Mathematical models of short-range dispersal in defoliating

lepidopterans

By

ELI BUSSEL

submitted as a THESIS toward the degree of BACHELOR OF SCIENCE

in

BIOLOGICAL SCIENCES

in

THE COLLEGE

of

THE UNIVERSITY OF CHICAGO

Approved:

Signature of thesis adviser/date:

Name of thesis adviser: GREG DWYER

6/1/2021g

Signature of faculty reader #1/date:

5/29/2021

Name of faculty reader #1: ALISON HUNTER

Signature of faculty reader #2/date:

5/30/2021

Name of faculty reader #2: DMITRY KONDRASHOV

Abstract

The gypsy moth (*Lymantria dispar*) is among the most destructive invasive species in North America, responsible for defoliating millions of hectares of oak forest. The spatial dynamics of defoliating lepidopteran populations, such as those of the gypsy moth, are thus of great interest to forestry and conservation efforts. Despite numerous studies on the long-range dispersal patterns of defoliators, there is comparatively little theoretical understanding or field research concerning short-range dispersal via ballooning. Previous studies of ballooning have assumed random diffusion, but these models cannot account for non-random biases, such as the effect of wind on the angle of dispersal.

Here, I develop models of short-range dispersal in larvae via ballooning, informed by methods from the seed dispersal kernel literature. I then fit models to field data of gypsy moth larvae dispersal using MCMC to perform Bayesian inference, and PSIS-LOO to perform model selection. I found that dispersal kernel models are able to reliably detect biases in angle of dispersal due to wind direction, and allow for testing of correlations between experimental variables and measures of dispersal. These modeling methods can help inform future studies into short-range larval dispersal and provide a novel framework with which to analyze dispersal data.

Introduction

Outbreaks of defoliating insects in temperate forests diminish the photosynthetic capacity of trees and increase the loss of nutrients such as nitrogen and carbon, causing interrupted growth and death. The gypsy moth (*Lymantria dispar*), an invasive species introduced in Massachusetts in 1869, is the most prevalent and destructive forest defoliator in North America. Since its introduction, the gypsy moth has spread across the Northeastern United States and into the Midwest, and has been responsible for defoliating more than 30 million hectares of oak forest since 1970 (de Beurs and Townsend 2008). In 2017 alone, a gypsy moth outbreak in Southern New England defoliated over 4000 km² of forest (Pasquarella et al. 2018). It is expected that the gypsy moth will continue to spread across the United States, eventually affecting the entire range of oak forests in North America (McManus and Csóka 2007). The gypsy moth thus poses a serious threat to the health of temperate forest ecosystems.

The mechanisms of dispersal are crucial to understanding the dynamics of invasion and outbreaks. The dispersal of egg masses over long distances, primarily caused by human activity such as logging, is a major mechanism of gypsy moth invasion (Biggsby et al. 2011). But gypsy moth larvae also disperse over shorter distances. Short-range dispersal occurs during the early instars of larval development; late-instar larvae are too heavy to balloon, and adults do not disperse, as females are flightless. Early-instar larvae can disperse either through crawling or gliding with the use of silken threads. The latter behavior, known as “ballooning,” allows larvae to travel tens of meters through the air to reach new trees, allowing populations to spread widely through forests. Larval crawling is much slower than ballooning as to be negligible when measuring dispersal within a season. Early-instar ballooning is therefore the primary mechanism of dispersal in gypsy moth (Sharov and Liebhold 1998).

Patterns of short-range dispersal can have a large impact on lepidopteran populations due to the spatial effects of parasitoid and infectious disease outbreaks. Dwyer 1992 showed that local population density greatly influenced the spatial structure of nucleopolyhedrovirus (NPV) outbreaks in Douglas-fir tussock moth (*Orygia pseudotsugata*), whose dispersal via ballooning is

similar to that of the gypsy moth. Dwyer's model assumed larvae dispersed via random diffusion, which is a suitable model for crawling, but may not accurately describe directional biases associated with ballooning. Thus, a better theoretical understanding of dispersal via ballooning could help to more accurately model the spatial spread of NPV in Douglas-fir tussock moth.

There are relatively few studies of short-range dispersal of defoliating lepidopterans within a single generation. McManus 1973 showed that environmental factors such as wind and ambient temperature affected rates of ballooning, but did not measure the spatial distribution of dispersing larvae. Mitchell 1979 conducted a study in Douglas-fir tussock moth which found that first-instar ballooning accounted for the majority of dispersal, and established a relationship between silk thread length and distance traveled in a laboratory setting. Although Mitchell gathered field data on the spatial distribution of ballooning larvae, he did not propose a theoretical model for field dispersal, and did not determine correlation between dispersal in the field and any biological or environmental variables. Some studies have investigated ballooning behavior in gypsy moth, such as McCormick et al. 2019, but studies in this field commonly measure only the percentage of ballooning larvae without considering the ultimate spatial distribution of ballooning insects. Overall, the literature on defoliators lacks studies on short-range seasonal dispersal, and has little theoretical basis to inform study design.

When it comes to building mathematical models and designing experiments to study lepidopteran dispersal via ballooning, the extensive literature on tree seed dispersal is highly informative. Mathematical models have been a common component of seed dispersal studies since Janzen 1970, which proposed an influential model describing the distribution of seedlings in tropical forests as a function of distance from the parent tree. A typical model of seed dispersal assigns a distribution function over space, called a *dispersal kernel*, that describes the density of seeds landing at a particular point in space. A study may then estimate kernel parameters in field experiments by placing traps near a shedding tree and measuring the density of seeds that fall into traps as a function of space. (Klein et al. 2013).

As the wind-borne dispersal of lepidopteran larvae via ballooning is mechanistically similar to the dispersal of falling seeds, dispersal kernels may provide a useful framework for analyzing

datasets of dispersal via ballooning, such as the data collected for Hunter and Elkinton 2000. Hunter and Elkinton's primary interest was in the synchrony of larval hatching and black oak budburst, and they considered several aspects of population dynamics in their study, including dispersal. To measure dispersal, they released larvae reared in laboratory at a field site, and sampled along transects from the release point in all directions to obtain larval counts from nearby trees. Their analysis compared mean displacement between treatment groups which were released at different times relative to budburst, assuming a bivariate normal distribution, and found that dispersal increased in groups released later. Though Hunter and Elkinton did not develop a mathematical model themselves, their dispersal data is well suited to a modeling approach using dispersal kernels.

