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

Carissa Mayo
University of Washington
Quantitative Ecology and Resource Management
Dr. Michael Metzger

UNIVERSITY of WASHINGTON

pnri Pacific Northwest Research Institute

1. Introduction
2. Research Goals
3. Methods
4. Results
5. Conclusions

Natural Transmissible Cancers

Tasmanian Devil Facial Tumor Disease (DFTD)

- Transmitted through biting
- ~ 40 years old
- Two lineages
- Almost always lethal

Canine Transmissible Venereal Tumor (CTVT)

- Transmitted sexually
- 4000 - 6000 years old
- Mostly harmless

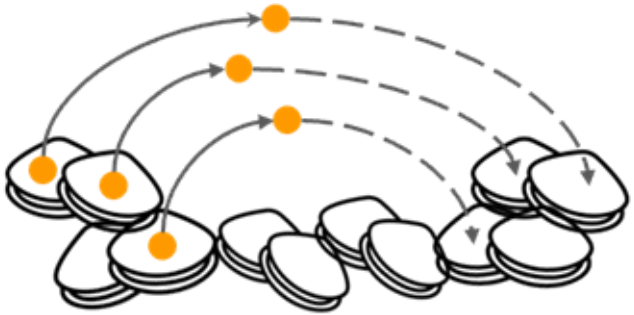
Bivalve Transmissible Neoplasia (BTN)

- Not transmitted through contact
- ~ 400 years old
- Many lineages
- Reported to be lethal

BTN

- Spreads through waterborne cancer cells
- Leukemia-like cancer classified as transmissible in 2015
- Found most notably in the hemolymph (invertebrate blood-like fluid)
- Neoplastic cells infiltrate into tissues in later stages of disease
- Cockles, Oysters, Clams

Soft-shell clam (*Mya arenaria*)



(Metzger et al. 2015)

Knowledge Gaps

- Due to its recent classification, much remains unknown about BTN
- No existing epidemiological models
- Lack quantitative estimates of the fundamental processes governing transmission, progression, and particle persistence
- Data-driven modeling and seasonal forecasts are essential components of disease response strategies

1. Introduction
2. **Research Goals**
3. Methods
4. Results
5. Conclusions

1) Core Disease Model

Estimate key transmission and progression rates using a Bayesian epidemiological framework.

2) Particle Component

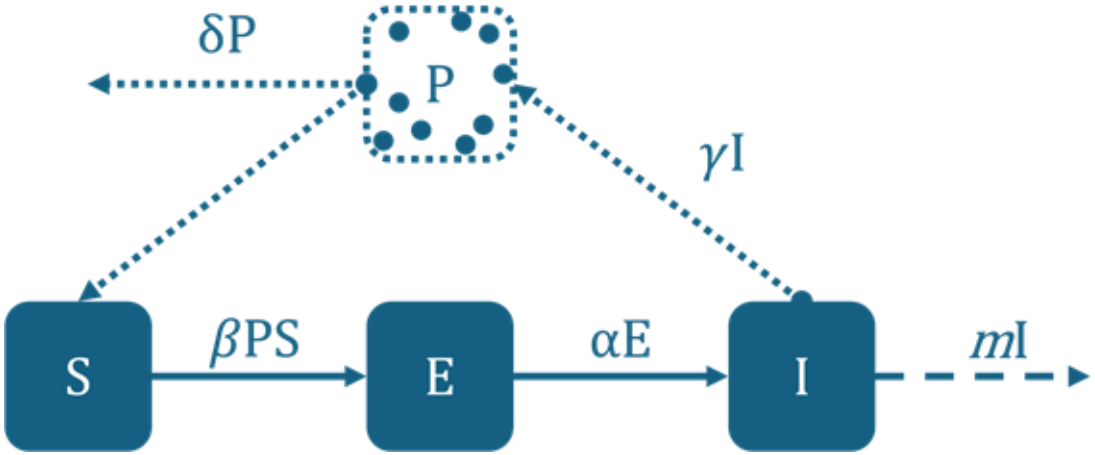
Integrate particle dynamics into a mechanistic disease model.

3) Seasonality

Test the hypothesis that incorporating seasonal temperature effects allows the model to reproduce observed trends in disease dynamics.

1. Introduction
2. Research Goals
3. **Methods**
4. Results
5. Conclusions

SEIP Model



- β : Rate at which susceptibles get infected from particles in the water
- α : Progression rate from exposed to infectious
- m : Mortality rate of infectious individuals
- γ : Particle shedding rate
- δ : Particle decay rate

Four Model Variants

$$dS/dt = -\beta*P*S + m*I$$

$$dE/dt = \beta*P*S - \alpha*E$$

$$dI/dt = \alpha*E - m*I$$

$$dP/dt = \gamma*I - \delta*P$$

1. No temperature dependence
2. Temperature dependence in particle decay (δ)
3. Temperature dependence in disease progression (β)
4. Both temperature dependent parameters (δ, β)

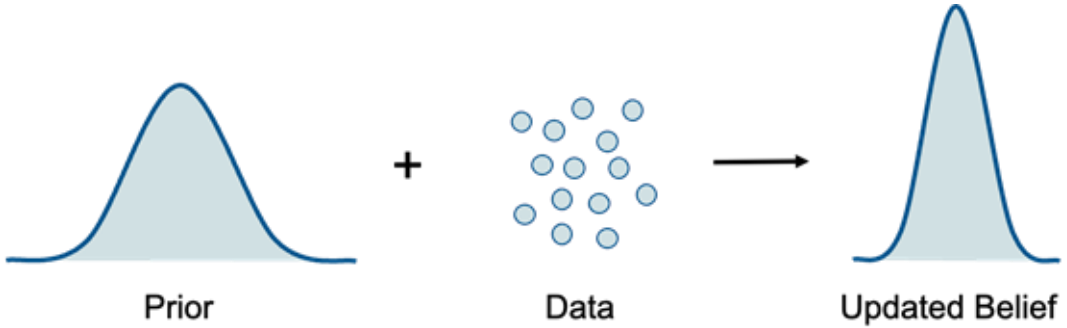
Model Assumptions

- No natural births/deaths unrelated to disease
- Constant population size
  - Total population is conserved over time

    $\Rightarrow S + E + I = 1$
  - Individuals return to the system into S at rate mI to maintain population balance
  - Modeling at an endemic state of the disease

Bayesian Modeling

- Combine prior knowledge (lab data) with new observations (survey data)
- Quantify uncertainty in parameters and predictions
  - Gives a range of likely values, not just one number
- Works well with limited or noisy data



$$p(\theta) * p(y|\theta) \propto p(\theta|y)$$

Bayes Formula

STAN

A programming tool for running Bayesian models that has an R interface (rstan, cmdstanr)

Process:

1) Define the system (ODEs) and what ranges are reasonable for each parameter (priors)
2) Stan explores parameter space
    a) Hamiltonian Monte Carlo to efficiently test many combinations of parameter values
    b) Prefers values that make the simulated data look like the observed data
3) Sampling setup
    a) 4 HMC chains, each run for 2000 iterations
        i) First 1000 iterations are a warm-up to adjust and tune the sampler
        ii) Remaining iterations are used for the posterior
4) The output is thousands of possible parameter combinations (posterior)
    a) Report medians and uncertainty ranges

