A Bayesian framework to model temperature-dependent transmissible cancer dynamics within soft-shell clam *(Mya arenaria)* populations

Carissa Mayo

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

Master of Science

University of Washington

2026

Committee:

Michael Metzger

Alison Feder

Robert Noble

Chelsea Wood

Program Authorized to Offer Degree:

Quantitative Ecology and Resource Management

University of Washington

Abstract

A Bayesian framework to model temperature-dependent transmissible cancer dynamics within soft-shell clam *(Mya arenaria)* populations

Carissa Mayo

Chair of the Supervisory Committee:
Michael Metzger
Quantitative Ecology and Resource Management

Bivalve Transmissible Neoplasia (BTN) is a recently recognized transmissible cancer that spreads among bivalve species worldwide through the transmission of living cancer cells, producing disease dynamics unlike those of typical pathogens. Its marine setting adds a layer of complexity to disease transmission that we account for by modifying standard infectious disease models. Many elements of BTN including transmission, seasonality, and environmental sensitivity are not well understood, which represents a challenge for managing this infectious cancer. In this study, we examined how BTN progresses within *Mya arenaria* populations and how temperature affects these dynamics. To accomplish this objective, we developed a Bayesian epidemiological model of BTN progression within clam populations. We used this model to quantify transmission and progression rates along with the uncertainties of these estimations. Comparisons among model variants that included different combinations of temperature dependencies indicated that the incorporation of temperature into the model improved its ability to reproduce the patterns observed in survey data. The results further indicate that temperature-dependent progression had the largest impact among the mechanisms evaluated in explaining the seasonal variation present in survey data. However, the results of this study also suggest that there are additional mechanisms, not included in this model, that contribute to the dynamics of BTN infection in clam populations. Overall, this study establishes a quantitative, mechanistic framework for describing the dynamics of BTN in natural clam populations and identifies key areas for future research in both laboratory and field work.

Table of Contents

Acknowledgments

I would first like to thank my adviser, Michael Metzger, for his guidance, support, and mentorship throughout this project. I am grateful for his insight and encouragement as this work developed. I would also like to thank the members of the Metzger Lab for processing and organizing much of the data used in this thesis and for providing valuable feedback throughout this work.

I would also like to thank my committee members, Alison Feder and Chelsea Wood, for their thoughtful feedback and support. I would especially like to thank Robert Noble for his generous time and guidance throughout this project. His help with the modeling and statistical components of this work was invaluable, and I am especially grateful for the many meetings and conversations he made time for despite being many time zones away.

Finally, I would like to thank my family for their constant support and encouragement. To my mom, dad, and sister, thank you for always being there for me throughout graduate school. I would also like to thank my significant other, Jack, for his patience, encouragement, and support during this process. This work would not have been possible without the people who helped me through it.

1 Introduction

Transmissible cancer is a rare phenomenon that naturally occurs in a few species, including Tasmanian devils, dogs, and several bivalve species (Metzger and Goff, 2016). Devil Facial Tumor Disease (DFTD) (Pearse and Swift, 2006) and Canine Transmissible Venereal Tumor (CTVT) were identified as transmissible in 2006 (Murgia et al., 2006). More recently, in 2015, a fatal leukemia-like cancer affecting bivalve species worldwide, including soft-shell clams, was confirmed to be transmissible (Metzger et al., 2015). Due to its recent classification, much remains unknown about BTN, particularly its mechanisms of spread among individual bivalves and populations. There are no existing epidemiological models for this system, leaving a major gap in understanding how BTN spreads in natural populations.

BTN is spread via transmissible cells derived from a single clam (Metzger et al., 2015). The cancer cells are released by infected clams and can survive for weeks in seawater, with colder temperatures being more suitable for their survivability (Giersch et al., 2022). Studies also show a seasonal pattern of BTN prevalence, with higher incidence rates occurring during late fall and winter compared to early summer each year (Leavitt et al., 1990). Therefore, tracking both the disease infections in clam populations and the presence of BTN cells in the water column are crucial to disease monitoring. However, we still lack quantitative estimates of the fundamental processes governing transmission, progression, and particle persistence.

While infectious diseases occur in all ecosystems, most epidemiological models are derived from terrestrial environments. Marine diseases, however, exhibit different dynamics, such as variations in transmission and recruitment patterns (McCallum et al., 2004). Adapting current epidemiological models to marine environments is critical for generating accurate disease predictions, especially as marine ecosystems continue to change rapidly (Lafferty, 2017). More generally, the approach taken in this research is relevant to many other types of aquatic disease systems because it explicitly represents pathogen persistence in the environment and allows processes affecting pathogen release, survival, and removal to be included in the model.

Various environmental stressors influence the conditions of an aquatic environment, which then affect disease outbreaks (Tracy et al., 2019). Temperature, in particular, can increase the rate of disease spread and can also potentially contribute to the survivability of pathogens in a

marine system (Aalto et al., 2020). Data-driven modeling and seasonal forecasts are critical elements of a disease management strategy to enable early intervention and effective disease surveillance, especially under the threat of climate change (Groner et al., 2016). For example, one study was the first to develop a disease model for Whirling Disease (WD) in trout and used its results to determine a more effective treatment and management strategy (Turner et al., 2014). The mechanisms behind observed seasonal peaks remain unidentified for BTN, and no epidemiological framework exists to test hypotheses or predict BTN dynamics under environmental change.

Bayesian analysis offers a powerful method for making inferences and forecasting disease spread, especially with limited data (Cook et al., 2023; Grinsztajn et al., 2021). BTN datasets are sparse, highly seasonal, and frequently include measurements at or below detection limits, creating uncertainty in progression and transmission estimates. By combining prior knowledge with available data, a Bayesian framework allows for probabilistic inference, providing posterior distributions that quantify uncertainty in model parameters and predictions (Gelman et al., 2020). There are also available computational methods that enable efficient exploration of wide prior distributions (Carpenter et al., 2017). This framework supports inference even under sparse and noisy data, a common feature of biological data and this unfamiliar disease system.

This research develops a numerical marine disease model of BTN in *Mya arenaria* (soft-shell clam) populations to study the temporal patterns of the disease using a Bayesian compartmental modeling approach. This approach integrates controlled laboratory experiments with environmental monitoring data to synthesize current knowledge of the disease. Cancer particles are specifically included in this model as they represent the unique mode of transmission associated with BTN and temperature dependent effects are included to simulate both progression and particle decay. Using this framework, key transmission and progression rates can be estimated, along with uncertainty in each parameter, and seasonal patterns can be evaluated. Together, these results improve our understanding of BTN transmission and provide a foundation for forecasting disease dynamics in changing marine environments.

Specifically, this study aims to (1) utilize a Bayesian epidemiological model to estimate key transmission and progression rates for BTN, (2) integrate particle dynamics into a mechanistic disease model, and (3) assess how temperature shapes seasonal BTN patterns.

2 Methods

2.1 Model Formulation

We modeled the dynamics of transmissible cancer in a marine bivalve population using a compartmental epidemiological model adapted from the traditional SIR model by Tillett (1992), to describe the BTN patterns in a *Mya arenaria* population. We modified this model to account for disease latency and environmental transmission (Bidegain et al., 2016). This model is shown in Figure 1. We included an exposed (E) stage of disease progression and a particle (P) compartment to represent the amount of cancer cells in the water that can cause new infections, expressed as qPCR-derived target-sequence copies per milliliter (copies/mL). Susceptible (S) individuals only become exposed, most likely, through uptake of disease particles from the water column during filter feeding, since marine bivalves are sessile organisms. In the exposed stage, individuals have a level of infection, but are not emitting cancer cells into the water column. As the disease progresses, clams become infectious (I) and release cancer cells, contributing to the particle concentration compartment. The system equations are provided below.

$$\frac{dS}{dt} = -\beta PS + mI \tag{1a}$$

$$\frac{dE}{dt} = \beta PS - \alpha E \tag{1b}$$

$$\frac{dI}{dt} = \alpha E - mI \tag{1c}$$

$$\frac{dP}{dt} = \gamma I - \delta P \tag{1d}$$

Parameter β represents the transmission rate between cancer cells in the water and susceptible individuals. The rate α governs the time from exposure to being infectious with particle emission. Infectious individuals shed particles into the water at rate γ, and δ denotes the decay rate of particles. We assume a closed population with no natural births or deaths unrelated to disease. To maintain a constant population size, we add an input term mI into the susceptible class, where m is the disease mortality rate. This balances the loss of infected individuals, ensuring the total population remains constant over time, N = S + E + I. Individuals that die

from the disease are not explicitly tracked beyond the infectious class. The removed class is therefore not included in the model without loss of information when the population balance is enforced, as is standard in models assuming constant population size (Tien and Earn, 2010). Tables 1 and 2 summarize the model variables and parameters.

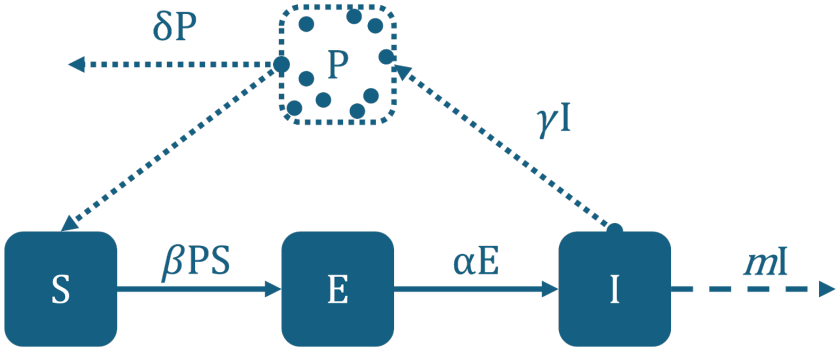


Figure 1: SEIP model structure.

Table 1: Model variables.

Symbol	Definition	Units
S	Proportion susceptible	unitless (fraction)
E	Proportion exposed (infected, not emitting)	unitless (fraction)
I	Proportion infectious (infected, emitting cancer cells)	unitless (fraction)
P	Cancer cell concentration (particles) in the water	copies /mL
N	Total population size (normalized to 1 in the model)	unitless (fixed at 1)

Table 2: Model parameters.

Symbol	Definition	Units
β	Disease transmission rate between susceptibles and waterborne cancer cells	$(\text{copies/mL})^{-1} \cdot \text{day}^{-1}$
α	Progression rate from exposed to infectious	day^{-1}
m	Mortality rate of infectious individuals	day^{-1}
γ	Particle shed rate	$\text{copies/mL} \cdot \text{day}^{-1}$
δ	Particle decay rate	day^{-1}

We hypothesized that disease dynamics are influenced by temperature, particularly in shaping seasonal patterns of transmission. To test this, we incorporated temperature-dependence into two key model parameters: the transmission rate β and the particle decay rate δ. Prior work demonstrated the thermal decay of cancer cells in saltwater (Giersch et al., 2022), and laboratory data suggested that temperature also affects disease progression in clams (see Figure 2. (d)). However, these relationships are contradictory to disease prevalence with an increase

in cell decay hindering disease transmission and increased disease progressions accelerating it in warmer temperatures. Thus, we explored and compared models where either or both of these parameters vary as a function of temperature to determine how the seasonality we observe in populations is shaped. These model variants are defined in Table 3.

Table 3: Model variants.

Model	**Description**
1	No temperature-dependence
2	Temperature-dependent decay only (δ_T)
3	Temperature-dependent transmission only (β_T)
4	Both temperature-dependent parameters

To account for seasonal temperature variation in the survey sites, we modeled the temperature as a sine wave with a period of one year (with t in days):

$$T(t) = \mu_T + A_T \cdot \sin(\frac{2\pi t}{365} + \phi)$$

where μ_T ($^\circ C$) is the mean annual temperature, A_T ($^\circ C$) is the amplitude, and ϕ is a phase shift. For the temperature-dependent components of the model, temperature was transformed to a standardized, dimensionless index, $z_T(t) = \frac{T(t)-\mu_T}{|A_T|/\sqrt{2}}$, so that coefficients were estimated on a centered and comparable scale. This improved interpretability and numerical stability (Gelman et al., 2020). Our laboratory data show that temperature affects cancer progression within clams, thus we modeled temperature as a modifier of the effective transmission rate β_{eff}. We modeled temperature-dependent progression as

$$\beta_{\text{eff}} = \beta \cdot \beta_T, \qquad \beta_T = \text{logit}^{-1}(a_\beta + b_\beta z_T(t)),$$

where β_T is a temperature-dependent progression modifier bounded between 0 and 1. We modeled particle decay as

$$\delta_T = \exp(a_\delta + b_\delta z_T(t)).$$

This captures the relationship of temperature effects on cancer cell decay in the water. Further explanation on these derivations is provided in Section 2.4 and Appendix A.

2.2 Structural Identifiability

To identify which parameters and combinations of parameters of the SEIP transmission model can be uniquely recovered from ideal, noise-free observations, independent of any particular dataset, we analyzed the structural identifiability of the model (Wieland et al., 2021). We utilized the full ODE system of Model 4 (both temperature-dependent parameters) so identifiability was evaluated in the most complex case to reveal all possible issues and ensure that identifiability can only improve under simplifications. We used the 'StructuralIdentifiability.jl' package in Julia (Dong et al., 2023) by declaring the observed outputs (S,E,I,P) and stating the model parameters and ODEs.

The results showed that most model parameters were structurally identifiable, including the progression rate (α), death rate (m), and particle shedding parameter (γ). However, the baseline transmission rate (β) and its temperature effect (β_T) were not individually identifiable and only their product, β_{eff} was identifiable. Similarly, a_β, b_β, a_δ and b_δ were not identifiable. Initial conditions for all compartments were also non-identifiable, which is common without early-time data or informative priors.

Accordingly, in the model, we parameterized and reported the effective transmission β_{eff} and δ_T directly. We also anchored initial values to baseline observations using informative priors, which improves identifiability.

2.3 Bayesian Framework

The SEIP disease transmission model was implemented within a Bayesian framework. Following Bayes' theorem, the posterior distribution of the unknown parameters, $P(\theta|y)$, is proportional to the product of the likelihood, $P(y|\theta)$, which represents the probability of the observed data given the parameters, and the prior distribution, $P(\theta)$, which encodes prior knowledge or belief about the parameters (Gelman et al., 2020):

$$P(\theta|y) \propto P(y|\theta) \cdot P(\theta)$$

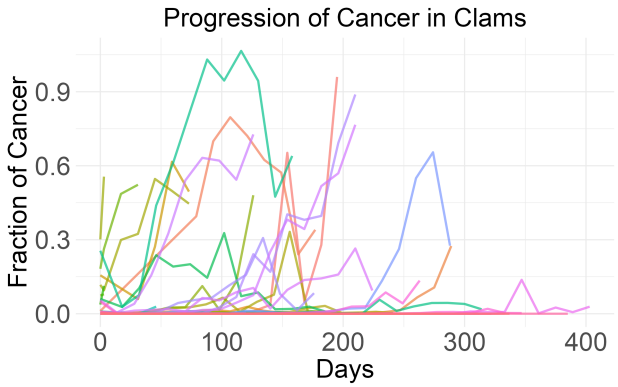
Prior distributions for each parameter reflect available biological information from laboratory experiments or plausible ranges. For parameters with strong empirical support, informative priors were used to convey precise biological constraints. For parameters with limited data, weakly informative priors were specified to encode broad, biologically plausible ranges without overly constraining inference, allowing the data to dominate the posterior. The following section details prior specifications for all parameters in the SEIP model.

2.4 Prior Distributions Informed by Laboratory Data

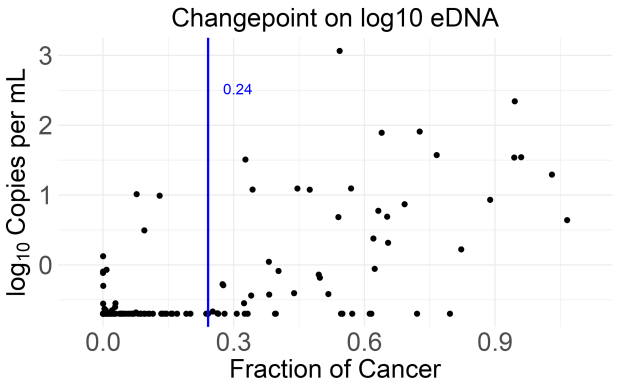
From laboratory data conducted by the Pacific Northwest Research Institute (Figure 2) and existing biological understanding, we derived informative or weakly informative priors for each parameter. These priors help enforce biologically realistic constraints and improve model stability, especially in the presence of limited or noisy field data. When sufficient data were available, we used time-to-event data and concentration measurements to fit probability distributions directly. Where data was sparse, we used weakly informative priors that encode plausible ranges without constraining inference. Following methods in classical epidemiological modeling, we define time-to-event parameters using positive-valued distributions. Gamma distributions are commonly used to represent waiting times such as latency and infectiousness periods due to their flexibility and interpretability (Anderson and Watson, 1980). In addition to gamma distributions, we use lognormal priors where greater skewness or heavier tails are needed, particularly for weakly informed parameters where prior knowledge is limited. For all weakly informed parameters and their hyperparameters, we followed a prior predictive checking approach (Schad et al., 2021), iteratively adjusting hyperparameters of tentative priors until simulated trajectories from the model were biologically plausible and model diagnostics were clean, while maintaining vague boundaries. Data collection and preprocessing follow Giersch et al. (2025) and Giersch et al. (2022). Here we focus on the SEIP modeling and prior derivation. Full details on data subsetting, regression procedures, and distribution fitting are provided in Appendix A. The prior distributions and their data sources are summarized in Table 4.