Here, I seek to develop mathematical models of larval dispersal via ballooning, informed by the literature on seed dispersal kernels, which are capable of describing the effects of environmental and biological parameters on dispersal. I then reanalyze the dispersal data from Hunter and Elkinton 2000 to perform model fits and select between models based on a variety of hypotheses on dispersal, to determine which factors are most relevant to modeling larval dispersal.

Methods

Study design

The field data I use here originated in Hunter and Elkinton 2000, from gypsy moth dispersal experiments in 1995 and 1996 on Otis Air Force Base in Cape Cod, Massachusetts. Release protocols differed between the two years due to availability of egg sources and new protocols to reduce variance in hatching time in 1996. In 1995 the eggs were sourced from laboratory stock at Otis Air Force Base, while in 1996 eggs were collected in West Virginia. The authors did not believe the two populations had significant phenological differences.

Eggs were reared in the laboratory until they began to hatch. In 1995, the unhatched eggs were released into the field site, but the authors noted high variance in hatch timing, so in 1996 larvae were released after $\geq 70\%$ had emerged. Since this difference in methodology only

affected release timing, and not dispersal, the dispersal results from the two years are still comparable.

Insects were released at three different times in each year, with four replicate plots per release time. The three release times in each year were chosen relative to the black oak budburst in each year, with one release before budburst, one a few days after budburst, and the third several weeks after budburst. Exact dates for each release are given in Table 1.

Release plots were ≥ 300 m apart to eliminate crossing between plots. In 1995, each release plot was a 20×20 m square and eggs were released on oak boles, and in 1996 larvae were released on the closest 20 trees within a 7 m radius of a central tree. All plots received 5×10^5 eggs/larvae distributed uniformly across the plot. However, due to differences in survival, each release date had a slightly different number of larvae hatching. Table 1 gives estimates of the number of larvae hatching in each release group in both years.

Table 1: Release dates and number hatched for release groups

Group	1995 date	1996 date	1995 # hatched	1996 # hatched
Early	2 May	28 April	5.0×10^5	4.1×10^5
On-time	10 May	12 May	5.7×10^5	5.1×10^5
Late	25 May	28 May	4.6×10^5	3.8×10^5

Release dates and estimated number of larvae hatching for the three release groups. Black oak budburst began on 4 May in 1995 and on 5 May in 1996. Adapted from Hunter and Elkinton 2000.

Larval density data was collected after ballooning, but before the second instar, by sampling along transects radiating out from the release area. Trees were sampled in 5-m intervals in eight compass directions, with additional NNE and ENE transects in 1996. Collection continued along transects in each direction until no larvae were found. Larvae and leaf counts were obtained by dropping foliage onto tarpaulins and counting all larvae and leaves that fell. Black oaks were sampled at each point along the transect, and a white oak was sampled if one was present within 1.5 m of the sampling point.

Dispersal kernel models

A typical model of seed dispersal in two dimensions assigns a distribution function over space, called a kernel, that describes the density of seeds landing at a particular point in space (Klein et al. 2013). Here, I adapt the dispersal kernel framework to model ballooning larvae.

Throughout, I consider larval dispersal in a two-dimensional polar coordinate space, with the source of dispersal at the origin, and let $k(r, \theta)$ be the kernel describing the density of ballooning larvae at coordinates (r, θ) . Assuming that distance traveled is independent of the angle of dispersal, the kernel may be expressed as $k(r, \theta) = k(r)\ell(\theta)$, the product of a one-dimensional kernel with a distribution on the angle. I will refer to $k(r)$ and $\ell(\theta)$ as the *radial* and *angular* distributions, respectively.

Unlike a seed which almost inevitably falls and disperses, a larva may never disperse, instead remaining on the tree where it hatched. Thus, we need to adjust the kernel to account for the probability of dispersal. If $q \in [0, 1]$ is the fraction of larvae dispersing, then for $r > 0$, the adjusted kernel is $qk(r, \theta)$. Here, we assume that all larvae at the origin have not dispersed.

Ideally, we would estimate $k(r, \theta)$ directly by taking measurements of larval density. But sampling larvae from trees as in Hunter and Elkinton 2000 necessarily gives an indirect measure of density, as it is not feasible to accurately sample the true density of insects on a given tree. It is therefore necessary to account for sampling in the field when modeling larval dispersal. Consider T trees at approximate coordinates (r_i, θ_i) for $i = 1, 2, \dots, T$. Then a larva lands on tree i with approximate probability $k(r_i, \theta_i)$. If N larvae are released from a point source with probability of dispersal q , and K_i is the number of larvae on tree i , then $K = (K_1, \dots, K_T)$ follows a multinomial distribution, with expectations given by $\mathbb{E}[K_i] = Nqk(r_i, \theta_i)$. But we would only expect to sample this many larvae if we were sure to count every insect on a tree. In practice, we can only sample a fraction of the branches on a given tree, so larger trees will be more sparsely sampled than smaller ones if we collect the same number of branches from each tree. We must therefore adjust our sampling expectation by the size of each branch. Leaf count is a convenient measure of branch size which Hunter and Elkinton recorded in their data. If sampling a branch

yields L_i leaves, then the expected number of larvae sampled should be $\mathbb{E}[K_i] = Nqk(r_i, \theta_i)/L_i$.

Model formulations

I formulated a total of 11 dispersal kernel models to fit the Hunter and Elkinton 2000 dataset. When building these models, I focused on three factors for which Hunter and Elkinton 2000 collected data: the timing of release, wind direction, and tree species. the Table 2 provides a summary of the different models I develop here.

The first major modeling decision is the choice of the radial distribution $k(r)$. A simple yet common choice in seed dispersal literature is to assign an exponential distribution to a one-dimensional kernel, giving $k(r) = \lambda e^{-\lambda r}$ for some rate parameter λ (Klein et al. 2013). But following Dwyer 1992, we might wish to model dispersal as random diffusion, i.e. as a Brownian motion. The distribution of position in a Brownian motion is a normal whose mean is the origin and whose variance is proportional to the amount of time passed. Thus, under random diffusion, the distance of a larva from the source point should follow a half-normal distribution:

$$k(r) = \frac{\sqrt{2}}{\sigma\sqrt{\pi}} \exp\left(-\frac{r^2}{2\sigma^2}\right),$$

where $\sigma > 0$ is proportional to the square root of the time since release. To test the predictions of a random diffusion model against more conventional dispersal kernels, Model 1, the “diffusion” model, assumes a half-normal radial kernel. All other models I develop here use an exponential radial kernel. Figure 1 give plots of the exponential and half-normal distributions for comparison.