(Gelman et al. 2020)

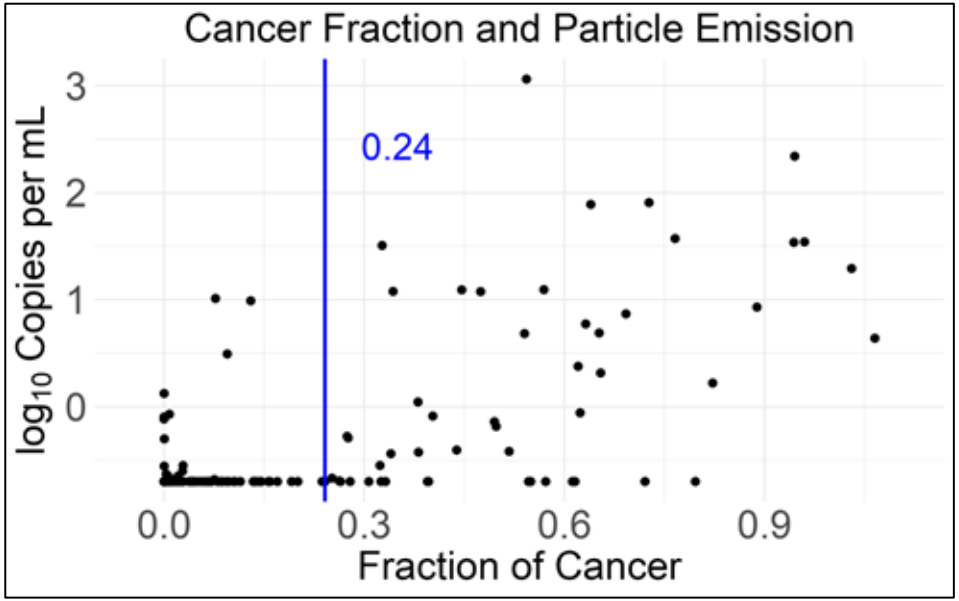
Lab Data - Cancer Cell Emission

Purpose: When and how much cancer is released into the water by infectious clams

Informs: Infectious cutoff, γ

- Infectious starting at 0.24 cancer fraction
- Average particle emission ≈ 10.11 copies per mL

Method: Change point analysis on emission



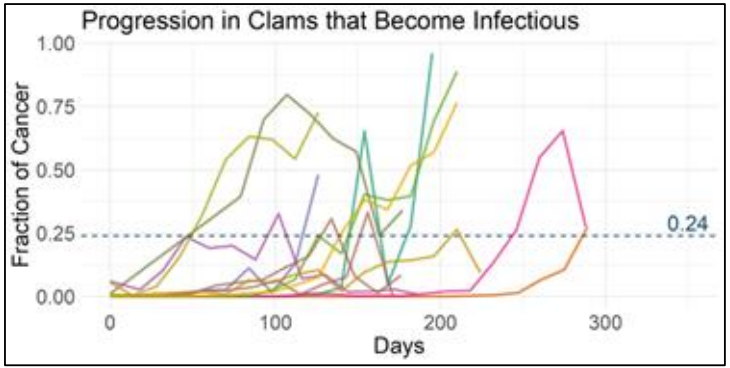
Lab Data - Cancer Progression

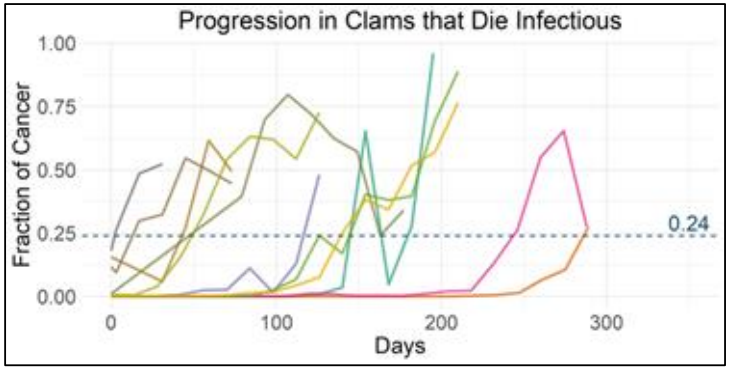
Purpose: Timing of cancer progression

Informs: *α, m*

- Average minimum time from exposed to infectious ≈ 80 days
- Average time from infectious to death ≈ 59 days

Method: Fit growth curves and derived times to infectious and death





Lab Data - Temp. Effect on Progression

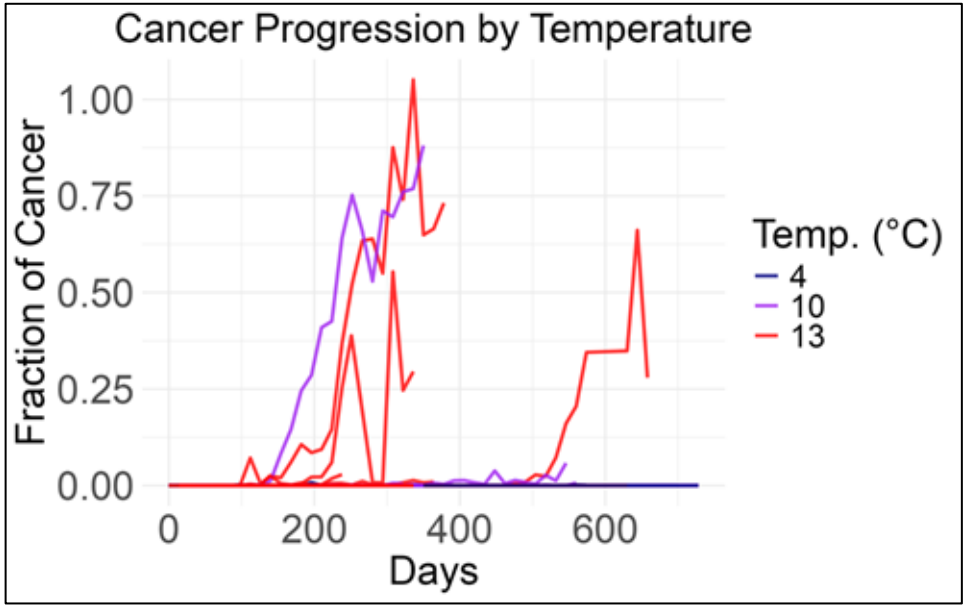
Purpose: How temperature affects progression