Initial conditions were treated as unknown parameters with informative priors anchored to the early field observations. The initial host-state values, $[S(0), E(0), I(0)]$, were assigned a Dirichlet prior centered on the first two observed SEI proportions, while the initial particle

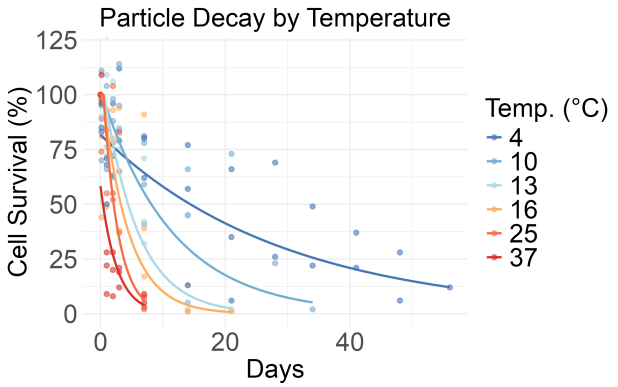
concentration $P(0)$ was assigned a lognormal prior informed by the first two eDNA survey measurements. These initial values were structurally unidentifiable (Section 2.2), so anchoring them was necessary to stabilize estimation and improve sampler behavior.



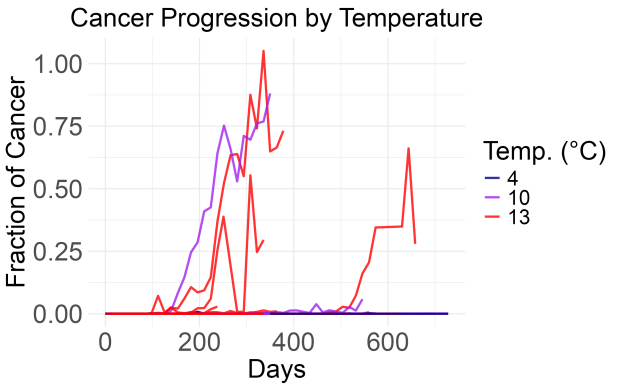
(a) Longitudinal cancer trajectories.



(b) Changepoint/infectiousness threshold.



(c) Temperature-driven decay of cancer cells.



(d) Temperature effects on progression.

Figure 2: Laboratory data used to inform prior distributions. **(a)** Longitudinal cancer progression trajectories of individual clams, showing the level of cancer over time. Each clam was held in a separate tank and repeatedly sampled to quantify cancer levels (Giersch et al., 2025). **(b)** Changepoint analysis identifying the infectiousness threshold (24.12% cancer fraction) used to classify infectious individuals. Clams were kept in individual tanks and water samples are taken to monitor the concentration of cancer particles released in a period of 24 hours in addition to the cancer fraction samples (Giersch et al., 2025). **(c)** Cancer cell decay rates measured in seawater across temperature treatments, demonstrating faster decay at higher temperatures. Each point represents a measurement and lines represent fitted regression lines for each temperature (Giersch et al., 2022). **(d)** Progression of cancer in clams across temperature treatments (n = 6 per temperature treatment), showing how temperature increases the likelihood of cancer progression. Susceptible clams were experimentally injected with BTN, maintained individually, and the cancer fraction was monitored over time.

As there is no data to directly inform β and it is a positive rate parameter, we assigned a weakly informative lognormal prior centered at 0.02, written as $\beta \sim \text{Lognormal}(\log(0.02), 1.0)$. Longitudinal cancer trajectories (Figure 2. (a)), previously described in Giersch et al. (2025), were used here to inform priors for progression (α) and mortality (m). We performed a change-

point procedure and found an infectiousness threshold of 24.12% hemolymph cancer fraction (Giersch et al., 2025). We used this value to define the infectious class (Figure 2. (b)). While fitted progression curves identified when clams crossed this threshold, the starting time at 0% cancer is unknown since we received these animals from the wild where their time of infection was not observed, meaning only a minimum time to infectiousness ($\sim$ 80 days) could be estimated. This timescale was used to biologically inform the center of the lognormal prior for the progression rate α, $\alpha \sim \text{Lognormal}(\log(0.01), 0.3)$, but was not imposed as a bound because doing so did not affect model performance. For both β and α, standard deviations are based on iterative model testing and chosen to balance between parameter flexibility and computational stability. Mortality rates were better supported, and m was modeled with an inverse gamma distribution (shape = 2.60, scale = 0.04) consistent with an average time to death of about 59 days.

The particle emission rate (γ) was estimated by eDNA measurements of cancer cell concentrations in the water from infectious clams (Figure 2. (b)). To account for the right-skewed distribution of nonzero values, we fit a lognormal model on the positive values, resulting in a lognormal(1.24, 1.93) prior for γ. The particle decay rate (δ) was estimated by exponential decay fits to cancer cell survival in saltwater (Figure 2. (c)), leading to a lognormal prior (log(0.05), SD = 0.1). This seawater survival dataset is described in Giersch et al. (2022) and reanalyzed here to construct priors.

Temperature effects on decay rate (δ_T), were estimated by fitting linear models to individual decay rates across temperature treatments (Figure 2. (c)). These hyperparameters were modeled with normal priors ($a_\delta \sim \mathcal{N}(-1.97, 0.3), b_\delta \sim \mathcal{N}(0.32, 0.06)$) and truncated to enforce the statistically significant positive relationship between cancer cell decay rate and temperature ($b_\delta > 0$). For temperature-dependent progression (β_T), a logistic regression on binary progression outcomes (Figure 2. (d)) returned a positive, but non-significant slope. Therefore, we did not enforce the relationship with truncation, but tighter priors were needed for model stability ($a_\beta \sim \mathcal{N}(-2.73, 2.0), b_\beta \sim \mathcal{N}(2.72, 6.0)$). Temperature covariates were standardized prior to fitting and all hyperparameters therefore correspond to the effect on the standardized temperature scale $z_T(t)$. Hyperparameter standard deviations were selected iteratively for model stability while preserving reasonable uncertainty.

Table 4: Parameter priors.

Parameter	Prior distribution	Data source
β	Lognormal(log(0.02), 1.0)	Figure 2A (weakly informed)
α	Lognormal(log(0.01), 0.3)	Figure 2A (weakly informed)
m	InvGamma(2.60, 0.04)	Figure 2A
γ	Lognormal(1.24, 1.93)	Figure 2B
δ	Lognormal(log(0.05), 0.1)	Figure 2C
β_T	$a_\beta \sim \mathcal{N}(-2.73,\ 2.0)$ $b_\beta \sim \mathcal{N}(2.72,\ 6.0)$	Figure 2D
δ_T	$a_\delta \sim \mathcal{N}(-1.97,\ 0.3)$ $b_\delta \sim \mathcal{N}(0.32,\ 0.06)$ truncated at $[0, \infty)$	Figure 2C

Note. Initial conditions were anchored to first two observations using informative priors (Dirichlet for $[S(0), E(0), I(0)]$, LogNormal for $P(0)$).

2.5 Field Data and Observation Models

Survey data was collected over a period of two years. We were sent soft-shell clams every other month from Gloucester Marine Genomics Institute (GMGI) from a site in Forest River, MA ($42.508043°N, -70.882746°W$) beginning in February 2023 and a site in Pavilion Beach, MA ($42.700431°N, -70.790456°W$) beginning in April 2023. Each shipment included approximately 50 animals per site, which we processed to estimate per-animal cancer fraction. Shipments also included three water samples per site for eDNA (environmental DNA) quantification of BTN in the water column. The site data is shown in Figures 3 and 4. We followed our lab's published protocols for cancer-fraction derivation and eDNA processing without modification (Giersch et al., 2022, 2025). The survey sites differ in observed particle dynamics and disease seasonal patterns, but there is no clear explanation for these differences.

We obtained continuous water temperature from HOBO U24 Conductivity logger (model U24-002-C, Onset Computer Corporation; temperature accuracy ±0.1°C) at each sample site. Using HOBOware, we processed the hourly temperature reads over the study period and fit a sine wave to the raw data (Figure 5) to derive the temperature parameters for each site as described in Section 2.1. Site temperature dynamics were similar, therefore, theoretical analyses used the Forest River parameterization, while individual parameters were used for fitting to the data. At each sampling time t, we classify animals into counts of Susceptible, Exposed, or Infectious:

$$\mathbf{y}_t = (y_{t,S},\ y_{t,E},\ y_{t,I}), \qquad n_t = y_{t,S} + y_{t,E} + y_{t,I}.$$

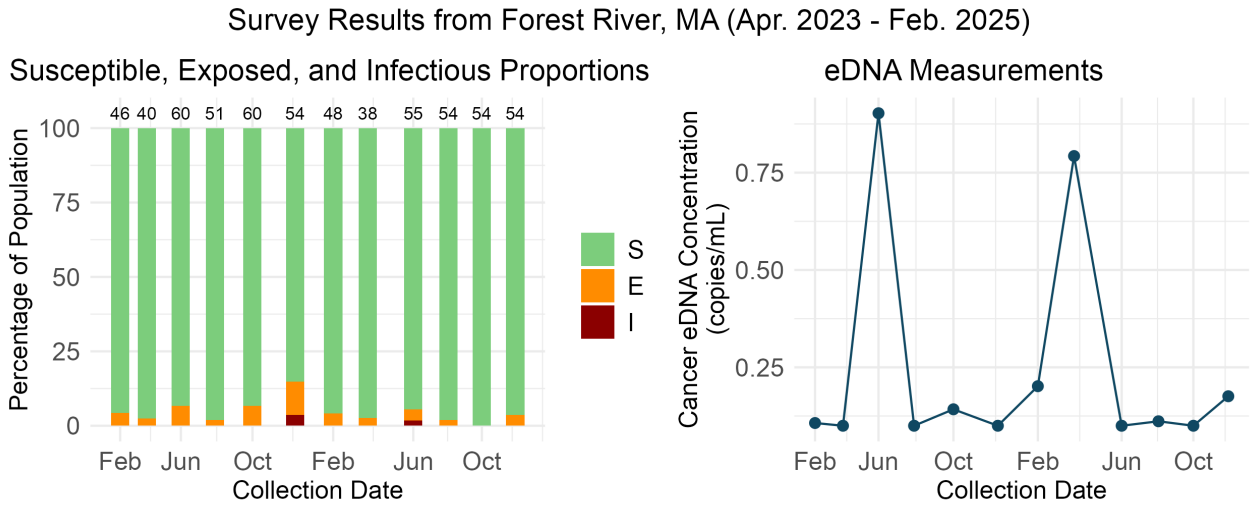


Figure 3: Survey data from Forest River, MA including SEI counts and eDNA measurements.

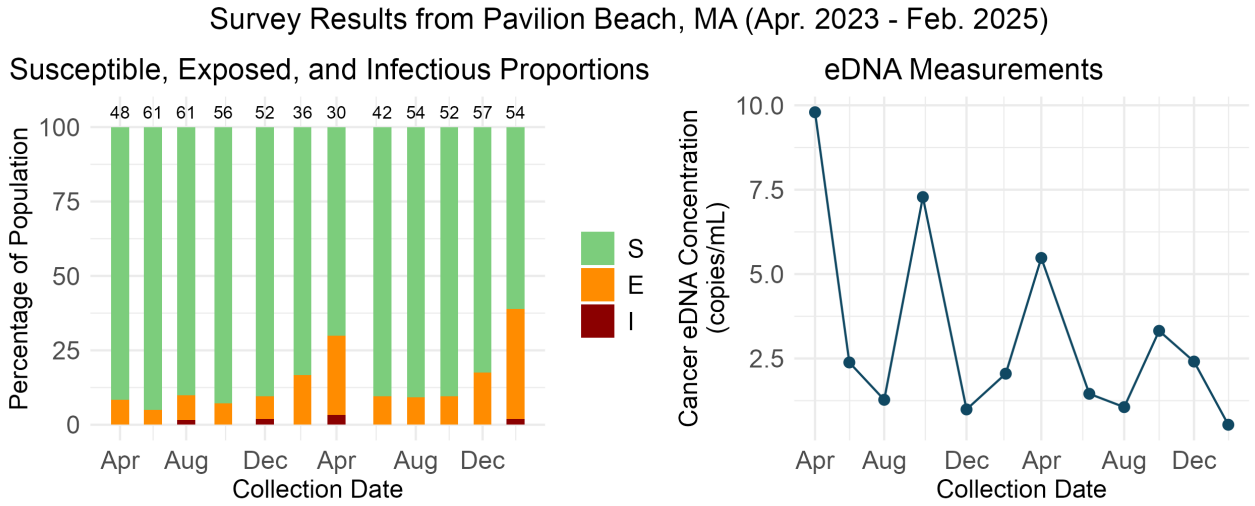


Figure 4: Survey data from Pavilion Beach, MA including SEI counts and eDNA measurements.

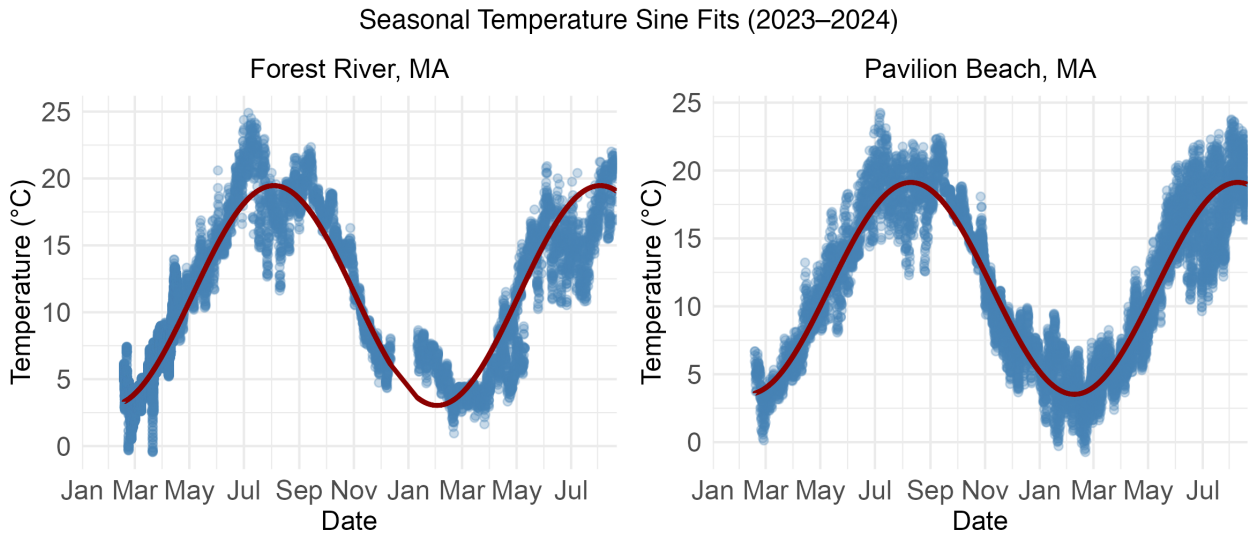


Figure 5: Temperature data from Forest River, MA and Pavilion Beach, MA sites including the sine wave fits. Data were unavailable 12/12/2023 to 1/11/2024 at Forest River, MA due to recording error.

We defined our observation models as follows. The observed counts were modeled using a multinomial likelihood. Let (S_t, E_t, I_t) be the model-implied compartment fractions:

$$\mathbf{y}_t \sim \text{Multinomial}(n_t,\ \mathbf{p}_t)\,, \qquad \mathbf{p}_t \equiv \frac{(S_t, E_t, I_t)}{S_t + E_t + I_t},$$

where $\mathbf{p_t}$ is normalized to sum to one to adjust for small numerical drift in the ODE solver.

For particle concentration (eDNA), let $y_{t,P}$ (copies/mL) be the measurement and P_t the model-implied value, where copies/mL denotes the number of target-sequence copies per milliliter of seawater as inferred from qPCR. Measurements at or below the assay limit of detection (LOD = 0.1, n = 5 from the Forest River site) are treated as left-censored and the likelihood $\ell(y_{t,P})$ is the probability mass below the LOD:

$$\ell(y_{t,P}) = \begin{cases} \text{Lognormal\_pdf}(y_{t,P} \mid \log(P_t),\, \sigma_{\text{eDNA}})\,, & y_{t,P} > \text{LOD}, \\ \text{Lognormal\_cdf}(\text{LOD} \mid \log(P_t),\, \sigma_{\text{eDNA}})\,, & y_{t,P} \le \text{LOD}. \end{cases}$$

where σ_{eDNA} is weakly informed variability ($\sigma_{eDNA} \sim \mathcal{N}(0, 0.35)$).

2.6 Seasonal Decomposition

To provide an analytical comparison of seasonal structure across sites, we applied classical seasonal decomposition using the `decompose` function in R (R Core Team, 2020) to the time series of the Susceptible data trajectory. These results were used as a qualitative diagnostic of site-specific seasonality to compare the timing and structure of these patterns as well as assessing how the model results represented the main seasonal patterns. Since the time series was sparse over the two year period, these results are interpreted cautiously and are used to contextualize the model outcomes among the survey sites.

2.7 Model Fitting and Inference

All of the models were implemented in Stan (Carpenter et al., 2017) using `RStan`, the R interface of Stan. Stan is a probabilistic programming language for Bayesian statistical inference via Markov Chain Monte Carlo (MCMC). The ODE systems were solved with the Runge-Kutta 4th and 5th order (RK45) solver. Four MCMC chains were run for each model, which included 1000 warm-up iterations and an additional 1000 iterations per chain, used for the posterior samples. These samples were then processed in R (R Core Team, 2020). The implementation code is included in the supplementary material.

