The role of host immunity and the environment in seasonal disease dynamics

Gabriel Kosmacher, Dillon Max, Zoi Rapti, Carla E. Cáceres and Tara E. Stewart Merrill

February 18, 2025

Abstract

In both human and wildlife disease systems, seasonal shifts in host immunity may shape the timing and severity of epidemics. Yet, the immune responses that govern infection, as well as seasonal patterns in their expression, are typically not observed. Rather, field studies collect phenomenological data on infection outcomes. Pairing epidemic data with models that directly parameterize immune metrics can then be a powerful approach for exploring the role of seasonally varying immunity on disease. This may be particularly successful when datasets include multiple outbreaks in the same host-pathogen system. Long-term data sets such as these can be used to determine how well a parameterized model follows trends in the field data, and whether deviations in the parameters can reproduce the differences observed among outbreaks.

Previous work in the *Daphnia dentifera-Metschnikowia bicuspidata* focal host-fungal pathogen disease system has not taken full advantage of coupling patterns in nature with mechanisms predicted by theory. Here, we study a mathematical model that takes into account host immunity in the form of resistance to and recovery from *M. bicuspidata* infections and seasonal variation in key aspects of the system's epidemiology and ecology. Specifically, host population birth rates, predation, transmission, and recovery rates, as well as the fungal spore release rate were allowed to vary within the season. Our simulations revealed that modification of a system's carrying capacity could produce good correspondence between observed and model-estimated densities. By modifying the transmission, spore release, and the fraction of recovering hosts, we were able to well capture the timing of disease outbreaks, as well as other qualitative features of outbreaks, such as the disparity between the prevalence of early and late infections. Our model results suggest that host immunological parameters are an important within-host constraint on disease dynamics.

1 Introduction

Parasites and pathogens are regulated by two environments – the host itself, and the environment in which the host lives. Both of these environments are subject to seasonal variation, which when combined can create temporal variation in the dynamics of disease [1]. For instance, in environmentally transmitted pathogens, free-living infective stages may be affected by seasonal variation in UV radiation and temperature [16, 17, 23, 24, 25]. The host environment, and the immune defenses that comprise it, can also shift seasonally with changes to resources, competitors [7, 20], and

predators [11, 7, 20] [25]. The contributions of the simultaneous seasonal changes in the within-host and external environment can be difficult to separate with field patterns, making model predictions an essential tool in disease ecology.

The interaction between *Daphnia* hosts and *Metschnikowia* fungal pathogens, has been a powerful system for bridging theory to empirical work [30, 14, 12, 24, 25]. Nearly two decades of data-theory coupling have explored many aspects of this system, often by varying one parameter at a time. As a result, we know that many, but not all, *Daphnia* hosts are capable of mounting immune responses that prevent *Metschnikowia* infection [30, 29, 31]. Hosts can recover from initial exposure to the pathogen, a mechanism termed barrier resistance, or may clear the infection later, a mechanism referred to as internal clearance. Additionally, the host may limit within-host pathogen growth, leading to reduced spore yield [36]. Resources strongly influence disease dynamics [13, 4], as does the presence or absence of particular predators or competitors [11]. Although all these ecological interactions vary seasonally, previous work has not specifically included the seasonal dynamics of the two environments faced by the parasite. Here, we investigate the role of seasonal variation in a suite of epidemiologically relevant parameters. In the host environment, we modify the pathogen's transmission rate, host immunity, and spore yield. In the extrinsic environment, we modify predation, host birth rate, and carrying capacity. All three of these immunological processes may vary seasonally, likely in response to changing resources, predation, and clonal selection.

In addition to the within-host processes that are governed by the immune response, transmission rate, which depends on both encounters with the pathogen per time and the probability of infection given sufficient contact, also greatly affects the progression of an outbreak [1, 26]. There are numerous examples of how seasonality can influence transmission [3, 15]. Contact rates in the case of *Daphnia* are affected by external factors, such as storms and the ensuing convection in lakes which increases the contact rate between spores and hosts [12] and temperature [25]. Predation, which varies seasonally, has long been known to affect *Daphnia* epidemics caused by a bacterial parasite (*Pasteuria ramosa*) [11]. More recently, it has also been shown to affect the competition of *P. ramosa* with *M. bicuspidata* [2]. Specifically, sloppy predators such as midge larvae (*Chaoburus spp.*) can enhance the contact between hosts and fungal spores, by releasing the spores into the water column while preying on infected hosts [6]. In the other extreme, some fish predators, like the bluegill sunfish, selectively prey on opaque infected *Daphnia* hosts that are visually conspicuous, thus effectively removing the fungal spores from the water column [11]. However, it does not mean that selective fish predation alone can eliminate the risk of epidemics since outbreaks occurred in a lake-enclosure experimental study under various predation intensities [10].

The goal of our study is to reproduce the distinct epidemic patterns observed in the *Daphnia-Metschnikowia* system and evaluate the importance of immunological parameters in generating them. In six Indiana lakes that were studied extensively through biweekly sampling [29], prior research documented unique temporal dynamics in disease outbreaks. In some lakes, there was an initial summer peak in the prevalence of late infections followed by a second and considerably higher peak in late fall, whereas in other lakes, a first peak was absent or negligible. These dynamics could be further differentiated by the extent to which early infections (from which the host is able

to recover) aligned with or were independent of, late infections (which are terminal). *Daphnia* exposed to fungal spores can clear early infections before they develop into late and transmittable infections, Fig. 1. Thus, these patterns suggest that in addition to seasonal changes in predation seasonally varying host immunity may play a role in generating the temporal dynamics of this disease.

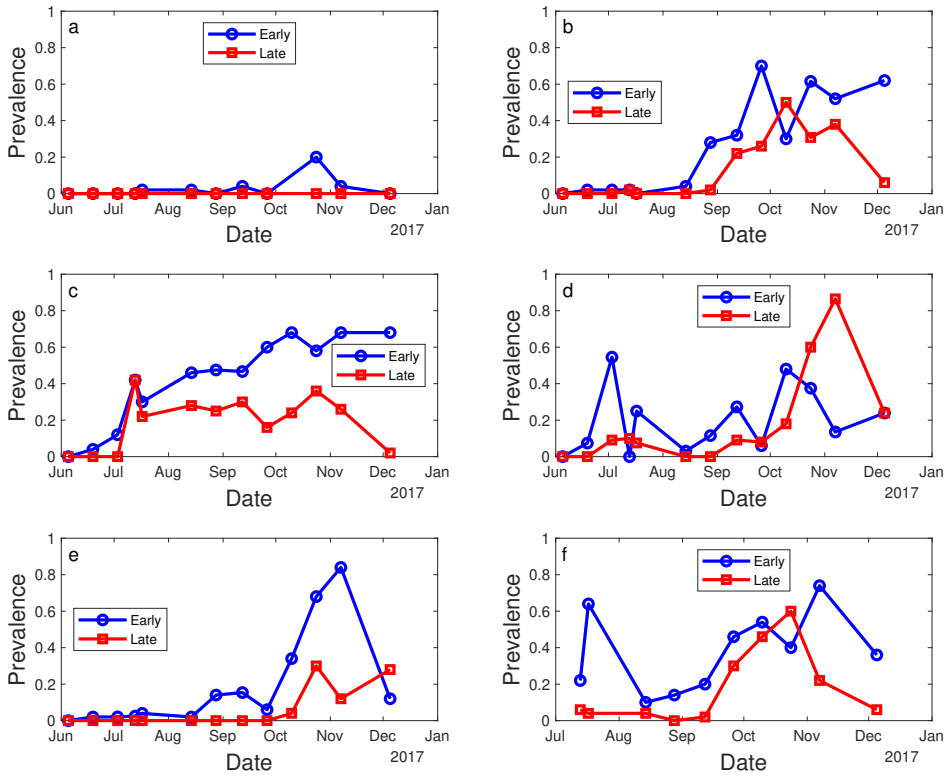

Figure 1: Field data for early (blue dots) and late (red squares) prevalence in Beaverdam (panel a), Benefiel (panel b), Downing (panel c), Hale (panel d), Midland (panel e), and Star (panel f) lakes from [29].

The article is structured as follows. In subsection 2 we describe our model with seasonally varying parameters. In section 3 we first explore the structural parameter identifiability properties of the model and then present results of model simulations where the role of the within-host environment and the extrinsic environment is teased apart. Finally, in section 4 we summarize our findings and outline future research directions. Additional details are included in the Appendix.

2 Model

We focus our investigation on a compartmental model for the densities of the host (susceptible, S, exposed (non-terminally infected), E, (terminally) infected, I) and free-living fungal spores, Z. A

diagram of the model is shown in Fig. 2. One main feature setting us apart from previous models is that due to recent findings [31], we allow non-terminally infected hosts E to recover and move back into the susceptible class, S. Moreover, we incorporate seasonal variation in environmental and within-host characteristics to better capture the qualitative transmission dynamics observed in the field. We model five parameters (Fig. 3) as time-dependent based on multiple studies showing seasonal variation in these parameters in the field [9, 26, 27, 29, 34].