Informs: Temperature effect on β

- Higher temperature seems to increase the likelihood of cancer progression

Method: Logistic regression on binary progression

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



Lab Data - Cell Decay

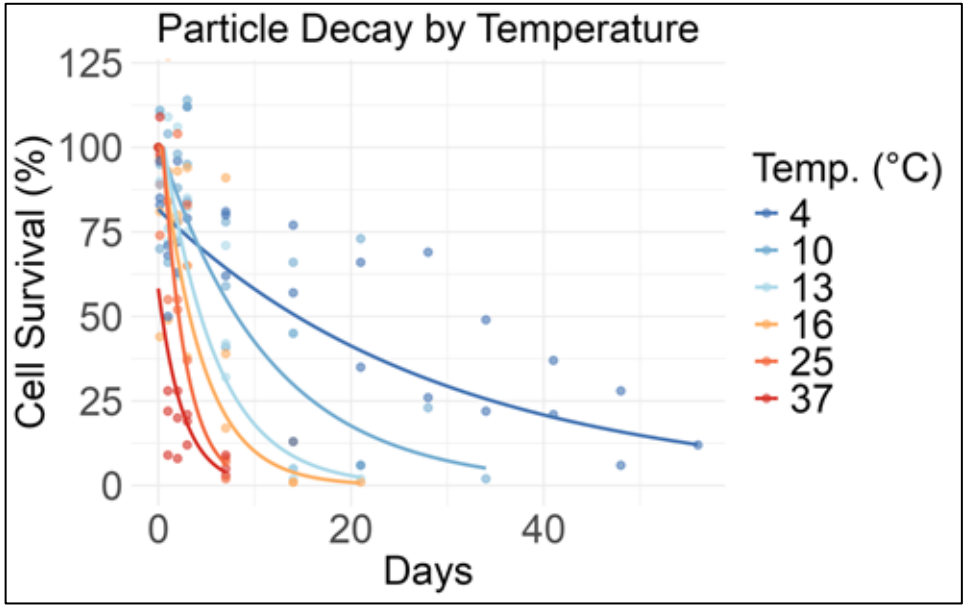
Purpose: How temperature affects particle decay

Informs: δ, temperature effect on δ

- Higher temperature increases cell decay rate

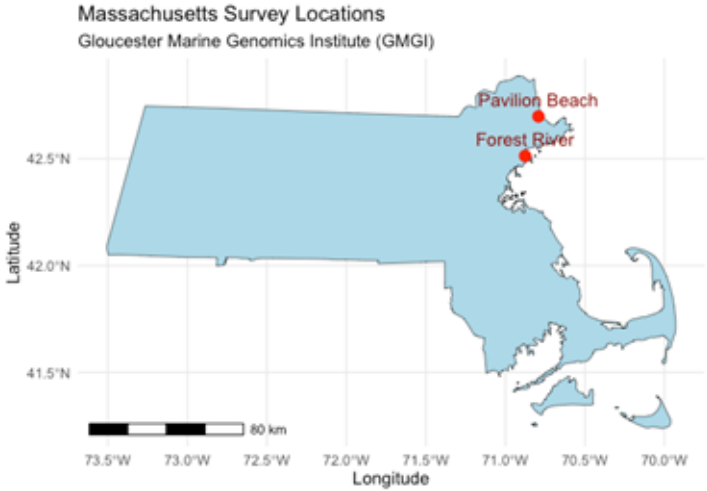
Method: Fit exponential decay curves and performed regression on rates across temperatures

$$\delta_T = \exp(a_\delta + b_\delta \cdot T(t))$$

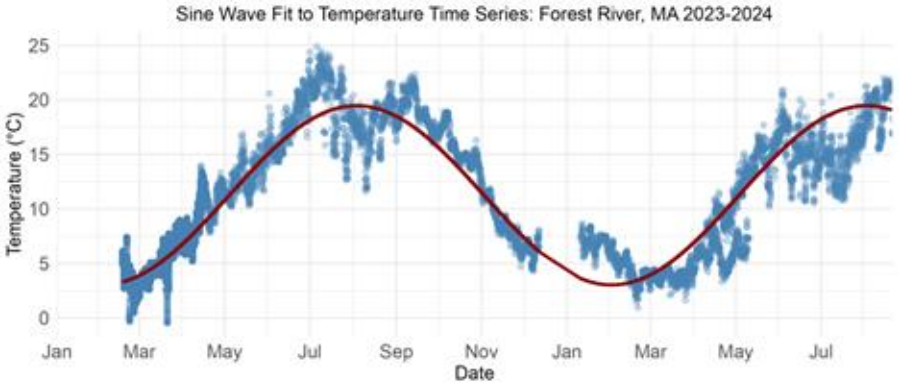


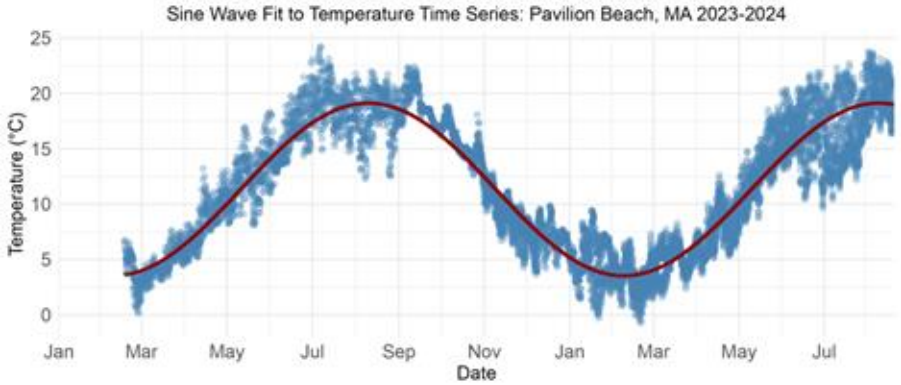
Study Sites


Survey Results from Forest River, MA (2023-2024)
Susceptible, Exposed, and Infectious Proportions
Percentage of Population
Collection Date
S
E
I
eDNA Measurements
Cancer Cell Concentration (copies/mL)
Survey Results from Pavilion Beach, MA (Apr. 2023 - Feb. 2025)


Massachusetts Survey Locations
Gloucester Marine Genomics Institute (GMGI)
Pavilion Beach
Forest River
Latitude
Longitude
80 km

