12-2023

Characterization of the Long-Distance Dispersal Kernel of White-Tailed Deer and Evaluating Its Impact on Chronic Wasting Disease Spread in Wisconsin

Mennatallah Gouda
Utah State University, a02368433@usu.edu

CHARACTERIZATION OF THE LONG-DISTANCE DISPERSAL KERNEL OF WHITE-TAILED DEER AND EVALUATING ITS IMPACT ON CHRONIC WASTING DISEASE SPREAD IN WISCONSIN

by

Mennatallah Gouda

A thesis submitted in partial fulfillment
of the requirements for the degree

of

MASTER OF SCIENCE

in

Mathematics

Approved:

James Powell, Ph.D. Major Professor	Brynja Kohler, Ph.D. Committee Member
Noelle Beckman, Ph.D. Committee Member	D. Richard Cutler, Ph.D. Vice Provost of Graduate Studies

UTAH STATE UNIVERSITY

Logan, Utah

2023

ABSTRACT

Characterization of the Long-distance Dispersal Kernel of White-tailed Deer and Evaluating its Impact on Chronic Wasting Disease Spread in Wisconsin

by

Mennatallah Gouda, Master of Science

Utah State University, 2023

Major Professor: Dr. James Powell

Department: Mathematics and Statistics

Chronic Wasting Disease (CWD) is a fatal untreatable neurodegenerative disease that infects cervids. It is highly contagious and caused by abnormal malfunction and assembly of the normal cellular prion proteins (PrP^{C}) into aggregation-prone prions (PrP^{Sc}). Centers for Disease Control and prevention (CDC) report that the prevalence of CWD in free-ranging deer in the US is still relatively low. However, in several states the infection rates exceed 1 deer in 10. Cervids may uptake CWD prions from direct interaction with infected individuals or from the environment. Infected individuals shed prions into the environment through feces, urine, saliva or carcass, and long-distance dispersal of infected deer poses a danger of spreading CWD to new regions.

We propose an Integrodifference Model (IDE) to capture CWD dynamics and the consequences of long-distance dispersal behavior of White-Tailed Deer (WTD, *Odocoileus virginianus*). Currently there are no dispersal kernels available to describe the long-distance dispersal behavior of WTD juveniles. Our aim is to characterize long-distance dispersal of WTD juveniles and assess how it may affect CWD spread. We introduce a long-distance dispersal model, based on diffusion-settling seed transport by vertebrates, accommodating a variety of hypothetical dispersal behaviors of WTD. Four kernels were

obtained by solving 2D diffusion-settling Partial Differential Equation (PDE) models and approximating using Laplace's method. We parameterized the kernels with GPS collar data collected in Wisconsin, US. Using a Maximum Likelihood Estimation (MLE) approach, we fitted the model parameters, and assessed model fits using the Bayesian Information Criterion (BIC).

A Holling type III settling rate function resulted in the most supported long-distance dispersal kernel reflecting deer preference to not settle down soon after they start dispersal. Sensitivity of results was determined using nonparametric bootstrapping and the impact of long-distance dispersal on CWD spread was quantified using the IDE model. Our results show that long-distance dispersal can magnify the CWD spread by a factor of 4. Therefore, controlling the total population density and fraction of long-distance dispersers will assist CWD management facilities in managing disease spread.

(87 Pages)

PUBLIC ABSTRACT

Characterization of the Long-distance Dispersal Kernel of White-tailed Deer and Evaluating its Impact on Chronic Wasting Disease Spread in Wisconsin

Mennatallah Gouda

Chronic Wasting Disease (CWD) is a fatal, untreatable neurodegenerative disease that infects deer and related species. It is highly contagious and caused by abnormal malfunction and assembly of normal cellular proteins into aggregation-prone proteins. The Centers for Disease Control and prevention report that the prevalence of CWD in free-ranging deer in the US is still relatively low. However, in several states the infection rates exceed 1 deer in 10. Deer may uptake CWD from direct interaction with infected individuals or from the environment. Infected individuals shed CWD into the environment through feces, urine, saliva or carcasses, and long-distance dispersal of infected deer poses a danger of spreading CWD to new regions. We propose a mathematical model to capture CWD dynamics and the consequences of long-distance dispersal behavior of White-Tailed Deer (WTD), since there are no dispersal models available to describe the long-distance dispersal behavior of WTD juveniles. Our aim is to characterize long-distance dispersal of WTD juveniles and assess how it may affect CWD spread. We introduce a long-distance dispersal model, based on models of seed transport which accommodate a variety of hypothetical WTD dispersal behaviors. Four dispersal models were obtained by finding the approximate solutions of the seed transport models tweaked to mimic deer dispersal and settling behavior. We used GPS collar data collected in Wisconsin, US to test the accuracy of the models we developed. By calculating the prediction errors made by the models, we adjusted the model parameters, and assessed the competency of models relative to each other. Sensitivity of results was estimated by changing the data randomly to

account for the effect of changing data on the results. The settling rate function that resulted in the most supported long-distance dispersal kernel reflects deer preference to not settle down soon after they start dispersal. Then, the impact of long-distance dispersal on CWD spread was quantified using the mathematical model for CWD dynamics we proposed earlier. Our results show that long-distance dispersal can magnify the CWD spread by a factor of 4. Therefore, controlling the total population density and fraction of long-distance dispersers will assist CWD management facilities in managing disease spread.

ACKNOWLEDGMENTS

First of all, I am grateful for being a Muslim for giving me a purpose in this life and light on the darkest days. Second, I would like to thank my beloved mom for always getting my back. No matter how much I say or do, it would be less than what you deserve. Third, I would like to say thank you, Dr. Jim, for placing trust in me. Thanks, Jake, for always giving all that you can do to help. Lastly, I would like to thank my whole support system, all family, and friends.

Thanks, Wisconsin Division of Natural Resources, for sharing GPS collar data from WTD, US Geological Survey for funding through the Cooperative Ecological Studies Unit program, and USDA/NIFA for funding through National Science Foundation's Ecology and Evolution of Infectious Disease program.

Mennatallah Gouda

CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

ACRONYMS

BIC	Bayesian Information Criterion
CDC	Centers for Disease Control
CI	Confidence Interval
CWD	Chronic Wasting Disease
CWDA	Chronic Wasting Disease Alliance
DNR	Division of Natural Resources
GDNSR	Georgia Department of Natural Resources
MLE	Maximum Likelihood Estimation
NIH	National Institute of Allergy and Infectious Diseases
NLL	Negative Log Likelihood
PDE	Partial Differential Equation
TSE	Transmissible Spongiform Encephalopathies
VDWR	Virginia Department of Wildlife Resources
WTD	White-Tailed Deer

CHAPTER I

INTRODUCTION

1.1 Chronic Wasting Disease (CWD)

1.1.1 CWD Biomedical Background

CWD is a fatal disease that infects cervids like elk (*Cervus canadensis*), deer (*Odocoileus* species in North America), caribou/reindeer (*Rangifer tarandus*) and moose (*Alces alces*). It is classified as one of the prion diseases or Transmissible Spongiform Encephalopathies (TSEs). TSEs are transmissible, untreatable rare neurodegenerative diseases that result in neuronal loss and inflammation deficiency (National Institute of Allergy and Infectious Diseases [NIH], 2019 & CDC, 2021). TSEs other than CWD include scrapie in goats and sheep, Bovine Spongiform Encephalopathy (BSE) in cattle (commonly known as "mad cow disease") and classic Creutzfeldt-Jakob disease (CJD) and variant Creutzfeldt-Jakob disease (vCJD) in humans (United States Department of Agriculture, 2020).

TSEs are caused by abnormal malfunction and assembly of the normal cellular prion proteins (PrP^{C}) into aggregation-prone prions (PrP^{Sc}). This can happen through spontaneous misfolding, a genetic mutation, or exposure to a prion from an external source (Wright et al., 2018). During this conversion, the PrP^{C}, rich in α-helices changes into the insoluble conformer PrP^{Sc}, rich in β-helices (Eghiaian et al., 2004). TSEs are contagious in the sense that a diseased prion protein induces the abnormal folding of normal proteins so that they become diseased as well. Prions are most abundant in the brain, causing fatal brain damage. Since prions may be shed in the environment, they represent a biological hazard that threatens susceptible host populations. (NIH, 2019 & CDC, 2021).

Chronic Wasting Disease Alliance (CWDA, 2019) and Georgia Department of Natural Resources (GDNSR, 2022) note that CWD disease exhibits prolonged periods in the incubation stage. Moreover, infected animals show various behavioral changes and

poor body condition. For example, they may lose coordination, walk in a repetitive pattern, and have wide-based stance. Also, they may develop slight head tremors, prefer to be near water sources, experience somnolent periods, and have droopy head and/or ears. In addition, their food consumption becomes lower, resulting in body weakness. In the last stages of the disease, they commonly show excessive drinking, salivation, drooling, and urination (CWDA, 2019; GDNSR, 2022). Unfortunately, science hasn't discovered a treatment or vaccine for CWD yet, and all CWD-positive animals inevitably die in the terminal stage of the disease (Virginia Department of Wildlife Resources [VDWR], 2022).

1.1.2 CWD prevalence

The CDC reports that the prevalence of CWD in free-ranging deer in the US is still relatively low. On the other hand, in several states the infection rates exceed 1 deer in 10. In some counties of Wisconsin and Colorado, localized infection rates have reached more than 25%. In Wisconsin, 37 counties have observed CWD positive deer. Moreover, among captive deer, the infection rates are reported to reach 79%. Until June 2022, 391 counties in 29 states in the US had CWD in free-ranging herds (Figure 1) (CDC, 2022). This poses an increasing threat to herd management, as Edmunds et al. (2016) have shown the deer infected with CWD were 4.5 times more prone to death each year in comparison with uninfected deer. Moreover, Jennelle et al. (2014) reported that CWD prevalence is higher among males and increases with age.

1.1.3 CWD transmission

CWD can be transmitted in cervids both directly and indirectly. Direct transmission could happen during interaction among conspecifics in reproduction, nurturing or competition. Environmental transmission of the disease could happen in various ways. Cervids may uptake prions from the environment via ingestion or inhalation. After infection with CWD prions, infected individuals shed CWD prions into the environment through feces, urine, or saliva. Moreover, when CWD-positive deer die, their

carcass decomposes into the environment or could be unsafely disposed to create a "hot spot" for infection (Haley et al., 2009; Jacobson et al., 2009; Tamgüney et al., 2009; Tamgüney et al., 2012). Tamgüney et al. (2009) showed that infected mule deer do not shed prions for around three months after infection. Then, they spend approximately nine months shedding the disease but show no symptoms. After that, they spend three more months both symptomatic and shedding till death (Tamgüney et al., 2009). Miller et al. (2004) and the VDWR (2022) stated that the prions remain infectious for at least 2 years in the environment, and likely they persist for longer. However, the infectious period of CWD prions has not been exactly quantified. This is because CWD prions are acted upon by several environmental factors like plant uptake, transport in the soil column, prion inactivation and attachment to particle surfaces, any at which alters prion distribution in space, bioavailability, and infectivity (Haley et al., 2009; Jacobson et al., 2009; Tamgüney et al., 2009; Tamgüney et al., 2012). The critical dose for environmental exposure was reported to be as low as 300 ng of homogenized infected CWD brain administered orally to White-Tailed Deer (WTD) a single dose; however, a dosage of 100 ng aliquots three times a week did not cause infection (Denkers et al., 2020).

1.1.4 CWD course of infection

The overall CWD course of infection has been hard to estimate in nature since the exposure time is not given. So, to estimate the duration from exposure to death as the end stage for CWD, Williams and Miller (2002) have infected mule deer orally with the disease. Their findings state that minimum incubation period was around 15 months, and 23 months for the average time taken from exposure to death. Moreover, they reported that the overall course of disease could exceed 25 months in CWD-positive deer (Williams and Miller, 2002). However, these estimates are not the most accurate as they were done in 2002 before the development and approval of immunohistochemistry (IHC) assays and Enzyme-Linked Immunosorbent Assay (ELISA) diagnostic methods for CWD prions detection (United States Department of Agriculture, 2020). Coming to a resolution,

CWDA (2019) and GDNSR (2022) stated that the clinical course of CWD (period of time from onset of symptoms) can range from few days to almost a year. They highlighted that most of the infected animals survive only a few weeks to several months from the start of clinical onset of the disease.

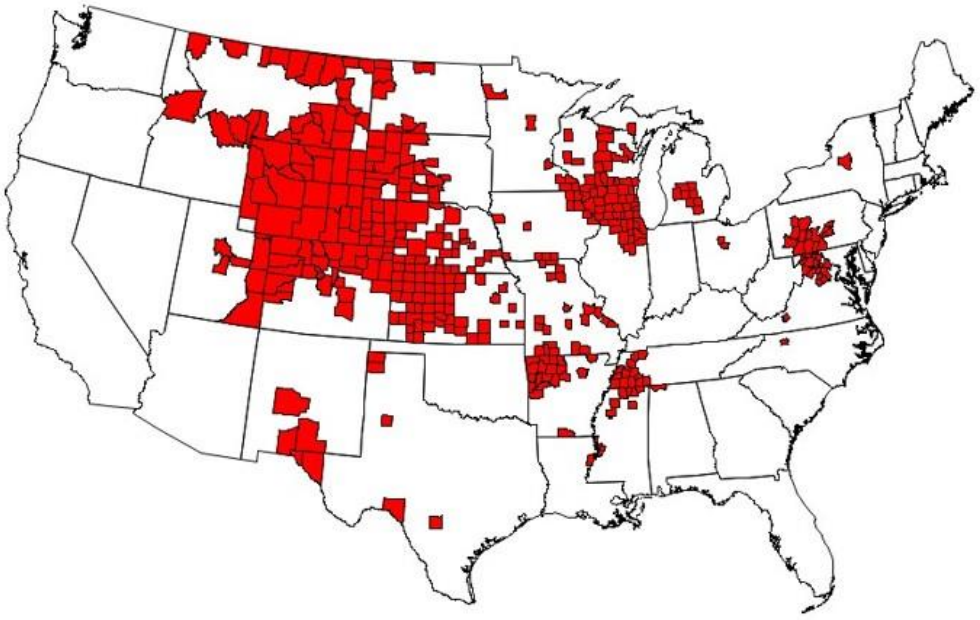

Figure 1. Counties witnessed CWD by June 2022. Adapted from CDC, retrieved from Occurrence | Chronic Wasting Disease (CWD) | Prion Disease | CDC.

1.2 Wisconsin State history with CWD

Among many US states with endemic CWD, Wisconsin has perhaps the most extensive documented CWD history. The first CWD cases in Wisconsin were detected in three WTD on February 28th, 2002, in the Mt. Horeb area of Iowa and Dane counties (CWDA, 2002). Since that time sampling has been performed yearly to test for CWD in various counties throughout the state (Figure 2) (Wisconsin Department of Natural Resources, 2022). Hence, CWD prevalence, history, extensive background research, and data availability in Wisconsin makes it an optimum place to investigate into CWD spread dynamics.

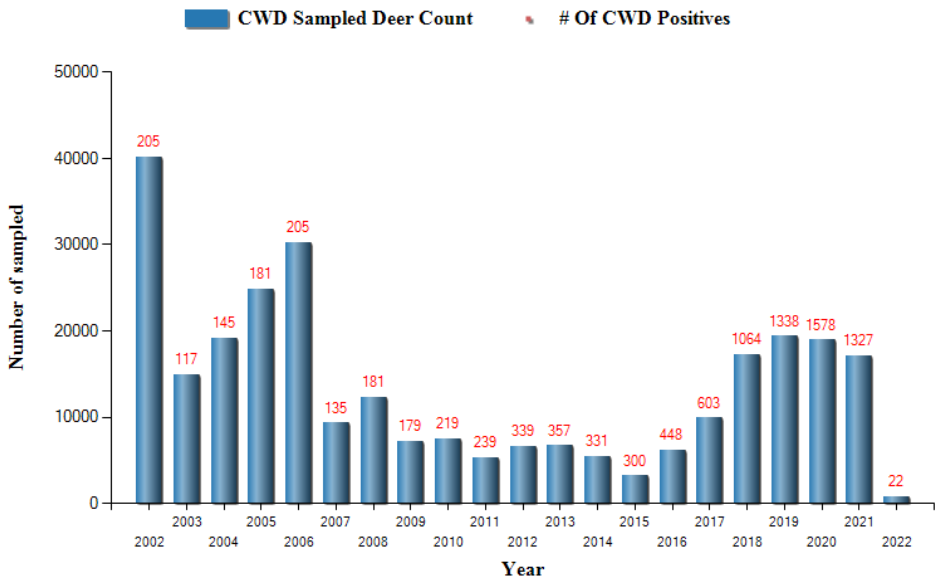

Figure 2. Incidence of CWD in Wisconsin in sampled population from 2002 to 2022. Adapted from Wisconsin Department of Natural Resources, by J. Pritzl & T. Marien, 2022, retrieved from Deer Statistics (wi.gov).

1.3 Role of long-distance dispersal in CWD spread

Long-distance deer dispersal away from home ranges in the CWD-positive portions of Wisconsin poses a danger of spreading CWD to new regions. Therefore, dispersal behavior of WTD is a key factor affecting how fast the disease will spread and how it will affect wildlife. According to the study done by Skuldt et al. (2008) in south-central Wisconsin, adult female WTD rarely exhibit dispersal. Most (80-97%) juvenile females stay in their natal home ranges. On the other hand, 46-80% of juvenile males as well as a nontrivial fraction of juvenile females travel to new home ranges. Deer of other ages and sexes rarely dispersed transiently among different home ranges. However, regardless of sex and age, deer may go for short, temporary walkabouts away from their home ranges. Migration is not common in south-central Wisconsin; however, any chance for movement poses a threat of increasing CWD incidence rate (Skuldt et al., 2008).