Next, we must decide on the angular distribution. The simplest model is a uniform distribution: $\ell(\theta) = \frac{1}{2\pi}$ for $\theta \in (-\pi, \pi]$. I assume a uniform distribution in Models 1 and 2. However, strong wind conditions are likely to produce biases in the angular distribution. To account for these circumstances, I follow Savage et al. 2011 and use a von Mises distribution:

$$\ell(\theta) = \frac{\exp[\kappa \cos(\theta - \bar{\theta})]}{2\pi I_0(\kappa)} \quad \theta \in (-\pi, \pi],$$

where $\bar{\theta}$ is the mean angle and κ is the “concentration” of the distribution, equal to the inverse

Figure 1: Exponential and half-normal radial kernels

Example plots of the exponential (top) and half-normal (bottom) distribution functions for a range of parameter values. In both plots, the black, blue, and red distributions have means of 10, 20, and 50, respectively.

of the variance. Here I_0 is the order-zero modified Bessel function of the first kind. The von Mises distribution is a well-behaved analogue of the normal distribution on the circle, with the concentration κ acting similarly to the standard deviation parameter σ of the normal distribution. When $\kappa = 0$, the von Mises is a uniform distribution, so the possibility of a uniform angle distribution model can be accounted for by a von Mises model. Figure 2 shows some example plots of the von Mises distribution.

To compare with a uniform angle model, Model 2 uses the exponential radial kernel with the uniform distribution, while Model 3 uses an exponential radial kernel with a von Mises angular kernel. I call Model 2 the “raw kernel” model, because further models I develop are based on modifications to it. Model 3 is called the “raw von Mises” model for the same reason. Each of the further models I develop are part of a pair, with one using a uniform angular kernel and the other using a von Mises angular kernel.

To test the effects of synchrony on the probability of dispersal q and the exponential rate λ , I assigned separate parameters to each of the release groups in some models, forming a set of

Figure 2: von Mises distribution

Example plots of the von Mises distribution with mean $\bar{\theta} = 0$ for a range of concentrations κ .

“release-dependent” models to contrast against “release-independent” models. In Models 4 and 5, the “three- q ” models, I fit one q parameter for each of the three release groups, and in Models 6 and 7, the “three- λ models, I fit one λ parameter for each release group. I label these parameters with indices 1, 2, 3 to correspond to the early, on-time, and late release groups, respectively.

The last four models are designed to model larval preference for tree species. The gypsy moth prefers to feed on black oak over white oak leaves, and Hunter and Elkinton 2000 found that larval densities were higher on black oaks than on neighboring white oaks in the first two release groups. However, in the last release group, the quality of black oak leaves had degraded, but the later budburst of the white oak meant that white oak leaves were still high quality, and so larvae preferred the younger white oak leaves instead. To examine the effects of tree species on dispersal, Models 8 and 9, the “tree species preference” (TSP) models, add a new parameter: a tree preference factor τ , which represents the ratio of larval density on white oak to density on a black oak at the same spatial coordinates. In these models, the expected density of larvae on a tree is $\mathbb{E}[K_i] = \tau N q k(r_i, \theta_i) / L_i$ if the tree is a white oak, and $\mathbb{E}[K_i] = N q k(r_i, \theta_i) / L_i$ if the tree is a black oak. To examine the effects of synchrony on tree preference, the “release-dependent”

Models 10 and 11 fit a separate τ parameter for each release group, analogously to Models 4-7. Since complete tree species data was only available in the 1996 dataset, I only fit the Models 8-11 to the 1996 data.

Table 2: Comparison of model formulations.

No.	Nickname	$k(r)$	$\ell(\theta)$	Parameters	Notes
1	Diffusion	Half-normal	Uniform	σ, q	Random diffusion
2	Raw kernel	Exponential	Uniform	λ, q	Basic kernel model
3	von Mises	Exponential	von Mises	$\lambda, q, \bar{\theta}, \kappa$	
4	Three- q	Exponential	Uniform	λ, q_1, q_2, q_3	One q per release group
5	Three- q VM	Exponential	von Mises	$\lambda, q_1, q_2, q_3, \bar{\theta}, \kappa$	Three- q + von Mises
6	Three- λ	Exponential	Uniform	$\lambda_1, \lambda_2, \lambda_3, q$	One λ per release group
7	Three- λ VM	Exponential	von Mises	$\lambda_1, \lambda_2, \lambda_3, q, \bar{\theta}, \kappa$	Three- λ + von Mises
8	TSP	Exponential	Uniform	λ, q, τ	Tree species preference
9	TSP-VM	Exponential	von Mises	$\lambda, q, \bar{\theta}, \kappa, \tau$	TSP + von Mises
10	3-TSP	Exponential	Uniform	$\lambda, q, \tau_1, \tau_2, \tau_3$	One M per release group
11	3-TSP-VM	Exponential	von Mises	$\lambda, q, \bar{\theta}, \kappa, \tau_1, \tau_2, \tau_3$	3-TSP + von Mises

Bayesian inference

Bayesian inference is a method of fitting model parameters to data. Given data Y , we want to find the most probable value of the parameter β ; that is, we want to maximize the *posterior distribution* $P(\beta | Y)$. We calculate the posterior using Bayes' theorem:

$$P(\beta | Y) = \frac{f(Y | \beta)\pi(\beta)}{\int f(Y | \beta)\pi(\beta)d\beta}.$$

Here, f is the *likelihood function* of our model, which is equal to the model density function at Y with parameter value β , but which we interpret as a function of β with fixed data Y . The *prior distribution* $\pi(\beta)$ represents our information about the distribution of β before collecting data.

Note that the marginal probability in the denominator is a constant, so to maximize the posterior it suffices to maximize $f(Y | \beta)\pi(\beta)$.

In all of my models, the vector of larval counts K follows a multinomial distribution. However, for the purposes of computing the likelihood function, it is convenient to approximate $K_i \sim \text{Poisson}(\mu_i)$, where $\mu = \mathbb{E}[K_i]$ is as described earlier. If we make this approximation, then since $\sum K_i = N$, the vector $K = (K_i)_{i \leq M}$ will follow a multinomial distribution. Thus, the resultant likelihood functions will yield equivalent results.