The model equations read

$$\frac{dS}{dt} = B_t b(t)(S + \rho(E + I)) \left(1 - \frac{S + E + I}{K}\right) - (d + P_t p(t))\, S - b_t \beta(t) SZ + k_t k(t) \alpha E \quad (1)$$

$$\frac{dE}{dt} = b_t \beta(t) SZ - (d + P_t p(t))\, E - \alpha E \quad (2)$$

$$\frac{dI}{dt} = (1 - k_t k(t))\alpha E - (d + v + \theta P_t p(t))\, I \quad (3)$$

$$\frac{dZ}{dt} = \Sigma_t \sigma(t)\, (d + v)\, I - \lambda Z - f\, (S + E + I)\, Z \quad (4)$$

Susceptible *Daphnia*, S, grow at a time-dependent rate $B_t b(t)$, which is driven by seasonality according to $b(t)$, where t denotes time and B_t is a constant parameter whose role is to modify the magnitude of the time-varying growth rate $b(t)$. Density-dependent effects on birth rates are well known to vary seasonally in *Daphnia* populations [34]. All hosts give birth to susceptible individuals, and the birth rate of the exposed and infected classes is reduced by ρ. Births by all classes are subject to density dependence via the carrying capacity K. Hosts die at a background mortality rate of d and due to predation by fish which varies throughout the year [27, 34] so it becomes time-dependent at a rate $P_t p(t)$. Again, we use $p(t)$ to denote the time-varying parameter and the constant parameter P_t to modify its magnitude. Hosts become infected at a rate $b_t \beta(t)$ after ingesting fungal spores, Z, existing in the water column. The infection rate $b_t \beta(t)$ depends on the contact rate between hosts and spores and the probability of infection given such a contact, and increases with foraging rate (i.e., exposure increases transmission) and decreases with resistance (i.e., immune barriers reduce transmission). This rate is also time-dependent due to the likelihood of seasonal variation in *Daphnia* immune responses [29]. Once hosts ingest the spores, they move, via $b_t \beta(t) SZ$ into the exposed *Daphnia* compartment, E. Finally, individuals are added back to the susceptible class via recovery from early infection at a rate α. The fraction that can recover likely varies in time, due to seasonally varying immunity, and is modified by $k_t k(t)$.

Exposed *Daphnia*, E, are assumed to suffer from the same death rate as susceptible hosts. A proportion $k_t k(t)$ of them recover at a rate α [31, 33]. The remaining proportion $1 - k_t k(t)$ moves into the infected class, I, at the same rate α.

Infected *Daphnia*, I, suffer an increased predation rate, denoted by the parameter $\theta > 1$ because infection renders them opaque and more conspicuous to visual predators. They also die at an increased rate, due to disease-induced mortality, v.

Fungal spores, Z, are released in the water column by dead infected hosts, I, according to $\Sigma_t \sigma(t)$ which accounts for the spores contained within each infected host. Due to seasonally varying

immunity, this spore yield is assumed to vary within the season. Spores inside hosts that are eaten by fish are assumed lost and do not contribute to this term. Spores are eliminated from the water column due to sinking and UV destruction at a constant rate λ and are also eliminated by *Daphnia* ingestion at a constant rate f. A diagram of the model is shown in Fig. 2.

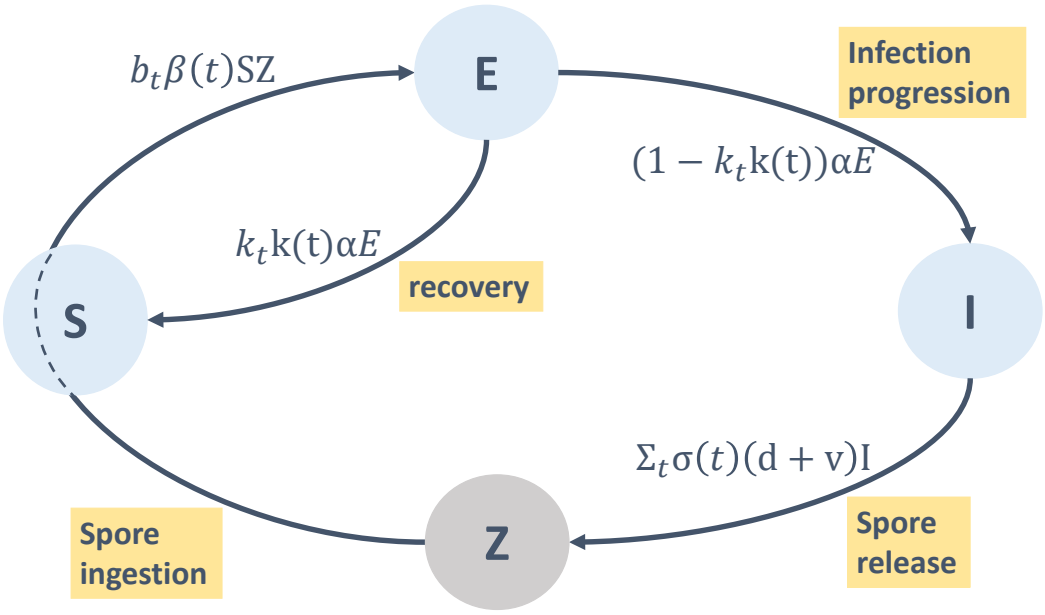

Figure 2: A diagram of the seasonal SEIZ model described by (1-4).

For the time-varying parameters, we assume that the birth rate $b(t)$, transmission rate $\beta(t)$, predation rate $p(t)$, and the spore release parameter $\sigma(t)$ increase at the beginning of the season, reach a peak and then decrease in the fall (Fig. 3, panels a-c and e). On the other hand, the fraction of hosts able to recover $k(t)$ exhibits the opposite trend (Fig. 3, panel d). The default time-varying parameters are modified by being multiplied by constant prefactors, B_t for $b(t)$, P_t for $p(t)$, b_t for $\beta(t)$, k_t for $k(t)$ and Σ_t for $\sigma(t)$. The benefit of this notation is that by varying the prefactor we vary the average of the time-varying parameter by the same factor, i.e., setting $B_t = 1.2$ increases the average of $B_t b(t)$ by 20%. Hence, it is a convenient way to make comparisons. Parameter values are based on previous publications as noted in Table 1.

3 Results

In this section, we present local parameter identifiability and simulation results for model (1-4) using the time-varying parameters shown in Fig. 3. We note that our ultimate goal is not to obtain the most accurate fit of the model to the data. Instead, we are interested in understanding the mechanisms driving the observed patterns. For this reason, we are only varying a few key parameters and studying how they affect the model output. It has been proved that SIR and SEIR type models can be fitted to almost any prevalence time series as long as the transmission coefficient is time-varying [19]. However, it was also shown that the inverse problem is under-determined which

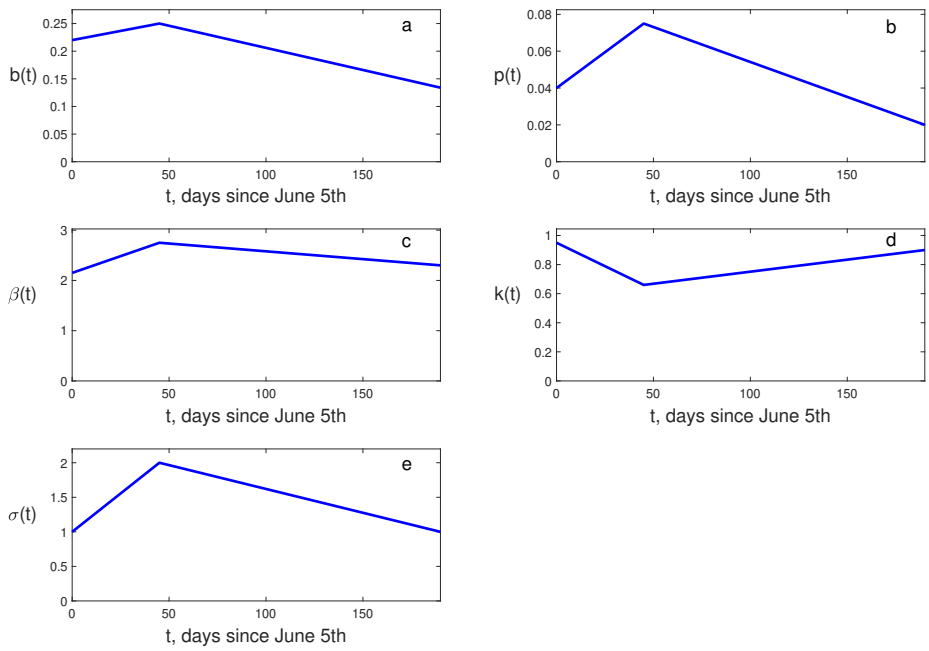

Figure 3: Time varying model parameters. Birth rate $b(t)$ (panel a), predation rate $p(t)$ (panel b), transmission rate $\beta(t)$ (panel c), fraction of hosts able to recover $k(t)$ (panel d), and spore release from dead infected hosts $\sigma(t)$ (panel f).

increases the risk of over-fitting. Hence, we focus on qualitatively matching the epidemic patterns observed in the field with the understanding that more than one parameter combination may result in qualitatively similar epidemic curves.

