 2017. Surveillance plan for highly pathogenic avian influenza in wild migratory birds in the United States. https://www.aphis.usda.gov/sites/default/files/2017-hpai-surveillance-plan.pdf.

USGS Bird Banding Laboratory. 2017. North American bird banding and band encounter data set. U.S. Geological Survey, Laurel, MD. <9/27/2017>

van Dijk J.G., J.H. Verhagen, M. Wille, and J. Waldenström. 2018. Host and virus ecology as determinants of influenza A virus transmission in wild birds. Curr Opin Virol 28: 26–36.

van Dijk, J. G. B., B. J. Hoye, J. H. Verhagen, B. A. Nolet, R. A. M. Fouchier, and M. Klaassen. 2013. Juveniles and migrants as drivers for seasonal epizootics of avian influenza virus. Journal of Animal Ecology 83: 266-275. https://doi.org/10.1111/1365-2656.12131.

VanderWall, K. L., and V. O. Ezenwa. 2016. Heterogeneity in pathogen transmission: mechanisms and methodology. Functional Ecology 30: 1606-1622. doi: 10.1111/1365-2435.12645.

Venkatesan, P. 2024. Human cases of avian influenza A(H5) in the USA. The Lancet Microbe 0: 0-0.

Ver Hoef, J. M., E. E. Peterson, M. B. Hooten, E. M. Hanks, and M. J. Fortin. 2017. Spatial autoregressive models for statistical inference from ecological data. Ecological Monographs 88: 36-59. https://doi.org/10.1002/ecm.1283.

Webster, R. G., W. J. Bean, O. T. Gorman, T. M. Chambers, and Y. Kawaoka. 1992. Evolution and ecology of influenza A viruses. Microbiological Reviews 56: 152 – 79.

Weins, J. A. 1989. Spatial scaling in ecology. Functional Ecology 3: 385-387.

Yang, Q., B. Wang, P. Lemey, L. Dong, T Mu, R. A. Wiebe, F. Guo, N. Sequeira Trovao, S. W. Park, N. Lewis, J. L. Tsui, S. Bajaj, Y. Cheng, L. Yang, Y. Haba, B. Li, G. Zhang, O. G. Pybus, and B. Grenfell. 2024. Synchrony of bird migration with global dispersal of avian influenza reveals exposed bird orders. Nature Communications 15: https://doi.org/10.1038/s41467-024-45462-1.

Appendix

Chapter 2 Supplemental Material

Prior Distributions

$$\mu_{ij} \sim t(1)$$

$$\tau \sim gamma(0.1, 0.1)$$

$$\beta_i \sim normal(0, \sigma^2 = 6.25)$$

$$\beta_{1,ik} \sim normal(0, \sigma^2 = 6.25)$$

$$\beta_{2,ik} \sim normal(0, \sigma^2 = 6.25)$$

$$\beta_{0,ik} \sim normal(0, \sigma^2 = 6.25)$$

$$\varepsilon_k \sim normal(0, \sigma^2 = 6.25)$$

Comparing Adjacency Matrices

It is possible that information about host migration could also be encoded in the local environmental adjacency matrix. Adjacent watersheds may have more host migration between them than watersheds that do not share borders. To ensure that the adjacency matrices were a clear representation of our hypotheses, we compared the local **W** matrix and the unweighted regional migratory host **W** matrix and evaluated their similarity using two different methods.

First, we used a Mantel test to check if the **W** matrices from each model were significantly correlated with each other. The local matrix was not significantly correlated with either the weighted or unweighted regional matrices ($r = 0.2$, $p > 0.05$; $r = 0.21$, $p > 0.05$). The weighted and unweighted versions of the regional matrices were highly positively correlated ($r =$

0.99, p = 0.001). This was expected since the same elements of the matrices are populated with non-zero values, differing only in how the matrices were weighted.

Second, each watershed pair in a local neighborhood was compared to the top ten connections between a focal watershed and all other watersheds in the unweighted network to see how many watershed connections overlapped between the two adjacency matrices (Figure S2.1). We found that only 13% of the unweighted migration matrix connections were between geographically adjacent watersheds. This shows that the local adjacency matrix and regional migration matrices encode different spatial phenomena.