Table 3 summarizes the priors for each of the model parameters I use here. In general, I set priors with relatively wide distributions to better ascertain the accuracy of model predictions in isolation, without outside information. For the diffusion model's standard deviation σ , I follow the recommendation of Gelman 2006 in using a half-normal with a large standard deviation of 100. I gave q and $\bar{\theta}$ uniform priors over their entire acceptable ranges. I chose a uniform prior for λ between 0.01 and 10, reflecting an acceptable range of mean dispersal between 0.1 m and 100 m. For the concentration parameter κ , I chose an exponential distribution to penalize high values, as I do not expect the distribution of larvae to be extremely directionally biased. However, I set the rate parameter relatively low at 1, to allow for a substantial range of potential concentrations. I gave the tree species preference factor τ a uniform distribution between 0 and 10, again to allow for a wide range of values.

In addition to the above parameters, I also fit priors on N_1, N_2, N_3 , the number of larvae released in each release group. Though Hunter and Elkinton 2000 estimated these parameters before their dispersal experiments, I wanted to reflect the uncertainty in these estimates by parametrizing and fitting them in my model. To simplify computation of priors, I fit $\tilde{N}_i = N_i \times 10^{-5}$ instead, multiplying by 10^5 to retrieve the number released. I set normal distributions on each $\tilde{N}_i$ with means based on the estimates provided in Table 1. Due to the differing numbers of larvae in 1995 and 1996, I used a different prior mean on $\tilde{N}_i$ for each release year. Hunter and Elkinton 2000 estimates were given with precision on the order of 10^4 , so I gave a standard deviation of 0.01 for each $\tilde{N}_i$ in each release year, giving a variance of 10^{-4} .

Table 3: Priors used in Bayesian inference.

Parameter	Prior
σ	Half-normal($0, 10^4$)
q	Uniform($0, 1$)
λ	Uniform($0.01, 10$)
$\bar{\theta}$	Uniform($-\pi, \pi$)
κ	Exponential(1)
τ	Uniform($0, 10$)
N_i	$\mathcal{N}(\mu_i, 10^{-4})^a$

^a Values of μ_i taken from estimates in Table 1.

Fit procedure

To perform fits with Bayesian inference, I used Markov chain Monte Carlo (MCMC) to compute posterior estimates. I provide a brief summary of the MCMC procedures I use here. Further can be found in Betancourt 2018, from which my summary is adapted.

In MCMC, our goal is to estimate a probability distribution $\pi(\beta)$, called the *target distribution*, by drawing random samples according to a Markov chain. That is, we begin at some arbitrary initial parameter value β_0 , and at time step n we choose β_n according to a conditional probability distribution dependent on β_{n-1} . We choose this conditional distribution, called the *Markov transition* $\mathbb{T}$, such that the distribution at each step is equal to the target distribution; that is,

$$\pi(\beta_n) = \pi(\beta_{n-1})\mathbb{T}(\beta_n | \beta_{n-1}).$$

Under ideal circumstances, due to the properties of Markov chains, $\{\beta_n\}$ will converge in distribution to the target distribution as $n \rightarrow \infty$. Thus, for a large enough chain length, $\{\beta_n\}$ is a representative sample from the target distribution, and so we can approximate $\pi(\beta)$ as the frequency of the value β in the chain.

The crucial element of MCMC is in using a transition algorithm such that the transition

probability $\mathbb{T}$ preserves the target distribution. Here, I use the Hamiltonian Monte Carlo (HMC) algorithm to accomplish this. In HMC, we treat the Markov chain as an “object” moving through parameter space according to Hamiltonian mechanics. At each step in the Markov chain, our coordinates in parameter space are given by the vector β . We further define a momentum vector ρ which controls the motion of the parameters. In each step of the Markov chain, we draw ρ from a multivariate normal distribution independent of β , and with covariance matrix given by a Euclidean metric matrix M on the parameter space. Then the Hamiltonian of the system is given by

$$\begin{aligned} H(\rho, \beta) &= -\log p(\rho) - \log p(\beta) \\ &= T(\rho) + V(\beta). \end{aligned}$$

Here, T and V are the kinetic and potential energy of the system. We may then solve Hamilton’s equations to determine the new state of the Markov chain. In practice, we approximate the solutions to Hamilton’s equations with a “leapfrog” method:

$$\begin{aligned} \rho' &= \rho - \frac{\epsilon}{2} \nabla V(\beta), \\ \beta^* &= \beta + \epsilon M^{-1} \rho', \\ \rho^* &= \rho^* - \frac{\epsilon}{2} \nabla V(\beta^*). \end{aligned}$$

Here ρ' is an intermediary calculation that is not used in later iterations, and ϵ is a “step-size” parameter. We may calculate the gradient of $V(\beta) = -\log p(\beta)$ numerically. The leapfrog algorithm is repeated a set number of times to generate a final candidate ρ^*, β^* . To adjust the transition probability so that we preserve the target distribution, we accept the candidate (ρ^*, β^*) with probability $\min(1, \exp[H(\rho, \beta) - H(\rho^*, \beta^*)])$, a procedure known as the Metropolis criterion. Iterating this procedure produces our Markov chain.

Convergence and mixing of Markov chains is essential to obtaining accurate estimates with MCMC. The *potential scale reduction factor* $\hat{R}$ and the *effective sample size* (ESS) n_{eff} are diagnostics

that can help identify whether chains exhibit adequate convergence and mixing. $\hat{R}$ is a convergence diagnostic, calculated as

$$\hat{R} = \sqrt{\frac{1}{W} \left(\frac{N-1}{N} W + \frac{1}{N} B \right)},$$

where W is within-chain variance, B is between-chain variance, and N is the number of draws per chain. If chains have converged well, then $B \rightarrow 0$ as N gets large, so we should have $\hat{R} \approx 1$. The ESS of M Markov chains of N draws each is defined as $n_{\text{eff}} = \frac{NM}{1 + 2 \sum_{t=1}^{\infty} \rho_t}$, where ρ_t is the autocorrelation with lag t , i.e. the correlation between two points of a chain separated by t steps. When n_{eff} is high, autocorrelation is low, indicating good mixing in chains (Vehtari et al. 2021).

I used Stan (Stan Development Team 2019) and the R package RStan (Stan Development Team 2020) to perform MCMC fits on my models. For each model, I ran four Markov chains of 2000 iterations each. The first 1000 iterations consisted of warmup, during which HMC parameters were tuned and the chains reached the target distribution, while the last 1000 iterations were used for posterior sampling. After runs completed, I ran diagnostics by checking Stan's estimates for $\hat{R}$ and n_{eff} , as well as plotting posterior draws and chain traces, to make sure that all chains converged and exhibited good mixing, and that a clear maximum was found. Following the recommendations of Stan Development Team 2019 and Vehtari et al. 2021, I consider $\hat{R} < 1.05$ to constitute acceptable convergence and $n_{\text{eff}} > 400$ to constitute acceptable mixing.