3.1 Local identifiability analysis

We use the software package STRIKE GOLDD [35] to study the structural identifiability properties of the time-varying parameter model (1-4). Parameters that are structurally identifiable can be determined uniquely (namely, are globally identifiable) or up to finite number of values (that is, are locally identifiable) from noise-free observations. Structural identifiability does not take into account the length and quality of the observations. In reality, fitting models to time series is even more challenging due to practical identifiability issues [21]. These arise when there is insufficient data, e.g., too short time-series, and/or noisy data.

We analyzed model (1-4) considering as observations the total density $N = S + E + I$, the early infection $P^E = \frac{E}{S+E+I}$ and late infection $P^I = \frac{I}{S+E+I}$ prevalence. We assumed that the values of parameters $\{\rho, d, \alpha, v, \theta, \lambda, f\}$ are known. This is consistent with published studies. For instance, the fecundity reduction ρ due to the infection and disease progression/recovery have been recently studied in [8, 28] and [31], respectively. The time-varying functions (model inputs) $b(t)$, $p(t)$, $\beta(t)$, $k(t)$ and $\sigma(t)$ were considered known, while $\{B_t, K, P_t, b_t, k_t, \Sigma_t\}$ are considered unknown. It was

Symbol	Definition	Value
S	susceptible population	- mg C/l
E	exposed population	- mg C/l
I	infected population	- mg C/l
Z	spores in water column	- mg C/l
t	time	- days
b	growth rate	0.134-0.25 day^{-1} [6]
ρ	decreased fecundity coefficient	0.85 [6, 7]
K	carrying capacity	2-20 mg C/l [7]
d	background mortality rate	0.03 day^{-1} [6]
p	predation rate	0.02-0.075 day^{-1} [6]
β	transmission rate	2.15-2.75 l $(\text{mg C day})^{-1}$ [7]
k	fraction of recovering hosts	0.66-0.95 [31]
v	disease-induced mortality rate	0.08 day^{-1} [6]
θ	fish predation intensity constant	3 [6]
α	infection progression rate	0.5 days^{-1} [31]
σ	spore biomass per dead infected host biomass	1-2 mg C/l l/mg C [7]
λ	spore loss rate	0.1 day^{-1} [6]
f	spore removal rate by hosts	0.2 l $(\text{mg C day})^{-1}$ [7]

Table 1: Model variables & parameters. For time-varying parameters, b, p, β, k and σ, their range is shown.

determined that parameters $\{B_t, K, P_t, k_t\}$ are locally structurally identifiable, while $\{b_t, \Sigma_t\}$ are not. Moreover, the initial condition for $Z(t)$, since it is considered unknown, is unidentifiable too. In reality, however, we are not able to know the model inputs, so we also run the analysis by considering $u_1(t) = B_t b(t)$, $u_2(t) = P_t p(t)$, $u_3(t) = b_t \beta(t)$, $u_4(t) = k_t k(t)$, and $u_5(t) = \Sigma_t \sigma(t)$, as unknown linear inputs and K as an unknown parameter. It was determined that K is identifiable, u_1, u_3 and u_4 are observable, but u_2 and u_5 are not observable, neither is the initial condition for the spore density, $Z(0)$.

3.2 Simulation results

In this subsection we present simulation results with parameter values and initial conditions for the ten best fits for each lake. Details of the fitting are in the Appendix. We used least squares fitting with a cost-function given in Eq. (5) and having $\{B_t, K, P_t, b_t, k_t, \Sigma_t\}$ as free parameters and Z_0 as a free initial condition. Parameter ranges are shown in Table 3 in the Appendix.

As shown in Fig. 1, late-stage infections (characteristic of an epidemic) were not detected in Beaverdam Lake even though early infections were present in up to 20% of observed hosts. Qualitatively, these patterns suggest that early-stage infections were not developing into late infections, or were so at a level below detection. This pattern, a moderate proportion of early-stage infections, E/N, where $N = S + E + I$ is the total host density, and a low proportion of late-stage infections, I/N, is

captured by our model when the scaling factor for the growth rate B_t ranges from $1-2.7$, carrying capacity K ranges between 11 and 23, predation rate is increased to $P_t = 2.1-5.4$, transmission rate b_t is between 0.7 and 2.9, the scaling factor for the fraction of recovering hosts is in the range of $k_t = 0.96-1.02$, the spore release factor Σ_t ranges from 0.65 to 2.6 and the initial spore density $Z_0 = 3.7 \cdot 10^{-6} - 1.5 \cdot 10^{-4}$. All other parameters are at their default values, shown in Table 1. In Fig. 4 we show ten simulated curves with the smallest mean squared error (ten best fits). The parameters clouds that generate these ten curves are shown in Fig. 5. The color coding in Figs. 4 and 5 is consistent. That is, each curve's color is the same as the color of the parameters that generated it.

Figure 5 suggests that to capture the field pattern, growth rate and predation rate must be balanced, namely there is a positive relationship between them (the Pearson coefficient is 0.83 for the scatter plot of panel a, Fig. 5 and the Spearman coefficient is 0.78). There also exists a negative relationship between spore yield and transmission rate in the 10 best fits (Fig. 5, panel c) with a Spearman coefficient equal to -0.74. We also note that the fraction of recovering hosts is high, $k_t \approx 1$. This is consistent with the hypothesis that a high fraction of recovering hosts does not lead to terminal infections, hence the peak in the early-stage infection prevalence can be larger than the peak in late-stage infection prevalence.

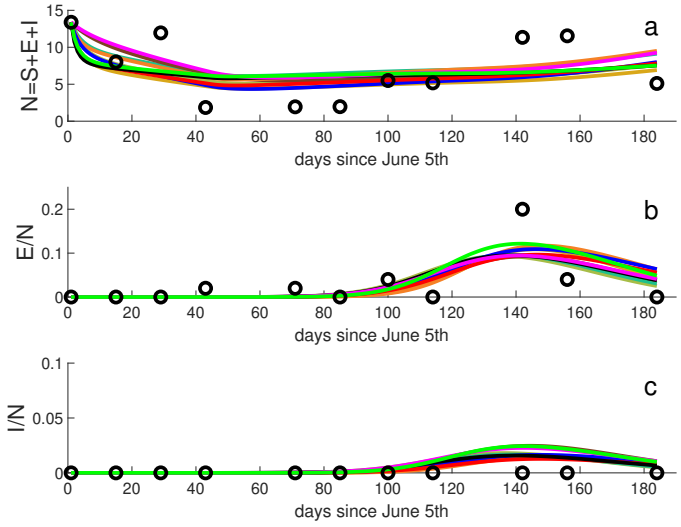

Figure 4: Model simulations capturing the Beaverdam Lake epidemic pattern. Results from the ten simulations with the smallest mean squared errors are shown for the total host density (panel a), early infection prevalence (panel b) and late infection prevalence (panel c).