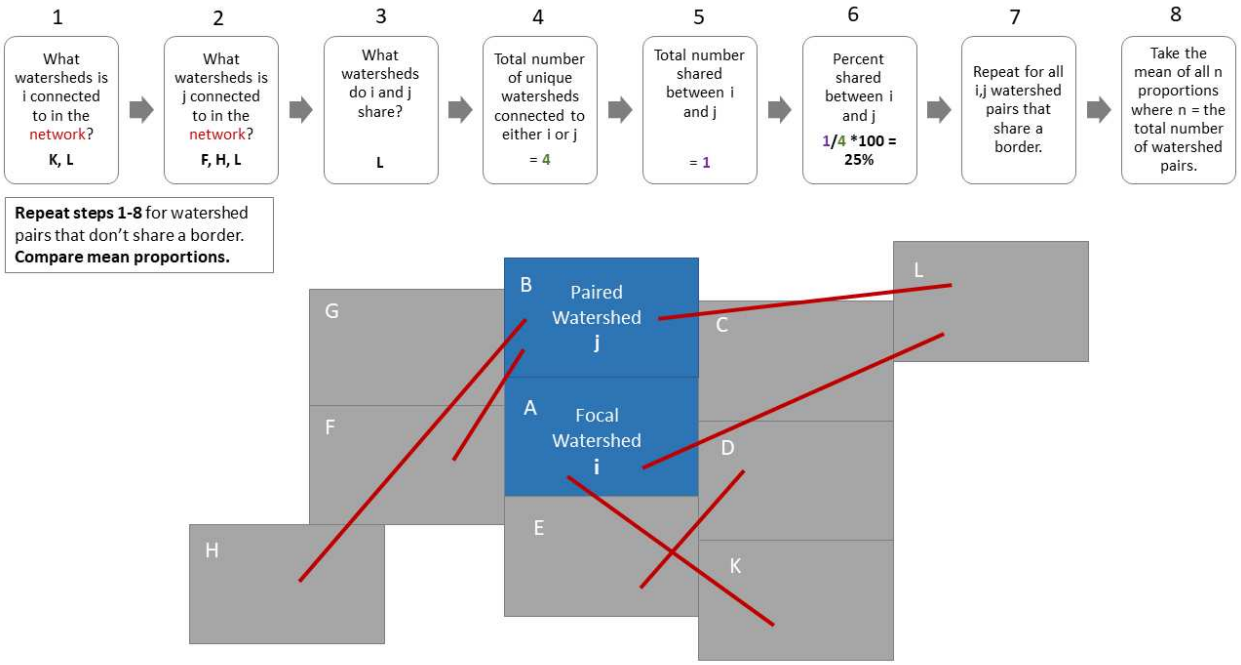

Figure S2.1: Diagram showing the method used to compare the spatial information encoded in the local and unweighted regional adjacency matrices. Blue rectangles represent watersheds arranged in geographic space (local environmental model) and red lines represent migratory connections in the unweighted regional model. The flowchart shows the stepwise process of calculating the proportion of connections shared between the two matrices using an example of two adjacent watersheds named A and B.

To ensure that the unweighted regional migration model and the weighted regional migration model encoded different spatial information, despite the high correlation between the

two matrices, we compared the community structure in each regional migration matrix with a community detection algorithm that grouped watersheds into migratory communities using a statistic called the consolidation factor (Buhnerkempe et al. 2016). The watersheds in the unweighted regional adjacency matrix were grouped primarily into four large communities that closely resemble currently recognized migratory flyways (Figure S2.2A). The weighed matrix showed similar community structure in the eastern United States but split up the western most community into three parts (Figure S2.2B). This shows that the weighted matrix encodes sufficiently different community structure to justify the inclusion of a weighted and unweighted model.