Divergences occur when the sampler struggles to explore certain regions of the distributions, indicating potential bias and unreliable results. Therefore, we aimed for low divergences of the sampler and further assessed trace plots and $\hat{R}$ values for sampler performance. Across runs for Forest River, $\hat{R}$ values indicated convergence ($\hat{R} < 1.01$) and effective sample sizes were adequate. We observed some divergences in the post-warm-up iterations. Increasing the target acceptance rate (0.99) removed these divergences without altering the posterior summaries. Pavilion Beach had persistent sampling diagnostic issues. While increasing the maximum tree depth (12) to allow the sampler to search longer in each iteration reduced divergences, tail effective sample size warnings remained. This indicates that posterior variances may be less reliable, but the medians and central mass may be more stable. We therefore interpret Pavilion Beach results cautiously.

For parameters and selected derived quantities, we report medians and 95% highest posterior

density intervals (HDIs) calculated with the `HDInterval` package (Juat et al., 2022). HDIs were computed both for prior predictive and posterior draws to summarize inference and to compare prior knowledge with updated beliefs.

To evaluate model fit, we performed posterior predictive checks (PPCs) by generating synthetic trajectories at each MCMC draw using the sampled parameter values. These simulations represent the model-implied dynamics for each compartment (S, E, I, and P). We also generated observation-level predictive trajectories that incorporate sampling error at each observation time. Both process and observation predictions were summarized with their median and 95% HDIs. These PPCs allow visualization of model performance, assessing how the mechanistic process and observation models reproduce the survey data measurements.

2.8 Model Comparison

We compared the four model variants (Table 3) using Pareto smoothed importance sampling with leave-one-out cross validation (PSIS-LOO) to estimate the predictive accuracy of the model without splitting the limited dataset (Vehtari et al., 2017). However, these estimates were indistinguishable across models (Δ ELPD ≤ 1.5 with SE $\approx$ 8-9; all Pareto-k ≤ 1.0). We also assessed fit using posterior predictive checks (PPCs). When observational error was included, PPCs showed similar agreement with the observed data across models (Figure 8; Appendix B), indicating that comparison through predictive performance was also indistinguishable. Therefore, we compared model results by biological interpretation and prioritizing seasonality results through visual comparison. Specifically, we prioritized smaller errors in the timing and amplitude of seasonal dips in susceptible clams (S) and particle concentration (P) compared to observation points. The chosen model best reproduced these seasonal features visually while maintaining clean diagnostics.

2.9 Global Sensitivity Analysis

To determine which parameters are most influential on key disease features and should be focused on for future data collection, we ran a sensitivity analysis based on our current prior knowledge of model parameters. We used the Morris elementary effect method (Morris, 1991)

implemented in Julia with `GlobalSensitivity.jl` (Dixit and Rackauckas, 2022). We analyzed the most developed model (Model 4) and varied the parameters independently and uniformly over their prior 5th - 95th percentiles with a fixed temperature forcing based on site temperature measurements. We computed Morris indices for four features including the peak of P, peak of I, timing of P peak, and timing of I peak to capture seasonality effects of parameters. For each parameter and feature, we summarized the mean absolute elementary effect (μ^*) to rank the overall influence.

2.10 Practical Identifiability

We assessed practical identifiability using parameter-recovery experiments on synthetic datasets, following a simulation-based calibration (SBC) approach (Talts et al., 2020). Specifically, we simulated 100 datasets from the Model 4 temperature-dependent SEIP by randomly drawing parameter values from their prior distributions and generating observations at the empirical sampling times using the same observation model (multinomial SEI counts and a lognormal P measurement), and refit the model under the same priors. Since temperature dynamics are consistent across populations, these values were kept constant among simulations. We then evaluated recoverability by comparing posterior estimates to the true parameter values using the 95% highest-density interval (HDI) coverage, and visualized by comparing posterior predictive latent trajectories to the simulated trajectories generated from the true parameter combinations. Under a well calibrated model, approximately 95% of the true parameter values should fall within the 95% HDIs. Using 100 simulations provides a stable estimate of HDI coverage.

3 Results

Results are presented first for Forest River, where seasonal structure and model behavior are most clearly shown, followed by Pavilion Beach, which is used to assess the generality of the model.

3.1 Seasonality in Site Data

Seasonal decomposition of the survey sites data shows the differing phase and structure between disease prevalence at Forest River and Pavilion Beach sites. While Forest River exhibits a coherent seasonal signal aligned with model-inferred mechanisms, Pavilion Beach shows a seasonal pattern that differs from Forest River in timing and shape. Forest River has two distinct seasonal spikes in the disease with one of smaller magnitude in June and a larger effect in December. Pavilion Beach exhibits one main seasonal disease spike in April. We refer to these seasonal patterns when presenting model results.

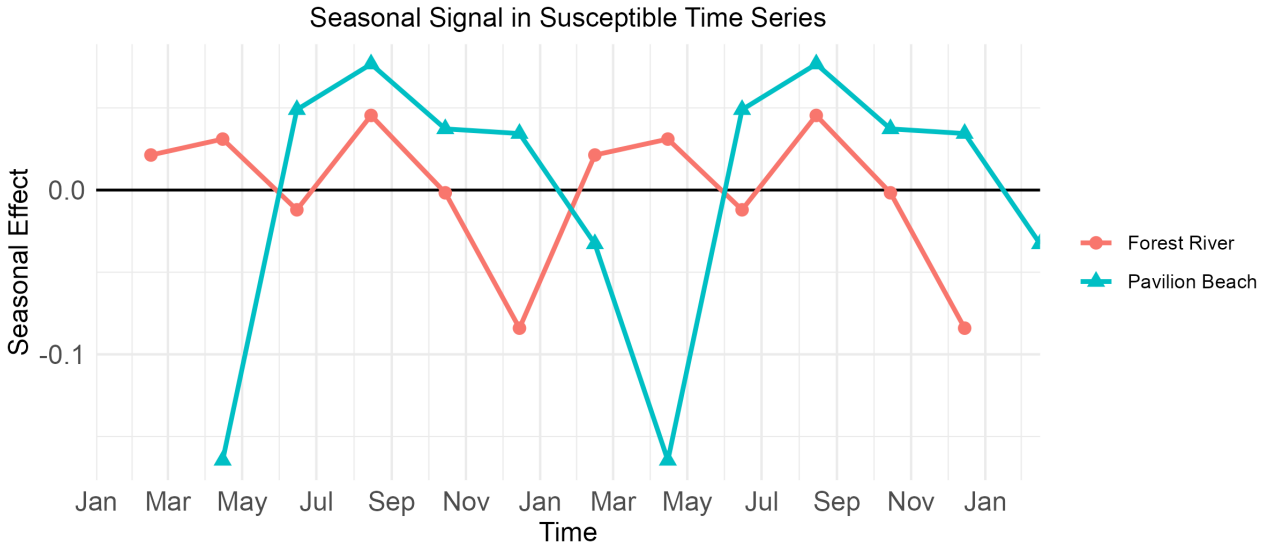


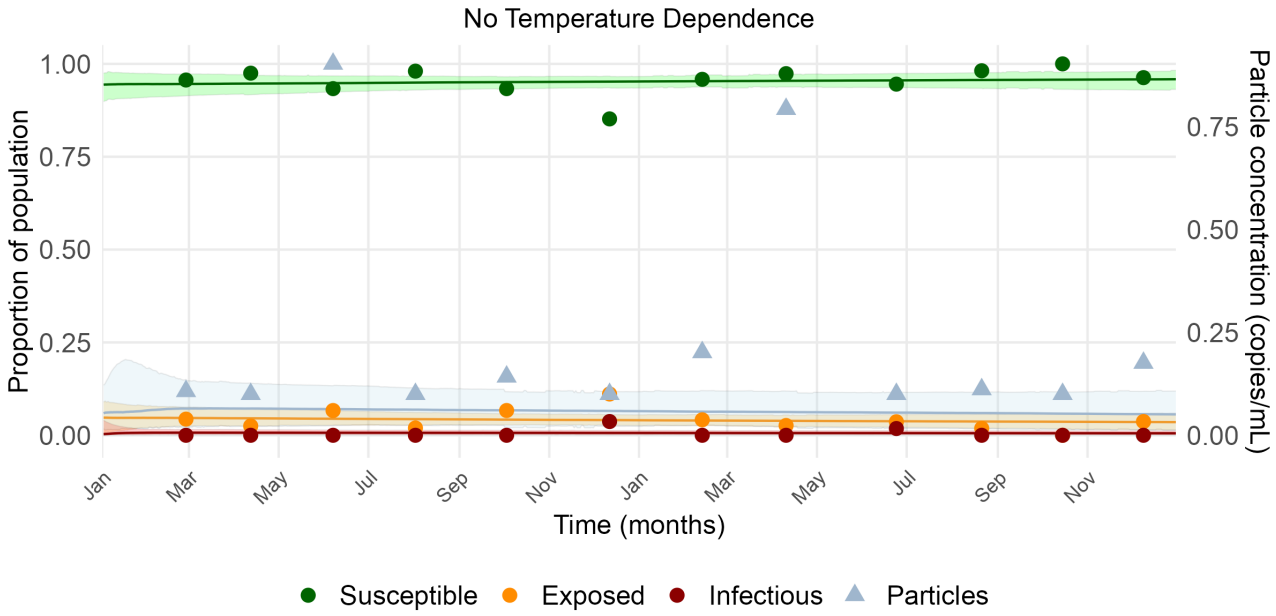
Figure 6: Seasonal components of the susceptible trajectories from Forest River and Pavilion Beach survey data over the two year survey period.

3.2 Model Trajectories

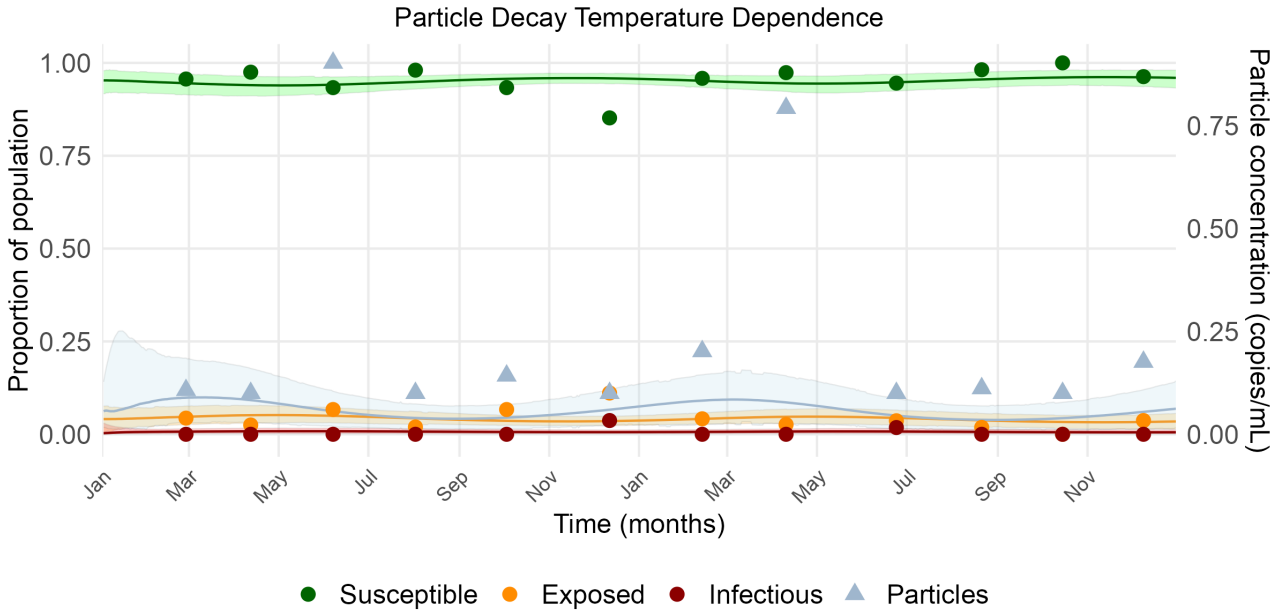
Posterior latent trajectories for the four model variants at the Forest River site are shown in Figure 7. (a) - (d). In each panel, the lines show the posterior median trajectories for each compartment susceptible (S), exposed (E), infectious (I) and particle (P). The bands around each line show the 95% HDI and the points represent the observation points for each compartment (P values at 0.1 indicate measurements at or below the limit of detection). The model variants 1-4 are shown, representing temperature dependencies in each plot. All models broadly capture the overall disease dynamics. The susceptible compartment (S) remains dominant while the exposed (E) and infectious (I) compartments are low. The particle (P) compartment shows

the greatest uncertainty in trajectory estimates. All models show a general increase in S while showing different seasonal patterns as discussed in the following section.

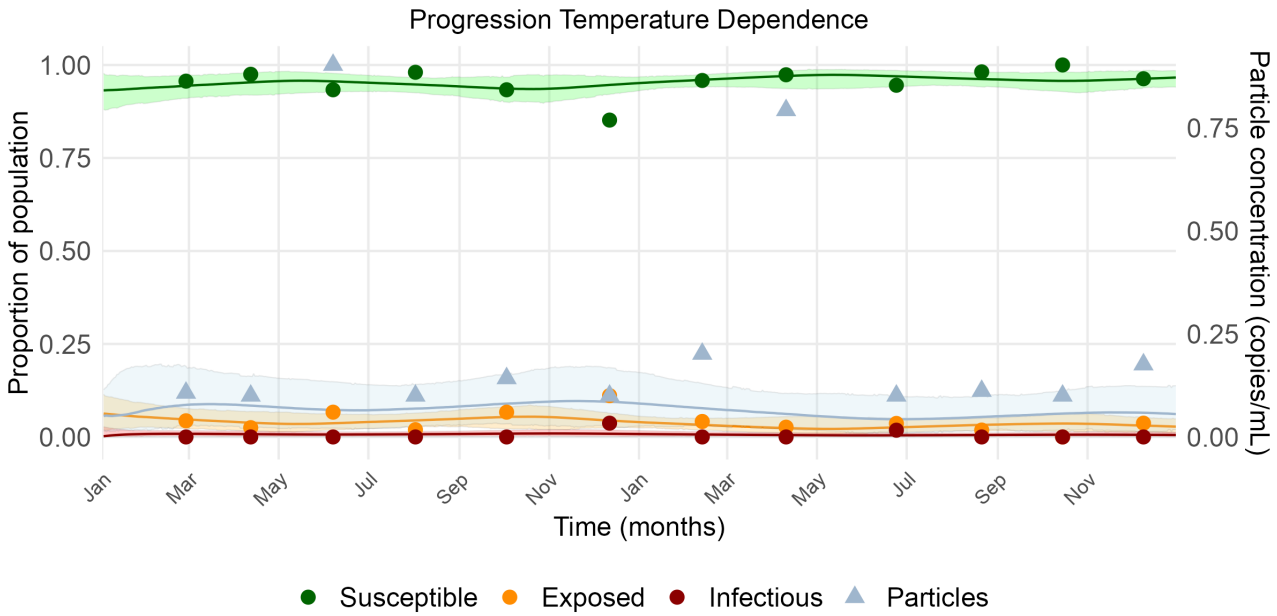
Observational predictive trajectories are shown in Figure 8 for model 4 (and in Appendix B for models 1-3), where simulated observation intervals are compared with the empirical data. Most observations fall within the observational intervals across all model runs.



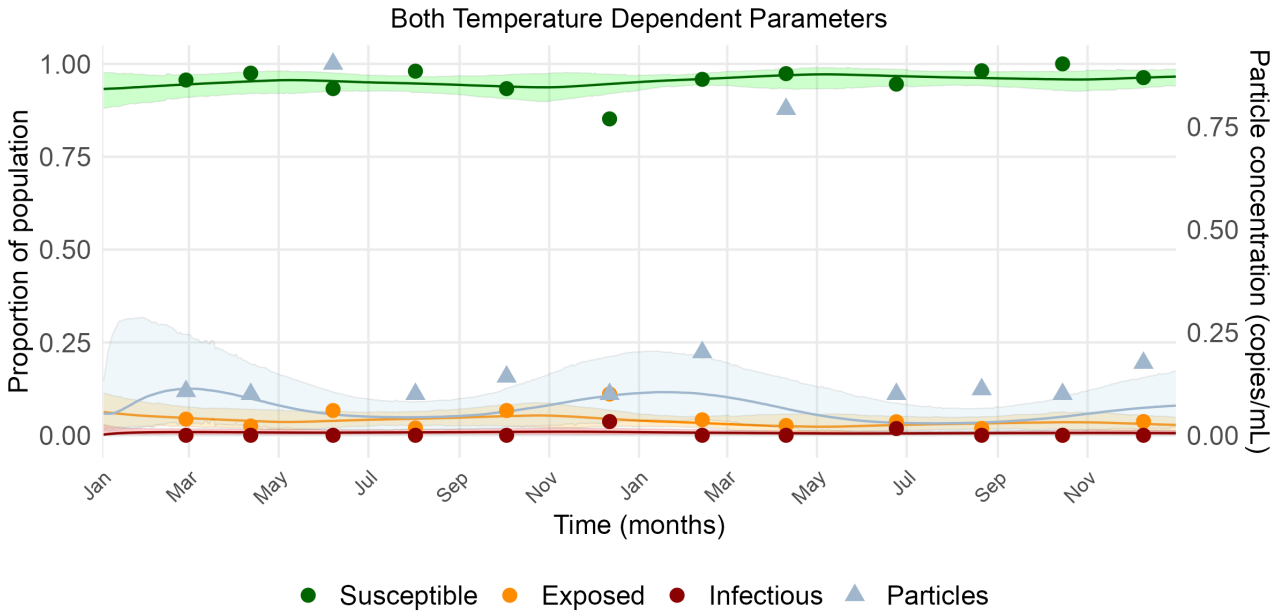
(a) Model 1 results without temperature dependent parameters.