In Benefiel Lake, early-stage infection prevalence reached almost 70% and late-stage infections followed closely in time (Fig. 1). This outbreak is characterized by a single peak (single wave in pandemic-era parlance) and occurred after mid-August. Our model results, Fig. 6, do not match quantitatively the observed patterns: early-stage infection prevalence is well below 70%. It should be noted that the initial density for the spores, Z_0, is several orders of magnitudes smaller than

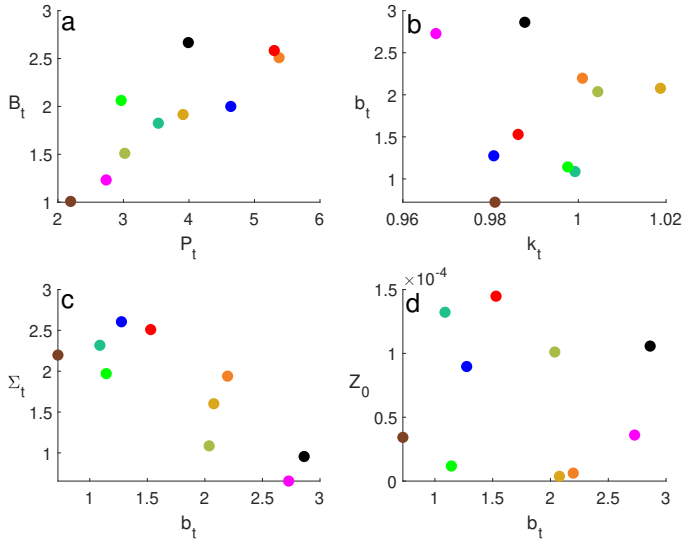

Figure 5: Parameter clouds generating the plots in Fig. 4 for the Beaverdam Lake. A wide range of scaling factors for the growth rate and predation rate (panel a), transmission rate and recovery fraction (panel b), spore yield and transmission rate (panel c), and initial spore density and the scaling factor for the transmission rate (panel d) produce comparable total density and prevalence curves. The color coding is consistent with that of Fig. 4.

in the Beaverdam simulations (Fig. 5, panel d) in order to delay the onset of the epidemic. In the parameter clouds for Beaverdam, there is a significant correlation (Pearson coefficient 0.9561, Spearman coefficient 0.9394) between the growth rate B_t and the predation rate P_t prefactors (Fig. 7, panel a) and a nonlinear relationship (Spearman coefficient is -0.9636) between the spore release, Σ_t, and transmission coefficient, b_t, prefactors (Fig. 7, panel c). For the ten best-fit simulations shown the carrying capacity varied from $K = 11$ to 23.4.

Downing Lake had a sustained epidemic that began in July and lasted until the end of the season. Early-stage infections were immediately followed by late-stage infections, both classes of which grew to high prevalence values. Specifically, early-stage infection prevalence reached almost 70%, and late-stage infection prevalence was higher than 20% from mid-July until mid-November. Our model captures this pattern, Fig. 8. The values of the carrying capacity K were between 6.3 and 7.3. In the parameter clouds we observe a negative relationship between the spore release Σ_t and transmission b_t factors (the Spearman correlation coefficient is -0.9394).

The model results for Hale Lake do not match the field data well, Fig. 10. The increase in the early infection prevalence between days 20 and 40 was missed by the model and the fluctuations in the field data between days 80 and 140 were also not present in the model simulations (panel b). The late infection prevalence trends were captured in a more acceptable fashion, since there was only a minor peak in the beginning of the season and a second peak later (panel c). Host total density experienced fluctuations (panel a) which may have affected the prevalence results. The parameter

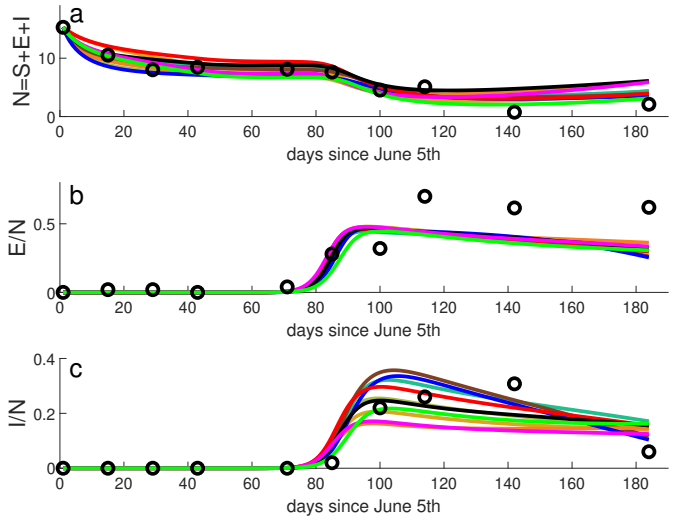

Figure 6: Model simulation capturing the Benefiel Lake epidemic pattern. Results from the ten simulations with the smallest mean squared errors are shown for the total host density (panel a), early infection prevalence (panel b) and late infection prevalence (panel c).

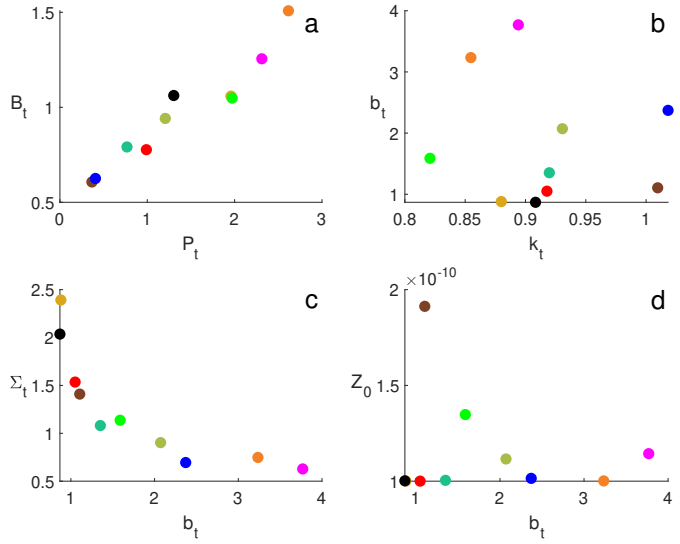

Figure 7: Parameter clouds generating the plots in Benefiel Lake, Fig. 6. A wide range of scaling factors for the growth rate and predation rate (panel a), transmission rate and recovery fraction (panel b), spore yield and transmission rate (panel c), and initial spore density and the scaling factor for the transmission rate (panel d) produce comparable total density and prevalence curves.

clouds generating these best fits are shown in Fig. 11. Host carrying capacity varied in the ten best fits from $K = 2.2$ to $K = 2.5$. There is a negative correlation between the prefactors of the spore release parameter Σ_t and transmission rate b_t (the Spearman correlation is -0.9273).

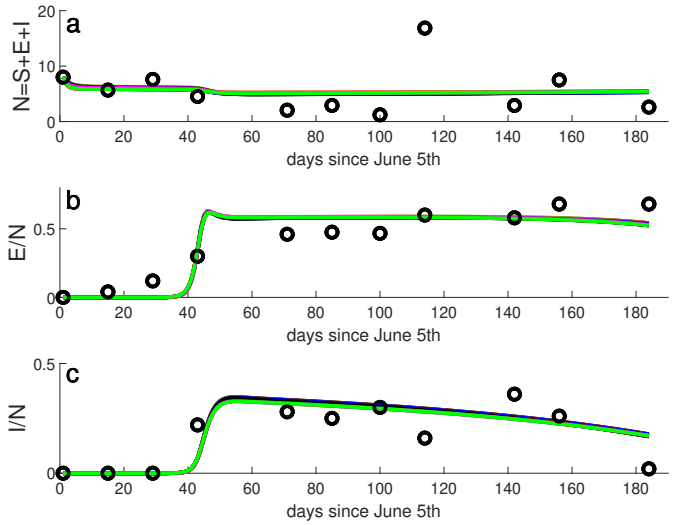

Figure 8: Model simulation capturing the Downing Lake epidemic pattern. The ten fits with the smallest errors are shown.

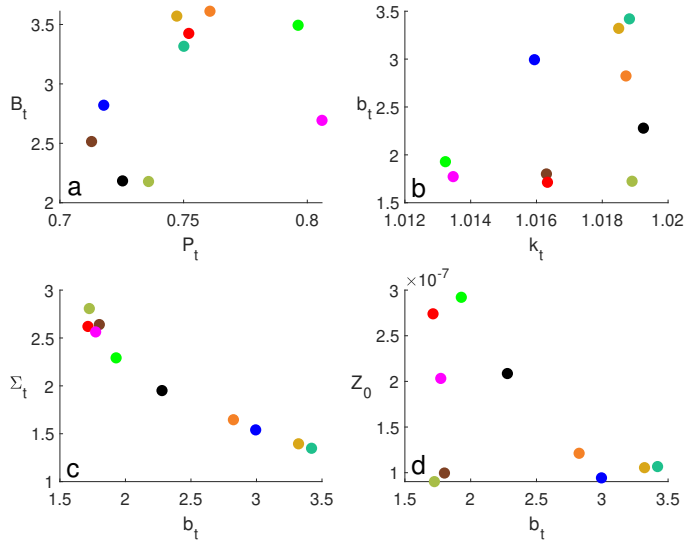

Figure 9: Parameter clouds generating the plots in Downing Lake, Fig. 8. A wide range of scaling factors for the growth rate and predation rate (panel a), transmission rate and recovery fraction (panel b), spore yield and transmission rate (panel c), and initial spore density and the scaling factor for the transmission rate (panel d) produce comparable total density and prevalence curves.