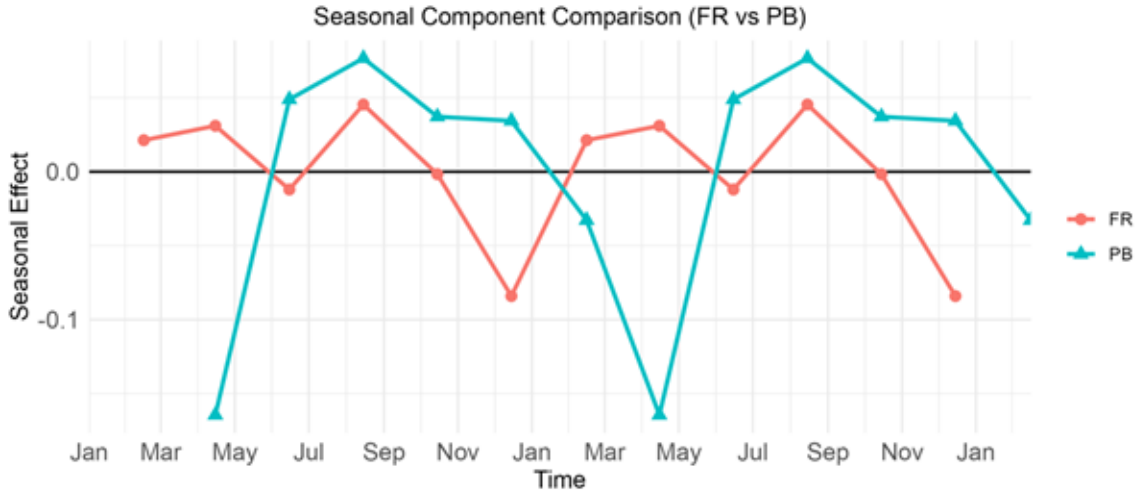
T(t) temp function



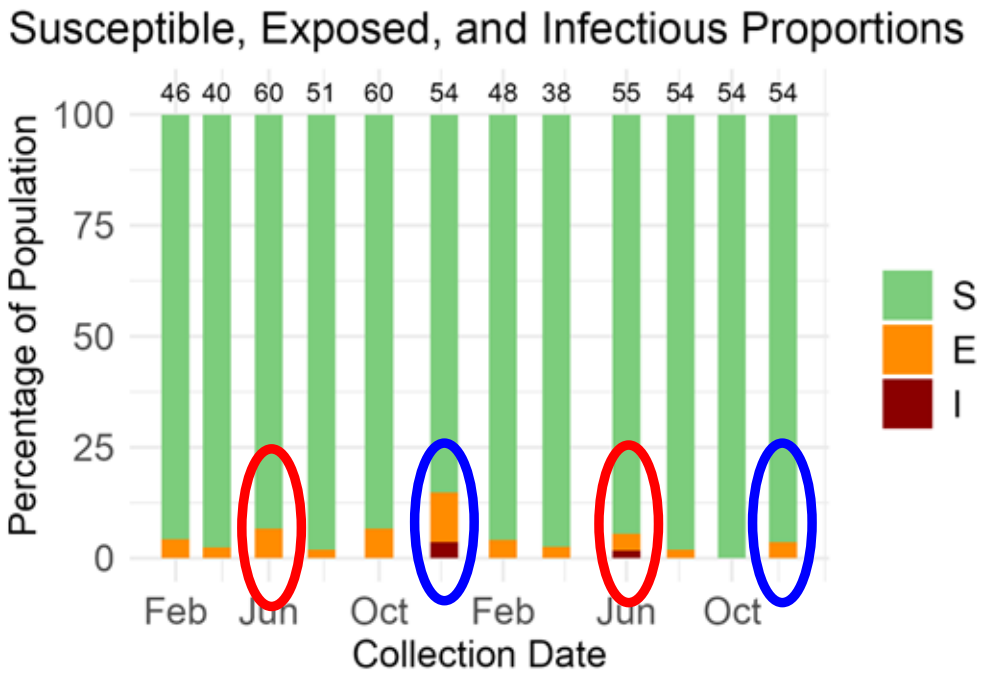


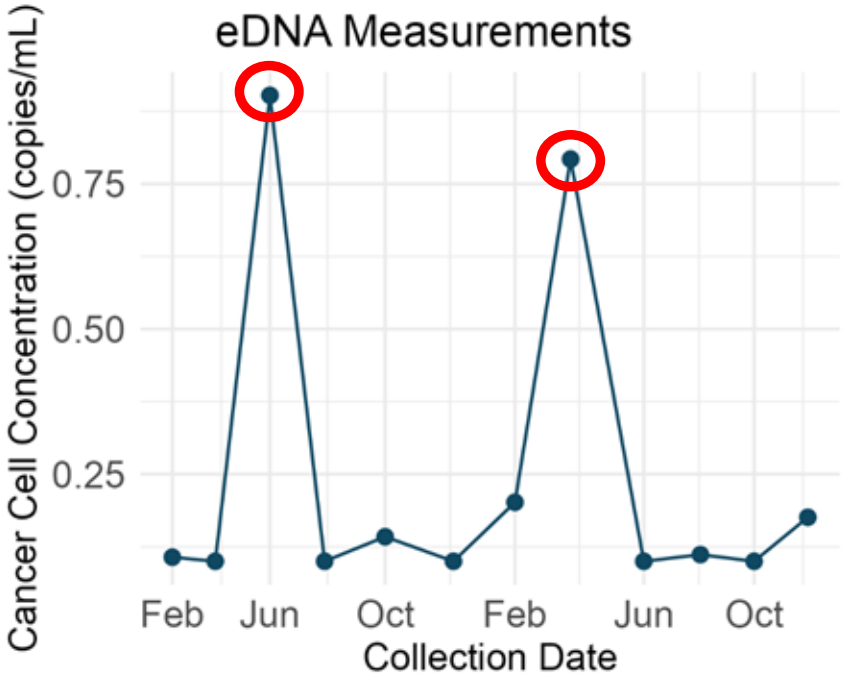
Observed Seasonality

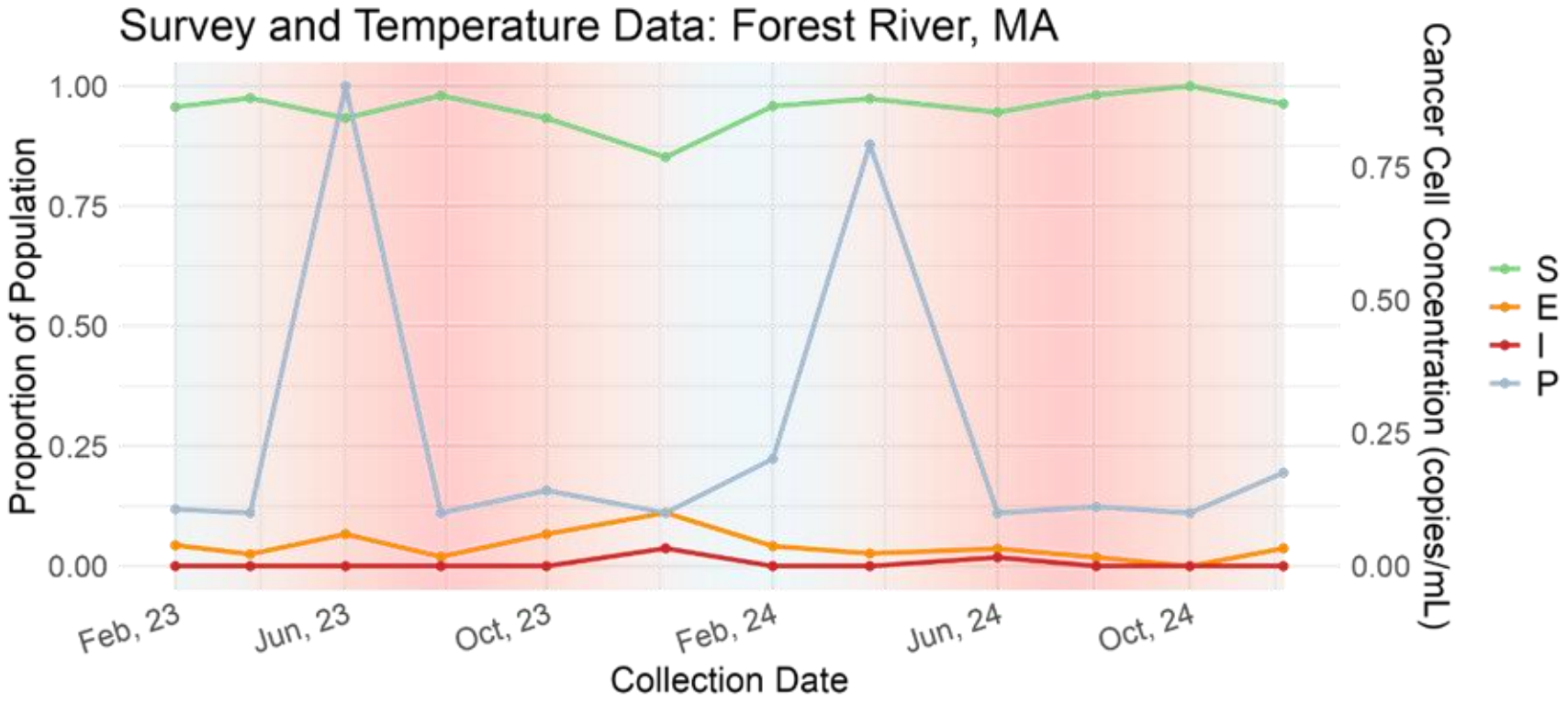
- Decomposition separates observed S into the repeating seasonal component
- Seasonal effect = how the data consistently rises or falls at certain times of year