(b) Model 2 results with temperature dependent particle decay rate δ_T.



(c) Model 3 results with temperature dependent progression rate β_T.



(d) Model 4 results with temperature dependent particle decay rate δ_T and progression β_T.

Figure 7: Variant model posterior latent trajectories for the Forest River site. Posterior medians and 95% HDIs are shown for S, E, I, and P with observations overlaid and represented by circles (SEI) and triangles (P). P observations at 0.1 (5 observations) represent points at or below the LOD.

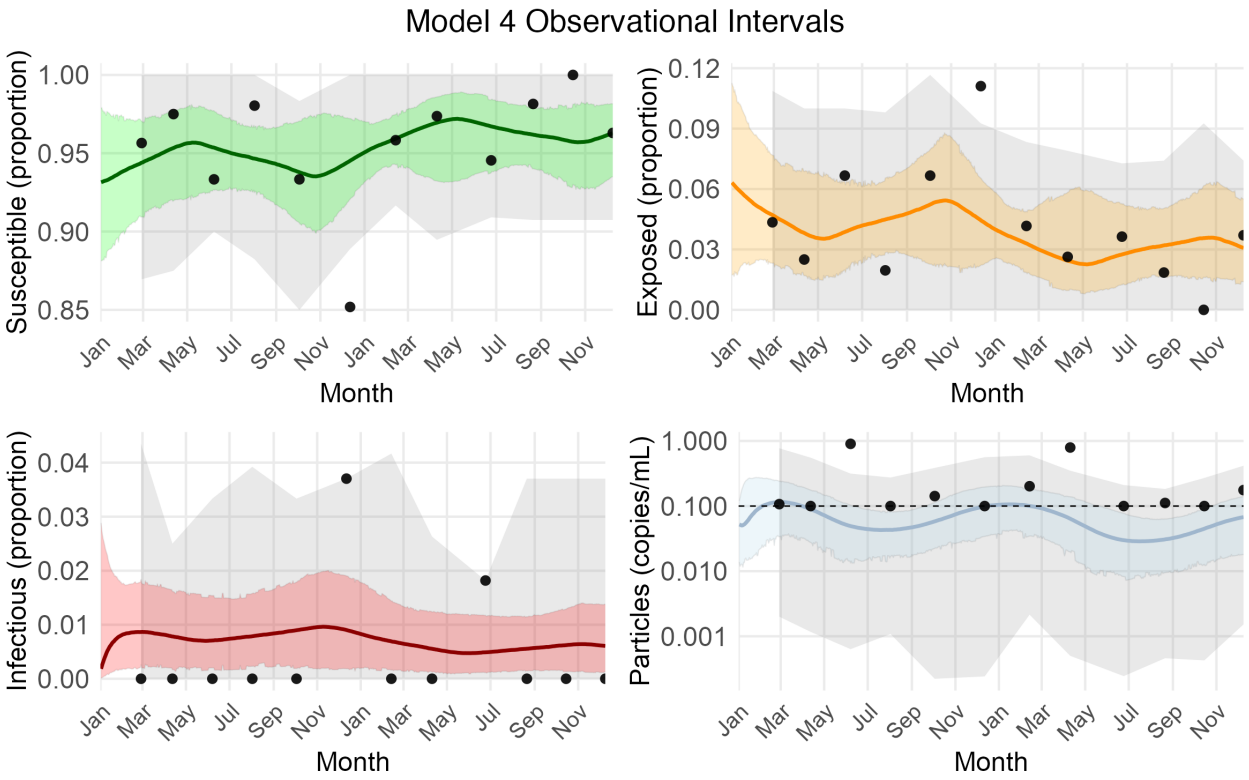


Figure 8: Model 4 trajectories with inner latent trajectory 95% HDI in color (as shown in the previous plots) and outer observational error 95% HDI in gray for each trajectory with survey data points for each compartment. The particle trajectory shows the limit of detection (LOD) value by a dotted line. All particle (P) points at the dotted line (5 observations) represent measurements at the LOD and can therefore be a value between 0 and 0.1.

3.3 Model Comparisons

As noted in section 2.7, PSIS-LOO model comparison was inconclusive and we instead assessed seasonality results of the model trajectories. To directly compare seasonal dynamics, we overlaid the trajectories of all four model variants with the observed data for each compartment and observed site temperature shown visually for the Forest River site (Figure 9).

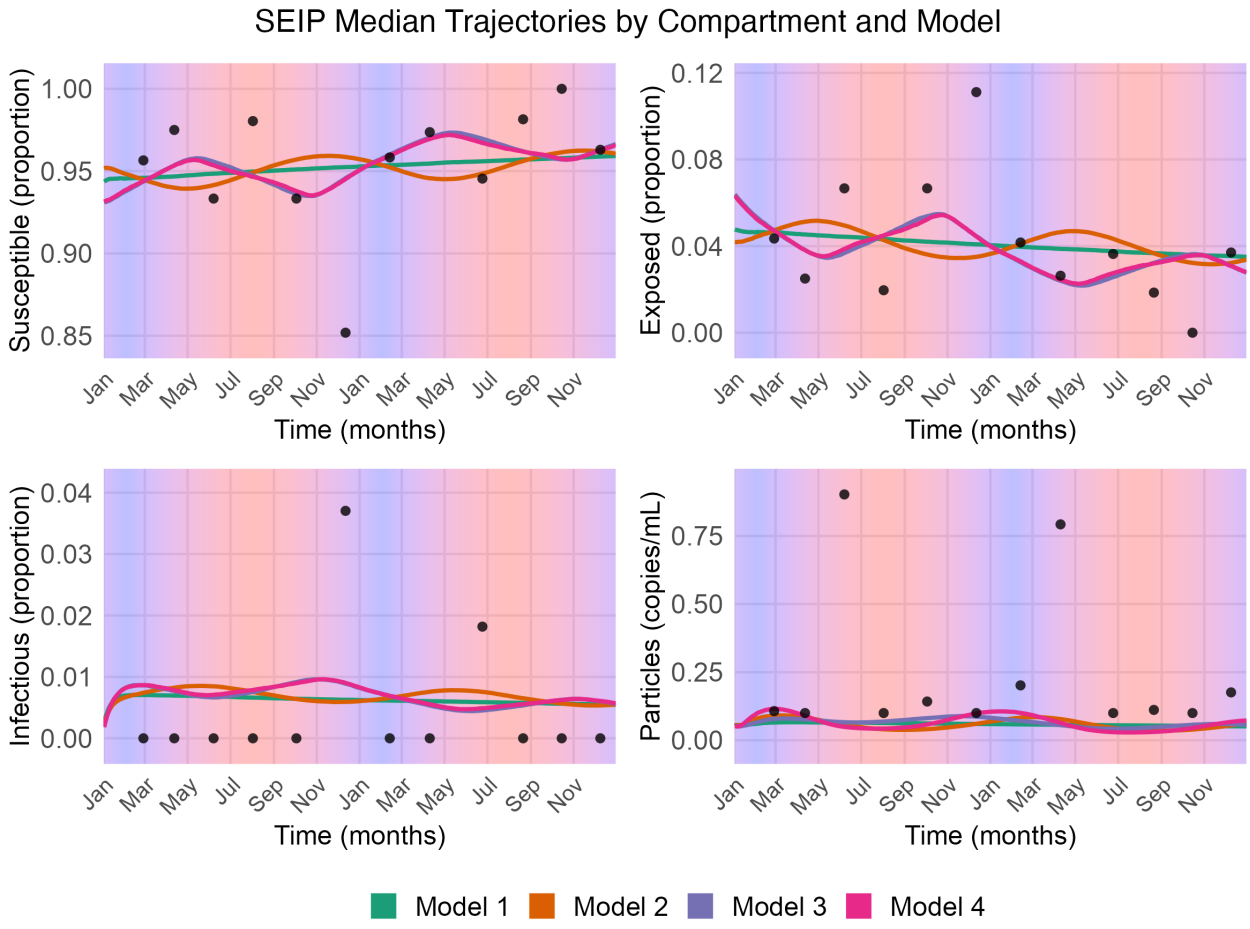


Figure 9: Mean trajectories from models 1-4 with Forest River data, separated by compartment S, E, I and P with a temperature grid for model performance comparisons, where red represents the highest temperature and blue represents the lowest temperature in the observation system. Survey data is represented by the data points. A zoomed in comparison of the trajectories is shown in Figure B.2.

Model 1 (no temperature dependence) produces baseline linear trends with no seasonal structure. Model 2 (temperature dependence in particle decay parameter, δ_T) shows seasonality in the S, E, and I compartments, with peaks in disease prevalence in the colder-to-warmer month transitions and produces minimal seasonal variation in P. Model 3 (temperature dependence in the progression parameter, β_T) and 4 (temperature dependence in both parameters, particle decay and progression) generate larger amplitude seasonal cycles with patterns opposite of model 2. These align more closely with the main observed peak in disease prevalence (Figure 6) in the warmer-to-colder month transitions. Model 3 and 2 produce similar amplitudes in P, though model 3 is shifted to the left, whereas model 4 shows the highest amplitudes in P.

Models 2 and 3 each capture different seasonal signals found within the empirical data. Model 2 reproduces one of the observed annual dips identified in the decomposition (Figure 6), while

model 3 reproduces the other. When both mechanisms are included (model 4), the resulting trajectories strongly resemble model 3, suggesting that, among the mechanisms examined, temperature-dependent progression (β_T) may be more influential on the fitted seasonal dynamics than temperature-dependent cell decay (δ_T).

3.4 Parameter Posteriors

For model 4 of the Forest River site, the comparison of the prior and posterior distribution HDIs for the identifiable parameters α, m, and γ is shown in Figure 10. We show the raw posterior and prior draws in Figure B.3. The posterior distribution of α shifted away from its prior upper bound, consistent with longer latent periods than initially assumed. Its reduced uncertainty relative to the prior demonstrates data-driven learning. The posterior of γ also has a narrowed distribution, with less variability than its broader prior, reflecting data support for a lower emission rate. In contrast, m shifted slightly upward, but retained a similar level of uncertainty, implying weak data influence.

Given that β, a_β, b_β, a_δ and b_δ are unidentifiable in the system (see section 2.2), we instead examine the temperature-dependent derived identifiable parameters β_{eff} and δ_T (Figure 11). As expected by laboratory studies, both exhibit an increasing trend with warmer temperatures. β_{eff} approaches zero in colder months, but increases to about 0.010 at peak temperatures, while δ_T fluctuates between 0.09 and 0.23 throughout the year. The survey data was informative for these quantities, as the posterior intervals are tighter than their defined priors. Notably, although we did not explicitly impose a positive temperature dependence for β_{eff}, the posterior distribution is consistent with a positive relationship.

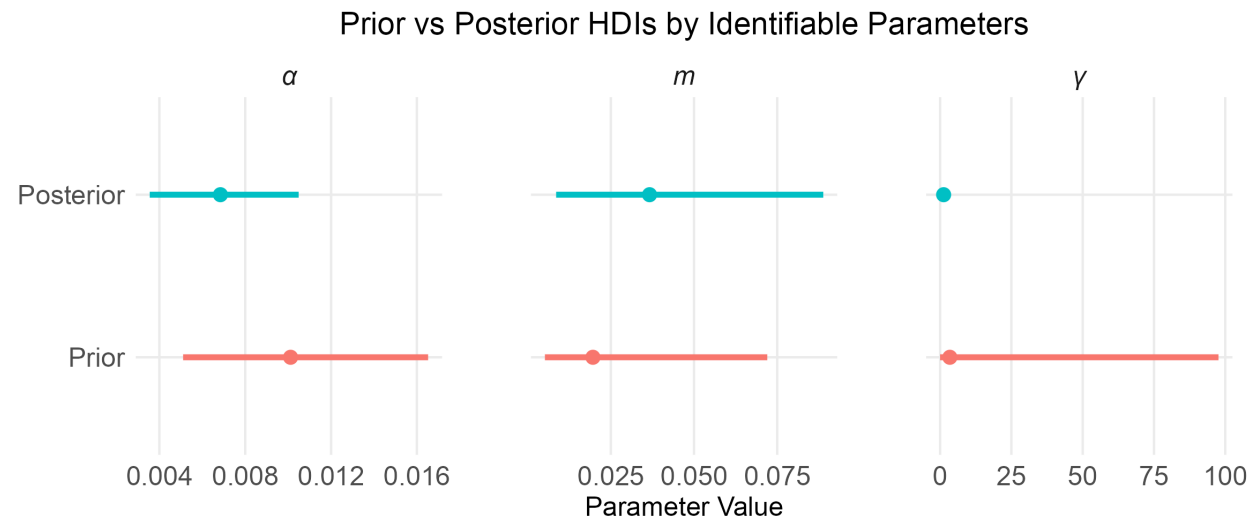


Figure 10: Posterior and prior distributions for identifiable model parameters, α (time to I from E), m (time to death from I), and γ (particle emission from I) with median values and 95% HDI intervals from model 4 results.

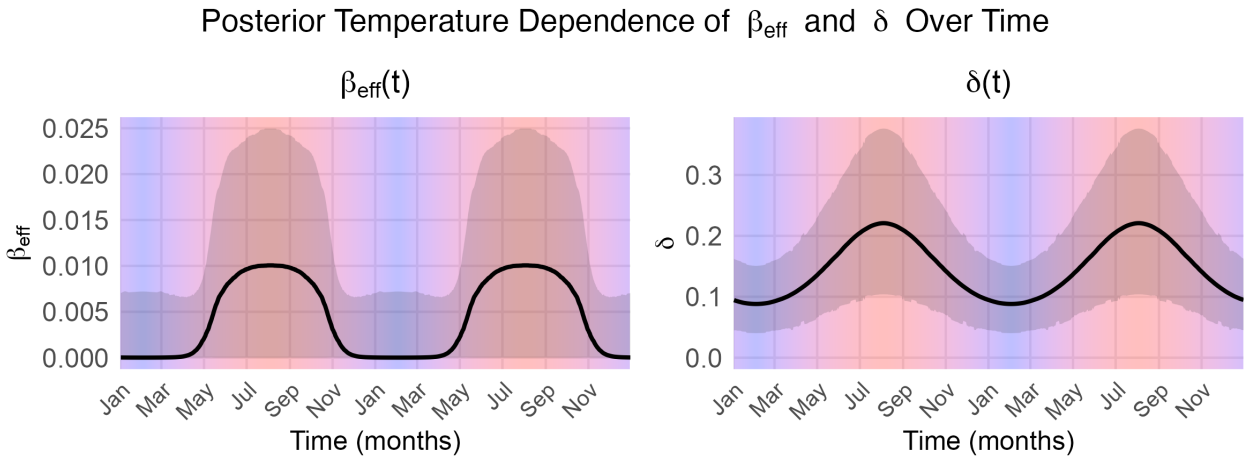


Figure 11: Posterior draws of $\beta_{\text{eff}}(t)$ and $\delta_T(t)$ with median and 95% trajectories against survey temperature grid from model 4 results.

3.5 Pavilion Beach

In contrast to Forest River, at Pavilion Beach (Figure 12) all four model variants produce broadly similar fits to the SEI observations, so the data do not strongly distinguish $\beta_{\text{eff}}(t)$ and $\delta_T(t)$ as a seasonal driver. Also, the inferred temperature effect on progression ($\beta_{\text{eff}}(t)$) in model 3 was not consistently positive across posterior draws. Therefore, we treat Pavilion Beach as a diagnostic site for model generality rather than a primary site for seasonal mechanism inference.

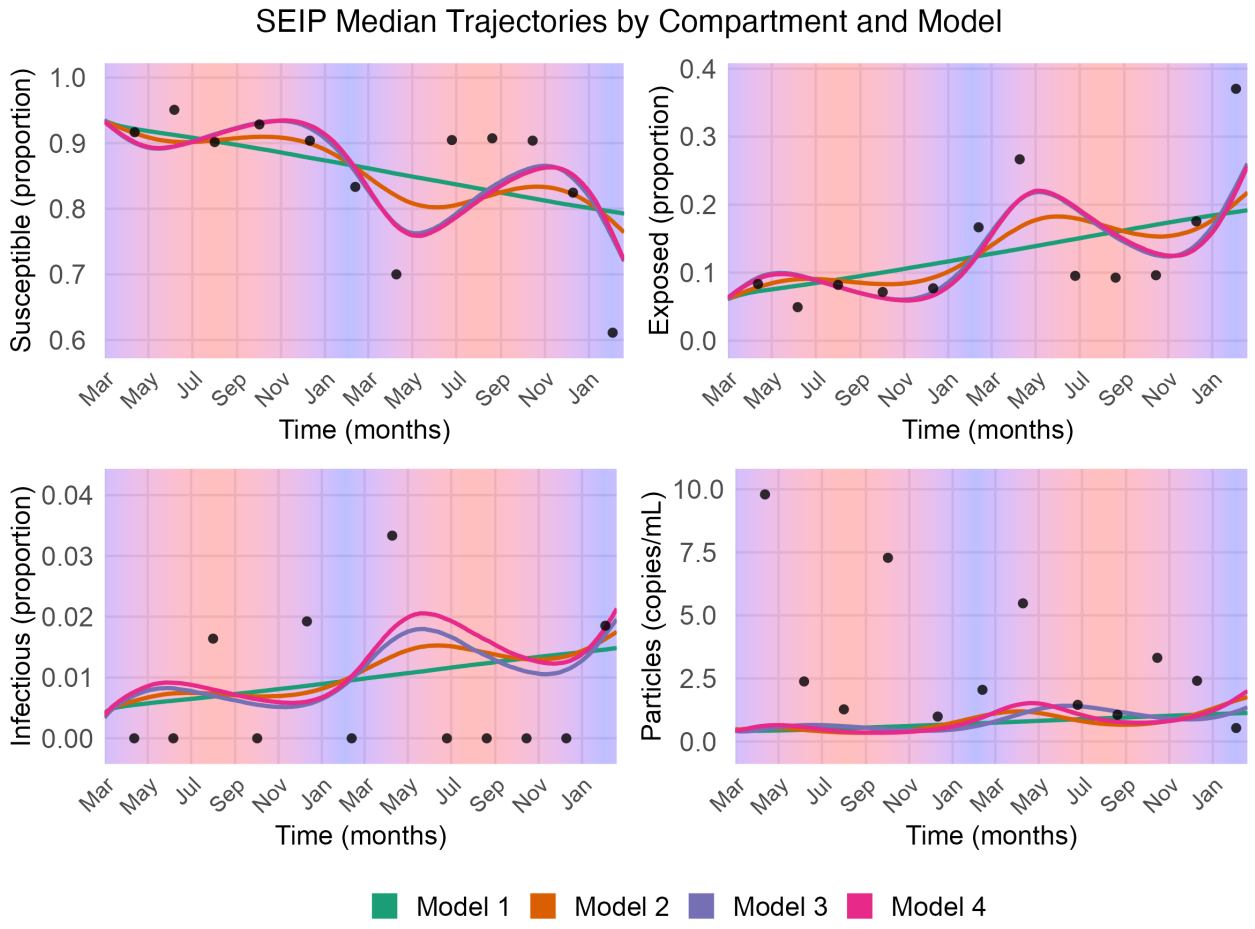


Figure 12: Mean trajectories from models 1-4 (lines) and Pavilion Beach data (points) separated by compartment S, E, I and P with a temperature grid for model performance comparisons, where red represents the highest temperature and blue represents the lowest temperature in the observation system.