Midland Lake experiences an epidemic late in the season. Early-stage infections occur at low prevalence for months before late-stage infections appear. Once the epidemic began, early-stage infection prevalence reached a maximum of more than 80%, while late-stage infection prevalence reached a maximum of about 30%. Host immune responses and/or high predation may be responsible for the

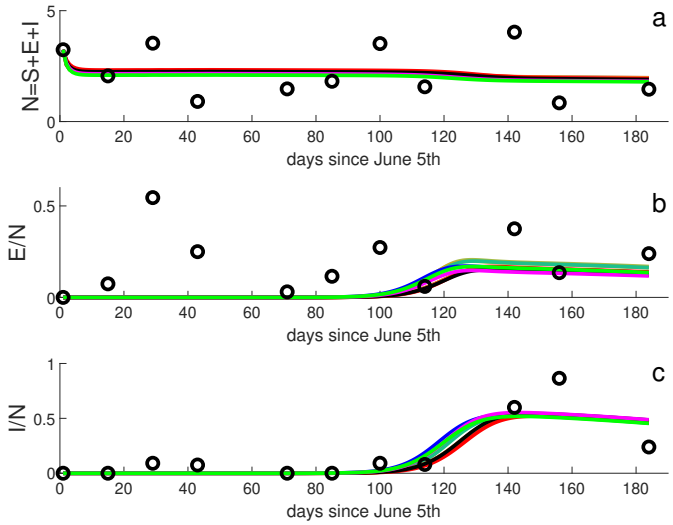

Figure 10: Model simulation for the Hale Lake epidemic pattern. The ten fits with the smallest errors are shown.

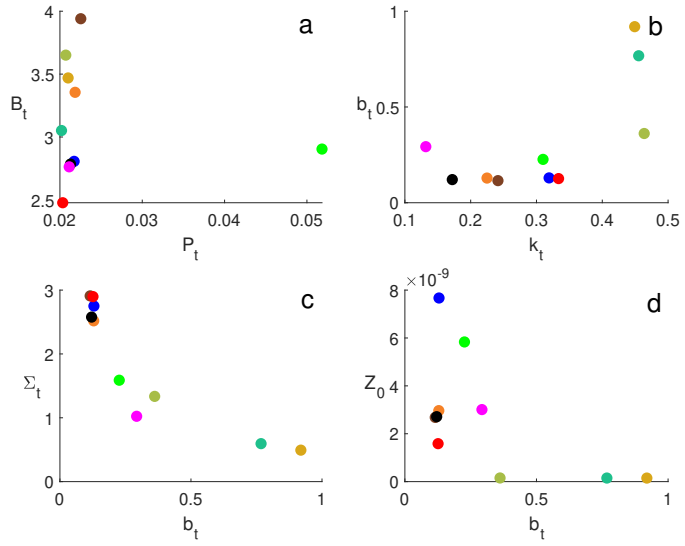

Figure 11: Parameter clouds generating the Hale Lake plots, Fig. 10. A wide range of scaling factors for the growth rate and predation rate (panel a), transmission rate and recovery fraction (panel b), spore yield and transmission rate (panel c), and initial spore density and the scaling factor for the transmission rate (panel d) produce comparable total density and prevalence curves.

difference in both the temporal dynamics and prevalences of these two classes. Our model is once again inadequate to match the observed field pattern. The simulated early infection prevalence is lower and starts later than the observed one (Fig. 12, panel b) and the simulated late infection prevalence stays at levels lower than the observed ones (Fig. 12, panel c). The ten best fits showed

large variation in the host carrying capacity K which ranged from 8 to 18.1. The remaining parameter clouds are shown in Fig. 13. There is again a correlation between the prefactors of the spore release parameter Σ_t and transmission rate b_t (the Spearman correlation is equal to -0.903).

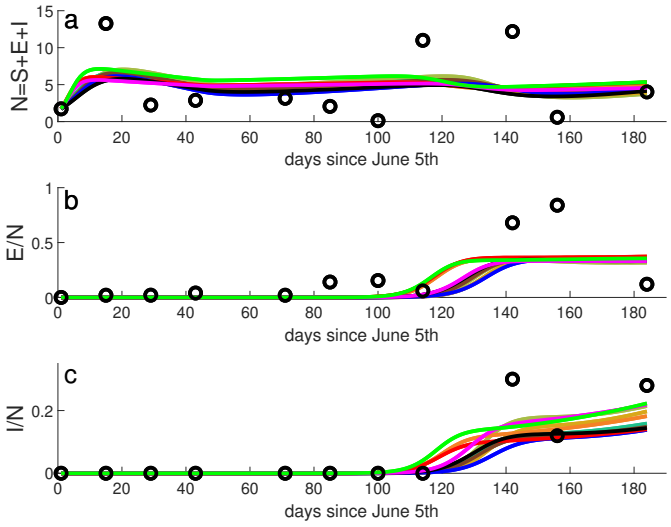

Figure 12: Model simulation for the Midland Lake epidemic pattern. The ten fits with the smallest errors are shown.

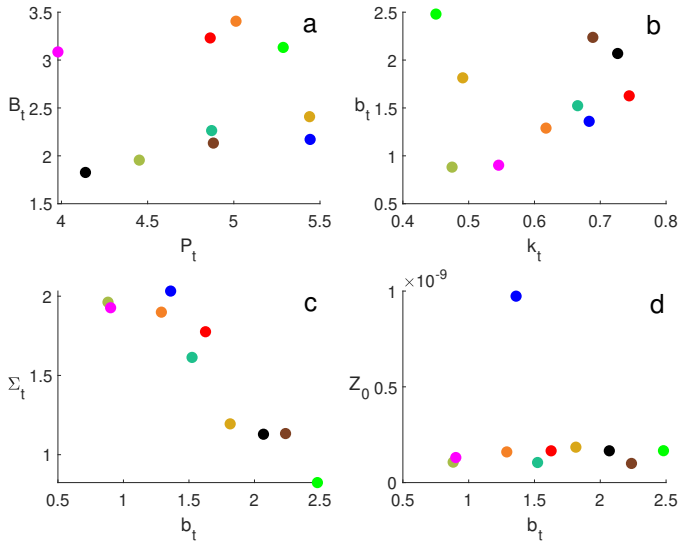

Figure 13: Parameter clouds generating the Midland Lake plots, Fig. 12. A wide range of scaling factors for the growth rate and predation rate (panel a), transmission rate and recovery fraction (panel b), spore yield and transmission rate (panel c), and initial spore density and the scaling factor for the transmission rate (panel d) produce comparable total density and prevalence curves.

Star Lake had a unique pattern of infections Fig. 14. While early-stage infections exceeded 60% in

July, they rapidly declined only to rebound later in the season. Late-stage infections were rare in the early season and did not map well to early-stage infections. However, later in the season, there was a close correspondence between the timing and prevalence of the two infection types. The early temporal mismatch suggests that immune responses may have been preventing late-stage infections in July, but that immunity may have eroded over the season, allowing late-stage infections to eventually become abundant. We could not recreate these unique dynamics with the time-varying parameters. The sampling period for Star Lake was shorter and this may affect the results of the numerical study. The host carrying capacity, K, for the curves in Fig. 14 was between 19 and 24. The rest of the fitted parameter values are shown in Fig. 15. There is a positive correlation between the prefactors for the growth rate, B_t, and predation rate, P_t, shown in panel a of figure 15 (the Spearman correlation is 0.6727). There is also a negative nonlinear relationship between the prefactors of the spore release parameter, Σ_t, and transmission rate, b_t, shown in panel c (the Spearman correlation coefficient is -0.9639).

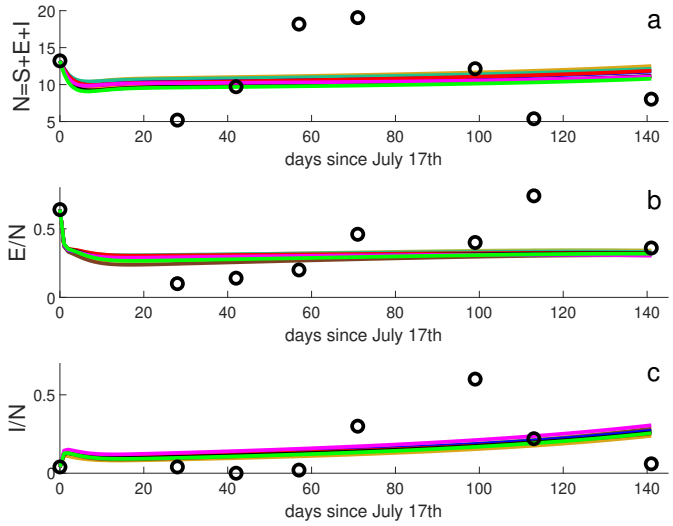

Figure 14: Model simulation for the Star Lake epidemic pattern. The ten fits with the smallest errors are shown.