Model selection

Model selection helps to determine the quality of fits made using Bayesian inference, and rank models based on how accurate the predictions they make are. Here, I give an overview of the philosophy behind model selection and the mathematical basis of cross-validation. Further reading on Bayesian model selection can be found in Gelman et al. 2013 and Vehtari et al. 2016, from which my summary is adapted.

Consider a model posterior fit $P(\beta \mid y)$ on data y sampled from a distribution $f(y)$. Suppose we were to sample an extra data point $\tilde{y}$. What predictions can we make about the value of $\tilde{y}$

based on our model fit alone? We can answer this by defining a *predictive density* which measures the posterior likelihood of new data given old data:

$$p_{\text{post}}(\tilde{y} | y) = \int p(\tilde{y} | \beta) P(\beta | y) d\beta.$$

For ease of computation, we often use the *log predictive density*, $\text{lpd} := \log p_{\text{post}}(\tilde{y} | y)$. If we have not seen future data yet, we want to compute the expectation of lpd , called the *expected predictive density*:

$$\text{elpd} = \log \int p_{\text{post}}(\tilde{y} | y) f(\tilde{y}) dy.$$

The elpd of a model is a measure of its predictive power. We should thus prefer models with higher elpd over other models. Calculating the elpd of various models to determine the model with maximum elpd is the essence of model selection.

Since precise calculation of elpd requires knowledge of the true distribution f , we must estimate elpd from the data we have available. One means of doing this is *cross-validation*, in which we repeatedly fit the model to subsets of the data and use those fits to predict the remaining data points. The simplest version of this is leave-one-out cross validation (LOO-CV), in which each of our model fits exclude a single data point. Mathematically, the estimate of elpd from LOO-CV based on n data points y is

$$\widehat{\text{elpd}}_{\text{loo}} = \sum_{i=1}^n \widehat{\text{lpd}}_{\text{loo},i} = \sum_{i=1}^n \log \int p(y_i | \beta) p(\beta | y_{-i}) d\beta,$$

where y_{-i} is the vector of all data points except the i th.

Since $\widehat{\text{elpd}}_{\text{loo}}$ is only an estimate of elpd , it is necessary to account for error. A convenient measure of this error is an approximation of the standard error, $\text{se}_{\text{loo}} = \sqrt{nV}$, where V is the sample variance of the $\widehat{\text{lpd}}_{\text{loo},i}$ terms. An approximate 95% confidence interval is then given by $\widehat{\text{elpd}}_{\text{loo}} \pm 2\text{se}_{\text{loo}}$. However, since se_{loo} is only an estimate of the true standard error, and performing selection on 18 model fits raises the issue of multiple testing, a significant difference in elpd between two models should ideally be many times the standard error. Here, I air on the side of

caution and set the threshold for a significant elpd difference at 4se_{loo} .

Another measure of error is $\hat{p}_{\text{loo}} := \widehat{\text{elpd}}_{\text{loo}} - \text{lpd}$, which is called the effective parameter size because it is no larger than the number of parameters under well-behaved circumstances. Large $\hat{p}_{\text{loo}}$ values are an indication that a model may be misspecified.

In practice, LOO-CV is computationally costly, as it requires running a separate model fit for every data point. This makes LOO-CV unfeasible for even moderately-sized data sets. However, sampling methods exist that provide close approximations to LOO-CV and are much more efficient, in much the same way that MCMC sampling improves the efficiency of Bayesian inference calculations. The method I use here is called Pareto smoothed importance sampling (PSIS), first applied to LOO-CV in Vehtari et al. 2016, and implemented in R under the `loo` package Vehtari et al. 2020. In PSIS, we first estimate the lpd by drawing S parameter samples β^s from the posterior distribution $p(\beta \mid y)$. We then assign a weight $r_i^s = 1/p(y_i \mid \beta^s)$, called an *importance ratio*, to each parameter draw, and approximate the predictive density as

$$p(y_i \mid y_{-i}) \approx \frac{\sum_s r_i^s p(y_{-i} \mid \beta^s)}{\sum_s r_i^s} = \left(\frac{1}{S} \sum_s \frac{1}{p(y_i \mid \beta^s)} \right)^{-1}.$$

This estimate often produces a longer-tailed distribution than desired due to the high variance of the importance ratios. To correct the tail, we assume the importance ratios follow a generalized Pareto distribution, given by the CDF $F(r) = 1 - (1 - kr)^{1/k}$ for some skew parameter $k \in \mathbb{R}$. We fit the skew k to the top 20% of importance ratios, and we smooth by replacing the biggest importance ratios with $F^{-1}\left(\frac{z-1/2}{M}\right)$. After smoothing, we truncate the weights, and use the truncated weights w_i^s to compute the elpd as usual:

$$\widehat{\text{elpd}} = \sum_{i=1}^n \log \left(\frac{\sum_s w_i^s p(y_{-i} \mid \beta^s)}{\sum_s w_i^s} \right).$$

In addition to the standard error of $\widehat{\text{elpd}}$ and the effective parameter size p_{loo} , PSIS-LOO also produces another diagnostic value: the Pareto skew parameters k . Larger k values indicate that the distribution of importance ratios is very long tailed, and thus elpd estimates may not converge. Theoretically, only $k < 0.5$ guarantees finite variance and thus convergence, but Vehtari

et al. 2016 notes that $k \leq 0.7$ provides a good guarantee of convergence in practice.

Following best practices outlined in Vehtari et al. 2020, I use Pareto k estimates and effective parameter size p_{100} together to determine whether elpd estimates are reliable. A model has reliable estimates if all Pareto k estimates are < 0.7 . If at least one Pareto k estimates are ≥ 0.7 , and p_{100} estimates are significantly larger than the true parameter size (i.e greater than twice the standard error of p_{100}), then I consider the model to be “misspecified”, and the elpd estimate should be considered unreliable. If $\geq 1\%$ of Pareto k estimates are ≥ 0.7 , and p_{100} estimates are significantly larger than the true parameter size, then the model is “severely misspecified.”

To perform model selection, I rank each of the models by elpd estimate, and select the model with the highest reliable $\widehat{\text{elpd}}_{100}$, along with any other models whose $\widehat{\text{elpd}}_{100}$ falls within 4 standard errors. If no model has a reliable estimate, I prefer misspecified models over severely misspecified models.