Long-distance dispersal is a density-dependent event; dispersal may take place to avoid inbreeding and intraspecific competition. Adult females exhibit intersexual

aggressive behavior towards male juveniles, leading to higher probabilities of male juvenile dispersal than females. Mature female WTD produce fawns in spring, and therefore dispersal probability increases in the spring season. Additionally, during the fall breeding season, intrasexual aggression takes place among male deer, leading to elevated dispersal probability (Gilbertson et al., 2022).

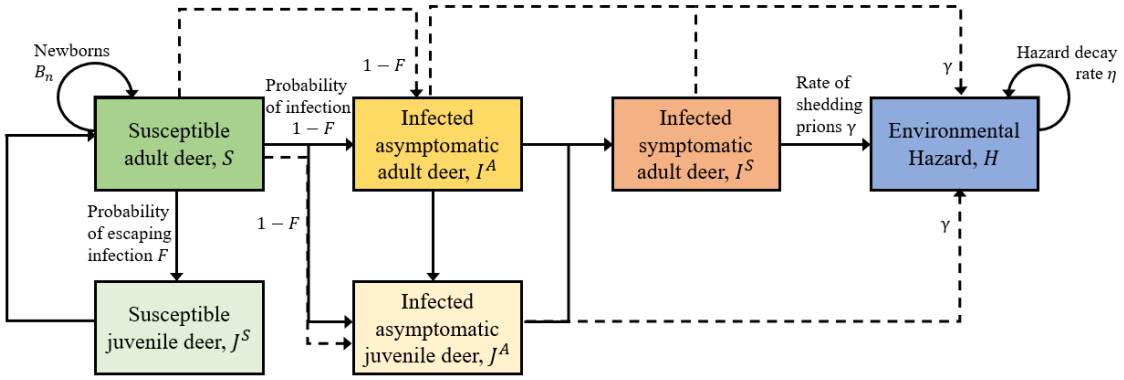

Figure 3. Schematic diagram of CWD transmission among deer population. Vertical transmission refers to transferring the disease from the mother deer to the newborn juvenile. Horizontal transmission takes place through interaction of adult deer with their conspecifics either in cooperation or competition. Susceptible adult deer may get infected but remain asymptomatic; infectious asymptomatic deer may have more energy to spread disease than symptomatic. All infected deer contribute to shedding CWD prions into the environment, posing a serious biological hazard for deer population.

1.4 Mathematical literature review

Mathematical modelling is a powerful tool to interpret the complexity of biological systems and predict the effects of changing one or more variables. Modelling long-distance dispersal of deer juveniles in this study aims at estimating the risk of dispersing far from home on CWD spread. Much efforts have been dedicated to modelling CWD transmission from both deterministic and stochastic approaches. For example, a stochastic modelling approach was adopted by Belsare & Stewart (2020) through developing a customizable agent-based modelling framework (OvCWD) for CWD spatial spread. Their model links WTD demography to CWD transmission accounts for heterogeneity and nonrandomness in the disease dynamics (Belsare & Stewart, 2020).

Miller et al. (2006) developed seven deterministic models for CWD transmission, neglecting spatial spread. They accounted for a simple direct transmission Susceptible-Infected (SI) model and added complexity to it to include a latency period (2nd model), or an incubation period (3rd model). Moreover, they developed a Susceptible-Infected-Hazard (SIH) model for indirect transmission through uptake or shedding of CWD prions into the environment, and two models combining direct and indirect transmission (one of which includes a latency period). The seventh model described maternal transmission but this one failed to mimic the epidemiology of the disease. Then, they tested the models against two CWD outbreak data in captive mule deer. Their results reveal that the indirect transmission of CWD hypothesis is 3.8 times more supported than the direct transmission one. However, they did not evaluate the effect of dispersal, especially for long distances, on CWD transmission (Miller et al., 2006).

Garlick et al. (2011) examined dispersal of CWD using a PDE model, derived a homogenization technique, and applied it to an ecological diffusion model of CWD to assess the effect of habitat on CWD spread. Their method proved to be much simpler than the multidimensional homogenization method. They considered a spatial model in mule deer based on a SIH model. The assumptions they have made included that the disease have not exceeded the initial stage, transmitted by direct contact with infected individuals, and is consistently fatal. Moreover, they assumed that the density of normal death cases is equal to the density of newborns, and sex, age, and season are independent of CWD spread. Their findings show that CWD increases at high-density natural deer habitats like riparian areas and forests. Moreover, their model shows less density of infected individuals in regions suitable for frequent dispersal behavior like rocky, barren regions and developed open spaces (Garlick et al., 2011).

Moreover, Garlick et al. (2014) used a homogenization approach with a more complex model of CWD ecological diffusion to link the local deer movement behavior to CWD spatial spread. Their approach shows low computational cost to simulate disease

spread. They included in their model deer dispersal and estimated motility coefficients from GPS collar data in mule deer. Also, they included the fidelity of mother deer to winter and summer periods, as well as the more tendency of male deer to disperse than females in the breading season. Their CWD model incorporated sex differences, environmental hazard of shed CWD prions, and ecological diffusion (Garlick et al., 2014).

1.5 Integrodifference equation model

Garlick et al. (2011, 2014) examined dispersal of CWD using a PDE model with SIH, derived a homogenization technique, and applied it for an ecological diffusion model of CWD to assess the effect of habitat and local dispersal on CWD spread. However, their models could not capture long-distance dispersal because the natural time scale for PDE SIH models is years while long-distance dispersal events take place in weeks. Long-distance dispersal events are not very frequent in juvenile deer and occur in seasonal pulses. It is hard to model those events mechanistically in a PDE SIH model, due to the difference in time scales between infection (years) and long-distance dispersal (days/weeks). On the other hand, Integro-Difference Equation (IDE) models consider the impacts of relevant events over a year. Hence, IDE models are a convenient choice to model CWD dynamics and the consequences of seasonal long-distance dispersal behavior of WTD, since dispersal events are resolved as discrete applications of probability kernels in space. Furthermore, it aligns with the fact that the infectious period of CWD (time from developing the disease till death) in WTD is approximately one year, roughly the time between big dispersal events and the maturation of juveniles.

We propose a simplified model for dispersal impacts on CWD spread:

$$S_{n+1} = (\rho K_L + (1-\rho)K_S) * [F(I_n, H_n)S_n] + B_n \quad (1)$$

$$I_{n+1} = (\rho K_L + (1-\rho)K_S) * \left[\left(1 - F(I_n, H_n)\right)S_n\right] \quad (2)$$

$$H_{n+1} = e^{-\eta}H_n + \gamma I_n \quad (3)$$

Here the variables S_n, I_n, and H_n represent the density of susceptible deer, infected deer, and the concentration of environmental prion hazard in the n^{th} year. The function F_n denotes the probability of susceptible deer escaping infection in the n^{th} year. The variable B_n indicates the density of newborn fawns. The parameters η represents the decay rate of prion, and γ represents the rate of shedding prions by infected deer into the environment. The parameter ρ represents the proportion of long-distance dispersing juveniles in the population, assuming that adult deer do not disperse for long distances. The long-distance and short-distance dispersal kernels K_L, and K_S are probability density functions representing the distribution of deer in space after long-distance and short-distance dispersals respectively relevant to their starting positions. The convolution of a kernel and a population density, $K * S$ or $K * I$, accounts for the potential dispersion of deer over a year.

Currently there are no dispersal kernels available which mechanistically describe the long-distance dispersal behavior of WTD juveniles, K_L. This study aims to characterize long-distance dispersal of WTD juveniles and assess how it may affect CWD spread. We introduce a long-distance dispersal model, based on diffusion-settling seed transport by vertebrates, which can accommodate a variety of hypothetical dispersal behaviors of WTD. Four kernels were obtained by solving the 2D diffusion-settling PDE models in integral form and then approximating using Laplace's method. We parameterized the four kernels with GPS collar data collected in Wisconsin, US. Data was filtered to remove unrealistic velocity spikes and extract long-distance dispersals from the uneven time sampling. Next, using a MLE approach, we fitted parameters for each model and assessed model fits using the BIC. We also tested models for separate female and male dispersal. Sensitivity of results was determined using nonparametric bootstrapping. Lastly, the impact of long-distance dispersal on CWD spread was quantified using the IDE model (1-3).

CHAPTER II

METHODS

2.1 Proposed PDE models for dispersal

We adopted the following PDE model in 2D based on a seed transport and settling by vertebrates model introduced by Neupane & Powell (2015) to describe long-distance deer dispersal. For $r = |\boldsymbol{x} - \boldsymbol{y}|$ where $\boldsymbol{y}$ is the initial position,

$$P_t(r,t) = D\nabla^2 P(r,t) - h(t)P(r,t), \quad P(r,0) = \delta(r) \tag{4}$$

$$S_t(r,t) = h(t)P(r,t), \quad S(r,0) = 0 \tag{5}$$

$$K(r) = \lim_{t\to\infty} S(r,t) \tag{6}$$

Here the variables $P(r,t)$ and $S(r,t)$ represent the density of long-distance dispersing juveniles, and juveniles which have settled in a new range, respectively. The long-distance dispersal kernel $K(r)$ is the end result after all dispersers have settled. Dispersers are assumed to move randomly with diffusion constant D and settling rate $h(t)$ over time. The Dirac delta function, $\delta(r)$, denotes the initial population density and indicates that the initial starting position of dispersers is known. Formulae for h are discussed below. The change over time in the dispersing deer density increases by diffusion of deer in 2D space $D\nabla^2 P(r,t)$ and decreases with their settling $h(t)P(r,t)$. Moreover, the change in settling deer density increases with time directly with the density of settling deer.

We introduce constant, Holling type II, III, and a new, strictly increasing type IV settling rate functions that represent different behavioral assumptions about juvenile WTD dispersal. Four different forms of the settling rate $h(t)$ were proposed as follows:

$$h_1(t) = ab,\ h_2(t) = \frac{abt}{a+t},\ h_3(t) = \frac{abt^2}{a^2+t^2},\ h_4(t) = \frac{bt^2}{a+t}\ . \tag{7}$$

Here the parameters a and ab are half asymptotic dispersal time and the asymptotic settling rate respectively. These functions capture differing behaviors of dispersing deer. For example, deer may settle down reflecting a constant settling rate function. This

assumes they have same tendency to settle down disregarding the amount of time they have been dispersing for. However, a Holling type II function assumes a more realistic scenario, assuming the probability of settling at the beginning of dispersal is zero and then increases gradually to a plateau at sufficiently enough time for deer to settle down. The type III Holling response function represents sigmoidal rates of setting to accommodate the fact that deer are less likely to settle close to their home ranges. Both behaviors align with adult females kicking male juveniles out so that they cannot settle down soon after they started dispersing. As they travel far enough, their tendency to settle down increases till it reaches a maximum settling tendency that doesn't depend on time anymore. However, the latter behavior assumes that deer do not get exhausted to the extent that they must settle down. Hence, a type IV settling function assumes that deer seldom settle down any time close to the dispersal starting time. However, when they spend long enough time dispersing, their tendency to settle down increases very rapidly to infinity.

2.2 Solving the PDE dispersal models

The PDE system (4-5) was solved analytically in integral form to get the general formula of a kernel. We introduce an integrating factor,

$$P = e^{-\int_0^t h(s)ds} u(r,t), \tag{8}$$

which gives $u_t = D\nabla^2 u,\ u(r,0) = \delta(r).$ (9)

Using the fundamental solution of the diffusion equation, we get:

$$u = \frac{1}{4\pi Dt} e^{\frac{-r^2}{4Dt}}, \tag{10}$$

and this in (8) gives

$$P = \frac{1}{4\pi Dt} e^{-\int_0^t h(s)ds - \frac{r^2}{4Dt}}. \tag{11}$$

Then, we have a solution for S in integral form,

$$S(r,t) = \int_0^t \frac{h(\tau)}{4\pi D\tau} e^{-\int_0^\tau h(s)ds - \frac{r^2}{4D\tau}}\, d\tau \tag{12}$$

Hence, a kernel would be:

$$K(r) = \lim_{t\to\infty} S(r,t) = \int_0^\infty \frac{h(\tau)}{4\pi D\tau} e^{-\int_0^\tau h(s)ds - \frac{r^2}{4D\tau}}\, d\tau \tag{13}$$

Using (13), the kernels K_2, K_3, and K_4 were derived.

2.2.1 Constant settling rate function PDE model

The kernel K_1 was previously derived by Neupane and Powell (2021) as a Bessel function,

$$K_1 = \frac{1}{2\pi\alpha^2} K_0\left(\frac{r}{\alpha}\right) \tag{14}$$

where $\alpha = \sqrt{\frac{D}{ab}}$, and K_0 is the modified Bessel function of the second kind.

2.2.2 Variable settling rate function PDE models

The kernels K_2, K_3, and K_4, resulting from the variable settling rate functions h_2, h_3, and h_4, are analytically approximated in their integral forms using Laplace's method.

2.2.2.1 Introducing Laplace's method

For the kernels K_2, K_3, and K_4, we employed the steepest descent method which is an extension to Laplace's method to approximate the kernel integral (Logan, 2013). The method approximates integrals of the form

$$I(\lambda) = \int_a^b f(t)e^{\lambda g(t)}dt, \;\; \lambda \gg 1, \tag{15}$$

assuming that f is continuous and g is sufficiently smooth with a unique maximum at the point $t = m \in (a,b)$, where $g'(m) = 0,\ g''(m) < 0$. This is because the main contribution to the value of the integral is expected to originate from the maximum of g. For large λ, we get the following approximation for I:

$$I(\lambda) \sim f(m)e^{\lambda g(m)}\sqrt{\frac{-2\pi}{\lambda g``(m)}} + O(\lambda^{-\frac{3}{2}}) \tag{16}$$

2.2.2.2 Using Laplace's method to approximate the kernels

2.2.2.2.1 Approximating K_2

Substituting $h_2(t)$ into (8) and using $D = \frac{c^2}{ab}$, we have

$$K_2(x) = \int_0^\infty \frac{ab^2\left(\frac{t}{a}+1\right)^{\beta-1}}{4\pi c^2} e^{ab\left(-t-\frac{x^2}{4c^2t}\right)}\, dt, \tag{17}$$

where $\beta = a^2b$. Letting $\lambda = ab$, and identifying

$$f(t) = \frac{ab^2\left(\frac{t}{a}+1\right)^{\beta-1}}{4\pi c^2} \quad \text{and} \quad g(t) = -t - \frac{x^2}{4c^2t}\,, \tag{18}$$

we find the critical point is $m = \frac{x}{2c}$ (details in appendix A). Application of Laplace's method then gives

$$K_2 \approx c_2\sqrt{\frac{r}{(ac)^5}}\left(1+\frac{r}{2ac}\right)^{a^2b-1} e^{-\frac{a^2b\,r}{ac}}. \tag{19}$$

Here c_2 is the constant of normalization, which must be calculated numerically in two-dimensional space.

2.2.2.2.2 Approximating K_3

Substituting $h_3(t)$ into (8) and using $D = \frac{c^2}{ab}$, we have

$$K_3(x) = \int_0^\infty \frac{b^2 t}{4\pi c^2\left(\frac{t^2}{a^2}+1\right)} e^{ab\left(-t+a\tan^{-1}\left(\frac{t}{a}\right)-\frac{x^2}{4c^2t}\right)}\, dt \tag{20}$$

Letting $\lambda = ab$, and identifying

$$f(t) = \frac{b^2 t}{4\pi c^2\left(\frac{t^2}{a^2}+1\right)} \quad \text{and} \quad g(t) = -t + a\tan^{-1}\left(\frac{t}{a}\right) - \frac{x^2}{4c^2t}, \tag{21}$$

we find the critical point is $m = \frac{r}{\sqrt{8}ac}\sqrt{1+\frac{\sqrt{r^2+16(ac)^2}}{r}}$ (details in appendix A).

Application of Laplace's method then gives

$$K_3 = c_3\sqrt{\frac{\tau}{\tau^2+2}}\, e^{ab\left(-\tau+\tan^{-1}(\tau)-\frac{r^2}{4(ac)^2\tau}\right)}\frac{1}{(ac)^2}\,, \tag{22}$$

where $\tau = \frac{m}{a}$. Here c_3 is the constant of normalization, which must be calculated numerically in two-dimensional space.

2.2.2.2.3 Approximating $\boldsymbol{K_4}$

Substituting $h_4(t)$ into (8) and using $D = \frac{c^2}{ab}$, we have

$$K_4(x) = \int_0^\infty \frac{b^2 t(\frac{t}{a}+1)^{-\beta-1}}{4\pi c^2} e^{ab(-\frac{t^2}{2a}+t-\frac{x^2}{4c^2 t})} dt \tag{23}$$

where $\beta = a^2 b$. Letting $\lambda = ab$, and identifying

$$f(t) = \frac{b^2 t\left(\frac{t}{a}+1\right)^{-a^2 b-1}}{4\pi c^2} \text{ and } \quad g(t) = -\frac{t^2}{2a} + t - \frac{x^2}{4c^2 t}, \tag{24}$$

we find the critical point is $m = \frac{a}{A+B}$, where

$$A = \sqrt[3]{2\frac{(ac)^2}{r^2}\left(1+\sqrt{1+\frac{16}{27}\frac{(ac)^2}{r^2}}\right)} \text{ and } B = \sqrt[3]{2\frac{(ac)^2}{r^2}\left(1-\sqrt{1+\frac{16}{27}\frac{(ac)^2}{r^2}}\right)} \tag{25}$$

(details in appendix A). Application of Laplace's method then gives

$$K_4 = c_4 \tau \frac{(1+\tau)^{-(a^2 b+1)}}{\sqrt{1+\frac{r^2}{2(ac)^2\tau^3}}} e^{\beta\left(-\frac{1}{2}\tau^2+\tau-\frac{r^2}{4(ac)^2\tau}\right)} \frac{1}{(ac)^2}, \tag{26}$$

where $\tau = \frac{m}{a}$. Here c_4 is the constant of normalization, which must be calculated numerically in two-dimensional space.