3.3 Possible mechanisms explaining observed patterns

We also created aggregated parameter clouds by combining the parameters from all six lakes. Our goal is twofold. First, we aim to look for and compare any signatures that specific lakes may have. Our second objective is to determine whether some of the trends (such as the positive correlation between B_t and P_t and the negative correlation between Σ_t and b_t) evident in isolated lakes persist and whether new correlations emerge. The parameter clouds generated from the 60 best fits for each of the seven fitted parameters are shown in Fig. 16. We note that the Star lake data set starts much later (July 17th) than the other five lakes (June 5th). Hence, when we compare the values for the initial spore density, $Z(0)$, time zero refers to different dates.

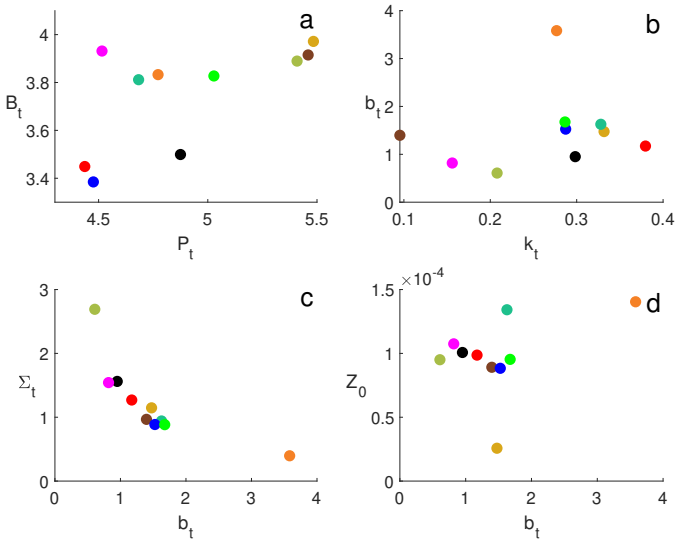

Figure 15: Parameter clouds generating the Star Lake plots, Fig. 14. A wide range of scaling factors for the growth rate and predation rate (panel a), transmission rate and recovery fraction (panel b), spore yield and transmission rate (panel c), and initial spore density and the scaling factor for the transmission rate (panel d) produce comparable total density and prevalence curves.

In the aggregated parameter clouds, there is a positive correlation (the Spearman coefficient is 0.5736) between the prefactor of the transmission rate, b_t, and the prefactor of the fraction of recovering hosts, k_t, as seen in Fig. 16, panel b. Our argument for explaining this positive correlation is that a higher transmission rate compensates for a higher fraction of recovering hosts. We must recall that only dead terminally infected individuals produce spores, hence as the fraction of recovering hosts increases (which translates into fewer free-living infectious spores) this must be compensated with higher transmission rate to achieve the same prevalence.

We also found a negative correlation (the Spearman coefficient is -0.4497) between Σ_t and b_t (panel c). We interpret this trend as follows. The parameters b_t and Σ_t are not identifiable and they both control transmission, b_t being the prefactor of the transmission rate and Σ_t being the prefactor of the spore yield. Hence, a negative correlation between them is not unexpected, as higher transmission rate can compensate for lower spore yield.

Another positive correlation (the Spearman correlation is 0.7174) exists between the prefactor of the predation rate, P_t, and the host carrying capacity, K, (panel e). While these parameters are locally identifiable, structural (theoretical) identifiability does not imply practical identifiability. Thus, in the absence of practical identifiability, high predation rate is compensated by high host carrying capacity to keep the host density at the required level. Another manifestation of this correlation is that when the variation in the fitted values for P_t is small, so is the variation in K. This can be seen in panels a, e and f of Fig. 16 for Downing (green dots) and Hale (gray dots).

Looking at the role of immunity, modeled through the transmission rate b_t and the fraction of

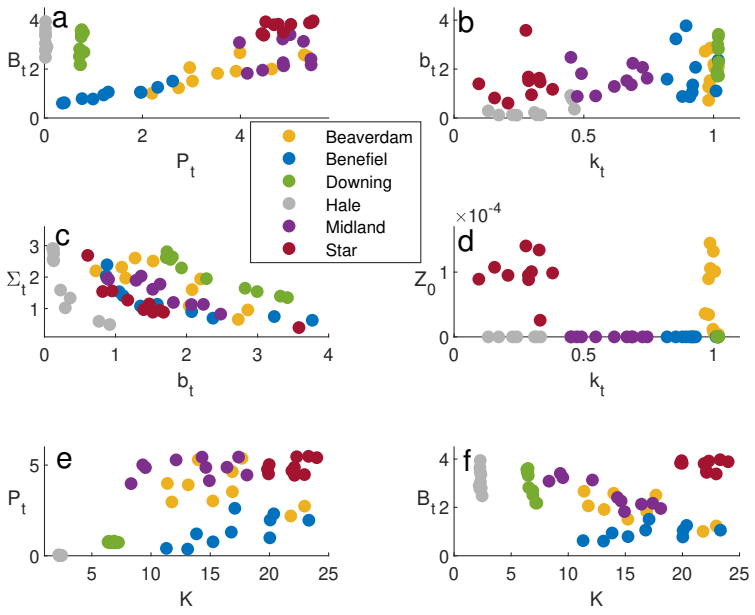

Figure 16: Parameter clouds for all six lakes. The clouds are aggregated from the 10 best fits for each individual lake.

	Beaverdam	Benefiel	Downing	Hale	Midland	Star
max P^E sim.	0.12	0.47	0.61	0.20	0.36	0.64
max P^E obs.	0.2	0.7	0.68	0.54	0.84	0.74
max P^I sim.	0.025	0.36	0.33	0.55	0.22	0.3
max P^I obs.	0	0.30	0.36	0.86	0.3	0.6
P_t	(2, 5.5)	(0.36, 1.5)	(0.71, 0.81)	(0.02, 0.052)	(4.1, 5.5)	(4.43, 5.49)
b_t	(0.72, 2.8)	(0.88, 3.7)	(1.72, 3.4)	(0.11, 0.92)	(0.88, 2.49)	(0.60, 3.58)
k_t	(0.96, 1.03)	(0.8, 1.03)	(1.01, 1.03)	(0.13, 0.47)	(0.45, 0.75)	(0.09, 0.38)
Σ_t	(0.65, 2.6)	(0.62, 2.4)	(1.34, 2.81)	(0.49, 2.91)	(0.82, 2.03)	(0.39, 2.69)

Table 2: Simulated and observed maximum early-and late-stage infection prevalence and fitted parameter values for the six lakes.

recovering hosts k_t, we notice the following. Hale has $\max P^I > \max P^E$ and low k_t, while the rest of the lakes which have $\max P^E > \max P^I$ also have higher k_t, Table 2.

Hypothesis I: We hypothesize that a higher fraction of recovering hosts causes the late-stage infection prevalence to be lower than the early-stage infection prevalence.

A second hypothesis is with regards to transmission rate b_t in combination with the spore yield Σ_t. We notice that as the transmission rate increases, the spore yield decreases, Fig. 16, panel c. We calculated the mean and standard deviation of the product $b_t\Sigma_t$ (for the ten best fits) for each of the six lakes. We found the values 2.78 ± 0.88 for Beaverdam, 1.86 ± 0.33 for Benefiel, 4.60 ± 0.13

for Downing, 0.37 ± 0.07 for Hale, 2.31 ± 0.39 for Midland, and 1.47 ± 0.13 for Star lake. We also fitted hyperbolas of the form $\Sigma_t = \frac{C}{b_t}$, where the constant C is equal to 2.42 for Beaverdam, 1.78 for Benefiel, 4.60 for Downing, 0.34 for Hale, 2.15 for Midland, and 1.49 for Star.

Hypothesis II: We hypothesize that a low product of transmission rate and spore yield suppresses early-stage infection prevalence more than it suppresses late-infection prevalence.

Finally, predation rate in Hale was also an outlier ($0.02 < P_t < 0.052$) compared to the rest of the lakes, Table 2. We recall that predators prey preferentially (by a factor of $\theta = 3$) on terminally infected hosts.

Hypothesis III: We hypothesize that low predation rate supports higher late-stage infection prevalence and lower early-stage infection prevalence.

To test these hypotheses, we ran 5000 simulations with fixed initial condition for the susceptible population, $S(0) = 5$, zero initial conditions for the exposed $E(0) = 0$ and the infected hosts $I(0) = 0$, and with the rest of the parameters and initial conditions being chosen randomly in the same way as when the fitting was performed. The ranges for these parameters and initial conditions are shown in Table 3 in the Appendix. We created two parameter clouds; the blue cloud for the runs with $\max P^E > \max P^I + 0.4$ and the red cloud for the runs for which $\max P^I > \max P^E + 0.4$. The difference between the maximum of the early- and late-stage infection prevalence was chosen to be 0.4 in order to obtain two distinct clusters, Fig. 17. Our first hypothesis, that a small fraction of recovering hosts k_t results in higher late-stage than early-stage prevalence is supported (panels b and d). Our second hypothesis that a low product of transmission rate and spore yield suppresses early-stage infection prevalence more than it suppresses late-infection prevalence is also supported (panel c) The cyan curve separating the two clusters is given by $\Sigma_t = \frac{3.3}{b_t}$. Finally, our third hypothesis that low predation supports higher late-stage infection prevalence is also supported (panels a and e). Panel f is as expected, since we have set no restrictions on the total host density which is mostly controlled by the growth rate B_t and the carrying capacity K.