3.6 Parameter Sensitivity

The sensitivity results (Figure 13) show that the most influential parameters on the magnitude and timing of I and P trajectory peaks include α and m. γ is the least influential parameter. The temperature effect on β and δ are of equal importance in timing of P and peak of I, but b_δ is slightly more influential on the amplitude of the P trajectory peak and the timing of the I trajectory peak.

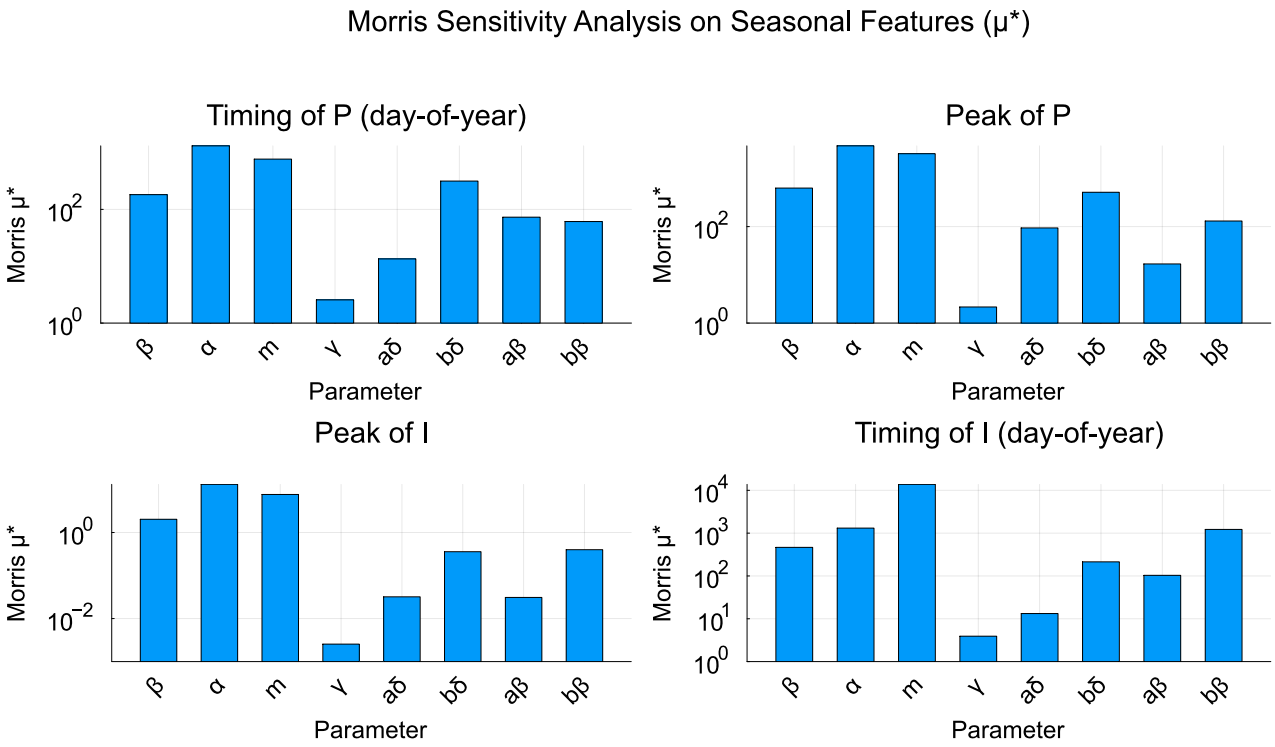


Figure 13: Morris sensitivity analysis with μ^* ranking for each parameter and four seasonal features of model results including the timing and amplitude of the peaks of I and P trajectories.

3.7 Parameter Recovery

Across 100 simulated datasets, the model's posterior intervals captured the true parameter values at rates close to their predefined levels, confirming good overall calibration. Coverage at the 95% credible level ranged from 0.80 for the progression rate (α) to 0.96 for the mortality rate (m) and mean decay rate ($\bar{\delta_T}$). The mean effective transmission rate ($\bar{\beta_{\text{eff}}}$) and emission rate (γ), were 0.93 and 0.95, respectively. The lower performing results for α suggests that this parameter remains weakly informed by the data under the current prior. In contrast, $\bar{\delta_T}$, $\bar{\beta_{\text{eff}}}$, γ, and m showed accurate performance, indicating that their priors and data likelihood provide sufficient information for accurate recovery of true values. Overall, these simulation-based calibration results demonstrate that the model performs well, with only mild under-coverage in the least identifiable parameter. Two of these simulations and model fits are shown in Figure 14 to visualize the latent recoverability performance of the model on data similar to the survey data.

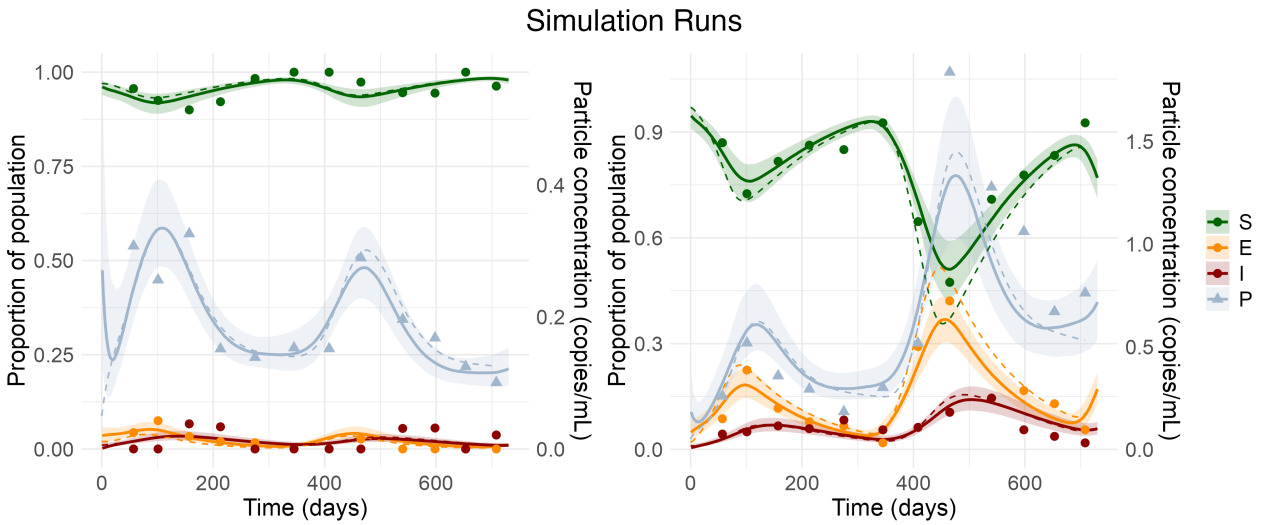


Figure 14: Two example simulations representing trajectories from randomly selected combinations of model parameters in the dotted lines and median model fits in the solid lines with the 95% HDI. The observed data points used to fit the model were points taken from the simulated trajectory with added sampling noise and are also shown on the plots.

4 Discussion

The results of this study not only provide a foundation for understanding the dynamics of Bivalve Transmissible Neoplasia and its effect on bivalve populations, but are also an example of the application of Bayesian inference to dynamic biological systems.

4.1 Site-Specific Characteristics

Forest River and Pavilion Beach display different seasonal signatures in both prevalence and particle measurements, and these site-level differences may contribute to the contrast in model behavior across sites. At Forest River, the survey time series shows two within-year peaks; the dominant seasonal peak is reproduced by the temperature-linked progression mechanism, while the smaller secondary peak is captured less consistently. At Pavilion Beach, the prevalence time series shows a single prominent spring spike (largely driven by the April observation), whereas the eDNA record exhibits multiple within-year pulses. This mismatch in structure suggests that additional site-specific processes affecting particle dynamics and/or detection (e.g., particle transport, episodic release, or unmodeled biological events) may be important at Pavilion Beach beyond what is represented in the current SEIP framework. Additionally, individual observations can have high leverage on inferred seasonal amplitude and timing due

to the sampling resolution, thus increased sampling density around expected peak windows would help determine whether the Pavilion Beach spring spike reflects a consistent seasonal feature or sensitivity to specific sampling dates.

4.2 Seasonality of Forest River Results

The results shown in Figure 9, are consistent with the hypothesis that incorporating temperature dependence improves the model's adaptability to reproduce observed seasonal patterns in the system. Allowing disease progression and particle decay to vary with temperature appeared to better reproduce the observed seasonal trends than the constant-parameter model. The model suggested that higher temperatures were associated with faster disease progression and faster particle decay. Despite this opposing dynamic, the inclusion of both effects produced the closest match to the main observed seasonal pattern (December peak). While dynamics in S, E, and I trajectories were nearly identical in Models 3 and 4 (progression temperature dependence and both temperature dependent parameters, respectively), the trajectory of P in Model 4 results aligned closer with observed patterns. Therefore, including both temperature dependent parameter dynamics produced the most plausible fit among the four proposed models. However, the results also suggest that there is a dominant mechanism of seasonality of the two proposed in this study. Since Model 4 results mostly aligned with Model 3 results, we note that temperature-dependent progression appeared to be the more influential seasonal mechanism among the two considered here of observed seasonal patterns in S, E, and I proportions, while decay rate variation refined the timing and amplitude of the particle concentration in the full temperature model.

While there is high uncertainty in estimates for magnitudes of seasonal parameters β_{eff} and δ_T, the fitted temperature relationships were positive over the observed temperature range as seen in Figure 11. Therefore, under this model, higher water temperature is associated with higher predicted BTN prevalence at this site. We performed a scenario analysis using model 4 parameter estimates and mechanistic structure (Figure 15) in which the site's seasonal temperature cycle was shifted upward by 1 to 3 $^{\circ}C$, consistent with projections of 1.1 to 2.4 $^{\circ}C$ increases in the year 2050 for sea surface temperatures in the Gulf of Maine where this site is located (Brickman et al., 2021).

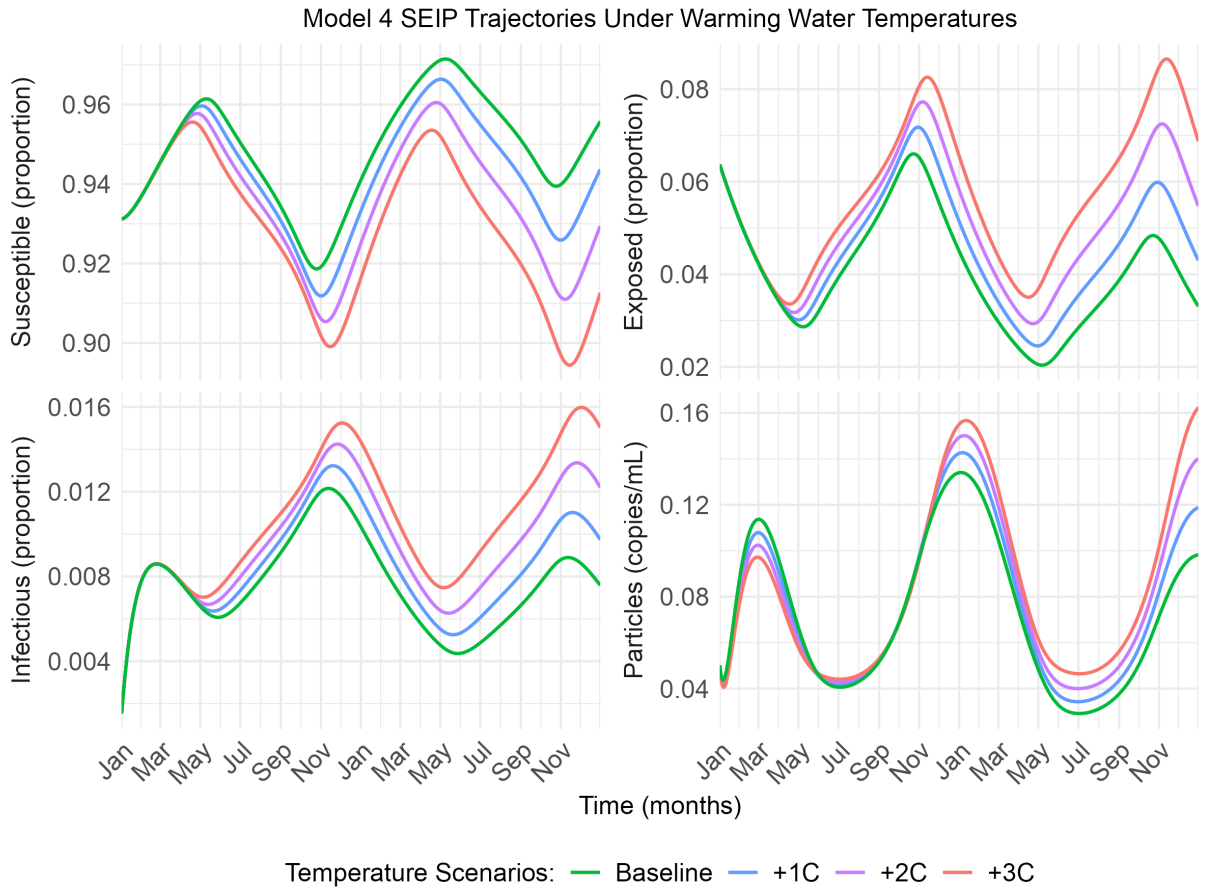


Figure 15: S, E, I, and P trajectories from model 4 parameter inference median values under 4 temperature scenarios: baseline, $+1°C$, $+2°C$, $+3°C$ mean annual temperature of the Forest River, MA site.

While these results suggest an important role for temperature in seasonal dynamics, the global sensitivity analysis (Figure 13) further clarifies which parameters most strongly influence model behavior. The analysis confirms the importance of temperature-dependent effects, yet shows that the progression variables (β, α, and m) have the most influence on seasonal attributes and may require prioritization in future laboratory experiments. Additionally, the sensitivity analysis demonstrates that while there appears to be significant variability in laboratory estimates of particle emission rate (γ), its effect on seasonal variation appeared to be minimal when compared to other factors, and is constrained by survey data.

4.3 Data Limitations

Simulation-based validation suggested that the underlying population trajectories (S, E, I, and P) can often be recovered and that most parameters were well estimated under the assumed data

generating process. Overall, most parameters were informed by the survey data, which reduced uncertainty. However, the available data (laboratory and field data) impose several limitations on model resolution and inference. Progression parameters are difficult to estimate in lab environments, but more information on β would help refine the estimates of the magnitude of the temperature effect (a_β and b_β). When eDNA observations are at the LOD, these values are treated as left-censored. This censoring reduces information about low particle concentrations and propagates as wider observational predictive intervals for P. While the model explicitly accounts for this measurement process, the censoring limits our ability to precisely estimate parameters governing decay and low-level transmission. Similarly, the low frequency of sampling restricts the ability to resolve sharp seasonal transitions in disease progression. Infectious clam samples are particularly sparse, which may weaken estimation of seasonal forcing and temperature-dependent effects.

The posterior predictive intervals in Figure 8 suggest that the model captures most of the observational error and measurement variability, especially in the eDNA particle data. While the simulated trajectories capture the general temporal patterns, there remains broad variation in observed values that cannot be fully explained by the mechanistic structure. This suggests that biological or environmental processes not included in the current model, such as unobserved covariates, may contribute to the observed spread. Nevertheless, by explicitly incorporating observation error, the model appropriately propagates this uncertainty into parameter estimates, providing inference that reflects uncertainty despite noisy and incomplete data.

4.4 Model Extension

The survey data captures a repeated seasonal observation of increased particle concentration in the summer months that also coincides with a jump in infected counts. However, this dynamic was not captured in any model results leading us to hypothesize a mechanistic element not included in our model formulations.