2.3 Data

The data used for this project was provided by the Wisconsin Department of Natural Resources (WDNR), collaboratively with more than 300 landowners, more than a thousand (1,157) WTD individuals were GPS collared from 2017 to 2020 in Dane, Grant, and Iowa counties in southwestern Wisconsin. The data was provided with potential erroneous readings screened out using dilution of precision and ruling out location changes due to GPS errors.

The data consists of time series of positions of deer juveniles, timestamp in Julian days at each position, difference between two consequent positions: difference between two consequent y positions: dy, juveniles' age, sex, collar ID, collar start and end dates, reason for removing the collar, x and y coordinates of mortality position if applicable, and simple cause of death like car accident, harvest, ... etc. if applicable. In addition, the data

was partitioned into male, female, 8-months-old, and 20-months-old deer. Moreover, dispersals were classified based on season of dispersal either spring or fall. Combinations of all these six categories were made to reveal the effects of sex, age, and dispersal season on the dispersal behavior.

2.4 Long-distance dispersal filtering

Long-distance dispersals are defined as the cumulative distance travelled from one home range to the other. To prepare for testing the proposed kernels against the long-distance dispersals from GPS collar data, we needed to extract the long-distance dispersals first. However, the datapoints were unevenly distributed in time, and this caused the cumulative distance and cumulative velocity curves to have sharp spikes up and down. Moreover, we needed to define home ranges to be able to calculate distances in between them. This required us to eliminate false home ranges to get a correct resolution on what the long distances are. Hence, extraction of long-distance dispersals was performed in two stages: Fourier smoothing, and filtering based on period and distance of dispersal.

2.4.1 Fourier smoothing

The Fast Fourier Transform (FFT) was used to remove transient spikes and identify sustained average velocities associated with long-distance dispersal. The velocities were smoothed based on frequencies with periods longer than 11 days. A subsequent velocity threshold of 300 m/day was set to eliminate movement not associated with travel in between home ranges. The 11 days and 300 m/day estimations were based on optimization of area under the receiver-operator curve. Potential home ranges were distinguished as intervals during which average velocity did not exceed the migration threshold.

2.4.2 Space and time filtering

After determining the potential home ranges, the starting and ending positions of potential long-distance dispersal movements were calculated based on displacement. Then, we ignored intervals when deer spent less than two weeks to avoid false dispersals as deer might be having a rest during travelling, not a settling into a new home range. In another

scenario those relocations could be interpreted as transient walkabouts from and to their home ranges, not long-distance dispersals due. Moreover, we ignored relocations that did not exceed twice the standard deviation of the average population home range, to avoid overlapping home ranges. This latter condition sets the boundaries of each home range in such a way that it could not be confused with a later home range.

2.5 Parameter fitting

2.5.1 Maximum Likelihood Estimation

We used Negative Log Likelihood (NLL) to estimate the parameters α and β in the four kernels. Let the observed dispersal distances be denoted by $\{r_j\}_{j=1}^{n}$, then the likelihood of model i is:

$$L_i = \Pi_j(K_i(r_j; \alpha, \beta) r_j \Delta r), \tag{26}$$

The NLL for each model was then

$$NLL_i = -\sum_{j=1}^{N} \log(K_i(r_j; \alpha, \beta) r_j \Delta r). \tag{27}$$

This was the objective function for minimalization using MATLAB's fminsearch.

2.6 Model competition & uncertainty quantification

2.6.1 Bayesian Information Criterion competition

BIC measure is used to estimate the competency of a model. The lower the BIC of a model, the better the model fit to the data, penalized by model complexity. Differences in BIC values assess how a model performs in comparison to another model. We used ΔBIC values to distinguish the best of all four hypothesized models for long-distance dispersal. Moreover, as the total juvenile population was subdivided into male and female juveniles of age 8 months, and age 20 months dispersing in spring and fall, the ΔBIC values were calculated for each category to select the kernel showing the best fit for it. Furthermore, we used BIC to determine which is most competitive between having one model for the whole juvenile dataset or using separate models for males and females, as we expected that male dispersal behavior favors a different dispersal model than females.

BIC values were calculated for all four models using

$$BIC = 2 * NLL + k * \log\left(\frac{N}{2\pi}\right). \quad (28)$$

Here the parameters k, N are the number of model parameters and number of observations respectively. Odds ratio is another measure that we are using to determine the probability that one model is better than another model, and hence detect if a model introduces significant improvement. The odds ratio between two models with a BIC difference of ΔBIC is $e^{\frac{1}{2}\Delta\text{BIC}}$. Significance is considered at ΔBIC = 4.6 as this yields odds ratio of 10 reflecting odds ten times in favor of the better model.

2.6.2 Nonparametric bootstrapping

We performed nonparametric bootstrapping to determine Confidence Intervals (CIs) and quantify uncertainty. One thousand datasets were generated by random sampling from a uniform distribution with replacement. For each dataset, models were fit and parameters were estimated using MLE. CIs were calculated for all parameters and BIC values with 95% confidence.

Moreover, the one thousand bootstrapped datasets were each partitioned into males and females. Each category was fit to the four models and BICs were calculated. The models showing the best fit for each category: females and males were determined as they had the lowest mean BIC value over the bootstrapped sample. Then, the ΔBICs were calculated relative to the lowest model. After that, to determine if there is an advantage to use different models for males and females than one model for the joint sample (males and females grouped), we added the BICs of male sample (calculated using the best model for males) to the BICs of female sample (calculated using the best model for females) and compared the summation to the BICs of the joint sample (calculated using the best model for joint sample). Also, we used ΔBIC as a measure of model competition considering that the better hypothesis is to fit different models to males and females, so they were subtracted from the joint BICs.

2.7 Assessing impact of long-distance dispersal on CWD spread

2.7.1 Specify IDE model variables

The short-distance dispersal kernel K_S was defined as follows:

$$K_S = \frac{e^{-\frac{r^2}{2\sigma^2}}}{2\pi\sigma^2} \tag{29}$$

where σ is the mean distance covered in dispersal within a home range. Moreover, a formula for the F function proposed,

$$F = P(\text{escape threshold environmental exposure}) \bullet P(\text{escape direct infection})$$

$$= e^{-\mu\lfloor H_n - H_c \rfloor}\left(1 + \frac{\alpha I_n}{\kappa}\right)^{-\kappa}. \tag{30}$$

Here the parameter μ is yearly probability per area/year of encountering prion hazard above critical threshold, H_c is the critical dose of environmental hazard, κ is the clumping parameter for negative binomial distribution, and $\lfloor x \rfloor$ notation means that if the value x is negative, a zero will be substituted for it. The probability of deer escaping CWD infection F can be estimated as a fraction of indirect to direct CWD transmission terms. Indirect transmission is estimated by the probability that the environmental hazard dose a deer is exposed to H_n does not reach H_c. So, P(escape exposure to hazard above threshold) assumes a Poisson distribution for a deer not to encounter an environmental hazard that exceeds critical threshold. On the other hand, P(escape direct contact among conspecifics) accounts for the probability that a deer escapes direct contact with an infected deer according to a negative binomial distribution.

2.7.2 Fitting IDE model parameters

The IDE model parameters in the absence of dispersal were fit using study data published by Miller et al. (2006). They provided cumulative death rates of captive mule deer from two CWD epidemics beginning in 1974 and 1992 respectively and held at the Colorado Division of Wildlife Foothills Wildlife Research Facility. The first epidemic (1974-1985) caused the second epidemic to happen (1992-2001), despite amelioration efforts (removal

of carcasses, exposure to free air and sunshine) which supports that CWD prions are very hard to eliminate.

The mortality rates provided by Miller et al. (2006) were used to determine five of the IDE model parameters: prion decay rate η, rate of shedding prions γ, indirect transmission coefficient μ, direct transmission coefficient ν, and clumping parameter κ. Parameters related to dispersal (either short or long) could not have been fit from Miller et al. (2006) data as the study was made on penned herds. To estimate other parameters, we have used the fminsearch function in MATLAB to minimize Sum Squared Error (SSE) between the IDE model predictions and observed mortality.

Theses parameters represent contact rates in penned herds; Miller et al. (2006) suggested that direct and indirect CWD transmission coefficients should be divided by 10 to reflect the crowdedness of the deer pen compared to the wild. We seek to determine a more specific scaling parameter. Following Hefley et al. (2017), we used their estimation of a net exponential force of infection rate $\lambda \approx 0.985$ per year. Let s be a contact scaling factor and the total population density $P_{total} \approx 40$ WTD/km^2 (approximate average density in southern WI). Then,

$$2.678 \approx e^{\lambda} = [1 - F(1,\, H_c + 1)]P \approx s[\nu + \mu]P \approx s\ (0.2)(40) \qquad (31)$$

Solving for s gives $s \approx 0.335$.

Moreover, through direct communication with Dr. Joel Pedersen, we learned that the amount of prions being shed by an infected deer during its lifetime, γ, is approximately the same as the amount of prions found in its carcass, so we doubled our estimate for γ. Furthermore, we estimated the standard deviation of an average home range is 819 m which will be used as the standard deviation, σ, for the short dispersal kernel, K_S.

2.7.3 IDE model simulations

The best performing long-distance dispersal kernel derived in this study was substituted in the IDE model to determine its effect on CWD spread. We have used FFT to solve the IDE model on a discretized space domain. Considering a one-thousand-kilometer spatial domain centered around the location $x = 0$ with step size $\frac{2*500}{2^{13}}$ km. As for the time domain, we have used a total time of 100 timesteps where each step is a year. We have set the initial infected deer density I_0 to 3 infected deer per km^2 in the space from $x = -10$ to 10. Then, to maintain the total population density P_{total} constant, which we assumed it to be 100 deer/km^2 (approximate deer density in the core CWD region of WI), the initial susceptible deer density S_0 was set to $100 - I_0$. The initial CWD hazard density H_0 was to set to be the quasi-steady state value of H which is the equilibrium ratio of hazard to infected deer density. Hence, $H_0 = \frac{\gamma}{(1-e^{\eta})* I_0}$. The fraction of long-distance dispersers ρ was set to 0.05. A critical threshold was chosen for the environmental hazard H_c and set to 1 infectious dose per km^2. Hence, all other hazard-related parameters were measured relative to it.

The density of infected deer was simulated over the space domain. A threshold for detectable level of CWD in a deer population was set to 5% of the total population density. Then, successive waves of infected deer density were graphed versus the space domain. The locations at which each wave equals the CWD detection threshold were determined. After that, the front velocity of CWD, c_{front}, is a linear fit estimated by the difference between locations over number of years. We calculated it over the 100 years to get a good grasp of how CWD is spreading over time. Also, we evaluated the impact of long-distance dispersal as total deer population density P_{total} and fraction of long-distance dispersers ρ vary in realistic parameter regimes by plotting the front velocity of CWD, c_{front}, versus total deer population density P_{total} using $\rho = 0$, 0.01, 0.05, and 0.1.

CHAPTER III

RESULTS

3.1 Characteristics of long-distance dispersal

Several measures of central tendency and ratios were calculated to characterize the empirical data. Males represented 51% of the total juvenile population in our dataset. Out of 363 deer, 113 exhibited long-distance dispersal. Mean dispersal distance for males was 10,325 m while that for females was 14,567 m. More juvenile males (49%) exhibited dispersal than females (12.36%). Moreover, 48% of deer died by the end of the study due to different reasons including CWD sickness. In about 42% of long-distance dispersals the deer dies in the end, which poses the danger of spreading infection to new regions if the deer was CWD positive. Percentage of males who were not reported to die by the end of the study (48%) was less than that of females (57%).

The age of juveniles was estimated at the time of collaring. All collaring events took place in winter, after the birth year. The fraction of 8-months-old juveniles was 71% while 20-months-old juveniles were only 29%. Juveniles of age 8 months old exhibited dispersal (40%) much more than juveniles of age 20 months old (10%). Almost the same fraction of juveniles died for each age 8 (53%) and 20 (50%) months years old. Twenty-month juveniles dispersed on average for slightly longer distances (12,837 m) than the 8-month juveniles (12,019 m).

We found that juveniles disperse in two episodes over the year, spring and fall (Figure 4). Frequency of long-distance dispersals in both episodes indicate that juveniles are slightly more likely to disperse in spring (58 long-distance dispersals) than fall (55 long-distance dispersals). Mean long-distance dispersals in fall (15,853 m) was almost double that in spring (8,124.7 m). Percentage of long-distance dispersal events were calculated at first spring, first fall, and second spring, where the order of the season is

according to which season juveniles encounter first after maturation (Figure 5). We found that most long-distance dispersal events take place at the age of one with a slight bias towards the spring season.

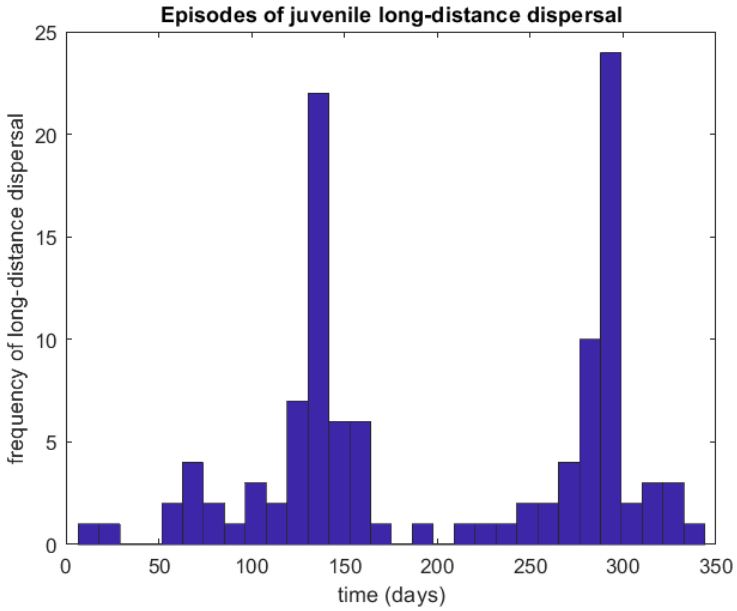

Figure 4. Long-distance dispersal in juveniles occurs mainly in two seasons. Two main peaks are shown for dispersal indicating spring and fall dispersals respectively.

3.2 Model competition and uncertainty quantification

3.2.1 Model competition with GPS collar data

The PDE kernels were multiplied by a factor of $2\pi r$ to account for observations in radidal bins, and plotted versus the dispersal distance histograms (Figure 6). The four kernels showed a good fit to the data with models II and III as the best two. On the other hand, model IV showed the least accurate fit. For more accurate model competition that differentiates between the performances of models II and III, we calculated BIC values for each using NLL. The best fit according to the BIC values was model III with BIC value of 4803.9 (lowest BIC value) while the second best was model II with BIC value of 4805.4. The third performing model was model I, and lastly model IV.

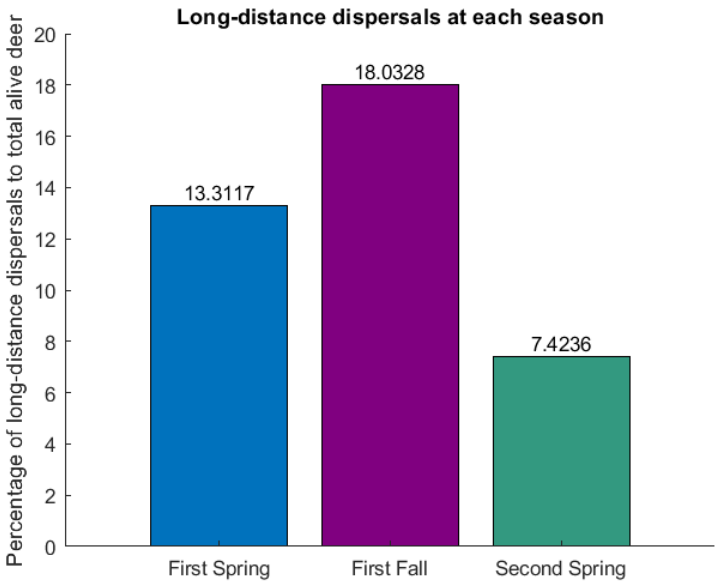

Figure 5. Percentage of long-distance dispersal events to total alive juvenile deer at a given season. The order of the season is according to which season juveniles encounter first after maturation.

Model parameters α and β were estimated for each of the four PDE models using the original dataset (Table 1) and using the bootstrapped dataset (Table 2).