Survey Results from Forest River, MA (2023-2024)






Survey and Temperature Data: Forest River, MA
Proportion of Population
1.00
0.75
0.50
0.25
0.00
Feb, 23
Jun, 23
Oct, 23
Feb, 24
Jun, 24
Oct, 24
Collection Date
Cancer Cell Concentration (copies/mL)
0.75
0.50
0.25
0.00
S
E
P

Model Set-up

p(θ) - Priors

Prior distributions for each parameter is defined based on laboratory data

p(y|θ) - Likelihood

How likely the data is given the chosen parameter values

1. SEI counts are a noisy sample of the underlying fractions S, E, I
   a. $y(t) \sim \text{Multinomial}(n(t), (S(t), E(t), I(t)))$
2. P is noisy eDNA sample of the underlying P fraction
   a. $\log(y(t)) \sim N(\log(P(t), \sigma)$

$$p(\theta) * p(y|\theta) \propto p(\theta|y)$$

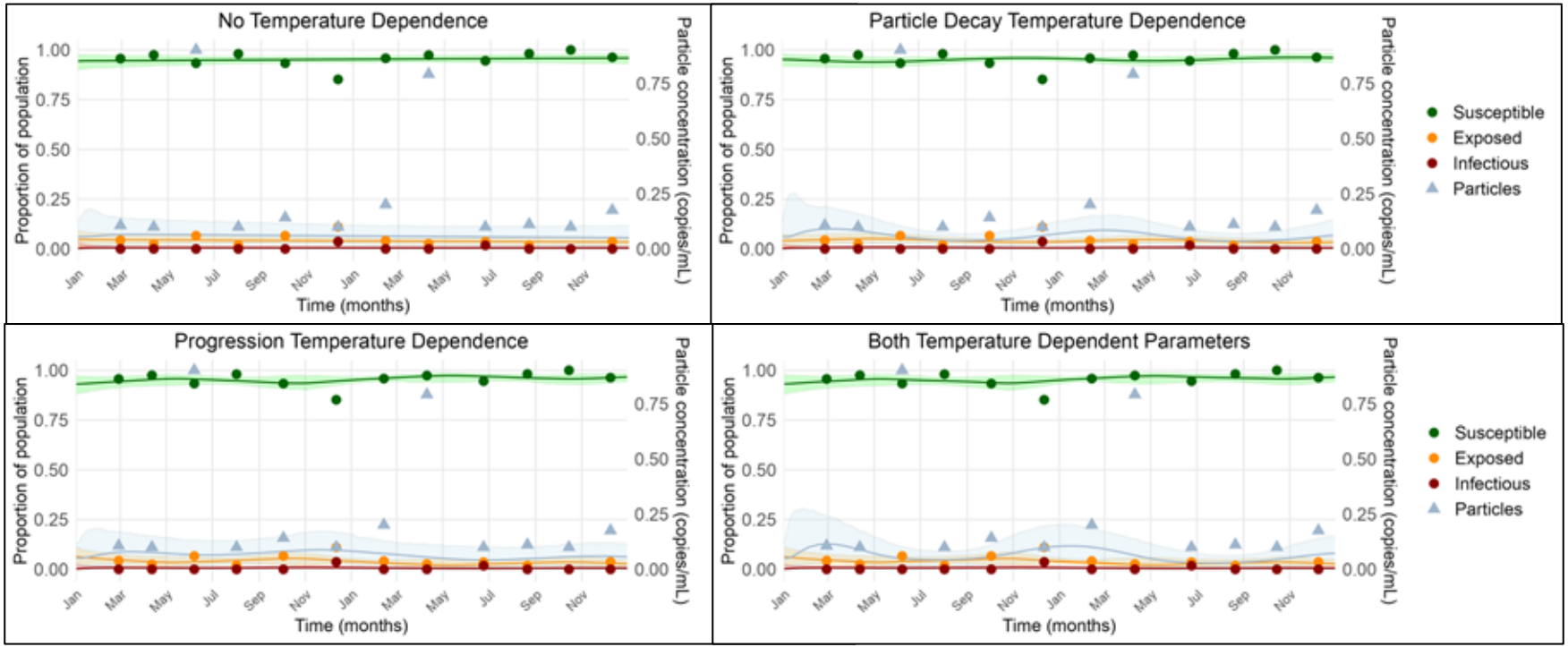
Bayes Formula

Structural Identifiability

- To distinguish data limitations from model misspecification
- Identifies which parameters and combinations can be uniquely recovered from ideal, noise-free observations
- Ran with Julia package
- Results:
  - Initial conditions of SEI proportions and P concentrations
  - Individual terms of temp dependent parameters (a_δ, b_δ, a_β, b_β)
  - Everything else is identifiable