4 Discussion

We considered a mathematical model that explores seasonal changes in the two environments that parasites can encounter: the host's external environment and the within-host environment. Given that environmental conditions and host immunity are known to fluctuate seasonally (with potential impacts on parasite dynamics), we used disease data from six natural populations spanning the time before and during when epidemics occur to ask how altering seasonally-dependent host parameters (birth, death by predation, immunity) and parasite parameters (spore number) influenced our ability to capture natural disease dynamics with our model.

With a basic time-dependent model, and without changing too many parameters, we were able to capture many of the epidemiological behaviors observed in the natural populations. In particular, our models well-replicated the timing of epidemic emergence, the mismatch in the mean prevalences of early- versus late-stage infections, and the decline of infections toward the end of the sampling

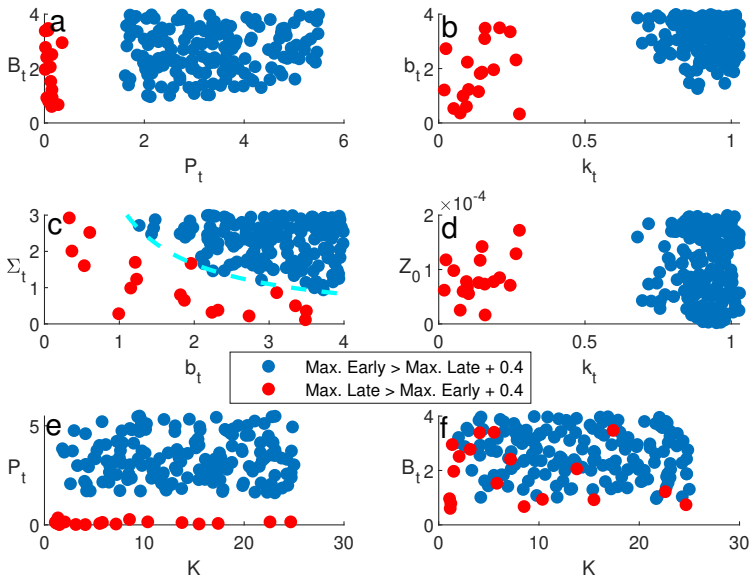

Figure 17: Parameter clouds from the random simulations. Red dots correspond to parameter combinations that resulted in the maximum of the late prevalence being more than 0.4 higher than the maximum of the early prevalence. The opposite holds for the blue dots, namely the maximum of early prevalence is more than 0.4 larger than the maximum of the late prevalence. In panel c, the cyan dashed curve is the hyperbola $\Sigma_t = 3.3/b_t$.

period. Predicting the next wave of disease is a shared goal in many disease systems. For several of our study populations, we achieved that goal, even in the face of substantial variation in host densities among lakes. However, there were some patterns observed in the lakes that our model was not able to replicate. Star and Hale lakes both showed a sizable pulse in early infections (over 50% in Hale and over 60% in Star) that occurred long before late infections became prevalent and exhibited epidemic growth. While our model could recreate the mismatch in the general prevalences of these two classes, the temporal mismatch in peaks between the two classes was not observed in the model simulations. Our model was able to achieve natural dynamics in lakes where a single wave was observed later in the season, such as in Beaverdam and Downing lakes. There was also a reasonably good fit with the pattern observed in Benefiel, which had no recorded late infections and about 20% prevalence of early stage infections.

The parameters that most critically affected the simulated disease dynamics were the transmission rate (b_t), recovery fraction (k_t), spore release rate (Σ_t), and predation rate (P_t). The reliance of disease dynamics on transmission, recovery, and spore yield hints at the importance of within-host processes for regulating epidemiological behavior. That seasonality can modify disease dynamics through its effects on host susceptibility and immunity is well appreciated. For example, the Panamanian rock frog (*Colostethus panamansis*), which can be infected by the fungal pathogen *Batrachochytrium dendrobatidis*, often excretes skin chemicals to inhibit infection. The chemical

composition of these anti-fungal skin secretions varies across seasons, being more effective during dry seasons than wet ones [22]. This seasonally varying immune defense likely contributes to observed seasonal variation in both the presence and intensity of *B. dendrobatidis* infections.

The effects of seasonality on disease can be complex because even simple changes to the extrinsic environment, like a shift in temperature, can manifest in multiple interrelated ways for a given host-parasite interaction. For instance, a *Daphnia* host's immune response can be directly reduced by temperature [33] while being indirectly increased by temperature-related changes to resource availability ([28]. At the same time, seasonal temperature changes can correspond with changes to the presence and abundance of predators, who can act as enhancers [6] or inhibitors [11] of epidemics depending on identity. Indeed, past research with *Daphnia* and its parasites has identified several important seasonal effects. Once an epidemic is underway, temperature influences both spore production and subsequent infectivity [25]. Reproduction, which generates susceptible hosts, is also temperature-dependent [33]. Cooler temperatures and storms [5] have been associated with increased turbulence and mixing in lakes which increases contact rates between susceptible hosts and infective spores. Even lake stratification, through modifying the size of the refuge from fish predators, can affect the severity of an outbreak [18, 32]. We embedded many of these factors into our model to allow for natural seasonal changes in our parameters and a better correspondence between the data and the model.

An important aspect of our approach is that fitting the model to the data was a tool and not the main focus of this work. While model-fitting can be a valuable approach for perfectly recreating the dynamics of a given system, it comes with the cost of being parameter-heavy. Logistically, this can result in unidentifiable parameters (as we observed with our model). And in many cases, the importance of parameters can be system-specific and can have idiosyncratic effects. In contrast to this approach, we used a simple model and modified a small set of parameters that were grounded in well-supported hypotheses about how the *Daphnia-Metschnikowia* system operates. By constraining our focus to these distinct parameters, we uncovered generality in processes across lakes. Namely, seasonality in parameters could account for the natural temporal changes in disease in these lakes, and the observed variation among lakes could be, in part, recreated through modifying key parameters such as those regulating transmission, recovery, spore yield, and predation. Such patterns inspire continued work on how the extrinsic environment, through modifying the within-host environment, can shape the severity of epidemics [1, 29, 30].

Acknowledgments

This work was supported by C3.ai Inc. and Microsoft Corporation. Additional support was provided through grants from the National Science Foundation (DMS 1815764 to Z.R., DBI 2022049 Biology Integration Institute Implementation Grant GEMS, and NSF 1701515 (awarded to C.E.C. and T.E.S.M.) and the Life Sciences Research Foundation (support provided to T.E.S.M.).

Appendix

Least Squares Fitting

The least-squares cost function that we used to fit the model to the data is:

$$C = \frac{1}{R}\sum_{i=1}^{n}(N(t_i) - N_d(t_i))^2 + \sum_{i=1}^{n}(P^E(t_i) - P_d^E(t_i))^2 + \sum_{i=1}^{n}(P^I(t_i) - P_d^I(t_i))^2, \tag{5}$$

where n is the number of sampling points, $R > 0$ is a scaling factor that differs among lakes and takes the values $R = 5000$ for Beaverdam, $R = 500$ for Benefiel, $R = 600$ for Downing, $R = 25$ for Hale, $R = 300$ for Midland, and $R = 1000$ for Star lake. The rationale behind these choices was to weigh the sum related to total densities N, since it is orders of magnitude higher than that of the early-stage P^E and late-stage P^I infection prevalence. The time-series data are denoted by N_d, P_d^E and P_d^I while the model results are denoted by $N = S + E + I$ (total density), $P^E = \frac{E}{N}$ (early prevalence) and $P^I = \frac{I}{N}$ (late prevalence).

For each lake we run 1000 simulations and find a "best fit" using the MATLAB "`fmincon`" function. Initial guesses for the parameters were drawn uniformly from the ranges shown in Table 3. Some of the runs were discarded, because either the algorithm was stuck at the boundary of the parameter range intervals or because they did not finish. Of the remaining runs, the 10 best fits were used to generate the model simulations and parameter clouds shown in the main text.

Parameter	Range
B_t	[0.5, 4]
K	[1, 25]
P_t	[0.01, 5.5]
b_t	[0.01, 4]
k_t	[0.01, 1.03]
Σ_t	[0.1, 3]
Z_0	[1e-10, 2e-04]

Table 3: Parameter ranges for the initial guesses of the parameters and initial conditions (Z_0) in the least-squares fitting.