Table 1. Nominal values of the PDE model parameters using joint sample

	Model I	Model II	Model III	Model IV
α (km)	7.7366	1.0891	1.0683	5.9993
β	--	0.1657	0.1504	1.0503

Table 2. Estimated parameters of PDE model among long-distance dispersal models using the bootstrapped joint sample

	Model I	Model II	Model III	Model IV
Mean α (km)	7.7306	1.3020	1.1577	6.3036
CI of α (km)	(6.6867, 8.7899)	(0.5149, 2.4795)	(0.7421,1.746)	(5.256,7.5696)
Mean β	--	0.2194	0.1712	1.1566
CI of β	--	(0.0548,0.4954)	(0.0902,0.3018)	(0.8337,1.7104)

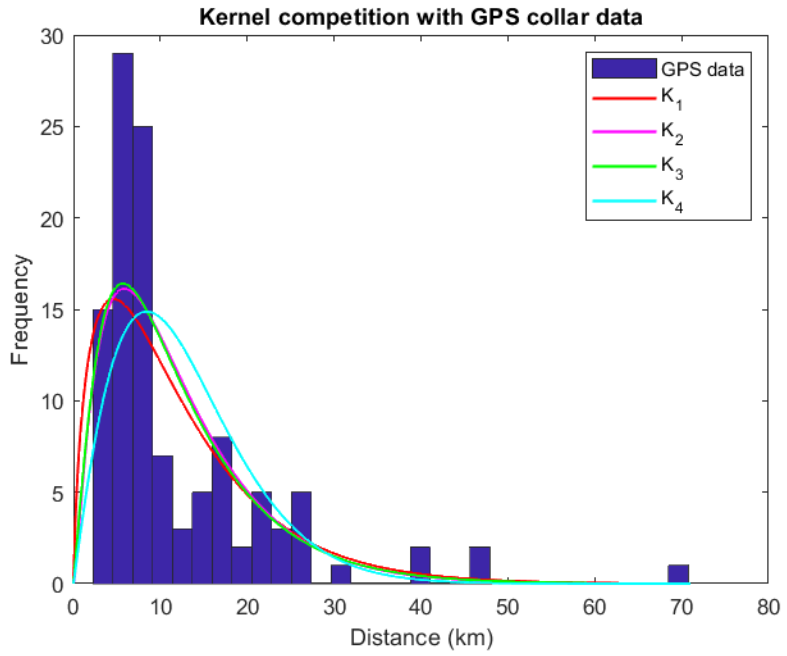

Figure 6. Fits of models I, II, III, and IV shown in red, magenta, green, and cyan respectively are plotted against long-distance dispersals from GPS data in km. Model I assumes deer will have equal tendency to settle down at any time from dispersal start. Model II considers that deer would not settle right away, and tendency increases with time. Model III adds that deer are motivated not to settle down any time soon after the dispersal start while model IV assumes motivation to settle down increases linearly with time. The four kernels show a good approximation to the data. However, K_4 is a little off the general behavior of data.

Nominal values of BIC values were near the peak in the CIs and their distributions behave well despite changing the data (Figure 7).

Table 3. Nominal BIC values competition among long-distance dispersal models using the joint sample

	Model I	Model II	Model III	Model IV
BIC	4807.8	4805.4	4803.9*	4828.1
ΔBIC	3.8846	1.4795	0	24.2398
Odds ratio	6.9748	2.0954	1	1.8349e+05

* Indicates the best model

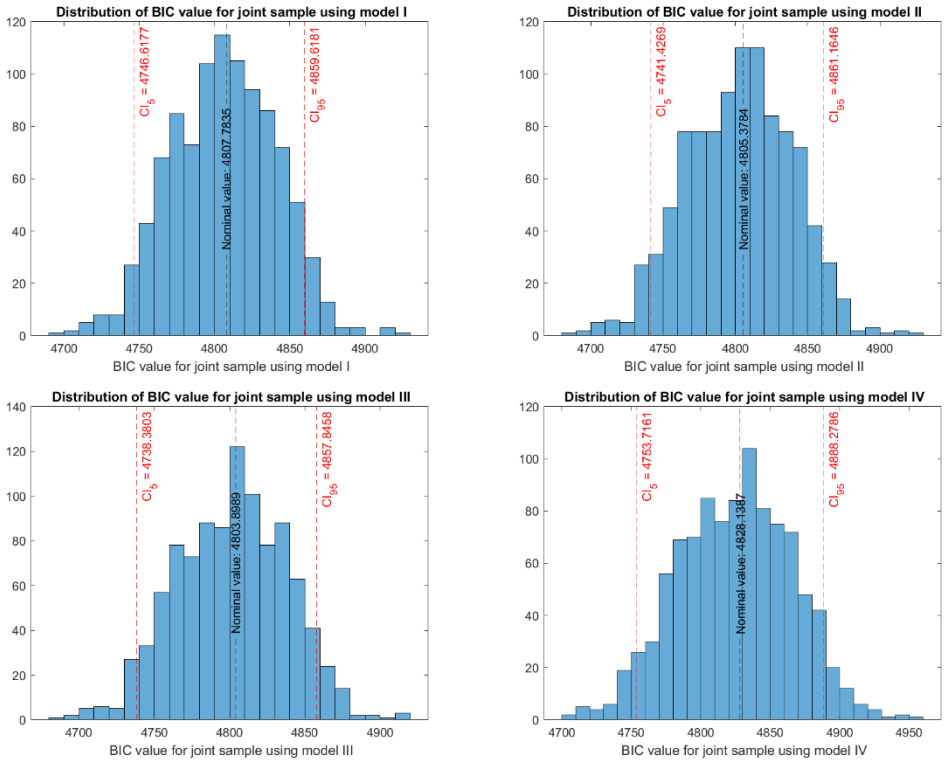

Figure 7. BIC distribution for all models using one thousand bootstrapped datasets. All four distributions are well-behaved and the nominal BIC values are close to the peaks and in the CI range which supports that the BIC values are valid.

3.2.2 Joint model selection based on ΔBIC

The ΔBIC values were calculated to evaluate the performance of long-distance dispersal models relative to the best. The BIC values showed that K_3 is the best kernel as it has the lowest BIC, and competency of the other models was assessed using ΔBIC relative to model III using the original dataset (Table 3) and using the bootstrapped dataset (Table 4 & Figure 8). Model II shows the second-best performance. The odds ratio shows that model III is almost seven times better than model I, and more than two times better than model II. On the other hand, model III is massively (1.8e+5) better than model IV.

3.2.3 Model selection across ages, sexes, and seasons

We assessed different kernels according to their ΔBIC performance with different subpopulations. Hence, using the GPS data, we learned that male juveniles showed better fit with Model III while females showed better fit with model I. Also, the 8-months-old

juveniles behaved best like model III with BIC value of 4376.8 and the second-best model for them was model II with BIC value of 4378.1, while those of 20 months old, favored model I with BIC value of 427.52 and the second-best model for them was model III with BIC value of 427.61. In the spring, juveniles preferred model II with BIC value of 2248, and their second best was model III with BIC value of 2248.5. However, in the fall, the best model for juveniles was model I with BIC 2530.4 and second best was model III with BIC 2530.4. Here, the comparison of BIC values across different genders is not applicable as the data modelled in the two cases is different.

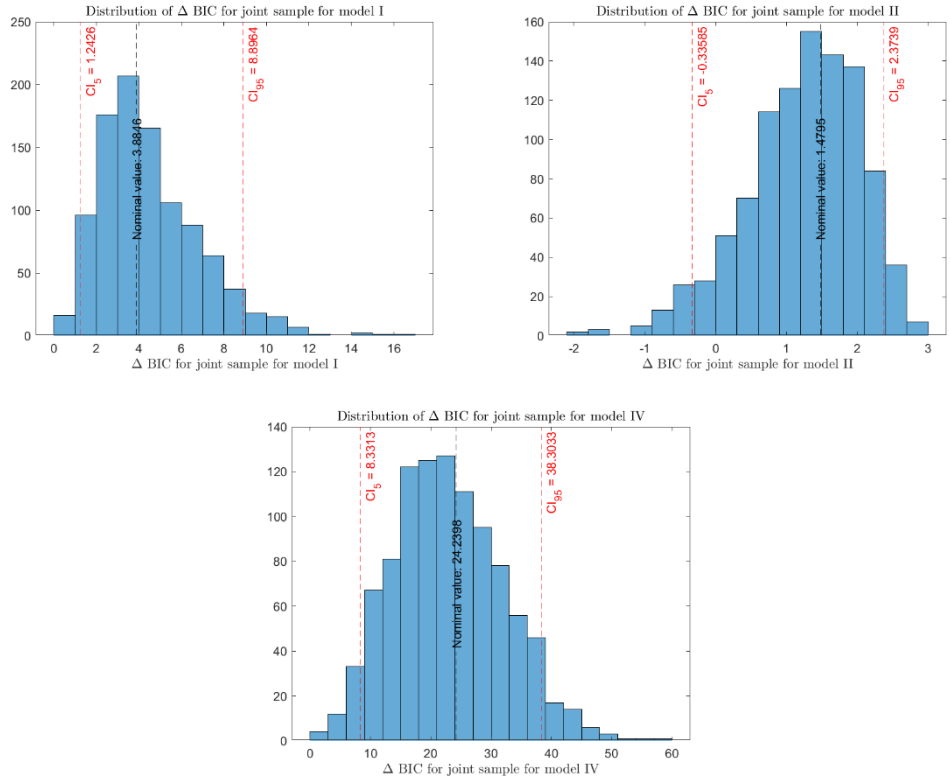

Figure 8. ΔBIC distribution for model I, II, IV relative to model III using one thousand bootstrapped datasets. All four distributions are well-behaved and the nominal ΔBIC values are close to the peaks and in the CI range which supports that the ΔBIC values are valid. Relative to model III, all models show positive mean ΔBIC values.

Table 4. Calculated BIC values among long-distance dispersal models using the bootstrapped joint sample

	Model I	Model II	Model III	Model IV
Mean BIC	4806.1	4802.8	4801.6*	4824.6
CI of BIC	(4747, 4860)	(4741,4861)	(4738,4858)	(4754,4888)
Δ Mean BIC	4.5	1.2	0	23
CI of ΔBIC	(1.2426,8.8964)	(-0.3358,2.374)	--	(8.3313,38.303)
Odds ratio	9.49	1.82	1	98715

* Indicates the best model

3.2.3.1 Competition between one model for both genders versus two models: males and females

We assessed the competition between setting up one model for the whole juvenile population or two different models for juvenile males and juvenile females. The best model was chosen for males and for females based on ΔBIC calculations using the original dataset (Tables 5 & 6 respectively) and the bootstrapped dataset (Tables 7 & 8 respectively). The results agree that the best fit for males was model II while for females it was model I in comparison to model III, the best fit for the joint sample. However, the odds ratio show that model III was very close to best represent the male sample using both the original and bootstrapped datasets. Also, it shows that model III was nearly the best fit for the female sample rather than model I using both datasets. Nevertheless, we found that there is an overall advantage of fitting different models to each category: males and females (Figures 10, 11) as the summation of mean BIC values for both male and females sample gave 4796.0 while the mean BIC value of the joint for the joint sample was 4801.6. This was also supported when the nominal BIC values were used as the summation of BIC values for male and female samples was 4799.77 while that of the joint sample was 4803.9.

Table 5. Nominal BIC values competition among long-distance dispersal models using the male sample

	Model I	Model II	Model III	Model IV
BIC	3849.5	3843.7	3843.6*	3852.6
ΔBIC	5.878	0.1054	0	9.0475
Odds ratio	18.897	1.054	1	92.18

* Indicates the best model

Table 6. Nominal BIC values competition among long-distance dispersal models using the female sample

	Model I	Model II	Model III	Model IV
BIC	956.17*	957.17	956.77	968.26
ΔBIC	0	0.9985	0.5972	12.092
Odds ratio	1	1.647	1.3479	422.42

* Indicates the best model

Table 7. Bootstrapped BIC values among long-distance dispersal models using the male sample

	Model I	Model II	Model III	Model IV
Mean BIC	3849.1	3842.4875*	3842.5185	3851.2
CI of BIC	(3518.4, 4135.9)	(3511.8, 4129.2)	(3512.2, 4129.4)	(3532, 4136.8)
Δ Mean BIC	6.6125	0	0.031	8.7125
CI of ΔBIC	(1.421,12.99)	--	(-1.077,1.1498)	(3.2265,15.752)
Odds ratio	27.28	1	1.0156	77.964

* Indicates the best model

Furthermore, we estimated the PDE model parameters α, and β for both males and females through minimizing the NLL using the original dataset (Tables 9 & 10) and bootstrapped dataset (Tables 11 & 12).

Table 8. Bootstrapped BIC values among long-distance dispersal models using the female sample

	Model I	Model II	Model III	Model IV
Mean BIC	953.55*	954.28	953.95	964.07
CI of BIC	(662.7,1265.3)	(662.9,1266.1)	(662.72,1265.7)	(666.94,1275.9)
Δ Mean BIC	0	0.73	0.4	10.52
CI of ΔBIC	--	(-0.087,1.1446)	(-0.208,0.733)	(1.5221,18.802)
Odds ratio	1	1.4405	1.2214	192.48

* Indicates the best model

Table 9. Nominal values of the PDE model parameters using the male sample

	Model I	Model II	Model III	Model IV
α (km)	7.2005	2.2097	1.6202	7.2426
β	--	0.41562	0.26548	1.5761

Table 10. Nominal values of the PDE model parameters using the female sample

	Model I	Model II	Model III	Model IV
α (km)	10.017	0.13304	0.4152604	6.8389
β		0.013	0.0408	0.8199

Table 11. Estimated parameters of PDE model among long-distance dispersal models using the bootstrapped male sample

	Model I	Model II	Model III	Model IV
Mean α (km)	7.2013	2.7765	1.8247	7.7837
CI of α (km)	(6.3095, 8.1396)	(0.995, 5.912)	(1.0288, 3.0529)	(5.899, 10.578)
Mean β	--	0.59487	0.32115	1.8199
CI of β	--	(0.1345, 1.424)	(0.148, 0.619)	(1.054, 3.124)

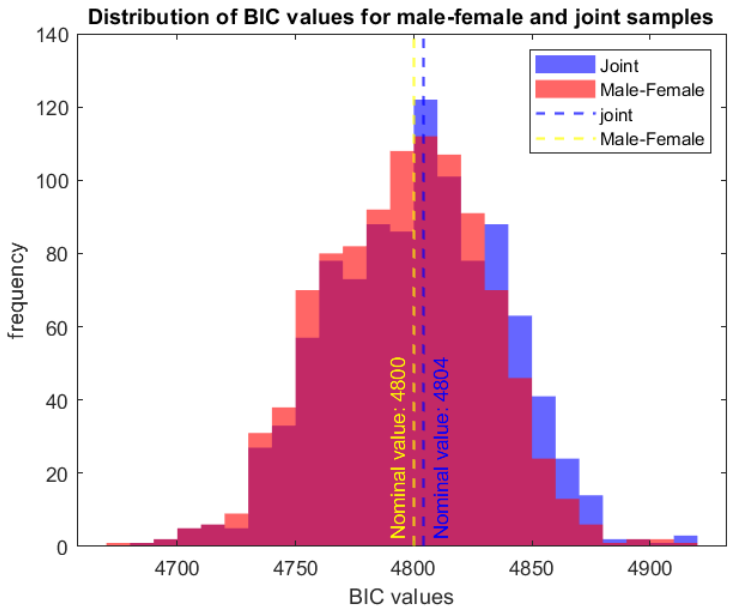

Figure 10. Distribution of BIC values for male-female and joint samples. The male bootstrapped sample was fit to model II, and the female bootstrapped sample was fit to model I whereas the joint bootstrapped sample (males and females) was fit to model III. The model selection was made based on the lowest BIC values. In this figure, the BICs (including nominal values) for males and females were added together versus the BICs for the joint sample. The trend shows that fitting different models to males and females gives overall lower BIC values than fitting one model for the joint sample. The summation of the nominal BIC values for male and female samples was 4800 while that of the joint sample was 4804. Hence, the ΔBIC is 4, which gives odds ratio of 7.39 (close to be significant) indicating that using separate models for each gender is more than seven times better than using one model for both genders.

Table 12. Estimated parameters of PDE model among long-distance dispersal models using the bootstrapped female sample

	Model I	Model II	Model III	Model IV
Mean α (km)	10.083	0.45192	0.54903	6.9654
CI of α (km)	(6.6548, 14.661)	(0.0028, 1.3577)	(0.2338, 1.112)	(5.3119, 9.001)
Mean β	--	0.09131	0.06574	0.89269
CI of β	--	(9.655e-5, 0.21)	(0.016, 0.151)	(0.716, 1.188)

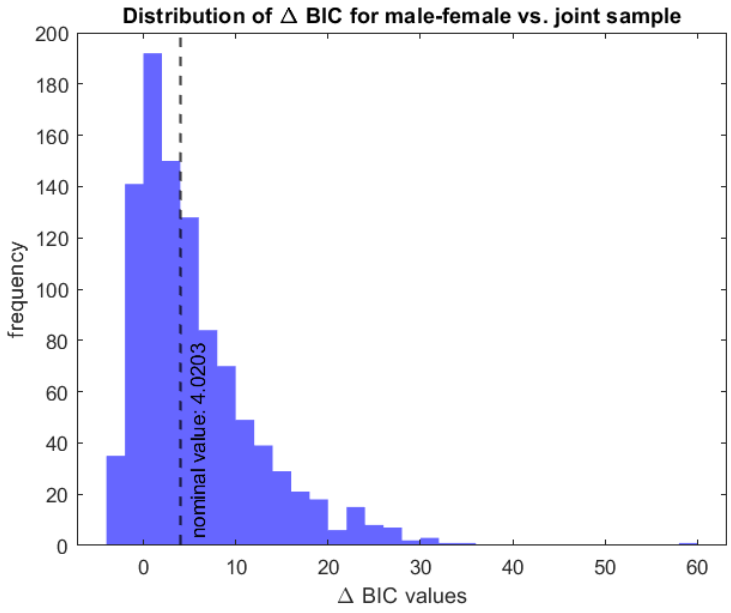

Figure 11. Distribution of ΔBIC values for male-female and joint samples. The BICs of male and female samples were subtracted from the BICs from the joint (male and female grouped). The graph shows a right tail (82.4% of ΔBICs are positive) which supports the hypothesis that fitting different kernels to males and females is better than fitting one kernel for both due to behavior variation between genders.