1. Introduction
2. Research Goals
3. Methods
4. **Results**
5. Conclusions

Model Results

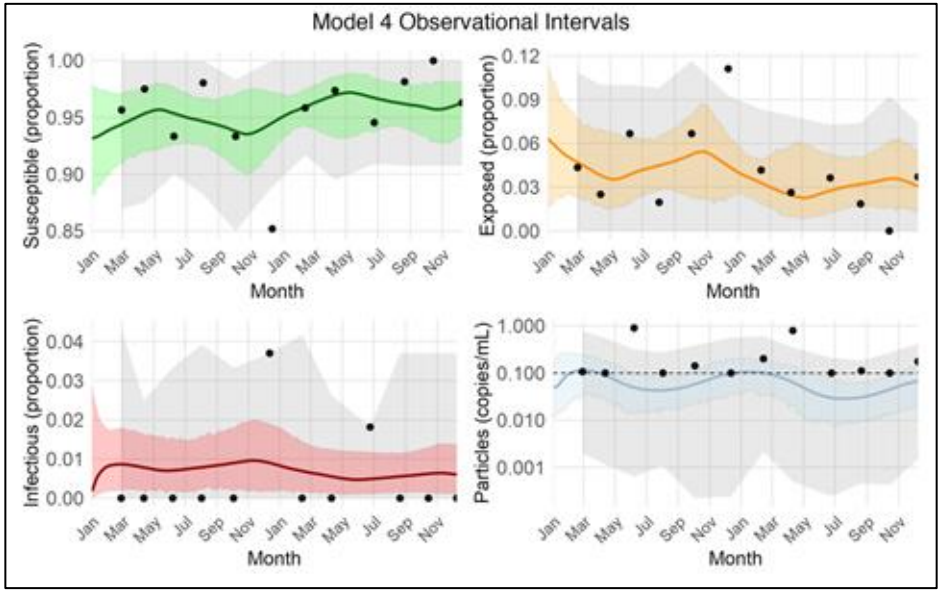


Model Comparisons

- **Quantitatively**: Pareto smoothed importance sampling with leave-one-out cross validation (PSIS-LOO) (Vehtari et al. 2017)
  - Rates model on how it predicts unseen data
  - Differences in comparisons were less than their standard errors
  - Models cannot be reliably compared by predictive accuracy
- **Quantitatively**: Posterior Predictive Checks (PPC)
  - Compute the 95% posterior predictive interval for SEIP using the observation models
  - Each model showed the same number of observations included in this interval
  - Models cannot be compared by posterior predictive accuracy
- **Qualitatively:**
  - Compare based on biological interpretation and visual seasonal pattern alignment

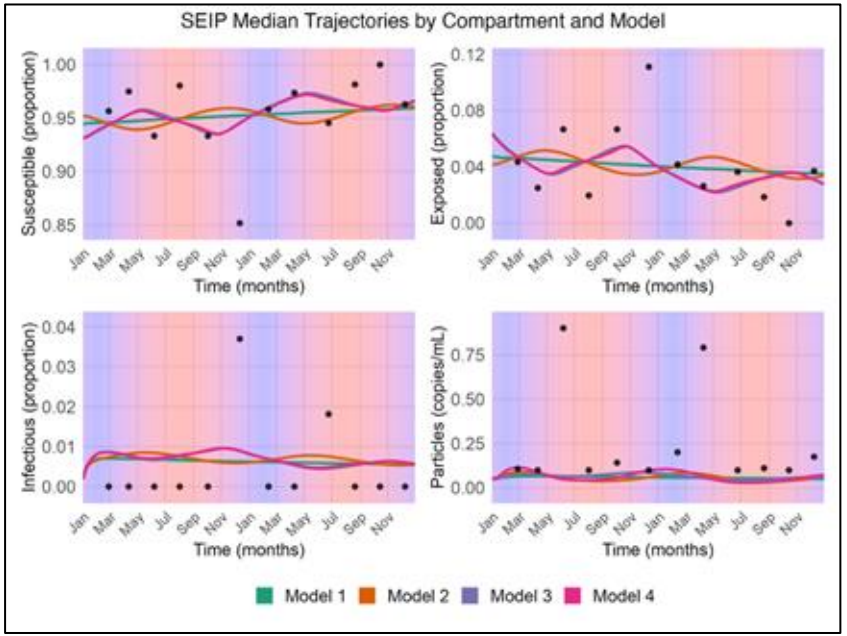
PPCs

All models show the same result: One or two observations outside of the interval for each trajectory.



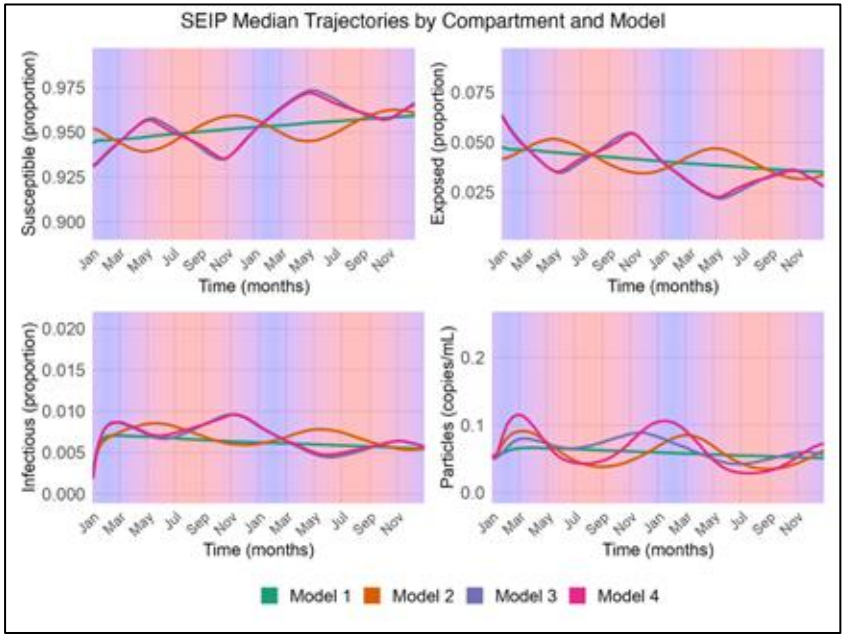
Visual Model Comparisons

- All models show a general decrease in disease prevalence over time
- Two opposing seasonal patterns
  - Model 2 - early summer peaks
  - Model 3/4 - winter peaks
- Model 4 is most similar to model 3
  - ⇒ progression temperature dependence is a driving force of seasonality



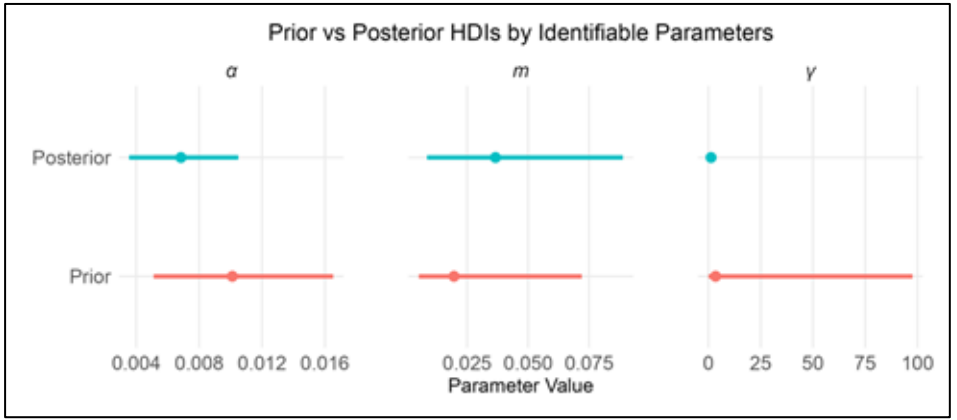
Visual Model Comparisons

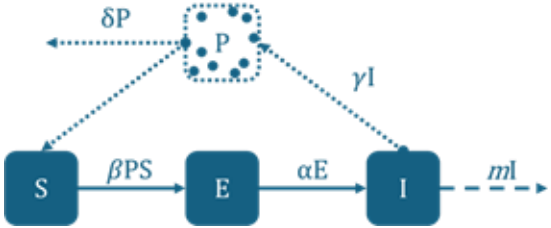
- Model 4 results in highest P trajectory peaks
- Temperature-dependent decay primarily refines the particle dynamics
- P is still not at the amplitude as seen in the data