Results

None of the MCMC runs had any divergent chains. Throughout, $\hat{R} < 1.05$, indicating reliable convergence. Trace plots indicated good mixing, and plots of posteriors reliably found unique optima. The estimated effective sample size n_{eff} was consistently > 2000 , indicating low overall autocorrelation.

Model selection

Table 4 shows the output of PSIS-LOO, along with standard errors and the percentage of acceptable Pareto k values. For the 1995 dataset, the three von Mises models (Models 3, 5 and 7) had the highest elpd estimates, though Model 7 was misspecified. There was no significant difference between the von Mises models or between the uniform models, but all three von Mises models had significantly higher elpd than every uniform model. For the 1996 dataset, the three- λ VM model (Model 7) was the clear frontrunner, exceeding all other models in elpd by at $> 5\text{se}_{100}$. Second place was a tie between the two TSP-VM models (Models 9 and 11). The models with

he lowest $\widehat{\text{elpd}}_{100}$ was the basic von Mises model (Model 3) and the basic TSP model (Model 8).

For the 1995 fits, diagnostic criteria were within acceptable ranges for all models except the basic von Mises model (Model 3). For the 1996 fits, all models were misspecified, and Models 3-6 and 8 showed severe misspecification. This is an indication that model fits and elpd estimates were far less reliable for the 1996 data than for the 1995 data.

Table 4: PSIS-LOO results.

1995					1996				
No.	Name	$\widehat{\Delta\text{elpd}}_{100}$	p_{100}	$k < 0.7$	No.	Name	$\widehat{\Delta\text{elpd}}_{100}$	p_{100}	$k < 0.7$
5*	Three- q VM	0	53.1 ± 6.5	100%	7* [†]	Three- λ VM	0	134.6 ± 31.0	99.7%
3*	von Mises	-5.4 ± 18.6	34.4 ± 4.2	100%	11 [†]	3-TSP-VM	-1579.6 ± 224.7	136.0 ± 18.0	99.5%
7 [†]	Three- λ VM	-15.1 ± 20.5	61.8 ± 8.4	99.7%	9 [†]	TSP-VM	-1603.5 ± 224.3	117.5 ± 15.5	99.9%
4	Three- q	-371.0 ± 80.5	41.8 ± 4.8	100%	6 [‡]	Three- λ	-1636.0 ± 341.8	138.5 ± 30.4	98.7%
6	Three- λ	-377.8 ± 82.2	42.2 ± 5.1	100%	4 [‡]	Three- q	-2311.2 ± 377.7	152.7 ± 34.4	98.7%
2	Raw kernel	-379.0 ± 87.1	21.5 ± 2.9	100%	5 [‡]	Three- q VM	-3800.1 ± 656.5	4289.8 ± 689.7	40.2%
1	Diffusion	-382.2 ± 86.7	20.4 ± 2.8	100%	10 [†]	3-TSP	-4205.3 ± 479.1	128.7 ± 20.1	99.5%
					2 [†]	Raw kernel	-4216.3 ± 499.9	104.0 ± 19.2	99.5%
					1 [†]	Diffusion	-5148.5 ± 544.7	136.8 ± 22.8	99.5%
					8 [‡]	TSP	-16087.6 ± 1480.4	14360.6 ± 1293.2	92.8%
					3 [‡]	von Mises	-20633.0 ± 2192.9	21502.2 ± 2118.7	92.7%

Models ranked from highest to lowest $\widehat{\text{elpd}}_{100}$. $\widehat{\Delta\text{elpd}}_{100}$ is the difference from the model with the highest $\widehat{\text{elpd}}_{100}$, p_{100} is effective parameter size, and the $k < 0.7$ column shows the percentage of Pareto k estimates falling below 0.7.

*Selected. [†] Misspecified. [‡] Severely misspecified.

Parameter fits

Figure 3 provides a visualization of parameter fits for the three selected models by plotting the predictions of selected models with dispersal data (cf. Hunter and Elkinton 2000, Fig. 3).

Tables 5 and 6 show the means and 95% credible intervals for fits of model without release-dependent parameters on the 1995 and 1996 datasets, respectively. Parameters varied widely between models and datasets. The standard deviation σ in the diffusion model had a mean of 30.06 in 1995 and a mean of 15.05 in 1996. Estimates of the fraction dispersing q exhibited the most drastic range, from $q = 0.07844, \text{CI}_{95\%} = (0.07408, 0.08376)$ in the 1995 diffusion model fit,

to $q = 0.6503$, $CI_{95\%} = (0.5961, 0.7220)$ in the 1995 raw kernel fit. Fits on the 1995 data showed more pronounced variation in q than in the 1996 data, which more reliably placed the mean of q between about 0.1 and 0.2, excluding the outlier of the diffusion model. The exponential rate parameter λ varied less drastically between means of approximately 0.025 and 0.13, indicating a mean displacement between 7 – 8 m and 40 m. Among the von Mises models, means for the mean angle $\bar{\theta}$ ranged from about 0.2559 to 0.4644, with means for the concentration κ between 0.7993 and 1.486, identifying a pronounced northeastern angle bias. Estimates for the tree preference factor τ were above 1 in mean, showing an overall preference for white oak in the 1996 dataset, though the credible interval in Model 8 extended well below 1.

Table 5: Parameter fits in release-independent models, 1995 dataset

Model		σ	q	$1/\lambda$	$\bar{\theta}$	κ
1	Mean	30.1	0.0784			
	$CI_{95\%}$	(27.5, 33.3)	(0.0741, 0.0838)			
2	Mean		0.650	37.6		
	$CI_{95\%}$		(0.596, 0.722)	(31.8, 45.7)		
3	Mean		0.489	23.5	0.465	0.799
	$CI_{95\%}$		(0.469, 0.511)	(21.1, 26.5)	(0.400, 0.534)	(0.740, 0.861)

Parameter fit means and 95% credible intervals from MCMC in RStan using the 1995 dataset, with models for which parameters did not depend on release time.