We explored an extended version of the previously described seasonal model with a hypothesized additional mechanism in order to improve the fit to the data. In particular, we propose that the spike in the particle concentrations during the summer months may have resulted from biological

events such as spawning. Since the timing of spawning in clam populations is uncertain, but generally occurs between May and August (Brousseau, 1978), we assigned the onset day t_{start} a uniform prior on [140, 174] with day 0 defined as January 1. The duration d_{pulse} has a lognormal prior centered near 10 days (truncated to 3-20 days). The pulse is implemented as a smooth introduction using the logistic function to avoid discontinuities that would destabilize Stan's ODE solver. Specifically,

$$\text{pulse}(t) = logit^{-1}(k(t - t_{\text{start}})) - logit^{-1}(k(t - [t_{\text{start}} + d_{\text{pulse}}])$$

where k is a steepness constant selected as a biologically plausible pulse in the time frame while ensuring numerical stability. This is scaled by a fixed amplitude of 0.25 and added to the equation describing how P changes over time:

$$\frac{dP}{dt} = \gamma I - \delta_T P + 0.25 \cdot \text{pulse}(t)$$

The amplitude was chosen to generate particle peaks comparable in magnitude to those seen in the eDNA data. Since the goal of this spawning pulse was to test whether adding this mechanism could generate the observed seasonal dip seen in the Forest River site rather than to estimate a biologically precise pulse magnitude, we treated the amplitude as a fixed scaling rather than a parameter to be estimated. We also fixed all other parameter values to median parameter inferences based on Forest River model 4 results. Therefore, in this model extension, the only parameters that were inferred were the timing and duration of the particle pulse.

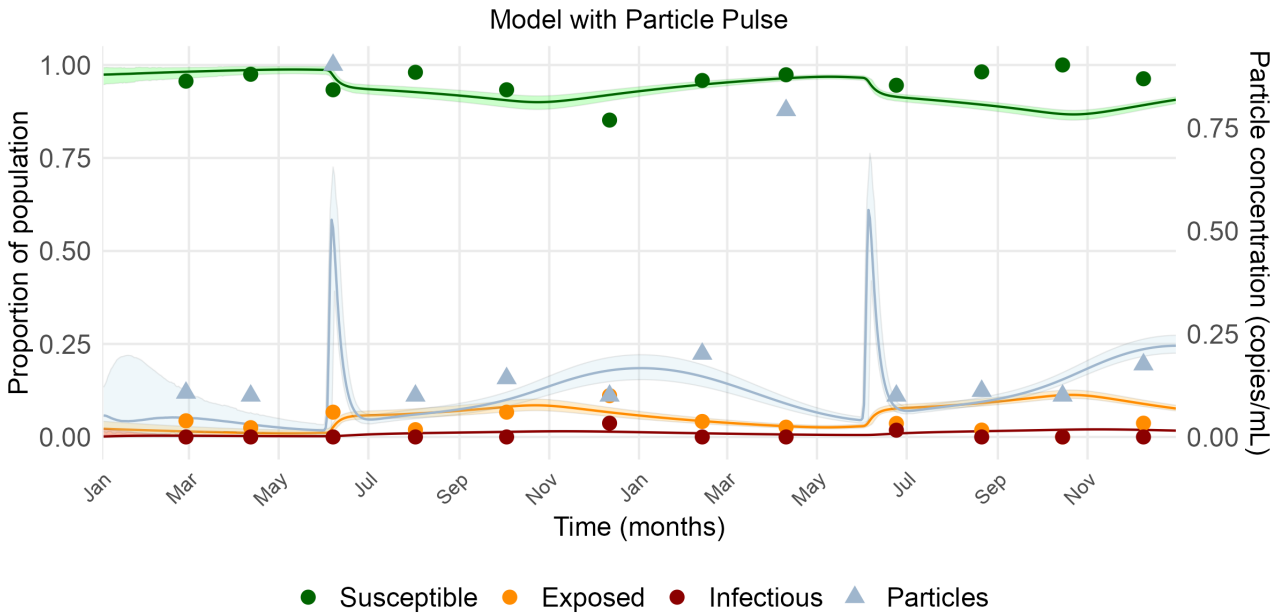


Figure 16: Adjusted Forest River model 4 results with a particle concentration pulse introduced in the summer months. Median trajectories and their respective 95% HDI are shown with data points for each compartment S, E, I, and P.

In the model extension, the added particle pulse produced a corresponding increase in the exposed population and helped reproduce the observed summer increase more closely than model 4. The values used in this extension are based on limitedly available knowledge and not well informed, therefore further work on spawning parameters is needed to make conclusive statements of this observed spike in cancer cells. This extension motivates further exploration of the hypothesis that spawning events produce more cancer particles in the water and thus, future data collection should be targeted to accurately monitor the particle concentration in the water during this time.

4.5 Future Directions

While the work presented captures key aspects of Bivalve Transmissible Cancer in clams, several biological and environmental processes require further investigation. Future research is needed to examine the role of spawning as a potential driver of key seasonal observations. Another factor that can be implemented to the model mechanism is tidal dynamics of sampling locations, which may influence cancer particle detection in sampling and impact disease transmission. An increase number of laboratory experiments investigating temperature effects are also needed to refine estimates of disease progression rates and the thermal sensitivity of the cancer spread.

Increase sampling resolution during summer months would better constrain temperature effects and validate pulse dynamics. In the field, increasing sampling frequency during summer months would help constrain temperature-dependent effects and increase the resolution of the particle pulse dynamics. Finally, extending surveys to additional sites spanning broader temperature gradients would help evaluate the model at additional sites and improve the generalization of results across populations. Field data collection continued at the sample sites through 2025, which will provide a validation dataset for model predictions of the disease and would be an important next step in this research.

More broadly, this modeling framework can also be extended to other disease systems in which pathogens are released into the environment, remain outside of the host for some period of time, and can be measured in environmental sampling. A useful feature of this model is that it links environmental pathogen measurements directly to transmission dynamics through an explicit environmental compartment, rather than treating those measurements as a separate value. This allows an in depth understanding of the system as a whole and how shedding, persistence, and removal of pathogen material in the environment may affect disease prevalence over time. Similar systems exist in which pathogens are shed into the water and environmental concentrations indicate exposure risk and shedding amount. For example, disease progression in zebrafish infected with *Vibrio anguillarum* has been investigated with eDNA, and eDNA experiments on white spot disease in shrimp have quantified viral shedding from infected hosts (Henard et al., 2024; Cox et al., 2023). These examples highlight systems that can be further explored with the mechanistic model used in this study due to their environmental pathogen measurements to improve inference on disease dynamics.

5 Conclusion

This study provides a mechanistic model of BTN dynamics in a natural population. By synthesizing laboratory experiments, environmental particle measurements, and field prevalence surveys, the model estimates progression and transmission parameters and their uncertainties. This model also supports the hypothesis that temperature-dependent progression contributes to seasonality in the disease prevalence observed in survey data of Forest River, MA popula-

tions. These results suggest that temperature-dependent progression is the dominant seasonal mechanism among the model variants considered, particularly explaining the fall disease spikes at Forest River. The thermo sensitive cancer cell decay rate adjusts the timing of disease infections and particle concentrations, but did not appear to be the primary seasonal mechanism in the Forest River fits.

This work provides a basis for exploring BTN dynamics under changing environmental conditions. As BTN continues to affect populations, this model may be useful to capture underlying dynamics and can be updated with new information of the disease. As shown through the model extension of including another mechanism to the disease model, this framework can be used to explore hypotheses and inform future data collection of the disease system such as the link between spawning events and an increased measurement of BTN in the water column. This modeling approach also contributes to the expanding use of Bayesian methods in disease modeling. We demonstrated that Bayesian modeling can be useful for evaluating plausible ecological mechanisms even with sparse, noisy, or limited data.

These results also highlight several directions for future work. Higher-frequency summer sampling and targeted laboratory experiments are likely needed to better refine parameter estimates and determine the role of spawning or other drivers of observed seasonality. Extending surveys across broader temperature gradients and incorporating additional environmental processes such as tidal dynamics will improve our assessment of the reliability and generality of the model. Together, these adjustments would build on this established model to deepen understanding of BTN transmission and improve predictions for bivalve populations.

As environmental DNA surveillance is increasingly popularized in aquatic systems, models that explicitly connect environmental pathogen signals to host-pathogen dynamics are likely to become increasingly valuable for disease monitoring, hypothesis testing, and predictions. In particular, models that represent pathogen persistence in the environment and account for processes affecting pathogen release, survival, and removal may be especially valuable in aquatic disease systems.

References

Aalto, E. A., Lafferty, K. D., Sokolow, S. H., Grewelle, R. E., Ben-Horin, T., Boch, C. A., Raimondi, P. T., Bograd, S. J., Hazen, E. L., Jacox, M. G., Micheli, F., and De Leo, G. A. (2020). Models with environmental drivers offer a plausible mechanism for the rapid spread of infectious disease outbreaks in marine organisms. *Scientific Reports*, 10:5975.

Anderson, D. and Watson, R. (1980). On the Spread of a Disease with Gamma Distributed Latent and Infectious Periods. *Biometrika*, 67(1):191–198. Publisher: [Oxford University Press, Biometrika Trust].

Bidegain, G., Powell, E. N., Klinck, J. M., Ben-Horin, T., and Hofmann, E. E. (2016). Marine infectious disease dynamics and outbreak thresholds: contact transmission, pandemic infection, and the potential role of filter feeders. *Ecosphere*, 7(4):e01286. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/ecs2.1286.

Brickman, D., Alexander, M. A., Pershing, A., Scott, J. D., and Wang, Z. (2021). Projections of physical conditions in the Gulf of Maine in 2050. *Elementa: Science of the Anthropocene*, 9(1):00055.

Brooks, M. E., Kristensen, K., Benthem, K. J. v., Magnusson, A., Berg, C. W., Nielsen, A., Skaug, H. J., Mächler, M., and Bolker, B. M. (2017). glmmTMB Balances Speed and Flexibility Among Packages for Zero-inflated Generalized Linear Mixed Modeling. *The R Journal*, 9(2):378–400.

Brousseau, D. J. (1978). Population dynamics of the soft-shell clam Mya arenaria. *Marine Biology*, 50(1):63–71.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., and Riddell, A. (2017). *Stan* : A Probabilistic Programming Language. *Journal of Statistical Software*, 76(1).

Cook, J. D., Williams, D. M., Walsh, D. P., and Hefley, T. J. (2023). Bayesian forecasting of disease spread with little or no local data. *Scientific Reports*, 13(1):8137. Publisher: Nature Publishing Group.

Cox, N., De Swaef, E., Corteel, M., Van Den Broeck, W., Bossier, P., Dantas-Lima, J. J., and Nauwynck, H. J. (2023). The Way of Water: Unravelling White Spot Syndrome Virus (WSSV) Transmission Dynamics in Litopenaeus vannamei Shrimp. *Viruses*, 15(9):1824.

Dixit, V. K. and Rackauckas, C. (2022). GlobalSensitivity. jl: Performant and Parallel Global Sensitivity Analysis with Julia. *Journal of Open Source Software*, 7(76):4561.

Dong, R., Goodbrake, C., Harrington, H., and G., P. (2023). Differential elimination for dynamical models via projections with applications to structural identifiability. *SIAM Journal on Applied Algebra and Geometry*, 7(1):194–235.

Gelman, A., Vehtari, A., Simpson, D., Margossian, C. C., Carpenter, B., Yao, Y., Kennedy, L., Gabry, J., Bürkner, P.-C., and Modrák, M. (2020). Bayesian Workflow. arXiv:2011.01808 [stat].

Giersch, R. M., Hart, S. F. M., Reddy, S. G., Yonemitsu, M. A., Orellana Rosales, M. J., Korn, M., Geleta, B. M., Countway, P. D., Fernández Robledo, J. A., and Metzger, M. J. (2022).

Survival and Detection of Bivalve Transmissible Neoplasia from the Soft-Shell Clam Mya arenaria (MarBTN) in Seawater. *Pathogens*, 11(3):283. Number: 3 Publisher: Multidisciplinary Digital Publishing Institute.

Giersch, R. M., Sevigny, J. K., Weinandt, S. A., Mayo, C., Garrett, F. E. S., Tindbaek, K., Yonemitsu, M. A., Hart, S. F. M., and Metzger, M. J. (2025). Variation in natural infection outcomes and cancer cell release from soft-shell clams (Mya arenaria) with bivalve transmissible neoplasia. *PLOS Pathogens*, 21(9):e1013537. Publisher: Public Library of Science.

Grinsztajn, L., Semenova, E., Margossian, C. C., and Riou, J. (2021). Bayesian workflow for disease transmission modeling in Stan. *Statistics in Medicine*, 40(27):6209–6234.

Groner, M. L., Maynard, J., Breyta, R., Carnegie, R. B., Dobson, A., Friedman, C. S., Froelich, B., Garren, M., Gulland, F. M. D., Heron, S. F., Noble, R. T., Revie, C. W., Shields, J. D., Vanderstichel, R., Weil, E., Wyllie-Echeverria, S., and Harvell, C. D. (2016). Managing marine disease emergencies in an era of rapid change. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371(1689):20150364. Publisher: Royal Society.

Henard, C., Blonç, M., Li, H., Duan, Y., and Jørgensen, L. v. G. (2024). Insights into the connection between pathogen water eDNA and disease progression in zebrafish infected with Vibrio anguillarum. *Comparative Immunology Reports*, 7:200160.

Juat, N., Meredith, M., and Kruschke, J. (2022). HDInterval: Highest (Posterior) Density Intervals.

Killick, R. and Eckley, I. A. (2014). changepoint: An R Package for Changepoint Analysis. *Journal of Statistical Software*, 58(3):1 – 19.

Lafferty, K. D. (2017). Marine Infectious Disease Ecology. *Annual Review of Ecology, Evolution, and Systematics*, 48(Volume 48, 2017):473–496. Publisher: Annual Reviews.

Leavitt, D. F., McDowell Capuzzo, J., Smolowitz, R. M., Miosky, D. L., Lancaster, B. A., and Reinisch, C. L. (1990). Hematopoietic neoplasia inMya arenaria: Prevalence and indices of physiological condition. *Marine Biology*, 105(2):313–321.

McCallum, H. I., Kuris, A., Harvell, C. D., Lafferty, K. D., Smith, G. W., and Porter, J. (2004). Does terrestrial epidemiology apply to marine systems? *Trends in Ecology & Evolution*, 19(11):585–591. Publisher: Elsevier.

Metzger, M. J. and Goff, S. P. (2016). A Sixth Modality of Infectious Disease: Contagious Cancer from Devils to Clams and Beyond. *PLoS Pathogens*, 12(10):e1005904.

Metzger, M. J., Reinisch, C., Sherry, J., and Goff, S. P. (2015). Horizontal Transmission of Clonal Cancer Cells Causes Leukemia in Soft-Shell Clams. *Cell*, 161(2):255–263. Publisher: Elsevier.

Morris, M. D. (1991). Factorial Sampling Plans for Preliminary Computational Experiments. *Technometrics*, 33(2):161–174. Publisher: ASA Website _eprint: https://www.tandfonline.com/doi/pdf/10.1080/00401706.1991.10484804.

Murgia, C., Pritchard, J. K., Kim, S. Y., Fassati, A., and Weiss, R. A. (2006). Clonal origin and evolution of a transmissible cancer. *Cell*, 126(3):477–487.

Pearse, A.-M. and Swift, K. (2006). Allograft theory: transmission of devil facial-tumour disease. *Nature*, 439(7076):549.

R Core Team (2020). R: The R Project for Statistical Computing.

Schad, D. J., Betancourt, M., and Vasishth, S. (2021). Toward a principled Bayesian workflow in cognitive science. *Psychological Methods*, 26(1):103–126. Place: US Publisher: American Psychological Association.

Talts, S., Betancourt, M., Simpson, D., Vehtari, A., and Gelman, A. (2020). Validating Bayesian Inference Algorithms with Simulation-Based Calibration. arXiv:1804.06788 [stat].

Tien, J. H. and Earn, D. J. D. (2010). Multiple Transmission Pathways and Disease Dynamics in a Waterborne Pathogen Model. *Bulletin of Mathematical Biology*, 72(6):1506–1533.

Tillett, H. E. (1992). Infectious Diseases of Humans: Dynamics and Control. R. M. Anderson, R. M. May, Pp. 757. Oxford University Press; 1991 (£50.00). *Epidemiology & Infection*, 108(1):211–211.

Tracy, A. M., Pielmeier, M. L., Yoshioka, R. M., Heron, S. F., and Harvell, C. D. (2019). Increases and decreases in marine disease reports in an era of global change. *Proceedings of the Royal Society B*. Publisher: The Royal Society.

Turner, K. G., Smith, M. J., and Ridenhour, B. J. (2014). Whirling disease dynamics: An analysis of intervention strategies. *Preventive Veterinary Medicine*, 113(4):457–468.

Vehtari, A., Gelman, A., and Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, 27(5):1413–1432. arXiv:1507.04544 [stat].

Wieland, F.-G., Hauber, A. L., Rosenblatt, M., Tönsing, C., and Timmer, J. (2021). On structural and practical identifiability. *Current Opinion in Systems Biology*, 25:60–69.

Zwietering, M. H., Jongenburger, I., Rombouts, F. M., and van 't Riet, K. (1990). Modeling of the bacterial growth curve. *Applied and Environmental Microbiology*, 56(6):1875–1881.

A Priors

We reuse the laboratory datasets shown in Figures 2. (a)- (b) and the associated preprocessing pipeline described in Giersch et al. (2025), as well as the seawater survival dataset for Figure 2. (c) from Giersch et al. (2022). Figure 2. (d) presents temperature-controlled longitudinal experiments (4, 10, 13, 16 °C) conducted with the same workflow and preprocessing as Giersch et al. (2025). Data at 16 °C were excluded from analysis because mortality was confounded by non-BTN causes. All time units are in days and rates are per day. Except where noted below, laboratory methods, qPCR processing, and cancer-fraction calculation follow those studies. Here we document only the analysis-specific filters of the data and analyses used for the SEIP model.