Model 4 Results

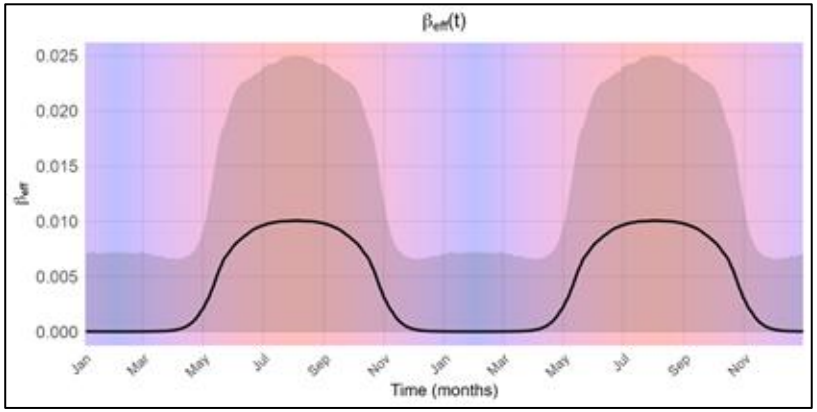
- α posterior shifted and is narrower
  - ⇒ longer latent periods than initially assumed
  - ⇒ strong data-driven learning
- γ posterior narrowed
  - ⇒ data support for a lower emission rate
- m posterior is similar to prior
  - ⇒ weak data influence

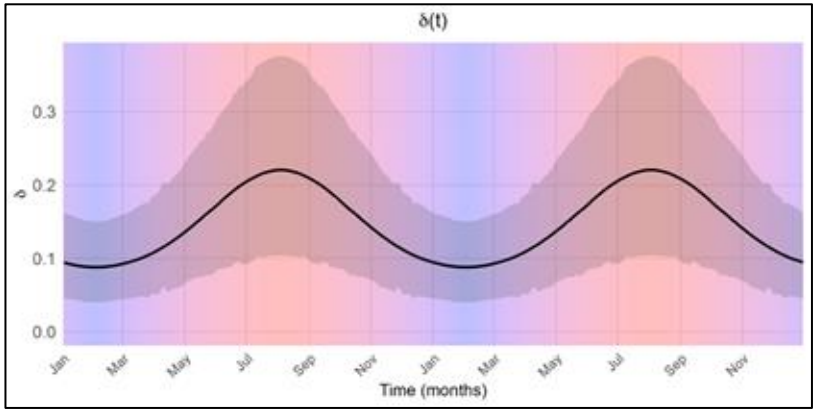




Model 4 Results

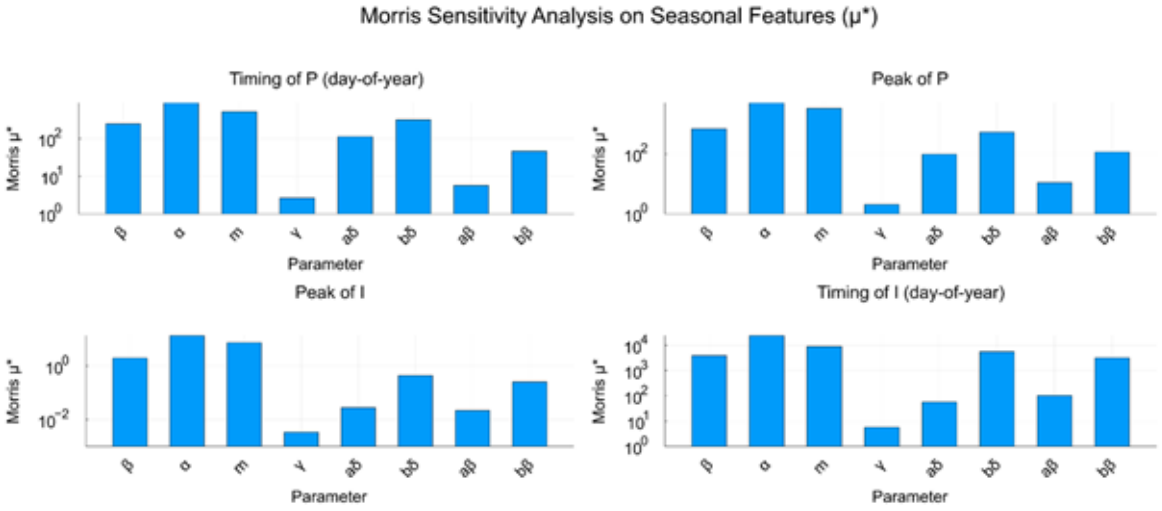
Beta temperature-dependence and delta rates increase with warmer temperatures, consistent with laboratory evidence.





Model Sensitivity

- Overall importance of each parameter on a given seasonal feature evaluated over 5-95% prior intervals
- α and m: Most influential parameters
- γ: Least influential parameter

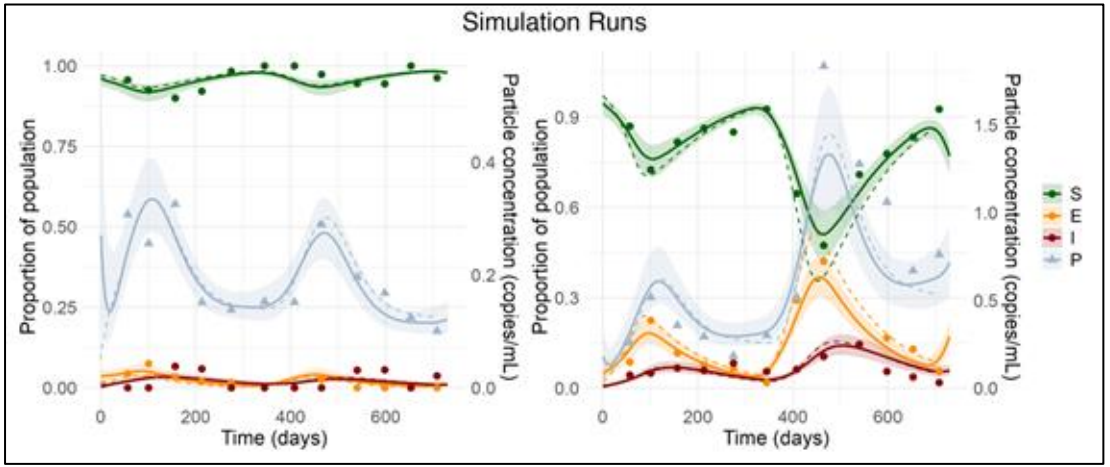


Model Validation

- Simulated trajectories from prior distributions 100 times (100 simulated datasets)
- Calculated posterior estimate 95% interval coverage of simulated parameters
- While overall calibration is good, α is weakly informed by data under the current prior

Results:

- α: 80% coverage
- m: 96% coverage
- δ(T): 96% coverage
- β(T): 93% coverage
- γ: 95% coverage



Model Extension - “Model 5”