3.3 Impact of long-distance dispersals on CWD spread

The IDE parameters: prion decay rate η, rate of shedding prions γ, indirect transmission coefficient μ, direct transmission coefficient ν, and clumping parameter κ were estimated as using SSE (Table 13). Using the fitted parameters, the IDE model showed a very good fit to the mortality rates reported by Miller et al. (2006) (Figure 12).

Using the scaling factor, $s = 0.335$, which matches our IDE model contact parameters with observed growth of infection. (Hefley et al., 2017) come to the values in Table 14. The direct and indirect transmission coefficients ν and μ respectively were divided each by s to account for changing the setting of deer from a pen as reported by Miller et al., (2006) to the wild. The rate of shedding prions γ was multiplied by two to include the number of prions added to the environmental hazard by corpses, since corpses were removed in the experiment done by Miller et al. (2006). Hence, the parameters γ, ν,

and μ were adjusted, the parameters H_c, and P were set, and we used the estimates from Miller (2006) for the parameters η, and κ, and the estimate from our GPS collar data for the parameter σ.

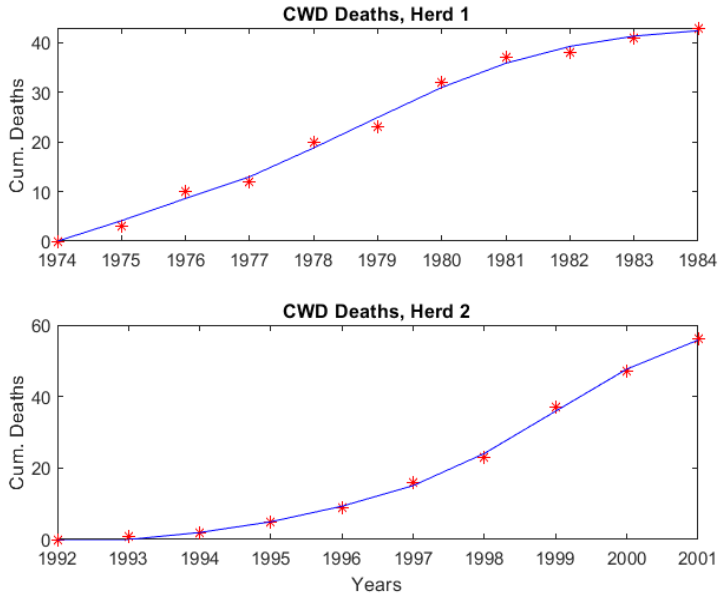

Figure 12. IDE model predictions for CWD mortality rates in two herds of WTD. The datapoints showing death rates from Miller et al. (2006) in red, were used to fit the IDE model parameters. Best-fit IDE model predictions are shown in blue for herd 1 in 1974, and herd 2 in 1992.

3.3.1 Effect of different long distance dispersal kernels on CWD spread

CWD infection waves were simulated over 100 years using the best performing long-distance dispersal model, model III (Figures 14 & 16) versus not accounting for any long-distance dispersal behavior (Figures 13 & 15) to assess the effect of long-distance dispersal on CWD spread. In Figures 14 & 16, the total deer population P_{total} was chosen to be 100 deer per km^2 and the fraction of long-distance dispersers ρ was chosen to be 0.05. While deer densities in the core WI CWD area approach this density, it is unrealistically high for most environments and was chosen purely for purposes of illustration. The figures support that long-distance dispersal enhances CWD spread into new regions as the

infected individuals spread over more than 400 km instead of only 100 km in the case of deer not travelling to new home ranges.

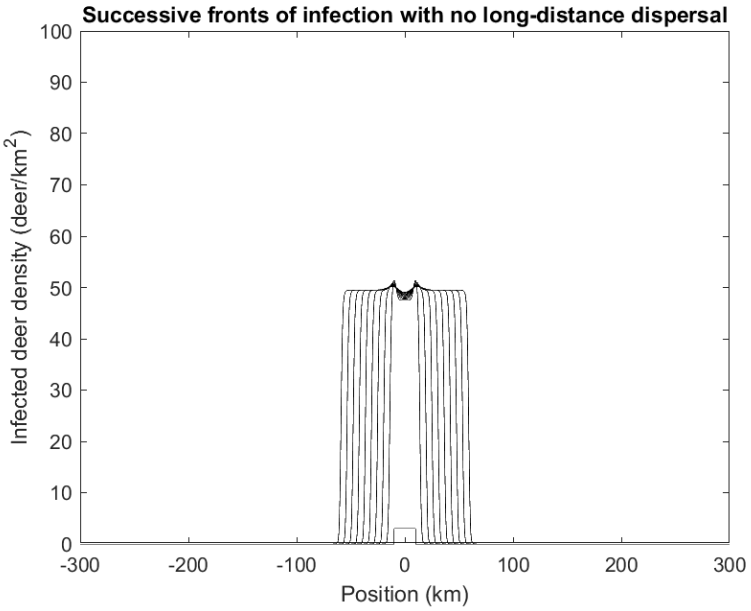

Figure 13. CWD wave simulations over space domain for 100 years with no long-distance dispersal ($\rho = 0$) and total population $P_{total} = 100$.

Table 13. Parameters estimated from Miller et al. (2006) data

Parameter	Value	Unit
Prion decay rate (η)	0.478	Year^{-1}
Rate of shedding prions (γ)	0.232	Infectious dose/deer
Indirect transmission coefficient (μ)	0.177	Area/infectious dose
Direct transmission coefficient (ν)	0.0313	Area/deer
Clumping parameter (κ)	0.769	--

CWD wave front positions show a linear trend with time. The starting jump is a transient due to the square-wave initial conditions accelerating to counter-propagating waves of invasion. (Figures 17 & 18).

Table 14. Final IDE model parameters

Parameter	Value	Unit
Standard deviation of short-dispersal distance (σ)	0.819	Km
Prion decay rate (η)	0.478	Year^{-1}
Rate of shedding prions (γ)	0.464	Infectious dose/deer
Direct transmission coefficient (ν)	0.010475	Area/deer
Indirect transmission coefficient (μ)	0.05925	Area/infectious dose
Clumping parameter (κ)	0.769	--
Prion hazard threshold (H_c)	1	Infectious dose/km^2
Initial population density (P)	100	Deer/area

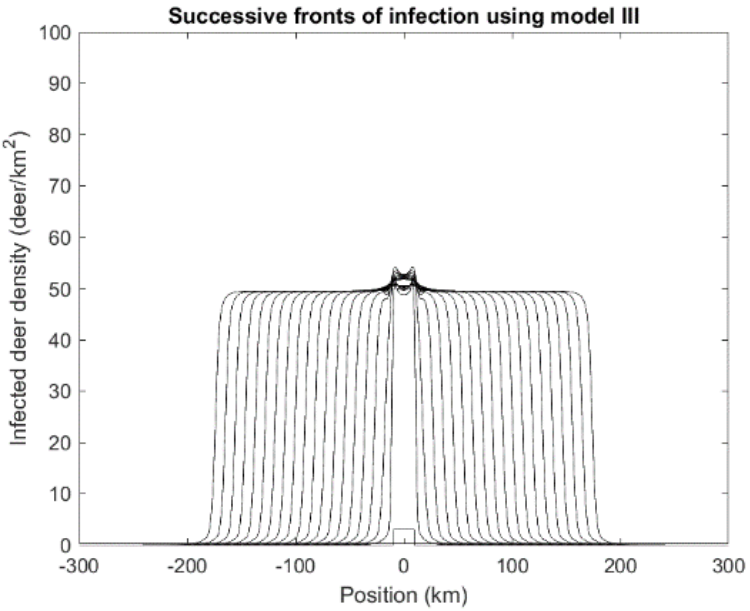

Figure 14. CWD wave simulations over space domain for 100 years using the third long-distance dispersal model with total population $P_{total} = 100$, and fraction of long-distance dispersers $\rho = 0.05$.

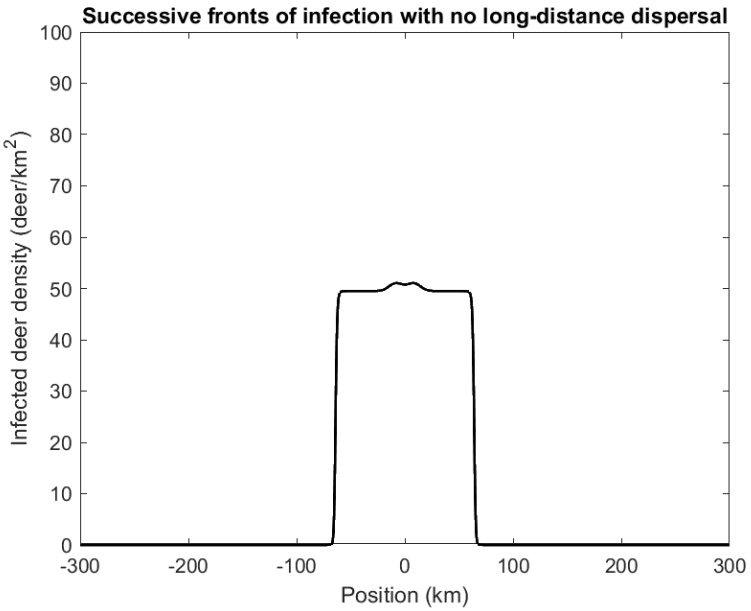

Figure 15. CWD 100$^{\text{th}}$ wave outline over space domain with no long-distance dispersal and total population $P_{total} = 100$.

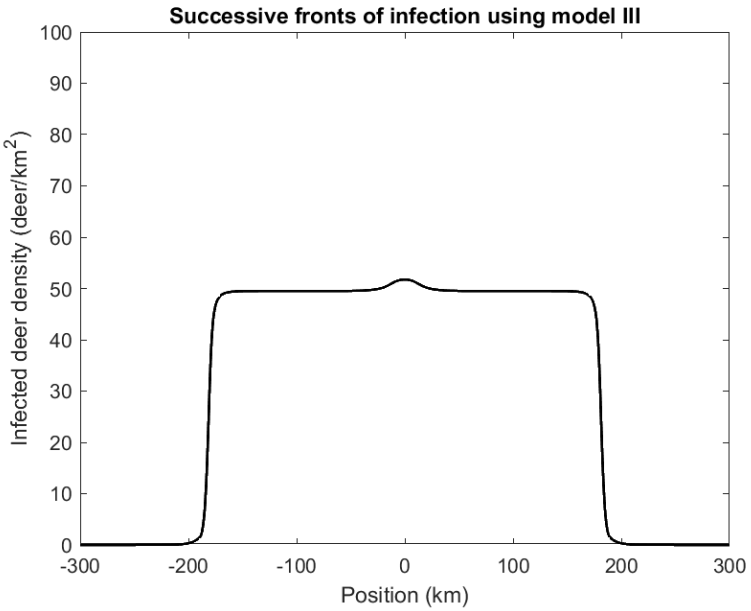

Figure 16. CWD 100$^{\text{th}}$ wave outline over space using the third long-distance dispersal model, total population $P_{total} = 100$, and fraction of long-distance dispersers $\rho = 0.05$.

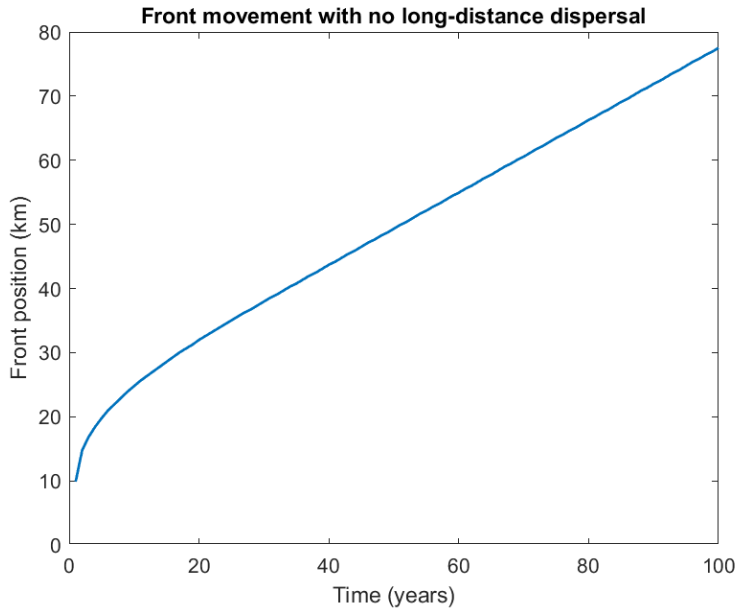

Figure 17. Front movement plotted over time domain with no long-distance dispersal, and total population $P_{total} = 100$. The figure shows slow and smooth disease spread in the first few years then it approaches a linear trend of changing the front position over time.

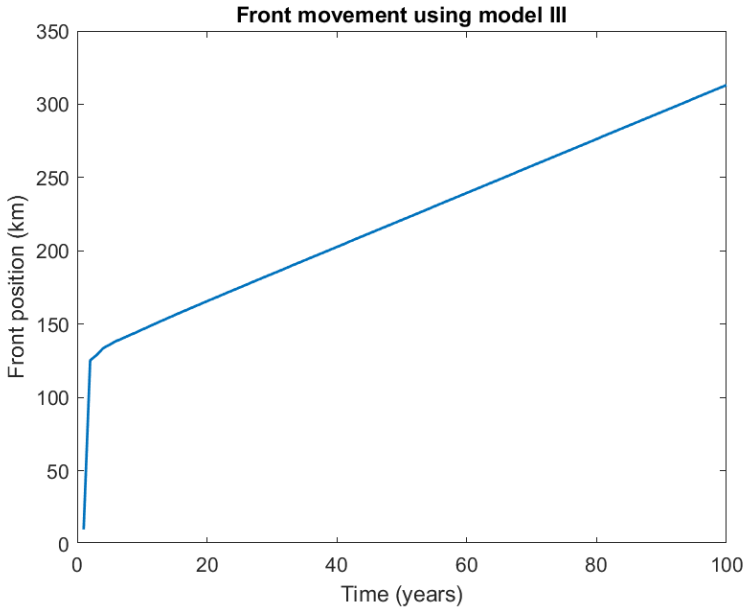

Figure 18. Front movement plotted over time domain using the third long-distance dispersal model with total population $P_{total} = 100$, and fraction of long-distance dispersers $\rho = 0.05$. Model III affects CWD with a big jump of disease spread in the first few years. Then, it increases at a steady rate but it is much faster than that in the case of no long-distance dispersal.

3.3.2 Role of population size and fraction of long-distance dispersers

The CWD front speed, c_{front}, appears to increase asymptotically linearly with higher total population density. Long-distance dispersal significantly influences CWD to spread more rapidly even for fractions of dispersing deer as low as $\rho = 0.01$ (Figure 19). The effect of long-distance dispersal on CWD spreading speed is most obvious after the total population density P_{total} becomes approximately 96 deer/km^2; this is the approach density of which an entirely susceptible population becomes unstable due to direct transmission alone.

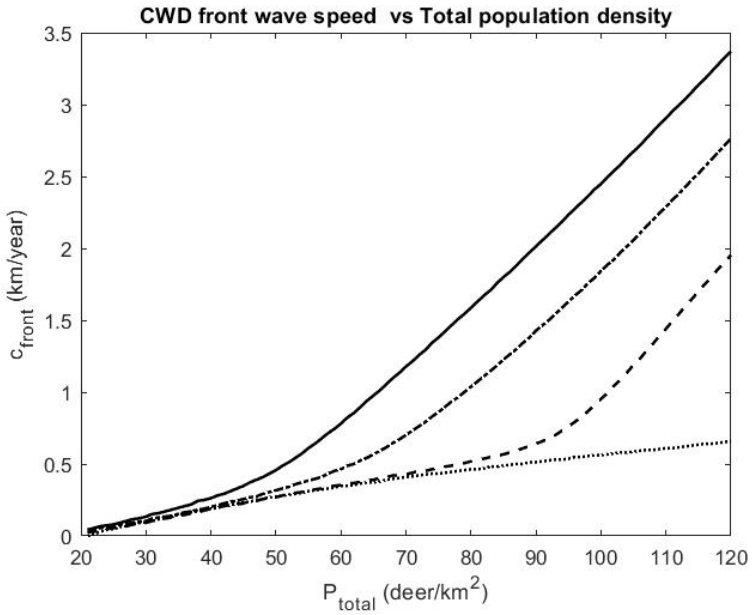

Figure 19. CWD front wave speed c_{front} simulated over total deer population density P_{total} using the third long-distance dispersal model with fraction of long-distance dispersers $\rho = 0$, 0.01, 0.05, and 0.1. The plot shows higher increases in CWD front wave speed as the fraction of long-distance disperses, ρ, increases in comparison to $\rho = 0$.

CHAPTER IV

DISCUSSION & CONCLUSION

In this thesis, we have considered long-distance dispersal events of juvenile WTD, as they may have huge impact on CWD spread. We have developed and parameterized new mechanistic long-distance dispersal kernels to describe this behavior probabilistically; this due to the key role that dispersal kernels play in assessing risk of spread. For this, we propose a novel IDE model capturing the big picture of how deer get infected with CWD by direct interaction with their conspecifics or indirect pathways through the environment. With these kernels and the IDE model, we can predict CWD spread to new regions and assist CWD management and surveillance.

A mechanistic PDE model was adapted from a seed transport and settling model presented by Neupane & Powell (2015). However, we proposed different settling rates to resemble animal movement more accurately by eliminating seeds digestion conditions affecting the settling distribution. Instead, we proposed setting rate functions denoting different mechanisms of juvenile long-distance dispersal. Then, by approximating solutions to the PDE model for long-distance dispersal kernels using Laplace's method, we introduce a novel class of dispersal kernels that captures the spatial distribution of deer after long-distance dispersal.