A.1 Defining Exposed vs Infectious Classes

To distinguish exposed (E) from infectious (I), we fit a single-changepoint model to a Bernoulli transformation of the emission measurements in Figure A.1 (1 = emission detected; 0 = not detected). After ordering observations by hemolymph fraction, we fit a single changepoint using binary segmentation in the R `changepoint` package (Killick and Eckley, 2014). The estimated changepoint occurred at 24.12% cancer, above which emission occurred. We therefore classify clams as infectious once their hemolymph cancer fraction exceeds 24.12% , susceptible at 0%, and otherwise they are exposed.

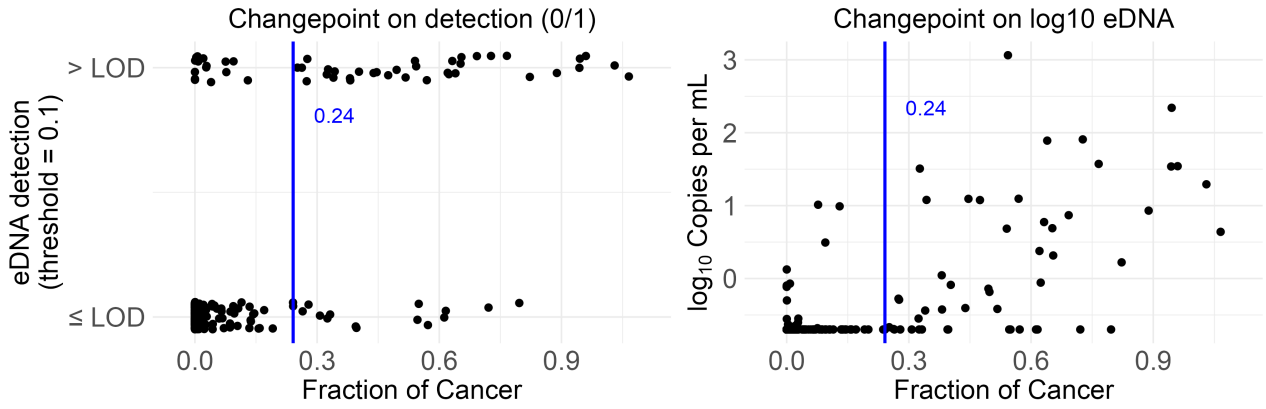


Figure A.1: Changepoint analysis identifying the infectiousness threshold (24.12% cancer fraction) used to classify infectious individuals. Clams were kept in individual tanks and water samples were taken to monitor the concentration of cancer particles released in a period of 24 hours in addition to the cancer fraction samples. The figure on the left shows how the Bernoulli data transformation with the identified changepoint and the figure on the right shows the original data with the labeled changepoint.

A.2 Progression (α) and mortality from the longitudinal dataset (m)

A.2.1 Hierarchical growth fits

We modeled clam cancer trajectories (Figure A.2) using a hierarchical modified Gompertz curve 2 (Zwietering et al., 1990). Shape parameters (asymptote (A) and growth rate(μ)) were shared at the group level and the timing parameter (λ_i) varied by individual clams (i). This structure assumes broadly similar cancer growth across clams while allowing heterogeneous onset times.

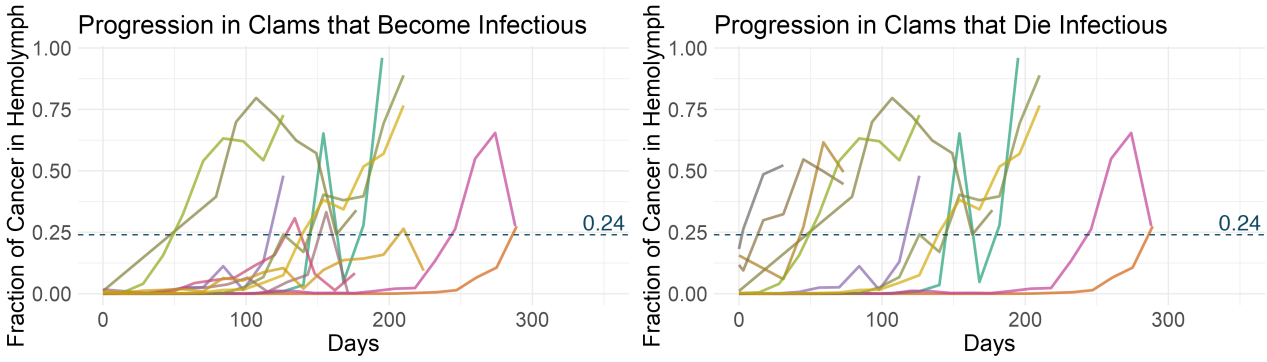


Figure A.2: Longitudinal data subsetted into clams that start at a low amount of cancer and progress to above the infectious level (left) and clams that begin below and die above the infectious level (right).

$$A * \exp\left(-\exp\left(\frac{\mu * e}{A} * (\lambda_i - t) + 1\right)\right) \tag{2}$$

A.2.2 Time to infectiousness (α)

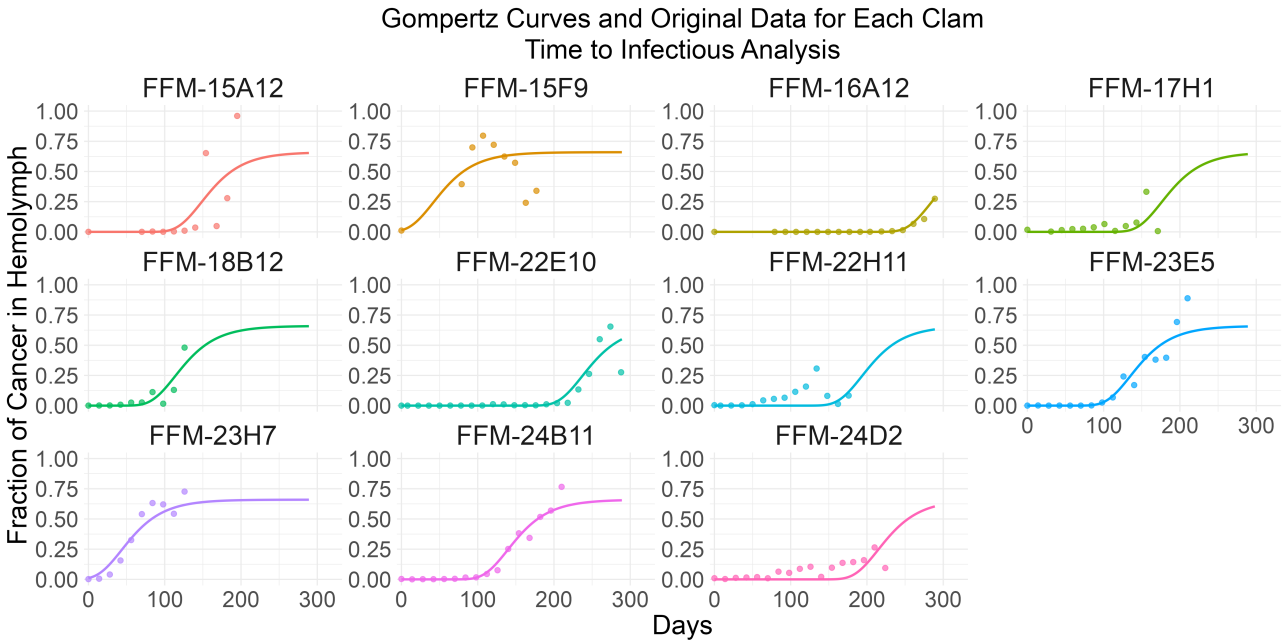


Figure A.3: Longitudinal data (points) of clams that have progressed to the 0.24 cancer fraction threshold, separated by each individual clam and fitted growth curves overlaid (lines).

For progression, we included clams whose hemolymph cancer fraction was $< 10\%$ at the first time point of the observation window and that progressed to exceed the infectiousness threshold of 24.12% (Figure A.2). Since the start of infection (0% cancer) is unobserved, we estimated a minimum time to infectiousness by using the assay's limit of detection as a reference fraction and computing the model-implied time to that level. Using the hierarchical modified Gompertz fits (Figure A.3), we computed the time the fitted curve first crossed $f_I = 0.2412$ and subtracted the time the curve would equal the minimum detectable fraction $f_{LOD} = \frac{10}{10000000} = 10^{-6}$. Summarizing across clams resulted in a group-level lower bound of about 80 days.

We represented the exposed-to-infectious rate with a lognormal prior centered around this timepoint:

$$\alpha \sim \text{LogNormal}(\mu = \log(0.01), \sigma = 0.3)$$

The dispersion (σ=0.3) is intentionally set to allow flexibility around this minimum by placing probability mass on both longer and shorter durations than 80 days. We do not truncate α at 80.6 days, as enforcing this bound did not change fit or performance.

A.2.3 Time to death and mortality rate (m)

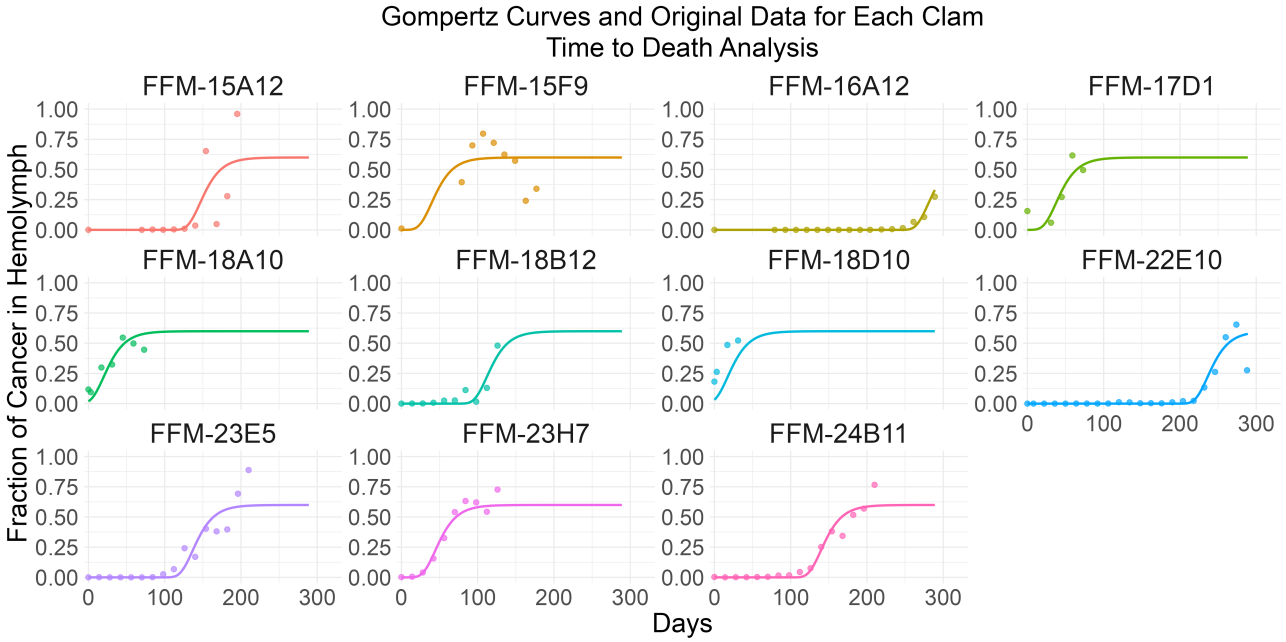


Figure A.4: Longitudinal data (points) of clams that have progressed to the 0.24 cancer fraction threshold and died with a cancer fraction above 0.24, separated by each individual clam and fitted growth curves overlaid (lines).

For mortality, we used trajectories with points below 0.24 cancer fraction and ending in death after passing 0.24 cancer fraction (Figure A.2). From the hierarchical Gompertz fits (Figure A.4), for each clam i and posterior draw we computed the time from first crossing 24.12% ($t_{infect,i}$) to death ($t_{death,i}$). We added 7 days to this term since $t_{death,i}$ is the last measured point before death and the real time of death is between that point and the next observed time point (14 days).

$$\Delta_i = t_{death,i} + 7 - t_{infect,i}$$

Draws with $\Delta_i \leq 0$ (rare cases where the fitted $t_{infect,i}$ occurs after $t_{death,i}$) were discarded. This affected $\sim 2\%$ of posterior samples.

For each clam we summarized the positive draws by their posterior mean μ_i and standard deviation σ_i, compared distributions (normal, lognormal, gamma) on fitting to the means by AIC/BIC, and chose the Gamma family (slightly better fit and correct support on $\mathbb{R}^+$). Under a Gamma rate parameterization for the means of time to death for each clam i,

$$\Delta_i \sim Gamma(\alpha_i, \beta_i), \quad \alpha_i = \frac{\mu_i^2}{\sigma_i^2}, \beta_i = \frac{\mu_i}{\sigma_i^2}.$$

We then aggregated to a population-level Gamma using the law of total variance (Figure A.5).

$$\mu_{pop} = \frac{1}{K}\sum \mu_i, \quad \sigma^2_{pop} = \frac{1}{K}\sum(\sigma_i^2 + \mu_i^2) - \mu^2_{pop}$$

$$\alpha_{pop} = \frac{\mu^2_{pop}}{\sigma^2_{pop}}, \quad \beta_{pop} = \frac{\mu_{pop}}{\sigma^2_{pop}}.$$

The ODE model uses the mortality rate $m = 1/\Delta_{pop}$. If $\Delta_{pop} \sim Gamma(\alpha_{pop}, \beta_{pop})$ with rate β_{pop}, then

$$m = 1/\Delta_{pop} \sim InvGamma(shape = \alpha_{pop}, scale = \beta_{pop}).$$

From our data this yields $\alpha_{pop} \approx 2.60, \beta_{pop} \approx 0.04$, thus

$$m \sim InvGamma(shape = 2.60, scale = 0.04).$$

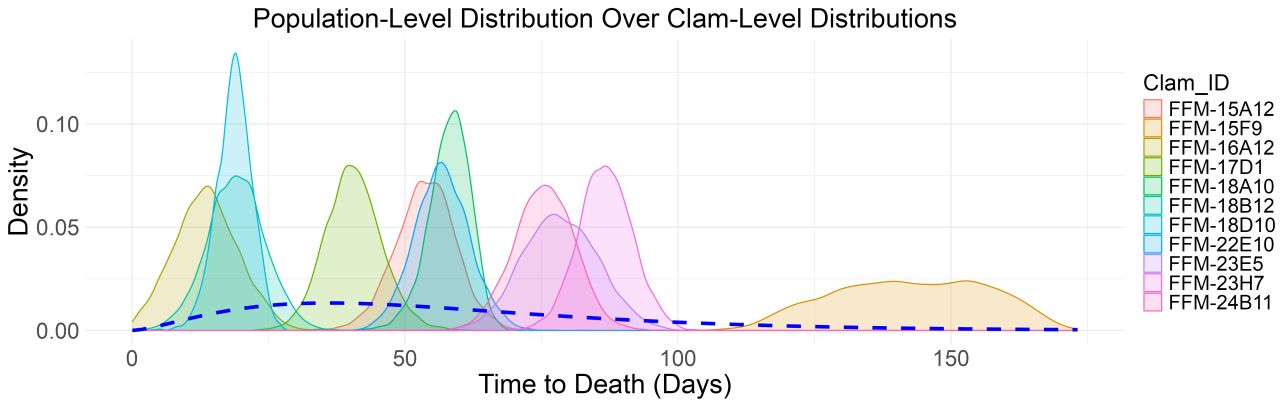


Figure A.5: Estimated distributions for time to death of each individual clam with the estimated population-level time distribution overlaid as the blue dotted line.

A.3 Transmission Rate (β)

We lack direct experiments on natural transmission. Therefore, β is assigned a weakly informative prior centered at a biologically plausible value and tuned via prior-predictive checks:

$$\beta \sim \text{LogNormal}(\mu = \log(0.02), \sigma = 1.0).$$

The larger σ maintains flexibility while preserving computational stability.

A.4 Particle Emission Rate (γ)

Using eDNA measurements of tank water from individual clams (Figure 2. (b)) that have progressed past the infectious threshold (24.12%), we found that the emission data were zero-inflated and right-skewed. We therefore subsetted the data for only positive emissions to fit a prior on the emission rate conditional on emission. We fit a Gaussian model to log(CopyPerMl) using the `glmmTMB` package in R (Brooks et al., 2017; R Core Team, 2020):

$$\gamma \sim LogNormal(\mu = 1.24, \sigma = 1.93).$$

A.5 Particle Decay (δ) and its Temperature Effect (δ_T)

We estimated the particle decay rate from survival trajectories of cancer cells emitted by clams across temperatures 4 °C, 10 °C, 13 °C, 16 °C, 25 °C, 37 °C (Figure 2. (c)). For each trial, we normalized $P_0 = 1$ and fit an exponential decay, estimating δ by log-linear regression of log(P(t)) on time.

$$P(t) = P_0 * e^{-\delta t} \Rightarrow log(P(t)) = -\delta t$$

Pooling across trials (ignoring temperature) supported a baseline prior:

$$\delta \sim LogNormal(\mu = \log(0.05), \sigma = 0.1).$$

To model temperature effects, we regressed log-decay rates on temperature (T):

$$\log(\delta_T) = a_\delta + b_\delta T$$

The slope was positive and statistically significant, so we used this as normal hyperpriors. The hyperparameter standard deviations were tuned for computational stability and the truncation of b_δ enforced faster decay at higher temperatures while allowing uncertainty in magnitude.