References

[1] S. Altizer, A. Dobson, P. Hosseini, P. Hudson, M. Pascual, and P. Rohani. Seasonality and the dynamics of infectious diseases. Ecology Letters, 9:467–484, 2006.

[2] S.K.J.R Auld, S.R. Hall, J. H. Ochs, and M. A. Sebastian, M. Duffy. Predators and patterns of within-host growth can mediate both among-host competition and evolution of transmission of potential parasites. The American Naturalist, 184:S77–S90, 2014.

[3] E. Burreson and L.R. Calvo. Epizootiology of perkinsus marinus disease of oysters in chesapeake bay, with emphasis on data since 1985. Journal of Shellfish Research, 15(1):17–34, 1996.

[4] C. E. Cáceres and Stewart Merrill T. E. The role of varying resources on Daphnia dentifera immune responses. Fundamental and Applied Limnology, 196:217–228, 2022.

[5] C. E. Cáceres, S. R. Hall, M.A. Duffy, A. J. Tessier, C. Helmie, and A. MacIntyre. Physical structure of lakes constrains epidemics in Daphnia populations. Ecology, 87:1438–1444, 2006.

[6] C. E. Cáceres, C. J. Knight, and S. R. Hall. Predator-spreaders: Predation can enhance parasite success in a planktonic host-parasite system. Ecology, 90(10):2850–2858, 2009.

[7] C.E. Cáceres, G. Davis, S. Duple, S.R. Hall, A. Koss, P. Lee, and Z. Rapti. Complex Daphnia interactions with parasites and competitors. Mathematical Biosciences, 258:148–161, 2014.

[8] T. Dallas, M. Holtackers, and J.M. Drake. Costs of resistance and infection by a generalist pathogen. Ecology and Evolution, 6(6):1737–1744, 2016.

[9] E.S. Davenport, Dziuba M.K., Jacobson L.E., Calhoun S.K., Monell K.J., and Duffy M.A. How does parasite environmental transmission stage concentration change before, during, and after disease outbreaks? Ecology, 105(2):e4235, 2024.

[10] M. A. Duffy. Selective predation, parasitism, and trophic cascades in a bluegill-Daphnia-parasite system. Oecologia, 153:453–460, 2007.

[11] M. A. Duffy, S. R. Hall, A. J. Tessier, and M. Huebner. Selective predators and their parasitized prey: Are epidemics in zooplankton under top-down control? Limnology and Oceanography, 50(2):412–420, 2005.

[12] S. R. Hall, C.R. Becker, M.A. Duffy, and C. E. Cáceres. Epidemic size determines population-level effects of fungal parasites on Daphnia hosts. Oecologia, 166:833–842, 2011.

[13] S. R. Hall, C.J. Knight, C. R. Becker, M.A. Duffy, A.J. Tessier, and C. E. Cáceres. Quality matters: resource quality for hosts and the timing of epidemics. Ecology Letters, 12:118–128, 2009.

[14] S. R. Hall, A. J. Tessier, M.A. Duffy, M. Huebner, and C. E. Cáceres. Warmer does not have to mean sicker: temperature and predators can jointly drive timing of epidemics. Ecology, 87(7):1684–1695, 2006.

[15] F.E. McKenzie, G.F. Killeen, J.C. Beier, and W.H. Bossert. Seaonality, parasite diversity, and local extinctions in plasmodium falciparum malaria. Ecology, 82(10):2673–2681, 2001.

[16] E. P. Overholt, M.A. Duffy, M.P. Meeks, T.H. Leach, and C. E. Williamson. Light exposure decreases infectivity of the Daphnia parasite Pasteuria ramosa. J. Plankton Res., 42:41–44, 2020.

[17] E. P. Overholt, S.R. Hall, C.E Williamson, C.A. Meikle, M.A. Duffy, and C.E. Cáceres. Solar radiation decreases parasitism in Daphnia. Ecology Letters, 15:47–54, 2012.

[18] R.M. Penczykowski, S. R. Hall, D.J. Civitello, and M.A. Duffy. Habitat structure and ecological drivers of disease. Limnol. Oceanogr., 59(2):340–348, 2014.

[19] M. Pollicott, H. Wang, and H. Weiss. Extracting the time-dependent transmission rate from infection data via solution of an inverse ODE problem. Journal of Biological Dynamics, 6(2):509–523, 2012.

[20] Z. Rapti, T. E. Stewart Merrill, L. E. Mueller-Brennan, J. H. Kavouras, and C. E. Cáceres. Indirect effects in a planktonic disease system. Theoretical Population Biology, 130:132–142, 2019.

[21] A. Raue, C. Kreutz, T. Maiwald, J. Bachmann, M. Schilling, U. Klingmueller, and J. Timmer. Structural and practical identifiability analysis of partially observed models by exploiting the profile likelihood. Bioinformatics, 25:1923–1929, 2009.

[22] C.M. Rosa, R. Perez, L.A. Richards, C.L. Richards-Zawacki, A.M. Smilanich, L.K. Reinert, L.A. Rollins-Smith, D.P. Wetzel, and J. Voyles. Seasonality of host immunity in a tropical disease system. Ecosphere, 13:e4158, 2022.

[23] C.L. Shaw, S. R. Hall, E. P. Overholt, C.E. Cáceres, C. E. Williamson, and M.A. Duffy. Shedding light on environmentally transmitted parasites: lighter conditions within lakes restrict epidemic size. Ecology, 101(11):e03168, 2020.

[24] M. S. Shocket, A. Magnante, M. A. Duffy, C. E. Cáceres, and S. R. Hall. Can hot temperatures limit disease transmission? A test of mechanism in a zooplankton-fungus system. Functional Ecology, 33:2017–2029, 2019.

[25] M. S. Shocket, A. T. Strauss, J. L. Hite, M. Sljivar, D. J. Civitello, M. A. Duffy, C. E. Cáceres, and S. R. Hall. Temperature drives epidemics in a zooplankton-fungus disease system: a trait-driven approach points to transmission via host foraging. The American Naturalist, 191(4):435–451, 2018.

[26] M. S. Shocket, D. Vergara, A.J. Sickbert, J.M. Walsman, A. T. Strauss, J. L. Hite, M. A. Duffy, C. E. Cáceres, and S. R. Hall. Parasite rearing and infection temperatures jointly influence disease transmission and shape seasonality of epidemics. Ecology, 99(9):1975–1987, 2018.

[27] U. Sommer, R. Adrian, L. De Senerpont Domis, J.¿ Elser, U. Gaedke, B. Ibelings, E. Jeppesen, M. Lürling, J.C. Molinero, W.M. Mooij, E. van Donk, and M. Winder. Beyond the plankton ecology group (peg) model: mechanisms driving plankton succession. Annual Review of Ecology, Evolution, and Systematics, 43:429–448, 2014.

[28] T. E. Stewart Merrill and C. E. Cáceres. The role of varying resources on Daphnia dentifera immune responses. Fundam. Appl. Limnol., page in press, 2022.

[29] T. E. Stewart Merrill, S. R. Hall, and C. E. Cáceres. Parasite exposure and host susceptibility jointly drive the emergence of epidemics. Ecology, 102(2):e03245, 2020.

[30] T. E. Stewart Merrill, S. R. Hall, L. Merrill, and C. E. Cáceres. Variation in immune defense shapes disease outcomes in laboratory and wild Daphnia. Integrative and Comparative Biology, 59(5):1203–1219, 2019.

[31] T. E. Stewart Merrill, Z. Rapti, and C. E. Cáceres. Host controls of within-host disease dynamics: insight from an invertebrate system. The American Naturalist, 198(3):317–332, 2021.

[32] A.T. Strauss, M.S. Shocket, D.J. Civitello, J.L. Hite, R.M. Penczykowski, M. A. Duffy, C.E. Cáceres, and S. R. Hall. Habitat, predators, and hosts regulate disease in Daphnia through direct and indirect pathways. Ecological Monographs, 86(4):393–411, 2016.

[33] S.-J.. Sun, M. K. Dziuba, R. N. Jaye, and M.A. Duffy. Temperature modifies trait-mediated infection outcomes in a *daphnia*–fungal parasite system. Philosophical Transactions B, 378:20220009, 2023.

[34] S.T. Threlkeld. The midsummer dynamics of two daphnia species in wintergreen lake, michigan. Ecology, 60(1):165–179, 1979.

[35] A. F. Villaverde, N. Tsiantis, and J.R. Banga. Full observability and estimation of unknown inputs, states and parameters of nonlinear biological models. Journal of the Royal Society Inference, 16:20290043, 2019.

[36] G.H. Westphal, L.M. Bradshaw, D.B. DuBose, N. Levy, D.K. Okamoto, M. Davis, M. Tarnowski, B. Preziosi, and T.E. Stewart Merrill. Differences in die-offs: Investigating variation in infection patterns and virulence across the landscape in an oyster – parasite system. In preparation, 2024.