Four settling kernels were proposed and evaluated to find the best approximation to the observed data. The first model supposed that juvenile deer have the same probability to settle down no matter time invested in finding a home range. The second model assumes deer would not settle right away, and tendency increases with time. As for the third model, it considered that deer do not prefer to settle down soon after they started dispersal as much as they do when they invest more time in dispersal. Lastly, the fourth

model adopted the same behavior of deer as model III but added that the tendency of deer to settle down increases as they disperse for longer time.

The four models were tested against real data from GPS collars attached to 363 juveniles of WTD between 2017 to 2020 in southwestern Wisconsin. For that, filtering home ranges was done on two stages. First, the FFT was applied to remove the erroneous unrealistic velocity spikes and identify sustained average velocities of long-distance dispersal. For this, two thresholds were set: a velocity of 300 m/day and a frequency threshold based on an 11-day interval as the time window for smoothing deer velocities. Then, the starting and ending locations of potential long-distance dispersal movements were calculated. Second, time and space filtering were applied as follows. We disregarded relocations that were more than twice the average population home range standard deviation to prevent overlapping home ranges. Also, we excluded ranges where deer spent less than two weeks to avoid rests during travel and short walkabouts. Then, model parameters were fitted using MLE and models competed using common data and BIC values. Furthermore, we performed uncertainty quantification through nonparametric bootstrapping.

Our findings showed that all nominal parameter values were close to the peak and within the 95% confidence intervals. For the joint sample, including both males and females, the third kernel K_3 indicated a general best fit to the data which supports that deer do not prefer to settle soon after they started dispersing as juveniles would be still close to their original home ranges. After they travel long enough, their tendency to stop increases till the time interval they spent in dispersal does not influence their decision to settle down. After that, deer reach their maximum tendency for settling. This hypothesis agrees with the social motive of juvenile deer to travel for long distances as adult female deer haze them to disperse away from their original home range. Hence, juvenile deer remember the stimulus for a period of time before they start settling down. However, K_3

was not significantly better than the second kernel K_2 in the case of joint sample as the odds ratio stated that K_3 is only twice better than K_2. This still supports that juvenile deer will not prefer to stop at the time they started dispersal.

The IDE model captured the CWD death rates in WTD reported by (Miller et al., 2006). Moreover, CWD waves of infection simulated over a period of 100 years showed that long-distance dispersal has increased the infected region four times (more than 400 km instead of only 100 km) in the case of 5% of the total population of deer disperse for long distances. Moreover, the speed of disease spread significantly increases with the total population density and the fraction of long-distance dispersers.

4.1 Relative merits of kernels across ages, sexes, and seasons

We found out that there is an advantage to fit different models to males and females separately as they exhibit different long-distance dispersal behaviors. This was supported by the histogram generated to compare the sum of BIC values that were separately calculated for males and females with the BIC values calculated for the joint sample consisting of both genders. Additionally, a further plot displaying the ΔBICs demonstrated that separate models are better suited for fitting each gender's data than one model for both genders combined.

Those results agree with the behavioral background of WTD long-distance dispersal as social motives behind dispersal vary depending on gender. For example, males are more likely to be hazed by adult females which makes the majority of long-distance dispersers are males. Moreover, due to the hazing, males tend to remember the stimulus of being hazed out of their home ranges for a longer time than females. So, they travel for longer distances and try to avoid settling down soon after they started dispersal. This explains the results regarding best kernel for each category where model II was the best fitting model for the male subsample and model I was the best for the female subsample.

For the IDE model to be applied on the male and female samples, more information would be needed. For example, the data we have lack the nominal fractions of males and females in the total WTD population as well as the fraction of juveniles relative to the total WTD population. Moreover, the IDE model parameters need to be specific to male and female juveniles according to their different behaviors in short-distance and long-distance dispersal, direct, and indirect interaction. Furthermore, the sample size of the female sample is not big enough to draw solid conclusions from it. Therefore, as demographic information and specific intersexual contact rates are currently unknown for WTD, it is hard to integrate them into a sex structured IDE model and derive the effect of male versus female long-distance dispersal on CWD spatial spread. Hence, this step is recommended for further research when more sex-specific data becomes available.

4.2 Higher levels of complexity to improve model accuracy

As deer always try to satisfy their motives behind travelling such as searching for food, water resources, cover, or mates, escaping danger like hunting, avoiding competition with conspecifics, being hazed by their mothers in case of juveniles, as well as their interaction with the surrounding landscape, all these factors shape deer movement and social behavior throughout the year. Furthermore, this might potentially impact CWD spread, but how this can happen exactly is still unknown.

4.2.1 Mortality dispersal kernel

The CWD infected individuals display a range of changes in behavior due to their health deterioration. This includes impaired coordination, walking in circles, and having a stance that is wide based. They can also have slight shakes in their head, seem to constantly seek out water sources, spend long parts of the day sleeping, and may have drooping heads or ears. Eating less due to these symptoms causes the animal to become fatigued and ultimately leads to excessive drinking, excess saliva production, dribbling, and urination at the last stages of the disease.

Hence, the poor body condition of CWD infected deer affects their movement pattern, and hence their long-distance dispersal kernel is expected to be different from that of healthy deer. In this case we recommend another kernel to be introduced: mortality dispersal kernel K_M that accounts for the dispersal pattern of CWD infected individuals.

4.2.2 Natural landscape variability

Long-distance dispersal patterns can be interpreted by understanding the nature of landscape as a driver for dispersal. In regions with low forest coverage, rates and distances of deer dispersal show an increase. Also, rivers and roads usually hinder deer movement (semipermeable barriers), and hence help prevent gene flow and pathogen transmission (Gilbertson et al., 2022). In Wisconsin, the habitat where CWD was first discovered (same as the habitat of juveniles collared for the data) exhibits a rolling topography with approximately 41.0% highly dissected deciduous and 4.2% mixed or evergreen forest patches. The patches are interspersed with agricultural land (approximately 19.9% pasture or hay and 21.8% cultivated crops). Furthermore, grassland and emergent herbaceous or woody wetlands cover few portions of the landscape (Gilbertson et al., 2022). Hence, we recommend that future long-distance dispersal models account for variability of dispersal patterns in geographic regions.

4.3 Potential applications

Prospective studies may benefit from quantifying the pattern of WTD long-distance dispersal in Wisconsin in various ecological and/or epidemiological research areas. For example, the long-distance dispersal kernel could be used to better estimate the spread of other diseases that affect deer populations than CWD such as Epizootic hemorrhagic disease or similar prion diseases like BSE, CJD, and vCJD. Also, it is a step to generalizing deer dispersal patterns in other landscapes than Wisconsin and to other deer species than WTD.

REFERENCES

Centers for Disease Control and Prevention (2022). *Occurrence of Chronic Wasting Disease*. Retrieved September 2. 2022, from https://www.cdc.gov/prions/cwd/occurrence.html#:~:text=Nationwide%2C%20the%20overall%20occurrence%20of,in%204)%20have%20been%20reported.

Centers for Disease Control and Prevention (2021). *Prion Diseases*. Retrieved September 2. 2022, from https://www.cdc.gov/prions/index.html

Chronic Wasting Disease Alliance (2002). *Arrival of CWD in Wisconsin Biggest Natural Resources Story of 2002*. Retrieved October 10. 2022, from Arrival of CWD in Wisconsin Biggest Natural Resources Story of 2002 – CWD-INFO.ORG

Chronic Wasting Disease Alliance (2019). *CWD Overview*. Retrieved October 10. 2022, from CWD Overview – CWD-INFO.ORG

Denkers, N. D., Hoover, C. E., Davenport, K. A., Henderson, D. M., McNulty, E. E., Nalls, A. v., Mathiason, C. K., & Hoover, E. A. (2020). Very low oral exposure to prions of brain or saliva origin can transmit chronic wasting disease. *PLoS ONE*, *15*(8 August). https://doi.org/10.1371/journal.pone.0237410

Edmunds, D. R., Kauffman, M. J., Schumaker, B. A., Lindzey, F. G., Cook, W. E., Kreeger, T. J., Grogan, R. G., & Cornish, T. E. (2016). Chronic wasting disease drives population decline of white-tailed deer. *PLoS ONE*, *11*(8). https://doi.org/10.1371/journal.pone.0161127

Eghiaian, F., Grosclaude, J., Lesceu, S., Debey, P., Doublet, B., Tréguer, E., Rezaei, H., & Knossow, M. (2004). Insight into the PrPC-->PrPSc conversion from the structures of antibody-bound ovine prion scrapie-susceptibility variants. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(28), 10254–10259. https://doi.org/10.1073/pnas.0400014101

Garlick, M. J., Powell, J. A., Hooten, M. B., & McFarlane, L. R. (2011). Homogenization of Large-Scale Movement Models in Ecology. *Bulletin of Mathematical Biology*, *73*(9), 2088–2108. https://doi.org/10.1007/s11538-010-9612-6

Garlick, M. J., Powell, J. A., Hooten, M. B., & MacFarlane, L. R. (2014). Homogenization, sex, and differential motility predict spread of chronic wasting disease in mule deer in southern Utah. *Journal of Mathematical Biology*, *69*(2), 369–399. https://doi.org/10.1007/s00285-013-0709-z

Georgia Department of Natural Resources (2022). *Chronic Wasting Disease (CWD).* Retrieved October 10. 2022, from Chronic Wasting Disease (CWD) | Department Of Natural Resources Division (georgiawildlife.com)

Gilbertson, M., Ketz, A., Hunsaker, M., Jarosinski, D., Ellarson, W., Walsh, D., Storm, D., & Turner W. (2022). *Agricultural land use shapes dispersal in white-tailed deer (Odocoileus virginianus)* [Unpublished manuscript]. Department of Forest and Wildlife Ecology, Wisconsin Cooperative Wildlife Research Unit.

Haley, N. J., Seelig, D. M., Zabel, M. D., Telling, G. C., & Hoover, E. A. (2009). Detection of CWD prions in urine and saliva of deer by transgenic mouse bioassay. *PLoS ONE*, *4*(3). https://doi.org/10.1371/journal.pone.0004848

Hefley, T. J., Hooten, M. B., Russell, R. E., Walsh, D. P., & Powell, J. A. (2017). When mechanism matters: Bayesian forecasting using models of ecological diffusion. In *Ecology Letters* (Vol. 20, Issue 5, pp. 640–650). Blackwell Publishing Ltd. https://doi.org/10.1111/ele.12763

Jacobson, K. H., Lee, S., McKenzie, D., Benson, C. H., & Pedersen, J. A. (2009). Transport of the pathogenic prion protein through landfill materials. *Environmental science & technology*, *43*(6), 2022–2028. https://doi.org/10.1021/es802632d

Jennelle, C. S., Henaux, V., Wasserberg, G., Thiagarajan, B., Rolley, R. E., & Samuel, M. D. (2014). Transmission of chronic wasting disease in Wisconsin white-tailed deer: Implications for disease spread and management. *PLoS ONE*, *9*(3). https://doi.org/10.1371/journal.pone.0091043

Logan, J. (2013). *Applied Mathematics (Fourth Edition)*. Hoboken, New Jersey: John Wiley & Sons, Inc.

Miller, M. W., Williams, E. S., Hobbs, N. T., & Wolfe, L. L. (2004). Environmental sources of prion transmission in mule deer. *Emerging infectious diseases*, *10*(6), 1003–1006. https://doi.org/10.3201/eid1006.040010

Miller, M. W., Hobbs, N. T., & Tavener, S. J. (2006). Dynamics of prion disease transmission in mule deer. *Ecological Applications*, *16*(6), 2208–2214. https://doi.org/10.1890/1051-0761(2006)016[2208:DOPDTI]2.0.CO;2

National Institute of Allergy and Infectious Diseases (2019). *Prion Diseases*. Retrieved September 2, 2022, from https://www.niaid.nih.gov/diseases-conditions/prion-diseases#:~:text=Prion%20diseases%20are%20transmissible%2C%20untreatable,that%20cause%20infectious%20brain%20disease.

Neupane, R. C., & Powell, J. A. (2015). Mathematical Model of Seed Dispersal by Frugivorous Birds and Migration Potential of Pinyon and Juniper in Utah. *Applied Mathematics*, *06*(09), 1506–1523. https://doi.org/10.4236/am.2015.69135

Neupane, R. C., Powell, J. A., & Edwards, T. C. (2021). Connecting regional-scale tree distribution models with seed dispersal kernels. *Applied Mathematics and Computation*, *412*. https://doi.org/10.1016/j.amc.2021.126591

Skuldt, L. H., Mathews, N. E., & Oyer, A. M. (2008). White-Tailed Deer Movements in a Chronic Wasting Disease Area in South-Central Wisconsin. *Journal of Wildlife Management*, *72*(5), 1156–1160. https://doi.org/10.2193/2006-469

Tamgüney, G., Miller, M. W., Wolfe, L. L., Sirochman, T. M., Glidden, D. v., Palmer, C., Lemus, A., Dearmond, S. J., & Prusiner, S. B. (2009). Asymptomatic deer

excrete infectious prions in faeces. *Nature*, *461*(7263), 529–532. https://doi.org/10.1038/nature08289

Tamgüney, G., Richt, J. A., Hamir, A. N., Greenlee, J. J., Miller, M. W., Wolfe, L. L., Sirochman, T. M., Young, A. J., Glidden, D. v., Johnson, N. L., Giles, K., DeArmond, S. J., & Prusiner, S. B. (2012). Salivary prions in sheep and deer. *Prion*, *6*(1), 52–61. https://doi.org/10.4161/pri.6.1.16984

United States Department of Agriculture (2020). *Cervids: Chronic Wasting Disease Specifics*. Retrieved 10 October 2022, from https://www.aphis.usda.gov/aphis/ourfocus/animalhealth/animal-disease-information/cervid/cervids-cwd/cervid-cws-specifics#:~:text=Two%20diagnostic%20te

Williams, E. S., & Miller, M. W. (2002). Chronic wasting disease in deer and elk in North America. *Revue scientifique et technique (International Office of Epizootics)*, 21(2), 305–316. https://doi.org/10.20506/rst.21.2.1340

Wisconsin Department of Natural Resources (2022). *Deer Health – Disease*. Retrieved October 10. 2022, from Deer Statistics (wi.gov)

Wright, C., Howard, A., Lim, S., Lakshman, P. and Loo, C. (2018). PrPc: The Normal Prion. *The FASEB Journal*, 32: 794.8-794.8. https://doi.org/10.1096/fasebj.2018.32.1_supplement.794.8

Virginia Department of Wildlife Resources (2022). *What are Prions?* Retrieved October 10. 2022, from What are Prions? | Virginia DWR

APPENDICES

Appendix A. Kernel Approximations

A.1 Solving the diffusion-settling PDE system

To generate the general formula for a dispersal kernel in the integral form, we solved the PDE system (4-5) as follows:

$$P = e^{-\int_0^t h(s)ds} u(r,t)$$

where $u_t = D\nabla^2 u$, $u(r,0) = \delta(r)$

Solving for the fundamental solution of the diffusion equation, we get:

$u = \frac{1}{4\pi Dt} e^{\frac{-r^2}{4Dt}}$, then substituting in

$$P = \frac{1}{4\pi Dt} e^{-\int_0^t h(s)ds - \frac{r^2}{4Dt}}$$

Then, $S(r,t) = \int_0^t \frac{h(\tau)}{4\pi D\tau} e^{-\int_0^\tau h(s)ds - \frac{r^2}{4D\tau}} d\tau$

Hence, a kernel would be:

$$K(r) = \lim_{t\to\infty} S(r,t) = \int_0^\infty \frac{h(\tau)}{4\pi D\tau} e^{-\int_0^\tau h(s)ds - \frac{r^2}{4D\tau}} d\tau$$

Using (42), the kernels K_2, K_3, and K_4 were derived as follows.