Figures 4, 5, and 6 show boxplots of parameter fits in release-dependent models. The probability of dispersal q varied widely between dispersal dates, but a definitive trend was not evident. In 1995, q was highest in the second release, with no distinct difference between 1 and 3. In 1996, however, the exact opposite occurred: the second release had the lowest q by far. The exponential rate λ also exhibited a mixture of trends in three- λ models. In 1995, the basic three- λ model showed no significant difference between the 3 releases groups, while the three- λ von Mises model showed a consistent, but not highly significant, decrease in λ in later releases. In 1996, however, both three- λ models gave the highest estimate for the exponential rate λ in the second release, followed by the first and third releases, with high significance. The 3- τ tree species pref-

Table 6: Parameter fits in release-independent models, 1996 dataset

Model		σ	q	λ	$\bar{\theta}$	κ	τ
1	Mean	15.1	8.96×10^{-3}				
	CI _{95%}	(14.7, 15.4)	$(8.65 \times 10^{-3}, 9.30 \times 10^{-3})$				
2	Mean		0.101	0.126			
	CI _{95%}		(0.0968, 0.106)	(0.1230, 0.130)			
3	Mean		0.219	0.0677	0.256	1.16	
	CI _{95%}		(0.154, 0.403)	(0.0100, 0.131)	(-0.311, 0.478)	(0.200, 1.52)	
8	Mean		0.184	0.0588			2.05
	CI _{95%}		(0.0977, 0.431)	(0.00674, 0.124)			(0.300, 2.89)
9	Mean		0.183	0.128	0.439	1.49	1.12
	CI _{95%}		(0.175, 0.191)	(0.126, 0.131)	(0.402, 0.475)	(1.44, 1.53)	(1.02, 1.23)

Parameter fit means and 95% credible intervals from MCMC in RStan using the 1996 dataset, with models for which parameters did not depend on release time.

erence (3TSP) model gave estimates for the tree preference factor τ as > 1 for all three releases, while the analogous model with a von Mises distribution (3TSP-VM) gave a large τ only to the third release.

Discussion

PSIS-LOO results indicate that the von Mises models reliably outperformed uniform-angle models, identifying a pronounced northeastern bias in angle. This bias corresponds with Hunter and Elkinton 2000's observation of a strong southwesterly wind at the field site. Modeling λ to depend on release time also seemed to significantly improve fits, although only in conjunction with a von Mises model. No other models improved elpd estimates significantly, indicating that angle distribution and release dependence on λ were the most influential factors. The consistent preference for the von Mises models in particular provides evidence that wind direction is a highly important factor in understanding dispersal which should not be ignored.

Parameter fits on the probability of dispersal q revealed a wide range of q values across treatment groups, with relatively large credible intervals. Although there seemed to be a significant

difference in q between release groups, the trends in the 1995 fits and 1996 fits were contradictory. Furthermore, PSIS-LOO did not reveal a significant difference in expected log predictive density between models which fit separate q parameters to different release groups and models which fit a single q to the entire dataset. Thus, although q exhibited high variance between release groups, there is not enough evidence to conclude that q is best modeled as a function of release time.

The three- q and three- λ models revealed unusual behavior of dispersal parameters in the 1996 dataset that was not observed in the 1995 dataset. Contrary to the findings of Hunter and Elkinton 2000, the second release group in 1996 displayed the lowest dispersal, with low q indicating a small fraction dispersing and a high λ indicating small mean displacement among dispersing larvae.

The plots of model predictions seem to provide close fits, and diagnostic criteria for MCMC were within acceptable ranges, ruling out the possibility of error in MCMC. However, the high percentage of anomalous Pareto k values among the 1996 models, especially the three- q models, indicates that the fits on that dataset are indeed highly misspecified. While the model fits themselves are not in error, there is likely a confounding factor in the 1996 data that none of the models are able to account for.

I suspect that the relatively high number of hatching larvae in the second release group of 1996 is responsible for this result. The second release of 1996 had approximately 10^5 more larvae hatching than the other two releases of that year, yet the total larval counts after dispersal were quite similar to other releases. Taken as a proportion of the total number hatching, the number of larvae dispersing in the second release was therefore far lower than the other two releases, leading q estimates in the second release to be much lower in model fits. Given that Hunter and Elkinton 2000 did not find decreased dispersal in the second release of 1996, however, the discrepancy between the number hatching and post-dispersal density counts is likely due to some other factor not accounted for in my models, such as post-hatch mortality.

Despite the issues with the 1996 dataset, the 1995 3- λ VM model fit provides some validation that dispersal increases with release time. It seems that this trend is mostly due to increased travel distance among dispersing larvae, rather than an increase in the fraction dispersing. This

suggests that phenology has some influence on dispersal, though the effect is not nearly as pronounced as the angle bias caused by wind.

Though the tree species preference models confirmed Hunter and Elkinton 2000's finding that preference for white oak increased with later dispersals, PSIS-LOO indicated that this effect was less important than the wind effect on angle. It seems that wind is the major determining factor in dispersal, as opposed to larval choice.

Conclusion

The models I have developed above provide a rigorous theoretical framework for conducting dispersal experiments and fitting field data. The von Mises models were able to reliably detect angle bias due to wind angle, and outperformed models which assumed a uniform angular distribution. Models accounting for differences in release time slightly outperformed models which did not account for release time, although the difference in elpd between release-dependent and release-independent models was much smaller than the difference between von Mises and uniform-angle models. This indicates that angle bias due to wind was more of a determining factor in dispersal than synchrony of hatch with budburst.

The importance of wind as a determining factor in the dataset I examine here provides evidence that environmental factors are highly important to understanding small-scale dispersal. It is therefore likely that patterns of larval dispersal vary widely based on factors such as local climate and forest structure. Studies of larval ballooning in a wide range of environments are needed to further clarify the relation between dispersal and these environmental factors. The models I develop here might then be used to compare and contrast dispersal under different environmental conditions.

Though I focus on the effects of wind and synchrony, the modeling approach I outline here can be easily adapted to model a wide variety of environmental or biological factors. Some variables of interest might be ambient temperature, effects of parasitoids or viruses, local insect density, and comparisons between populations or subspecies. This modeling approach thus creates many new potential avenues of research into small-scale dispersal, which may improve

our understanding of small-scale effects in populations of defoliating lepidopterans.

The models I develop here may be especially valuable to understanding the spatial structure of NPV outbreaks in Douglas-fir tussock moth, or other similar viral systems. A study similar in design to Dwyer 1992 might adapt one of the models I develop here to predict the density-dependent spread of an epizootic through a dispersing population. Such a study could provide insight into the effects of ballooning behavior on the spread of NPV in Douglas-fir tussock moth.

Acknowledgments

I would like to thank Will Koval, Greg Dwyer, Alison Hunter, and Dmitry Kondrashov for their advice and guidance on this project. I would also like to thank Alison Hunter for generously sharing her data.