A. Unweighted **B. Weighted**

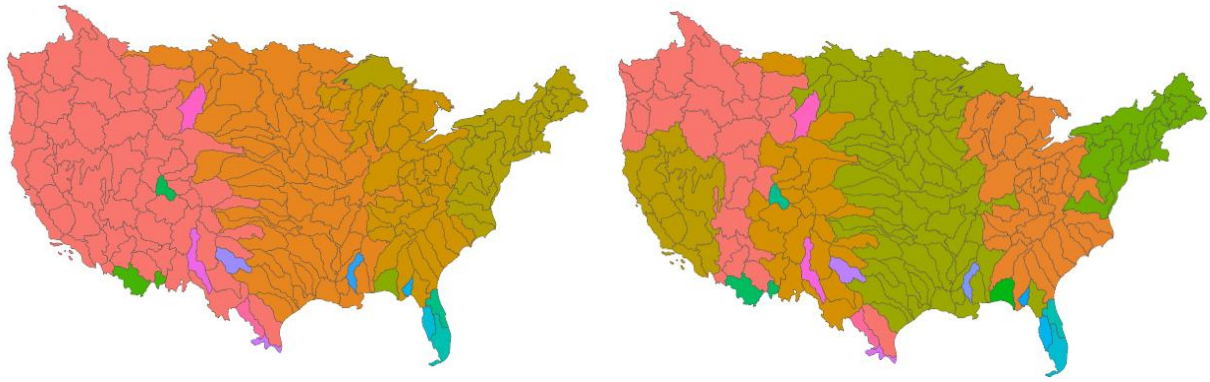

Figure S2.2: Community structure encoded in the unweighted regional migration adjacency matrix (A) and the weighted regional migration adjacency matrix (B). Watersheds in the same community are the same color, but community colors differ between maps.

Chapter 3 Supplemental Material

Prior Distributions

$$\beta_{X,p} \sim normal(0, 6.25)$$

$$\beta_{X\varphi,} \sim normal(0, 6.25)$$

$$\alpha_{i,\varphi} \sim Normal(0, \sigma_i)$$

$$\alpha_{j\varphi} \sim Normal(0, \sigma_j)$$

$$\alpha_{i,p} \sim Normal(0, \sigma_i)$$

$$\alpha_{j,p} \sim Normal(0, \sigma_j)$$

$$\sigma_i \sim Half - Normal(0,2)$$

$$\sigma_j \sim Half - Normal(0,2)$$

Posterior Predictive Checks

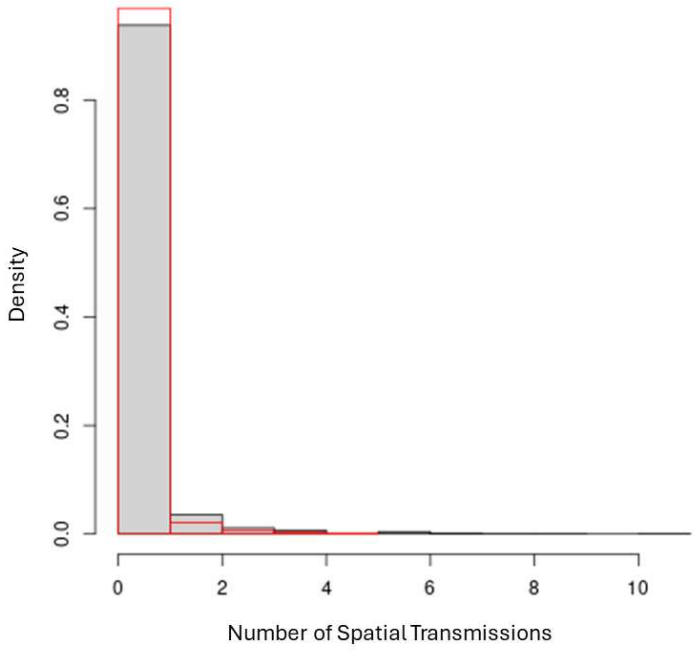

Figure S3.1: Histogram showing the distribution of data simulated from the fitted model (red) plotted over a histogram of the observed data used to fit the model (grey).

Detection Probability Results

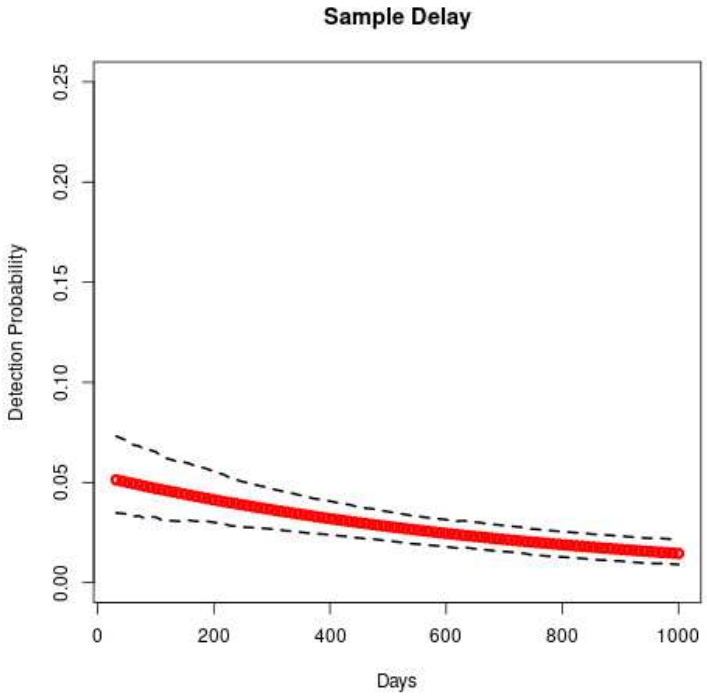

Figure S3.2: The probability of detecting spatial transmission of IAV for different values of sample delay in days. Dashed lines are the 95% Bayesian credible interval.