A.2 Approximation of $\boldsymbol{K_2}$

Considering $h_2(t) = \frac{abt}{a+t}$

$$\int_0^t h_2(s)ds = \int_0^t \frac{abs}{a+s} ds = ab \int_0^t \frac{s}{a+s} ds$$

$$= ab \int_0^t \frac{s+a}{s+a} - \frac{a}{s+a} ds = ab \int_0^t 1 - \frac{a}{s+a} ds$$

$$= ab[s - a\ln(s+a)]|_0^t = ab[t - a\ln(s+a) - 0 + a\ln(0+a)]$$

$$= ab[t - a\ln(t+a) + a\ln a] = abt - a^2 b\ln(t+a) + a^2 b\ln a$$

$$= abt - a^2 b \ln\left(\frac{t}{a} + 1\right)$$

For $\beta = a^2 b$,

$$\int_0^t h(s)ds = abt - \beta \ln\left(\frac{t}{a} + 1\right)$$

Substituting in the general formula of kernels, and let $D = \frac{c^2}{ab}$

$$K_2(r) = \int_0^\infty \frac{abt}{(t+a) * 4\pi \frac{c^2}{ab} t} e^{-abt + \beta \ln\left(\frac{t}{a}+1\right) - \frac{r^2}{4\frac{c^2}{ab}t}} dt$$

$$K_2(r) = \int_0^\infty \frac{ab^2 \left(\frac{t}{a} + 1\right)^{\beta - 1}}{4\pi c^2} e^{ab\left(-t - \frac{r^2}{4c^2 t}\right)} dt$$

Applying Laplace's rule to approximate the integral, we get

$$f(t) = \frac{ab^2 \left(\frac{t}{a} + 1\right)^{\beta - 1}}{4\pi c^2}$$

$$\lambda = ab$$

$$g(t) = -t - \frac{r^2}{4c^2 t}$$

To maximize g, $\frac{dg(t)}{dt} = -1 + \frac{r^2}{4c^2 t^2}$

Setting the time derivative of $g(t)$ to zero we get,

$m = \frac{r}{2c}$, where m is the time value at the maximum

Getting the second time derivative, $\frac{d^2 g(t)}{dt^2} = \frac{-r^2}{2c^2 t^3}$

Evaluating the second time derivative at m, $\frac{d^2 g(m)}{dt^2} = \frac{-4c}{r}$

Then, approximating the kernel, $K_2(r) \approx \frac{ab^2\left(\frac{r}{2ac}+1\right)^{\beta-1}}{4\pi c^2} e^{ab\left(\frac{-r}{2c}-\frac{2cr^2}{4c^2r}\right)} \sqrt{\frac{-2\pi}{ab*\frac{-4c}{r}}}$

$$K_2(r) \approx \frac{a^{\frac{1}{2}}b^{\frac{3}{2}}}{4\sqrt{2}\sqrt{\pi}c^{\frac{5}{2}}}\sqrt{r}\left(\frac{r}{2\alpha}+1\right)^{\beta-1} e^{\frac{-\beta}{\alpha}r}, \text{ where } \alpha = ac$$

A constant c_2 is introduced to account for the approximated constant value for $K_2(r)$

Therefore, $K_2(r) \approx c_2\sqrt{r}\left(\frac{r}{2\alpha}+1\right)^{\beta-1} e^{\frac{-\beta}{\alpha}r}$

A.3 Approximation of K_3

Considering $h_3(t) = \frac{abt^2}{t^2+a^2}$

$$\int_0^t h_3(s)\,ds = \int_0^t \frac{abs^2}{s^2+a^2}\,ds = ab\int_0^t \frac{s^2+a^2}{s^2+a^2} - \frac{a^2}{s^2+a^2}\,ds$$

$$= ab\int_0^t 1\,ds - a^3b\int_0^t \frac{1}{s^2+a^2}\,ds = ab[s|_0^t] - a^3b\left[\frac{1}{a}\tan^{-1}\frac{s}{a}|_0^t\right]$$

$$= abt - a^2b\tan^{-1}\frac{t}{a}$$

Substituting in the general formula of kernels,

$$K_3(r) = \int_0^\infty \frac{abt^2}{(t^2+a^2)*4\pi\frac{c^2}{ab}t} e^{-abt+a^2b\tan^{-1}\frac{t}{a}-\frac{r^2}{4\frac{c^2}{ab}t}}\,dt$$

$$K_3(r) = \int_0^\infty \frac{b^2t}{4\pi c^2\left(\frac{t^2}{a^2}+1\right)} e^{ab\left(-t+a\tan^{-1}\frac{t}{a}-\frac{r^2}{4c^2t}\right)}\,dt$$

Applying Laplace's rule to approximate the integral, we get

$$f(t) = \frac{b^2t}{4\pi c^2\left(\frac{t^2}{a^2}+1\right)}$$

$$\lambda = ab$$

$$g(t) = -t + a\tan^{-1}\frac{t}{a} - \frac{r^2}{4c^2t}$$

To maximize g, $\frac{dg(t)}{dt} = -1 + \frac{a^2}{t^2+a^2} + \frac{r^2}{4c^2t^2}$

Setting the time derivative of $g(t)$ to zero we get,

$$\frac{a^2}{t^2 + a^2} + \frac{r^2}{4c^2t^2} = 1$$
$$\frac{4a^2c^2t^2 + a^2r^2 + r^2t^2}{4a^2c^2t^2 + 4c^2t^4} = 1$$
$$4c^2t^4 - r^2t^2 - a^2r^2 = 0$$

Let $y = t^2$, therefore we have a second-degree polynomial as follows

$$4c^2y^2 - r^2y - a^2r^2 = 0$$

Using the quadratic formula, and considering the definite positive root only, we get

$$y = \frac{r^2}{8c^2}\left(1 + \frac{\sqrt{r^2 + 16a^2c^2}}{r}\right)$$

Therefore, $m = \frac{r}{\sqrt{8}c}\sqrt{1 + \frac{\sqrt{r^2+16a^2c^2}}{r}}$

Getting the second time derivative, $\frac{d^2g(t)}{dt^2} = \frac{-2t}{(\frac{t^2}{a^2}+1)^2} - \frac{r^2}{2c^2t^3}$

Evaluating the second time derivative at m, $\frac{d^2g(m)}{dt^2} = \frac{-2m}{(\frac{m^2}{a^2}+1)^2} - \frac{r^2}{2c^2m^3}$

Then, approximating the kernel $K_3(r)$,

$$K_3(r) \approx \frac{b^2m}{4\pi c^2\left(\frac{m^2}{a^2} + 1\right)}\, e^{ab\left(-m + a\tan^{-1}\frac{m}{a} - \frac{r^2}{4c^2m}\right)} \sqrt{\frac{-2\pi}{ab * \left(\frac{-2m}{\left(\frac{m^2}{a^2} + 1\right)^2} - \frac{r^2}{2c^2m^3}\right)}}$$

Simplifying the kernel using $\frac{m^2}{a^2} + 1 = \frac{4c^2m^4}{a^2r^2}$ from the quadratic equation in y

A.4 Approximation of K_4

Considering $h_4(t) = \frac{bt^2}{t+a}$

$$\int_0^t h_4(s)\,ds = \int_0^t \frac{bs^2}{s+a}\,ds$$

Let $y = s + a$, then

$$\int_a^{a+t} by + 2ab + \frac{a^2 b}{y}\,dy = \left(\frac{1}{2}by^2 - 2aby + a^2 b \ln y\right)|_a^{a+t}$$

$$= \frac{1}{2}bt^2 - abt + a^2 b(\ln(t+a) - \ln a)$$

Substituting in the general formula of kernels,

$$K_4(r) = \int_0^\infty \frac{b^2 t}{4\pi c^2}\left(\frac{t}{a}+1\right)^{-(1+\beta)} e^{ab\left(\frac{-1}{2a}t^2 + t - \frac{r^2}{4c^2 t}\right)}\,dt$$

Applying Laplace's rule to approximate the integral, we get

$$f(t) = \frac{b^2 t}{4\pi c^2}\left(\frac{t}{a}+1\right)^{-(1+\beta)}$$

$$\lambda = ab$$

$$g(t) = \frac{-1}{2a}t^2 + t - \frac{r^2}{4c^2 t}$$

To maximize g, $\frac{dg(t)}{dt} = -\frac{t}{a} + 1 + \frac{r^2}{4c^2 t^2}$

Setting the time derivative of $g(t)$ to zero we get,

$$\frac{t-a}{a} = \frac{r^2}{4c^2 t^2}$$

$$ar^2 = 4c^2 t^3 - 4\mathrm{a}c^2 t^2$$

$$t^3 - at^2 - \frac{ar^2}{4c^2} = 0$$

$$1 - \frac{a}{t} - \frac{ar^2}{4c^2 t^3} = 0$$

$$\frac{a^3 r^2}{4\alpha^2 t^3} + \frac{a}{t} - 1 = 0$$

where $\alpha = ac$. Then,

$$\frac{a^3}{t^3} + \left(\frac{4\alpha^2}{r^2}\right)\frac{a}{t} - \left(\frac{4\alpha^2}{r^2}\right) = 0$$

Let $u = \frac{a}{t}$, therefore we have a third-degree polynomial as follows

$$u^3 + \left(\frac{4\alpha^2}{r^2}\right)u - \left(\frac{4\alpha^2}{r^2}\right) = 0$$

Using the formula for simple cubic polynomials, and considering the positive real roots only, we get

$$u = A + B$$

where, $A = \sqrt[3]{\frac{2\alpha^2}{r^2} + \sqrt{\frac{4\alpha^4}{r^4} + \frac{64}{27}\frac{\alpha^6}{r^6}}}$ and $B = \sqrt[3]{\frac{2\alpha^2}{r^2} - \sqrt{\frac{4\alpha^4}{r^4} + \frac{64}{27}\frac{\alpha^6}{r^6}}}$

Hence, $m = \frac{a}{u} = \frac{a}{A+B}$

Getting the second time derivative, $\frac{d^2g(t)}{dt^2} = -\frac{1}{a} - \frac{r^2}{2c^2t^3}$

Evaluating the second time derivative at m, $\frac{d^2g(m)}{dt^2} = -\frac{1}{a} - \frac{r^2}{2c^2m^3}$

Then, approximating the kernel $K_4(r)$,

$$K_4(r) \approx \frac{b^2m}{4\pi c^2}\left(\frac{m}{a} + 1\right)^{-(1+\beta)} e^{ab\left(\frac{-1}{2a}m^2 + m - \frac{r^2}{4c^2m}\right)} \sqrt{\frac{-2\pi}{ab * \left(-\frac{1}{a} - \frac{r^2}{2c^2m^3}\right)}}$$

Simplifying the kernel, we get

$$K_4(r) \approx c_4 m\left(\frac{m}{a} + 1\right)^{-(1+\beta)} (1 + \frac{r^2}{2\alpha^2t^3})^{-\frac{1}{2}} e^{\beta(-\frac{1}{2}m^2 + m - \frac{r^2}{4\alpha^2m})}$$

Appendix B. Long-distance dispersal filtering

A frequency threshold ω was set based on the 11-days interval as follows:

$$\omega = \frac{4\pi}{11}$$

Here the frequency number for a wave 11-week long has an additional factor of two because of Nyquist sampling and to resolve bumps of window duration and another factor of two because we will reflect the Fourier series to minimize oscillations at both ends. Then, the indices of the starting and ending locations of potential home ranges were determined as a home range should not have rapid average movements.

Appendix C. MATLAB Code

C.1 Extracting long-distance dispersals from data

```
% Menna Gouda
load("meta_data.mat")
load("new_data.mat")
%% Searching for Long distances
Long_dists=[];
for n=1:363
    % Calling the fourier function
    x=new_data(IDindx(n,1):IDindx(n,2),1);
    y=new_data(IDindx(n,1):IDindx(n,2),2);
    T=new_data(IDindx(n,1):IDindx(n,2),3);
    dx=new_data(IDindx(n,1):IDindx(n,2),5);
    dy=new_data(IDindx(n,1):IDindx(n,2),6);
    [Isegs, Tsegs]=FT_traj(T,dx,dy,11,300);
    % Removing false home ranges
    [Isegs_new, Tsegs_new]=time_filter(Isegs,Tsegs,14);
    % Getting Long distances
    age=meta_data{n,"ageatcol1"};
    if meta_data{n,"sex"} == "Male"  % 0 for Male
        sex=0;
    elseif meta_data{n,"sex"} == "Female"  % 1 for Female
        sex=1;
    end
    id=new_data(IDindx(n,1),8);
    if meta_data{n,"reasonoff1"}=="mortality"
        status=1;  % alive or collar lost
    else
        status=0; % dead
    end
    if (height(Isegs_new) ~= 1)
    Long_dists=[Long_dists;
get_long_dists(Isegs_new,T,dx,dy,age,sex,id,status,n)]; % 165
```

```
    end
end
%% Filtering (using Average home range distance)
mean_std=819; % calc_std(new_data,IDindx,11,300);  %
818.6989974547785
final_indicies=find(Long_dists(:,1)>=2*mean_std);
final_long=Long_dists(final_indicies,:);  % 113
%% Partitioning data
% males
male_indicies=find(final_long(:,4)==0);
male_dists=final_long(male_indicies,:);       % 91 are males out
of 113
m=male_dists(:,1);
% females
female_indicies=find(final_long(:,4)==1);
female_dists=final_long(female_indicies,:);   % 22 females
f=female_dists(:,1);
% 8 yrs old
age8_indicies=find(final_long(:,5)==8);       % 103 out of 113
age8_dists=final_long(age8_indicies,:);
a8=age8_dists(:,1);
% 20 yrs old
age20_indicies=find(final_long(:,5)==20);     % 10 only
age20_dists=final_long(age20_indicies,:);
a20=age20_dists(:,1);
% spring
spring_indicies=find(final_long(:,6)==1);     % 46 deer
spring_dists=final_long(spring_indicies,:);
s=spring_dists(:,1);
% fall
fall_indicies=find(final_long(:,6)==3);       % 50 deer
fall_dists=final_long(fall_indicies,:);
fl=fall_dists(:,1);
%%
```

C.1.1 Fourier smoothing

```
function [Isegs, Tsegs] = FT_traj(T, dx, dy, days2filter,
speed_threshold)
%FT_TRAJ This is a function which uses Fourier filtering to find
signficant
%   excursions from home range based on GPS collar data (based
on WTD
%   collar data from upper Midwest).
%
%   Jim Powell, Dec. 23, 2021,  jim.powell@usu.edu
%
%   INPUTS (all vectors of the same length):
%       T           vector of location times, in Julian days
%       dx          W->E changes in location at each time,
meters
%       dy          S->N changes in location at each time,
meters
%
%   CONTROL PARAMETERS (scalars)
%       days2filter         time window of filter, days
%                               -- duration of a `week'
%       speed_threshold     average speed threshold for big
move, meters/day
%
%   OUTPUTS
%
%       Isegs       two columns which are starting, ending
indices for the
%                   portion of trajectories which stay below the
filtering
%                   threshold (i.e. bracketing potential home
ranges)
%       Tsegs       time individual spent in potential home
range (useful
%                   for further filtering)
```

```
%

%   Calculate cumulative changes from initial location
cumx=cumsum(dx); cumy=cumsum(dy);
%   Total square displacement from original location
cumd=sqrt(cumx.^2+cumy.^2);

%   Find indices so that averaging doesn't fall out of data
window
NT=length(T);
daze=days2filter/2;
idxwk_min=find(T-T(1)>daze, 1);
idxwk_max=find(T(NT)-T>daze, 1, 'last');
Ts=T(idxwk_min:idxwk_max);
N=idxwk_max-idxwk_min+1;    % length of smoothed trajectory,
subtracting a
                            % half week on each end

%   Sample velocities over a week (initial smoothing)
v_smooth=zeros(N,1);
for n=1:N
    % Find indices for beginning and end of weekly filter window
(times
    % steps not equal!)
    iwk_left=find(T-T(n+idxwk_min-1) > -daze,1);  %find relative
first index in half week interval to left
    iwk_right=min(NT-1,find(T-T(n+idxwk_min-1) > daze,1)); %
find last in half week on right
    deltat=T(iwk_right)-T(iwk_left);           % actual time
interval ~ 1 week
    v_smooth(n)=(cumd(iwk_right)-cumd(iwk_left))/deltat;
end

%   Set up for Fourier smoothing
%   First, figure out number of modes to keep based on filtering
window and
```

```
%   calculate a mask of 0/1s to keep those modes
Nsmooth=int16(ceil(4*(Ts(N)-Ts(1))/days2filter)); % modes to
keep; window duration/length of week
%       is freq. number for a wave one `week' long.  Addtl
factor of two because
%       of Nyquist sampling and we want to resolve bumps of
window duration
%       and another factor of two because we will reflect the
series to minimize
%       oscillations at beginning and end

if (rem(Nsmooth,1)==0 && Nsmooth > 0)
    mask=zeros(2*N,1);
    mask(1:Nsmooth+1)=1;
    mask(2*N-Nsmooth:2*N)=1;    % freqs to keep are in initial
and final positions of FFT vector

    %   Use FFT to generate smoothed data in wave space
    fv=fft([v_smooth; v_smooth(N:-1:1)]);   % reflect data
    fv2=fv.*mask;           % filter out high frequencies
    v_s2=real(ifft(fv2));   % invert the FFT
    v_s=v_s2(1:N);          % smoothed velocities; keep first
half of reflected results

    %   Find indices when smoothed velocites exceed threshold
    idx_normal=find(abs(v_s)<speed_threshold);   % indices
corresponding to home range moves
    idx_jump=find(diff(idx_normal)>1);      % indices of
idx_normal at which there
    %       is a jump between ranges, i.e. indices when rapid
average movement
    %       begins

    %   Set up output variables
    Nsegs=length(idx_jump)+1;   % number of distinct home range
segments
```

```
    Tsegs=zeros(Nsegs,1);       % to contain the temporal
duration of segments
    Isegs=zeros(Nsegs,2);       % start and end of relevant
indices for each segment
    idx_hr=Isegs;               % working variable to keep track
of home range indices
                                % relative to original inputs

    if (Nsegs==1)       % no jumps, individual stays at home
        idx_hr(1,1)=0; idx_hr(1,2)=idxwk_max-1;
    elseif (Nsegs==2)   % one jump, before/after ranges
        idx_hr(1,1)=0;
idx_hr(1,2)=idxwk_min+idx_normal(idx_jump(1));
        idx_hr(2,1)=idxwk_min+idx_normal(idx_jump(1)+1);
idx_hr(2,2)=idxwk_max-1;
    else                % multiple jumps among ranges
        idx_hr(1,1)=0;
idx_hr(1,2)=idxwk_min+idx_normal(idx_jump(1));
        idx_hr(Nsegs,1)=idxwk_min+idx_normal(idx_jump(Nsegs-
1)+1);
        idx_hr(Nsegs,2)=idxwk_max-1;
        for j=2:Nsegs-1
            idx_hr(j,1)=idxwk_min+idx_normal(idx_jump(j-1)+1);
            idx_hr(j,2)=idxwk_min+idx_normal(idx_jump(j));
        end
    end
    Isegs=idx_hr+1;    % starting index is 1, not zero

    % calculate the duration of residence in potential range
segments
    for j=1:Nsegs
        Tsegs(j)=T(Isegs(j,2))-T(Isegs(j,1));
    end
else
    disp("no");
    Isegs=[];
```

```
    Tsegs=[];
end

end
```