A.6 Temperature Effect on Transmission (β_T)

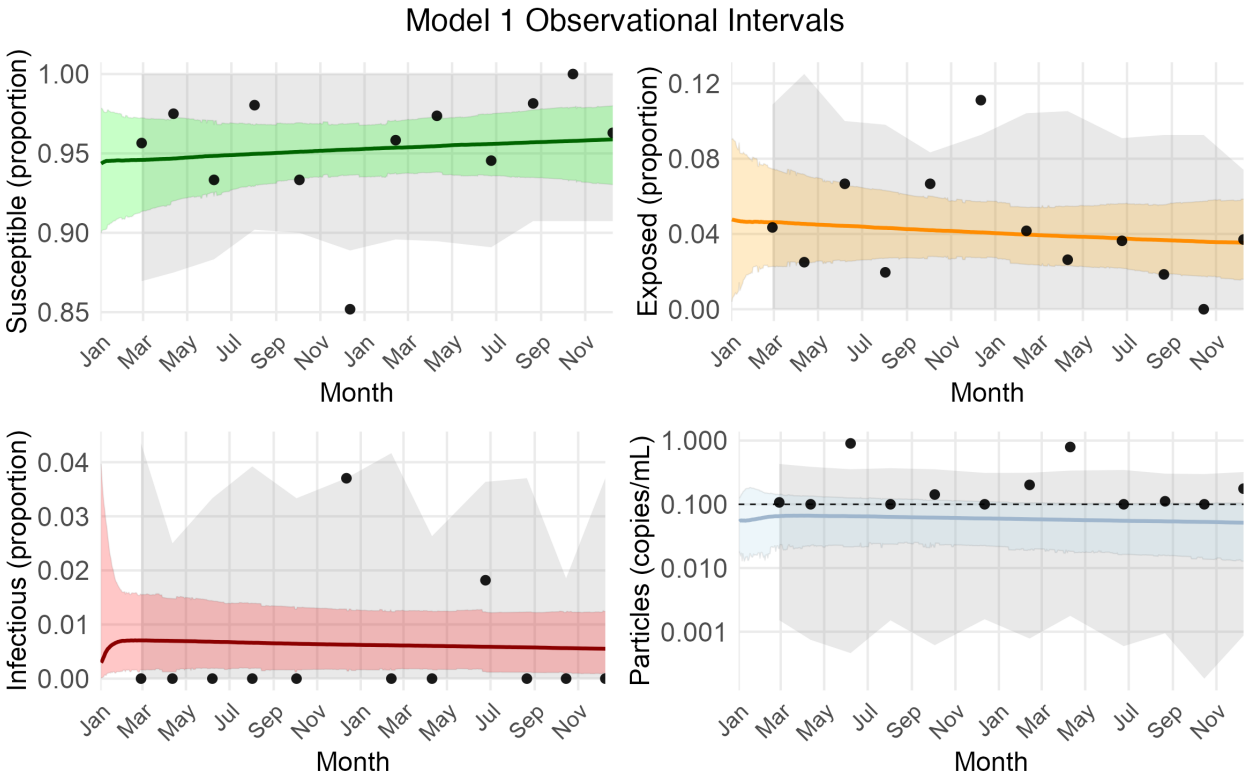
Using laboratory progression trials, we constructed a binary outcome for each clam at each temperature. For each clam trajectory, Progress = 1 if any hemolymph measurement reached at least 10% cancer fraction (i.e., ≥ 0.1), and Progress = 0 otherwise. We then fit a logistic regression with Progress as the response and temperature T as the predictor:

$$\beta_{\text{prob}}(T) = \text{logit}^{-1}(a_\beta + b_\beta T), \qquad \beta_{\text{eff}}(T) = \beta \cdot \beta_{\text{prob}}(T),$$

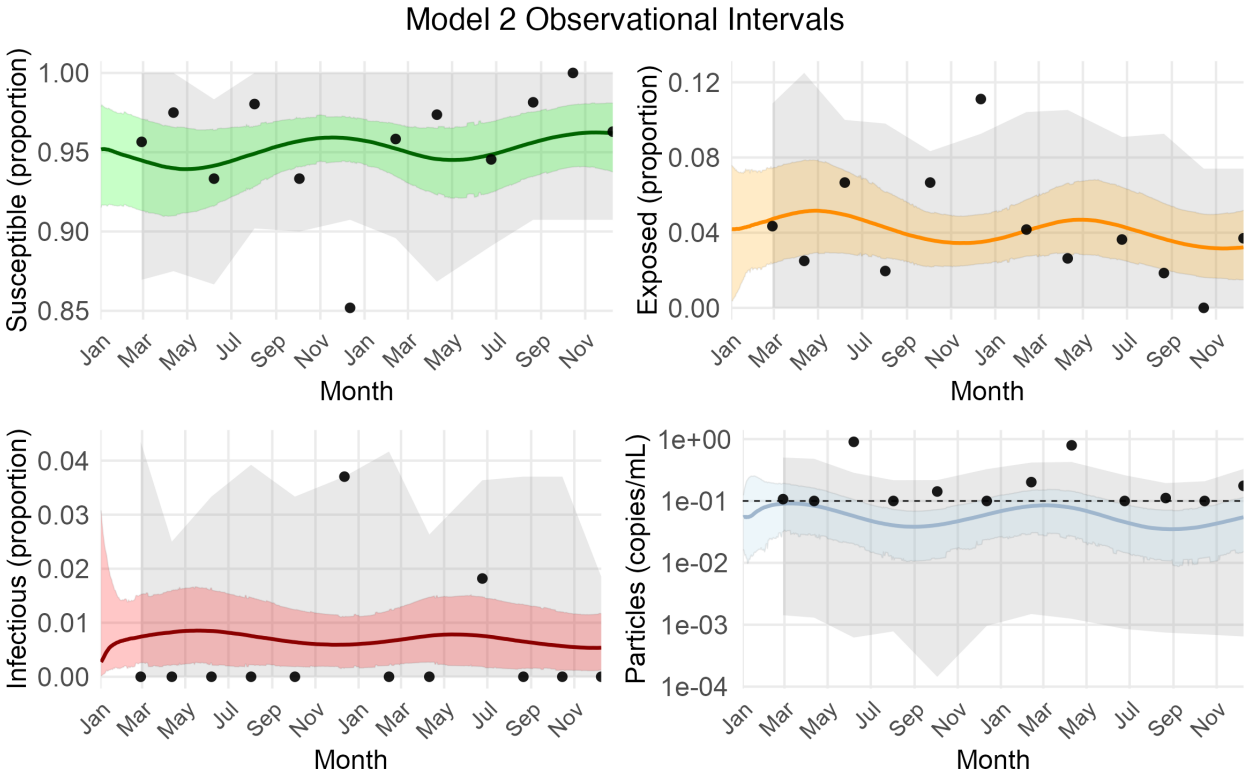
where $\beta_{\text{prob}}(T) \in (0, 1)$ represents the temperature-dependent probability of progression to an infectious state.

The fitted slope b_β was positive but not statistically distinguishable from zero, so we did not impose a positivity constraint (in contrast to b_δ, where the relationship was supported and truncated to enforce faster decay at higher temperatures). We used the regression estimates to define moderately informative priors that improved computational stability while allowing uncertainty in the temperature effect.

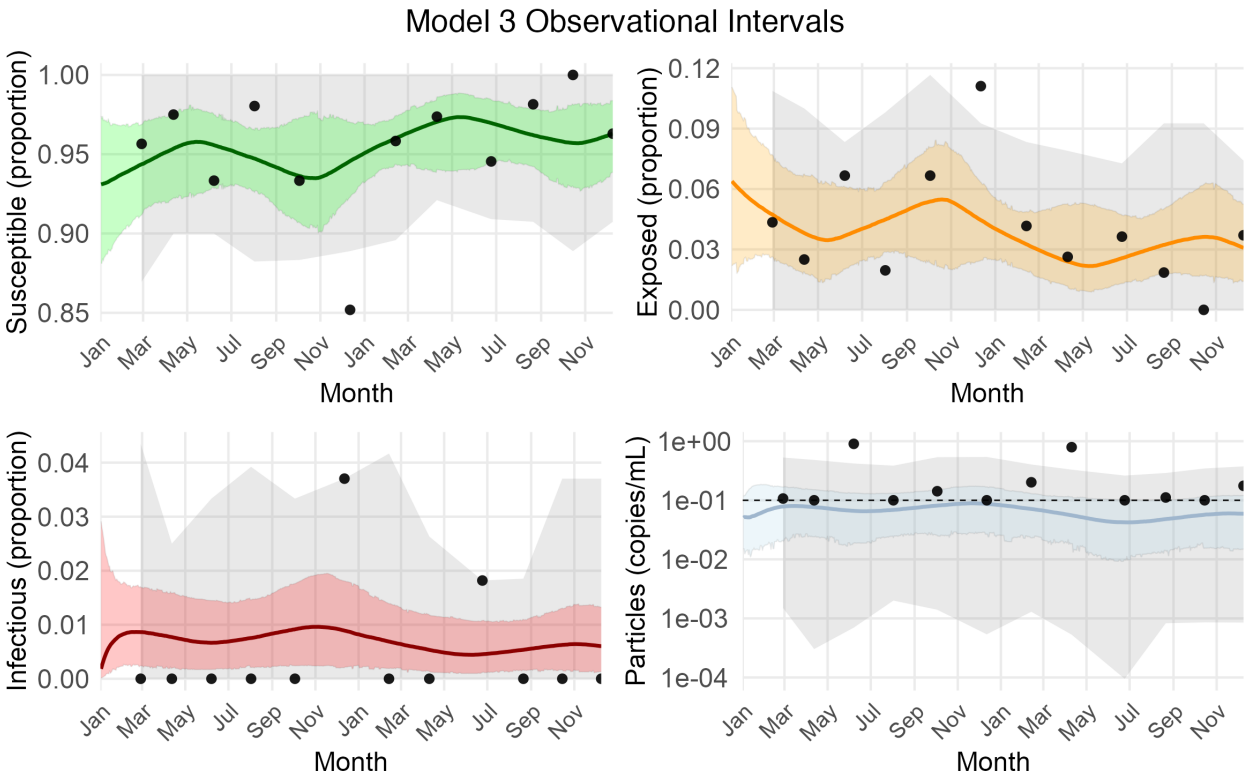
B Model Results



(a)



(b)



(c)

Figure B.1: Model 1-3 trajectories with observational error in 95% high density intervals as the grey shaded region and observed data points as the black dots.

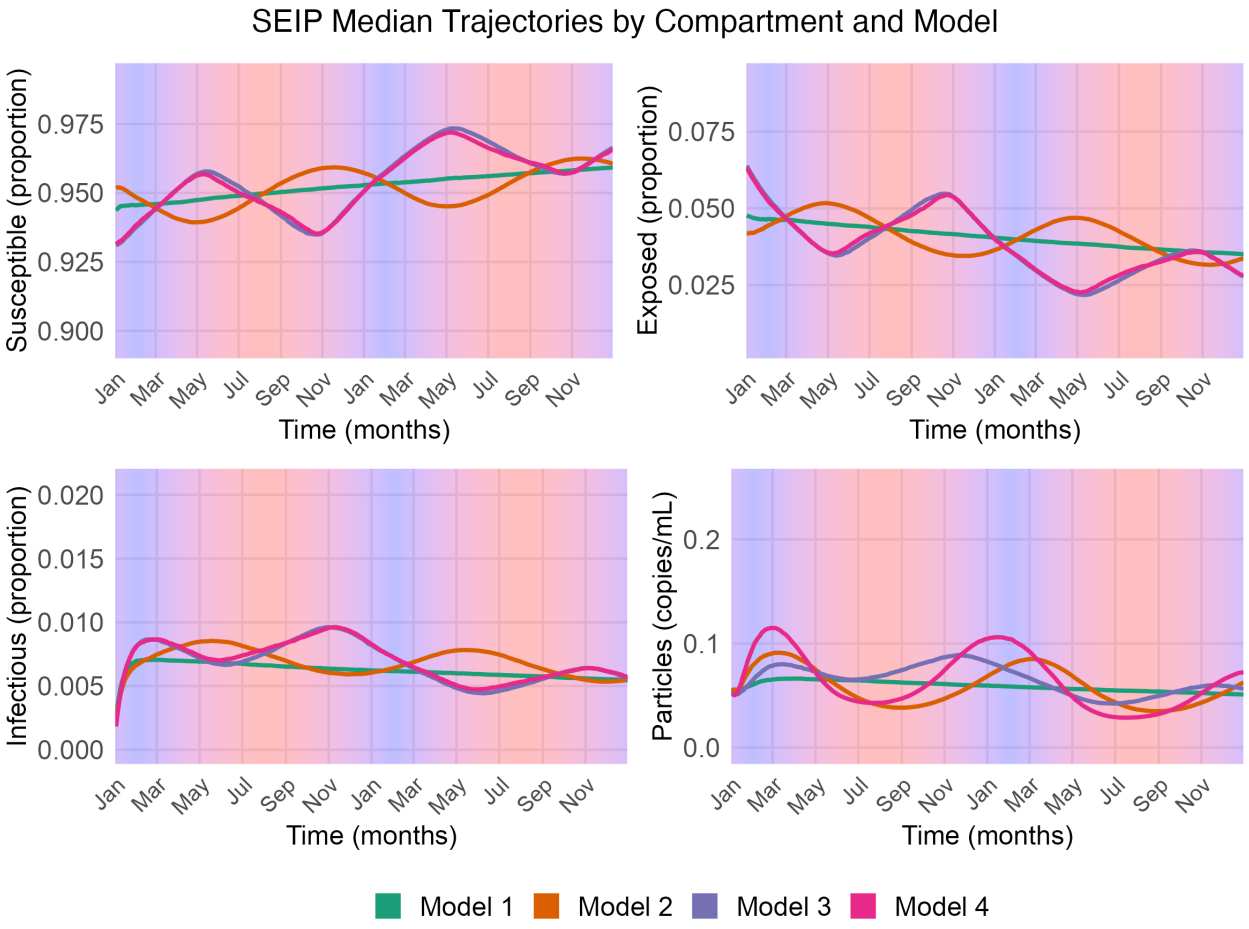


Figure B.2: A zoomed in look at the mean trajectories from Models 1-4 and Forest River data separated by compartment S, E, I and P with a temperature grid for model performance comparisons, where red represents the highest temperature and blue represents the lowest temperature in the observation system.

B.1 Priors vs. Posterior

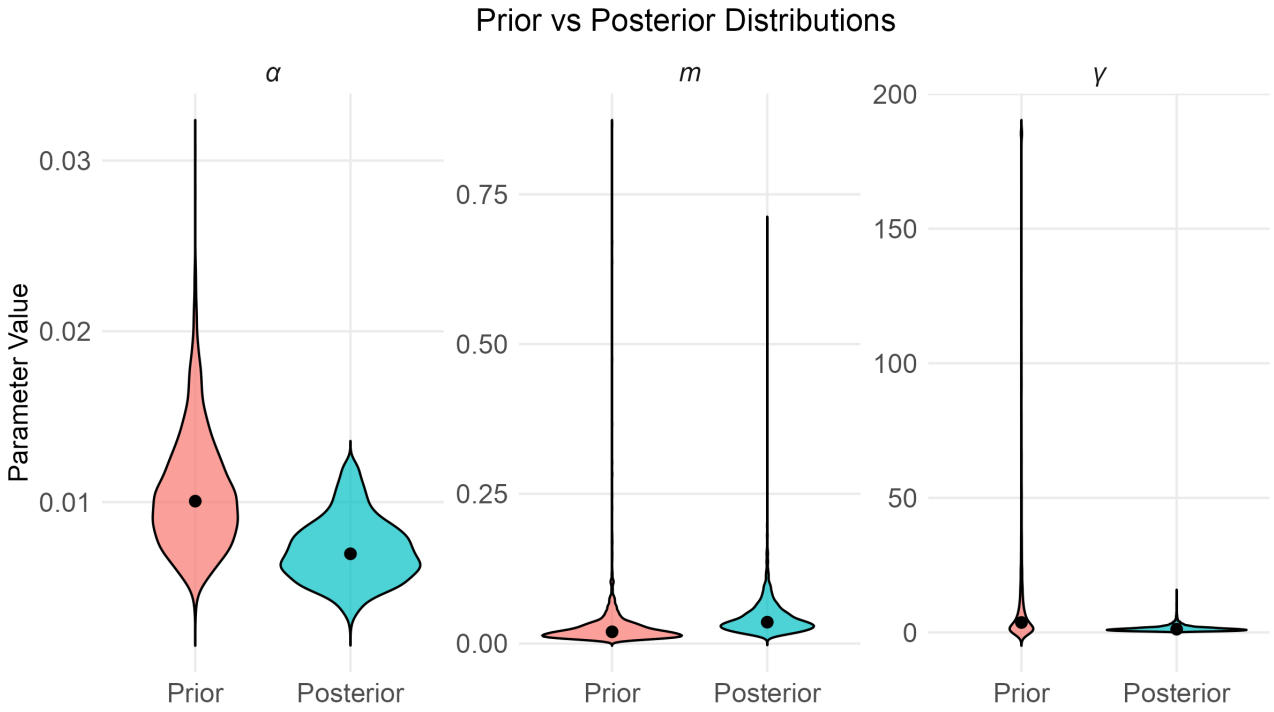


Figure B.3: Prior and model 4 posterior distributions of α, m, and γ with the γ prior showing the 99th quantile for easier comparison due to the high variability in the dataset.

C Software and Reproducibility

Analyses were conducted in R, Julia, and Stan. Code to reproduce all analyses and figures will be available on GitHub.