Chapter 4 Supplemental Material

Sources of Error

Argos satellite telemetry data are subject to varying degrees of observation error. At a minimum, it is important to remove all suspect observations and ensure that the spatial scale for location simulation is larger than the size of the measurement error. Several methods exist for dealing with observation error of Argos data. See Patterson et al. 2010, Douglas et al. 2012, Boyd and Brightsmith 2013, Buderman et al. 2016 for a selection of approaches. Additional simulation error can also be introduced through the choice of map projection. Because our simulation method relies on travel speed, we recommend equidistant projections that preserve distance.

Choosing Simulation Number and Buffer Size

There is no set optimal number of simulations to choose. The optimal number balances tradeoffs among the desired accuracy, the number of individuals to simulate for, the number of time-steps per individual, and computation time. The more time-steps and individuals, the greater the accuracy, but the longer the computation time. A similar tradeoff exists for the choice of "buffer zone". Too small of a buffer and very few simulations will be successful; too large, and posterior distributions of simulated locations will have variance larger than the desired scale of inference. We recommend choosing a buffer distance that balances computation time and accuracy.

Making Weighted Networks

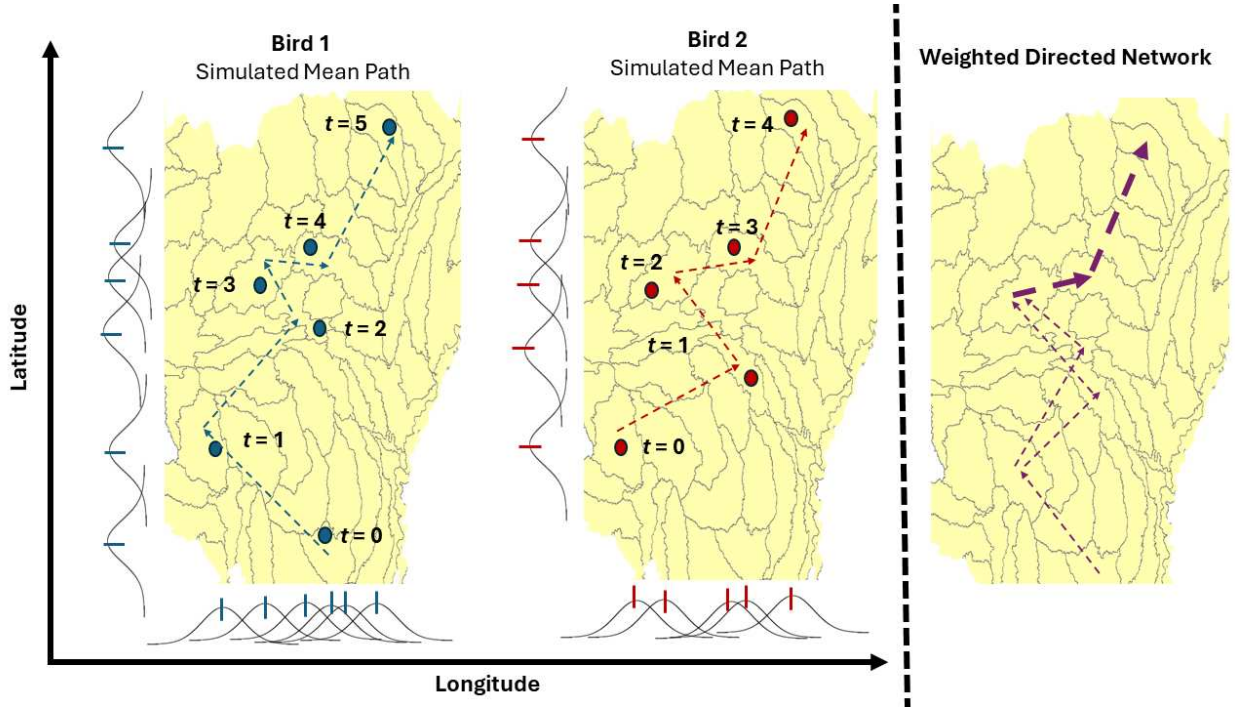

Figure S4.1: Conceptual diagram showing how simulated locations of two birds were mapped to a single weighted spatial network. For each bird, the mean latitude and mean longitude at each time-step were assigned to a watershed (yellow polygons). Dashed arrows show the resulting spatial networks for each bird. The two networks were added together resulting in a single weighted movement network shown in purple. The thickness of the arrow indicates the number of birds moving between the watersheds.