C.1.2 Time filtering

```
function [Isegs Tsegs]=time_filter(Isegs,Tsegs,time_threshold)

[p,q]=size(Isegs);
flag=[];
for i=1:p
   if Tsegs(i)<time_threshold
       flag=[flag 1];
   else
       flag=[flag 0];
   end
end
 to_del=find(flag==1);
 Isegs(to_del,:)=[];
 Tsegs(to_del,:)=[];
end
```

C.1.3 Calculating average standard deviation of home ranges

```
function
mean_std=calc_std(new_data,IDindx,days2filter,speed_threshold)

% This function calculates the average std of home range
std_d=[];
for n=1:363
    x=new_data(IDindx(n,1):IDindx(n,2),1);
    y=new_data(IDindx(n,1):IDindx(n,2),2);
    T=new_data(IDindx(n,1):IDindx(n,2),3);
    dx=new_data(IDindx(n,1):IDindx(n,2),5);
    dy=new_data(IDindx(n,1):IDindx(n,2),6);
```

```
    % Calling the fourier function
    [Isegs, Tsegs]=FT_traj(T,dx,dy,days2filter,speed_threshold);
    % Find average stdiance across home range
    [p,q]=size(Isegs);
    for i=1:p
        if Tsegs(i)>14  % filtering non home ranges
            new_x=x(Isegs(i,1):Isegs(i,2));
            new_y=y(Isegs(i,1):Isegs(i,2));
            mean_x=mean(new_x);
            mean_y=mean(new_y);
            new_std=sqrt(sum((new_x-mean_x).^2+(new_y-
mean_y).^2)./(length(new_x)-1)); % calculating std in 2D
            std_d=[std_d; new_std];
        end
    end
mean_std=mean(std_d,'omitnan'); % averaging all std
end
```

C.2 Fitting the four models to the GPS data

```
%% Fitting
r=final_long(:,1);
alpha=mean(r); N=length(r); d=1; % d is beta - 1

format short g;
% nll modified bessel function of the second kind
[parms1, fval1, iflag1]= fminsearch(@(x) nll_bessel(r, x),
[alpha])
% nll type II
[parms2, fval2, iflag2]= fminsearch(@(x) nll_typeII(r, x),
[alpha d])
% nll type III
[parms3, fval3, iflag3]= fminsearch(@(x) nll_typeIII(r, x),
[alpha d])
% nll model 4
[parms4, fval4, iflag4]= fminsearch(@(x) nll_model4(r, x),
[alpha d])
```

```
% nll model 4 new
[parms5, fval5, iflag5]= fminsearch(@(x) nll_model4_new(r, x),
[alpha d])
```

C.2.1 Model I

```
function [p,c1]=bessel(x)
alpha=x(1);
rc=linspace(1,71001,71001);
drc=rc(2)-rc(1);
K=besselk(0,rc/alpha);
p=2*pi.*K.*(rc./(alpha.^2));
c1=1/trapz(drc*p);
end
```

C.2.2 NLL calculation for model I

```
function err=nll_bessel(r,x)
alpha=x(1);
[p,c1]=bessel(x);
K=besselk(0,r/alpha);
err=length(r)*(2*log(alpha)-log(c1))+sum(-log(K));
end
```

C.2.3 Model II

```
function [p,c1]=typeII(x)
alpha=x(1);
d=x(2);
rc=linspace(1,71001,71001);
drc=rc(2)-rc(1);
p=2*pi.*sqrt(rc./alpha).*((1+0.5*rc./alpha).^d).*exp(-
(d+1).*rc./alpha).*(rc./(alpha.^2));
c1=1/trapz(drc*p);
end
```

C.2.4 NLL calculation for model II

```
function err=nll_typeII(r,x)
alpha=x(1);
d=x(2)
[p,c1]=typeII(x);
```

```
err=length(r)*(2*log(alpha)-log(c1)) +sum((d+1).*(r/alpha)-
0.5.*log(r/alpha)-d.*log(1+0.5.*(r/alpha)));
end
```

C.2.5 Model III

```
function [p,c1]=typeIII(x)
alpha=x(1);
d=x(2);
rc=linspace(1,71001,71001);
drc=rc(2)-rc(1);
t=(rc./(sqrt(8).*alpha)).*sqrt(1+(1+16*(alpha./rc).^2));
p=2*pi.*sqrt(t).*(((t.^2)+2).^-0.5).*exp((d+1).*(-t+atan(t)-
((1./(4.*t)).*((rc./alpha).^2)))).*(rc./(alpha.^2));
c1=1/trapz(drc*p);
end
```

C.2.6 NLL calculation for model III

```
function err=nll_typeIII(r,x)
alpha=x(1);
d=x(2);
[p,c1]=typeIII(x);
t=(r./(sqrt(8).*alpha)).*sqrt(1+(1+16*(alpha./r).^2));
err=length(r)*(2*log(alpha)-log(c1))+...
    sum(-0.5*log(t)+0.5*log((t.^2)+2)+(d+1).*(t-
atan(t)+((1./(4.*t)).*((r./alpha).^2))));
end
```

C.2.7 Model IV

```
function [p,c1]=model4(x)
alpha=x(1);
d=x(2);
rc=linspace(1,71001,71001);
drc=rc(2)-rc(1);
m=alpha./rc;
A=(m.^(2/3)).*nthroot(2+2*sqrt(1+((8/27).*m.^2)),3);
B=(m.^(2/3)).*nthroot(2-2*sqrt(1+((8/27).*m.^2)),3);
t=1./(A+B);
```

```
p=2*pi.*t.*((t+1).^(-d-
2)).*((1+(((rc./alpha).^2)./(2.*(t.^3)))).^-0.5).*exp((((-d-
1)/2).*t.^2)+((d+1).*t)-
(((rc./alpha).^2).*(d+1)./(4.*t))).*(rc./(alpha.^2));
c1=1/trapz(drc*p);
end
```

C.2.8 NLL calculation for model IV

```
function err=nll_model4(r,x)
alpha=x(1);
d=x(2);
rc=linspace(1,71001,71001);
drc=rc(2)-rc(1);
A=(m.^(2/3)).*nthroot(2+2*sqrt(1+((8/27).*m.^2)),3);
B=(m.^(2/3)).*nthroot(2-2*sqrt(1+((8/27).*m.^2)),3);
t=1./(A+B);
p=2*pi.*t.*((t+1).^(-d-
2)).*((1+(((R./alpha).^2)./(2.*(t.^3)))).^-0.5).*exp((((-d-
1)/2).*t.^2)+((d+1).*t)-
(((R./alpha).^2).*(d+1)./(4.*t))).*(rc./(alpha.^2));
c1=1/trapz(drc*p);
err=length(r)*(2*log(alpha)-log(c1))+sum(-
log(t)+(d+2).*log(t+1)+0.5.*log(1+((r./alpha).^2)./(2.*(t.^3)))+
(d+1).*(0.5.*(t.^2)-t+(1./(4.*t)).*((r/alpha).^2)));
end
```

C.3 Visualization of the four models versus GPS data

```
%% Visualization
r=final_long(:,1); N=length(r);
[counts,centers]=hist(r,30);
delta_r=(centers(2)-centers(1));  % bin width
hist(r/1000,30)
format short g;

rc=linspace(1,71001,71001);
drc=rc(2)-rc(1);
```

```
% bessel NLL
alpha=parms1(1);
[p1,c1]=bessel(alpha);
hold on, plot(rc/1000,N*delta_r*p1*c1,"r"),

% grad descent type II NLL
alpha=parms2(1);d=parms2(2);
[p2,c2]=typeII(parms2);
hold on, plot(rc/1000,N*delta_r*c2*p2,"m"),

% grad descent type III NLL
alpha=parms3(1);d=parms3(2);
[p3,c3]=typeIII(parms3);
hold on, plot(rc/1000,N*delta_r*c3*p3,"g"),

% grad descent model 4 NLL
alpha=parms4(1);d=parms4(2);
[p4,c4]=model4(parms4);
hold on, plot(rc/1000,N*delta_r*c4*p4,"c"), hold off

xlabel('Distance (km)'), ylabel('Frequency')
legend("Raw data", "K_1","K_2", "K_3", "K_4")
```

C.4 BIC Calculations

```
%% BIC Calculations
N=length(r);
format short g;
% bessel
BIC1= 2*fval1 + 1*log(.5*N/pi)   % 4807.8
% grad type II
BIC2=  2*fval2 + 2*log(.5*N/pi)  % 4805.4
% grad type III
BIC3=  2*fval3 + 2*log(.5*N/pi)  % 4803.9
% grad model 4
BIC4=  2*fval4 + 2*log(.5*N/pi)  % 4828.1
% grad model 4 new
```

```
BIC5=  2*fval5 + 2*log(.5*N/pi)  % 4825
% Comparisons
BIC13=exp(0.5*(BIC1-BIC3)); % 6.9748
BIC23=exp(0.5*(BIC2-BIC3)); % 2.0954
BIC43=exp(0.5*(BIC4-BIC3)); % 1.8349e+05
BIC53=exp(0.5*(BIC5-BIC3)); % 38525
```

C.5 Bootstrapping & fitting bootstrapped data

```
%% Bootstrapping while accounting for differences in sex
all_BICs_j=[];all_Parms_j=[];all_Flags_j=[];  % joint sample
all_BICs_m=[];all_Parms_m=[];all_Flags_m=[];  % male sample
all_BICs_f=[];all_Parms_f=[];all_Flags_f=[];  % female sample

while (sum(all_Flags_j)~=1000) | (sum(all_Flags_m)~=1000) |
(sum(all_Flags_f)~=1000)
    % disp("hello");

    rand_idx=ceil(113*rand(113,1));
    final_dists=final_long(rand_idx,:);

    % males
    male_indicies=find(final_dists(:,4)==0);
    male_dists=final_dists(male_indicies,:);
    m_btstrp=male_dists(:,1);
    % females
    female_indicies=find(final_dists(:,4)==1);
    female_dists=final_dists(female_indicies,:);
    f_btstrp=female_dists(:,1);

    [BICs_j,Parms_j,Flags_j]=fit(final_dists(:,1));
    [BICs_m,Parms_m,Flags_m]=fit(m_btstrp);
    [BICs_f,Parms_f,Flags_f]=fit(f_btstrp);

    % ignoring samples if their fit doesn't converge or BIC goes
complex
```

```
    if (Flags_j==1) & (Flags_m==1) & (Flags_f==1) &
(isreal(BICs_j)) & (isreal(BICs_m)) & (isreal(BICs_f))
       % joint sample
       all_BICs_j=[all_BICs_j;BICs_j];
       all_Parms_j=[all_Parms_j;Parms_j];
       all_Flags_j=[all_Flags_j;Flags_j];
       % male sample
       all_BICs_m=[all_BICs_m;BICs_m];
       all_Parms_m=[all_Parms_m;Parms_m];
       all_Flags_m=[all_Flags_m;Flags_m];
       % female sample
       all_BICs_f=[all_BICs_f;BICs_f];
       all_Parms_f=[all_Parms_f;Parms_f];
       all_Flags_f=[all_Flags_f;Flags_f];
    end
end
```

C.6 Visualization of model parameters and BICs using bootstrapped data

```
%% Boostrap histograms again (for males and females version)
nominal_BICs_j=[BIC1,BIC2,BIC3,BIC4];
nominal_Parms_j=[parms1,parms2,parms3,parms4];
nominal_BICs_m=[BIC1_m,BIC2_m,BIC3_m,BIC4_m];
nominal_Parms_m=[parms1_m,parms2_m,parms3_m,parms4_m];
nominal_BICs_f=[BIC1_f,BIC2_f,BIC3_f,BIC4_f];
nominal_Parms_f=[parms1_f,parms2_f,parms3_f,parms4_f];
CIs_BICs_j=CIs(1:4,:); CIs_BICs_m=CIs(5:8,:);
CIs_BICs_f=CIs(9:12,:);
CIs_Parms_j=CIs(13:19,:); CIs_Parms_m=CIs(20:26,:);
CIs_Parms_f=CIs(27:33,:);
CIs_delta_BICs_j=CIs(34:37,:); CIs_delta_BICs_m=CIs(38:41,:);
CIs_delta_BICs_f=CIs(42:45,:);
n=1;
for i=1:4
```

```
    create_bootstrap_hists(all_BICs_j(:,i),nominal_BICs_j(i),i,"
BIC for joint sample using model
",CIs_BICs_j(i,1),CIs_BICs_j(i,2));
    create_bootstrap_hists(all_BICs_m(:,i),nominal_BICs_m(i),i,"
BIC for male sample using model
",CIs_BICs_m(i,1),CIs_BICs_m(i,2));
    create_bootstrap_hists(all_BICs_f(:,i),nominal_BICs_f(i),i,"
BIC for female sample using model
",CIs_BICs_f(i,1),CIs_BICs_f(i,2));

create_bootstrap_hists(all_Parms_j(:,n),nominal_Parms_j(n),i,"
Parameter $\alpha$ for joint sample model using model
",CIs_Parms_j(n,1),CIs_Parms_j(n,2));

create_bootstrap_hists(all_Parms_m(:,n),nominal_Parms_m(n),i,"
Parameter $\alpha$ for male sample model using model
",CIs_Parms_m(n,1),CIs_Parms_m(n,2));

create_bootstrap_hists(all_Parms_f(:,n),nominal_Parms_f(n),i,"
Parameter $\alpha$ for female sample model using model
",CIs_Parms_f(n,1),CIs_Parms_f(n,2));
    create_bootstrap_hists(all_BICs_j(:,i)-
all_BICs_j(:,3),nominal_BICs_j(i)-nominal_BICs_j(3),i," $\Delta$
BIC for joint sample for model
",CIs_delta_BICs_j(i,1),CIs_delta_BICs_j(i,2));
    create_bootstrap_hists(all_BICs_m(:,i)-
all_BICs_m(:,2),nominal_BICs_m(i)-nominal_BICs_m(2),i," $\Delta$
BIC for male sample for model
",CIs_delta_BICs_m(i,1),CIs_delta_BICs_m(i,2));
    create_bootstrap_hists(all_BICs_f(:,i)-
all_BICs_f(:,1),nominal_BICs_f(i)-nominal_BICs_f(1),i," $\Delta$
BIC for female sample for model
",CIs_delta_BICs_f(i,1),CIs_delta_BICs_f(i,2));
    n=n+1;
    if i > 1
```

```
create_bootstrap_hists(all_Parms_j(:,n)+1,nominal_Parms_j(n)+1,i
," Parameter $\beta$ for joint sample model
",CIs_Parms_j(n,1),CIs_Parms_j(n,2));

create_bootstrap_hists(all_Parms_m(:,n)+1,nominal_Parms_m(n)+1,i
," Parameter $\beta$ for male sample model
",CIs_Parms_m(n,1),CIs_Parms_m(n,2));

create_bootstrap_hists(all_Parms_f(:,n)+1,nominal_Parms_f(n)+1,i
," Parameter $\beta$ for female sample model
",CIs_Parms_f(n,1),CIs_Parms_f(n,2));
       n=n+1;
    end
end
```

C.7 Calculating the best model for males, females, and the joint sample

```
%% Deciding which is the best model for each group: joint,male,
and female
best_model=[0 0 0];
for j=1:3
    if j==1
        myBIC=all_BICs_j;
    elseif j==2
        myBIC=all_BICs_m;
    else
        myBIC=all_BICs_f;
    end
    min_sum=inf;
    for i=1:4
        if mean(myBIC(:,i))<min_sum  % should I use sum?! or
mode? or mean?
           min_sum=mean(myBIC(:,i));
           best_model(j)=i;   % [3 2 1] using sum
        end                   % [3 2 3] using mode
    end                       % [3 2 1] using mean
```

```
end
```

C.8 Testing whether one model for joint data is better or one for males and one for females

```
%% Comparing BICs male + female vs joint
% map = brewermap(2,'Set1');
combined_BICs=all_BICs_m(:,2)+all_BICs_f(:,1);
figure,
h1=histogram(all_BICs_j(:,3),'facecolor',[0 0
1],'edgecolor','none'), hold on
h2=histogram(combined_BICs,'facecolor',[1 0
0],'edgecolor','none'), hold off
title('Distribution of BIC values for male-female and joint
samples','interpeter','latex')
xl=xline(BIC3, '--k',strcat("Nominal value:
",num2str(round(BIC3))));
xll=xline(BIC2_m+BIC1_f, '--k',strcat("Nominal value:
",num2str(round(BIC2_m+BIC1_f))));
xl.LabelVerticalAlignment = 'bottom';
xl.LabelHorizontalAlignment = 'right';
xl.Color=[0 0 1];
xl.LineWidth=1.3;
xll.Color=     [1 1 0];
xll.LineWidth=1.3;
xll.LabelVerticalAlignment = 'bottom';
xll.LabelHorizontalAlignment = 'left';
xlabel('BIC values','interpeter','latex');
ylabel('frequency','interpeter','latex')
legend("Joint","Male-Female","joint", "Male-Female")
%% getting delta BICs for male+female vs joint
combined_BICs=all_BICs_m(:,2)+all_BICs_f(:,1);
figure,
h1=histogram(all_BICs_j(:,3)-combined_BICs,'facecolor',[0 0
1],'edgecolor','none')
title('Distribution of \Delta BIC for male-female vs. joint
sample')
```

```
xl=xline(BIC3-BIC2_m-BIC1_f, '--k',strcat("nominal value:
",num2str(BIC3-BIC2_m-BIC1_f)));
% xll=xline(BIC2_m+BIC1_f, '--k',strcat("male-female sample
nominal value: ",num2str(round(BIC2_m+BIC1_f))));
xl.LabelVerticalAlignment = 'bottom';
xl.LabelHorizontalAlignment = 'right';
xl.LineWidth=1.3;
% xl.Color=[0 1 1];
xlabel('\Delta BIC values'); ylabel('frequency')
```