Literature Cited

- Michael Betancourt. A conceptual introduction to Hamiltonian Monte Carlo. 2018.
- Kevin M. Bigsby, Patrick C. Tobin, and Erin O. Sills. Anthropogenic drivers of gypsy moth spread. *Biological Invasions*, 13(2077), 2011. doi: 10.1007/s10530-011-0027-6.
- K.M. de Beurs and P.A. Townsend. Estimating the effect of gypsy moth defoliation using MODIS. *Remote Sensing of Environment*, 112(10):3983–3990, 2008. ISSN 0034-4257.
- Greg Dwyer. On the spatial spread of insect pathogens: Theory and experiment. *Ecology*, 73(2): 479–494, 1992.
- Andrew Gelman. Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Analysis*, 1(3):515 – 534, 2006. doi: 10.1214/06-BA117A. URL <https://doi.org/10.1214/06-BA117A>.
- Andrew Gelman, Jessica Hwang, and Aki Vehtari. Understanding predictive information criteria for Bayesian models, 2013.
- Alison F. Hunter and Joseph S. Elkinton. Effects of synchrony with host plant on populations of a spring-feeding lepidopteran. *Ecology*, 81(5):1248–1261, 2000.
- Daniel H. Janzen. Herbivores and the number of tree species in tropical forests. *The American Naturalist*, 104(940):501–528, 1970. ISSN 00030147, 15375323. URL <http://www.jstor.org/stable/2459010>.
- Etienne K. Klein, Aurore Bontemps, and Sylvie Oddou-Muratorio. Seed dispersal kernels estimated from genotypes of established seedlings: does density-dependent mortality matter? *Methods in Ecology and Evolution*, 4(11):1059–1069, 2013.
- Andrea Clavijo McCormick, Luca Arrigo, Helen Eggenberger, Mark C. Mescher, and Consuelo M. De Moraes. Divergent behavioural responses of gypsy moth (*Lymantria dispar*) caterpillars

- from three different subspecies to potential host trees. *Scientific Reports*, 9(8953), 2019. doi: 10.1038/s41598-019-45201-3.
- Michael McManus. The role of behavior in the dispersal of newly hatched gypsy moth larvae. (Res. Pap. NE-267), 1973.
- Michael McManus and György Csóka. History and impact of gypsy moth in North America and comparison to recent outbreaks in Europe. *Acta Silvatica et Lignaria Hungarica*, 3:47–64, 01 2007.
- R.G. Mitchell. Dispersal of early instars of the Douglas-fir tussock moth. *Annals of the Entomological Society of America*, 72(2):291–297, 1979.
- Valerie Pasquarella, Joseph Elkinton, and Bethany Bradley. Extensive gypsy moth defoliation in southern new england characterized using Landsat satellite observations. *Biological Invasions*, 20, 11 2018. doi: 10.1007/s10530-018-1778-0.
- David Savage, Martin J. Barbetti, William J. MacLeod, Moin U. Salam, and Michael Renton. Can mechanistically parameterised, anisotropic dispersal kernels provide a reliable estimate of wind-assisted dispersal? *Ecological Modelling*, 222(10):1673–1682, 2011. ISSN 0304-3800. doi: <https://doi.org/10.1016/j.ecolmodel.2011.03.003>. URL <https://www.sciencedirect.com/science/article/pii/S0304380011001104>.
- Alexei A. Sharov and Andrew M. Liebhold. Model of slowing the spread of gypsy moth (Lepidoptera: Lymantriidae) with a barrier zone. *Ecological Applications*, 8(4): 1170–1179, 1998. doi: [https://doi.org/10.1890/1051-0761\(1998\)008\[1170:MOSTSO\]2.0.CO;2](https://doi.org/10.1890/1051-0761(1998)008[1170:MOSTSO]2.0.CO;2). URL <https://esajournals.onlinelibrary.wiley.com/doi/abs/10.1890/1051-0761%281998%29008%5B1170%3AMOSTSO%5D2.0.CO%3B2>.
- Stan Development Team. Stan modeling language users guide, 2019. URL <http://mc-stan.org/>. version 2.26.
- Stan Development Team. RStan: the R interface to Stan, 2020. URL <http://mc-stan.org/>. R package version 2.21.2.

Aki Vehtari, Andrew Gelman, and Jonah Gabry. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, 27(5):1413–1432, Aug 2016. ISSN 1573-1375. doi: 10.1007/s11222-016-9696-4. URL <http://dx.doi.org/10.1007/s11222-016-9696-4>.

Aki Vehtari, Jonah Gabry, Mans Magnusson, Yuling Yao, Paul-Christian Bürkner, Topi Paananen, and Andrew Gelman. loo: Efficient leave-one-out cross-validation and WAIC for Bayesian models, 2020. URL <https://mc-stan.org/loo/>. R package version 2.4.1.

Aki Vehtari, Andrew Gelman, Daniel Simpson, Bob Carpenter, and Paul-Christian Bürkner. Rank-normalization, folding, and localization: An improved $\hat{R}$ for assessing convergence of MCMC. *Bayesian Analysis*, -1(-1), Jan 2021. ISSN 1936-0975. doi: 10.1214/20-ba1221. URL <http://dx.doi.org/10.1214/20-BA1221>.

Figure 3: Fits of selected models compared to dispersal data
Model 3, 1995 data

Model 5, 1995 data

Model 7, 1996 data

Plots are grouped by release date. The size of black circles is proportional to observed mean larval counts at each point, while the size of red circles is proportional to expected larval counts, calculated as $E[K_i] = Nqk(r_i, \theta_i)/L_i$. The thickness of red circles extends two standard deviations above and below the mean to give an approximate 95% confidence interval.

Figure 4: 3- q model comparison

Boxplots of fits for q_1, q_2, q_3 in each year, with corresponding single- q models for reference. Boxes represent 50% credible intervals and whiskers represent 95% credible intervals, with bars indicating means.

Figure 5: 3- λ model comparison

Boxplots of fits for $1/\lambda_1, 1/\lambda_2, 1/\lambda_3$ in each year, with corresponding single- λ models for reference.. Boxes represent 50% credible intervals and whiskers represent 95% credible intervals, with bars indicating means.

Figure 6: 3- τ model comparison

Boxplots of fits for τ_1, τ_2, τ_3 for 3TSP models, with single- τ TSP models for reference. Boxes represent 50% credible intervals and whiskers represent 95% credible intervals, with bars indicating means.