Calculating 95% Credible Intervals for Network Statistics

The equivariance property of Markov chain Monte Carlo (MCMC) states that any quantity calculated from random variables are random variables themselves with their own posterior distributions. We mapped the random variables of latitude and longitude to the larger watershed scale to construct a spatial network (Figure S4.1). Each network and its properties are thus also random variables described by posterior distributions. We took the upper and lower quantiles of the distributions of network properties to get upper and lower bounds on our uncertainty of the point estimate of network statistics. Figure S4.2 shows how credible intervals were estimated from these distributions. We compared credible intervals calculated from 100,

500, 1000, and 5000 samples and found that beyond 500 samples the credible intervals did not shift appreciably. The 0.025 and 0.975 quantiles of each of these distributions is the 95% CI of their respective network statistics.

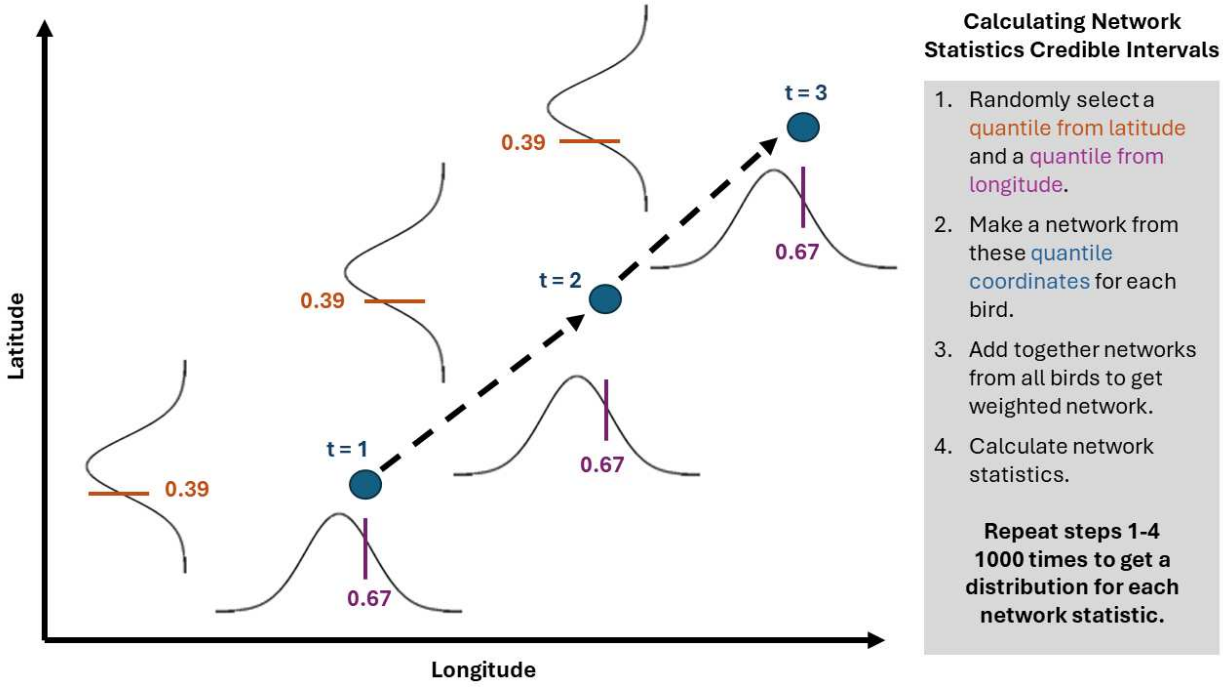

Figure S4.2: Conceptual diagram showing the sampling process used to calculate 95% credible intervals for network statistics of simulated movement networks. The process is shown for one bird with three simulated time-steps. Blue dots show the simulated locations taken from the 0.39 quantile of latitude and the 0.67 quantile of longitude at each time-step. Dashed arrows show the resulting directed network for a single bird. Repetition of steps 1-4 1000 times results in a posterior distribution with 1000 samples for each network statistic. For a detailed explanation of step 2, see figure S4.1.