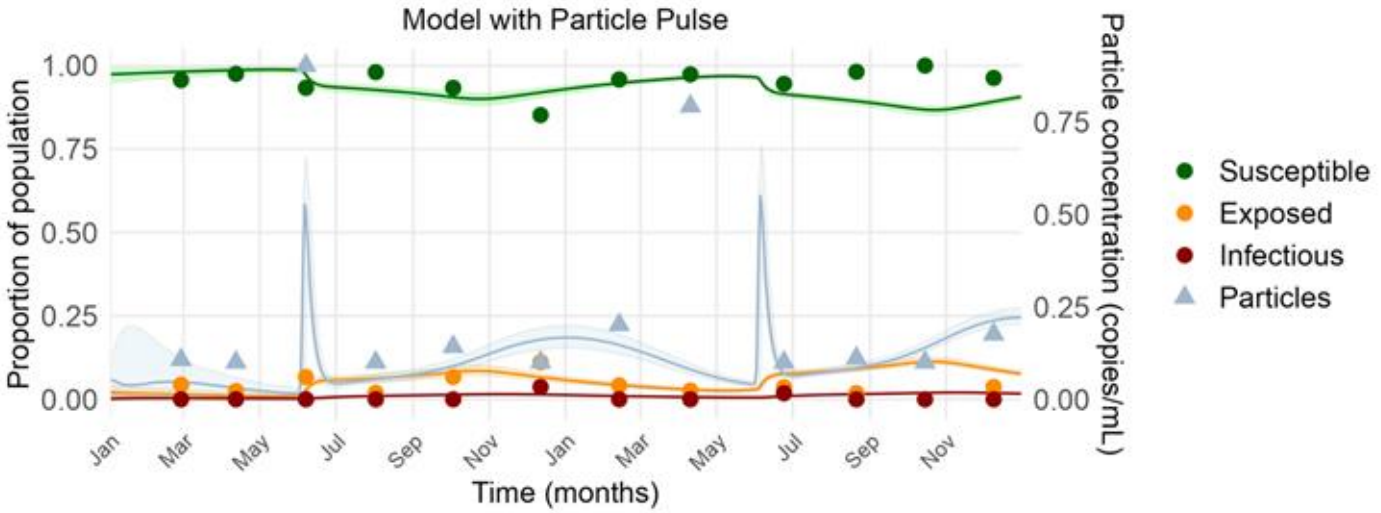
- Spawning could be a reason to why there are summer spikes that the model cannot capture mechanistically
- Introducing a particle spike in summer spawning months is a possible model extension
- Limited knowledge on spawning and cancer particle emission, and timing of spawning
- But still a good theoretical experiment

- Define a pulse equation to introduce an increase in P for a duration of 3-10 days and in the months of May/June (timing and duration inferred by model)

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

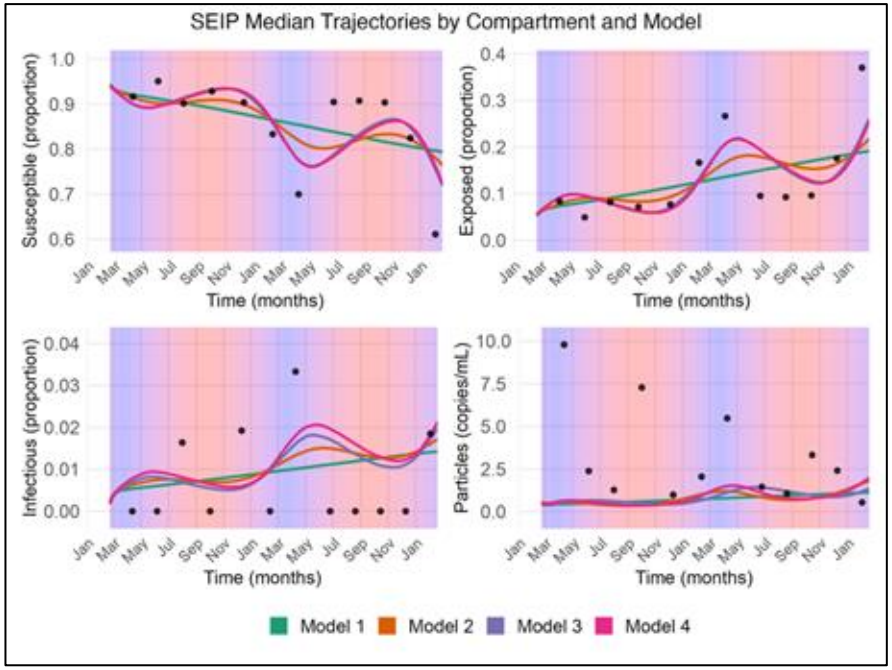
Model Extension - “Model 5”

Ran with fixed parameters from Model 4 and inferring timing/duration of pulse



Second Site - Pavilion Beach

- All model trajectories are showing similar patterns
- Seasonal drivers cannot be distinguished
- PB does not contain enough information to identify the seasonal mechanism
- Tests generality of the Forest River results
- It seems that the April prevalence point is highly influential in the model



Data Limitations

- Available data (laboratory and field data) impose several limitations on model inference
  - Progression parameters
  - Limits of detection (S and P)
  - Survey data resolution
  - Low levels of Infectious clams
- Mechanistic piece of the model is missing
  - Spawning
  - Tidal dynamics

1. Introduction
2. Research Goals
3. Methods
4. Results
5. **Conclusions**

Conclusions

- **Epidemiological Model of BTN -** Synthesized laboratory experiments and field surveys to quantify progression and transmission parameters.
- **Disease Seasonality and Driving Force in Forest River -** Temperature-dependent progression is the driving force of BTN seasonality in Forest River. Under these model assumptions, increasing ocean temps will increase prevalence.
- **Model Generality -** Seasonality is identifiable in Forest River, but not in Pavilion Beach under the same model, indicating differences in the site-specific processes governing BTN dynamics.
- **Foundational Model -** Can be updated with new information of BTN dynamics.
- **Bayesian Disease Modeling -** Demonstrated Bayesian modeling in an epidemiological system to uncover underlying processes even with sparse, noisy, and limited data.

Future Directions

- Compare model predictions to 2025 collection for validation
- Apply to different sites to assess differences in transmission across populations and test with other temperature gradients
- Apply to other species
  - Cockles

Puget Sound, WA

East Coast, MA & ME

Thank You

Adviser: Dr. Michael Metzger, University of Washington

Committee

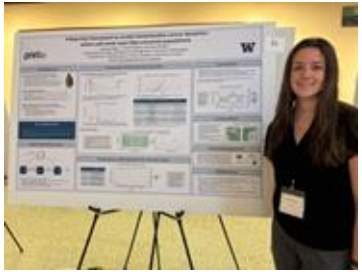
- Dr. Rob Noble, City St. George's, University of London
- Dr. Chelsea Wood, University of Washington
- Dr. Alison Feder, University of Washington

Metzger Lab

- Fiona Garrett
- Peter Smith
- Zach Child

Past lab members

- Sydney Weinandt
- Marisa Yonemitsu
- Rachael Giersch
- Sam Hart
- Jordana Sevigny



GLOUCESTER MARINE GENOMICS INSTITUTE

UNIVERSITY of WASHINGTON

NIH

pnri Pacific Northwest Research Institute

NSF