bioRxiv preprint doi: https://doi.org/10.1101/285080; this version posted March 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Using biological control data to understand host-pathogen dynamics

Joseph R. Mihaljevic[1,5*]

Karl M. Polivka[2]

Constance J. Mehmel[3]

Chentong Li[1,6]

Vanja Dukic[4]

Greg Dwyer[1]

1. Department of Ecology and Evolution, University of Chicago, Chicago, IL 60637;

2. Pacific Northwest Research Station, USDA Forest Service, Wenatchee, WA 98801;

3. Okanogan-Wenatchee National Forest, USDA Forest Service, Wenatchee, WA 98801;

4. Department of Applied Mathematics, University of Colorado, Boulder, CO 80309;

5. School of Informatics, Computing, and Cyber Systems, Northern Arizona University, Flagstaff, AZ 86011;

6. School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an Shaanxi P.R.China 710049.

∗ Corresponding author; e-mail: joseph.mihaljevic@nau.edu.

Manuscript elements: Figure 1, figure 2, figure 3, figure 4, figure 5, figure 6, table 1, table 2, online appendices A (including figure A1 through figure A16 and table A1 through table A8).

Keywords: Disease ecology, epidemiological modeling, Bayesian inference, microbial control.

Manuscript type: Article.

bioRxiv preprint doi: https://doi.org/10.1101/285080; this version posted March 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Prepared using the suggested LaTeX template for *Am. Nat.*

bioRxiv preprint doi: https://doi.org/10.1101/285080; this version posted March 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Abstract

Mathematical models have provided important insights into infectious disease spread in animal populations, but are only rarely used in environmental management. Part of the problem may be that direct tests of the models have focused on long-term model predictions, whereas short-term epizootics (= epidemics in animals) can sometimes be more informative about mechanisms of disease transmission. To illustrate this point, we tested models of density-dependent disease transmission and host variation in infection risk using multiple, single-host-generation epizootics of a fatal baculovirus of the Douglas-fir tussock moth (*Orgyia pseudotsugata*). The tussock moth baculovirus causes epizootics naturally, but because of its narrow host range, it is used in biological control to mitigate defoliation, an approach known as "microbial control". Using a combination of experiments, observational data, and nonlinear-fitting algorithms, we fit a set of stochastic epidemiological models to data from tussock moth microbial control programs. We then used formal model comparisons to show that there are strong effects of host and pathogen density and host variation on epizootic severity, and that transmission events at small scales help explain epizootic severity at large scales. When host density is high, even very low initial virus densities can lead to high cumulative infection rates, suggesting that, in some cases, baculovirus sprays may have been applied to populations that would have collapsed from natural epizootics, even without intervention. Our study illustrates the usefulness of linking epidemiological modeling and nonlinear fitting routines in understanding animal diseases, and shows that simple models can provide important guidance for microbial control programs.

bioRxiv preprint doi: https://doi.org/10.1101/285080; this version posted March 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Introduction

Efforts to use epidemiological theory to understand animal host-pathogen interactions often focus on long-term dynamics (Keeling and Rohani, 2008), following Anderson and May's championing of host-pathogen models that allow for host reproduction (Anderson and May, 1979; May and Anderson, 1979). Anderson and May's point that host reproduction is important over the long term is inarguable, but for many pathogens, seasonality prevents transmission for part of the year (Altizer et al., 2006). Seasonality may prevent transmission if there are seasons when the vector is not present (Ruder et al., 2015), or when the pathogen does not co-occur with its host (Mihaljevic et al., 2017), or when the host is in a life stage that the pathogen cannot infect (Cory and Myers, 2003). In such cases, models of single epizootics (= epidemics in animals) may be more useful than models of long term disease dynamics.

Comparison of single-epizootic models to short-term data may also be more useful for answering basic questions in disease ecology (O'Dea et al., 2014). Here we ask, how do host and pathogen density and host heterogeneity in infection risk affect disease spread? Simple models have shown that both factors can strongly affect disease spread (Keeling and Rohani, 2008), but the importance of either factor in nature is unclear (McCallum, 2016). Part of the problem may be that, over the long term, the effects of density and host heterogeneity are manifest mostly through changes in equilibrium host and pathogen densities (Anderson and May, 1980), but inferring mechanisms of population change from equilibrium densities is notoriously difficult (Bolker, 2008). Over the short-term in contrast, changes in host and pathogen densities can dramatically alter infection rates (Dwyer and Elkinton, 1993), while heterogeneity in host infection risk can strongly modulate the effects of density (Anderson and May, 1992; Dwyer et al., 1997).

Here we use data on single epizootics to infer the effects of host and pathogen density and host heterogeneity on the spread of an insect baculovirus. Because baculoviruses have narrow host ranges and are obligately lethal (Hunter-Fujita et al., 1998; Moscardi, 1999), baculovirus insecticides have minimal effects on non-target species, a type of biological control known as "mi-

bioRxiv preprint doi: https://doi.org/10.1101/285080; this version posted March 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

crobial control" (Lacey et al., 2015). The data that we use come from microbial control programs aimed at the Douglas-fir tussock moth, *Orgyia pseudotsugata* (Otvos et al., 1998, 1987), a major pest of Douglas-fir, *Pseudotsuga menziesii*, and true firs, *Abies* spp., in western North America (Shepherd et al., 1988). Because managers often monitor infection rates in both sprayed treatment populations and unsprayed control populations, and because microbial sprays can sharply raise pathogen densities, microbial control data allow for a broader range of infection rates and pathogen densities than data from natural epizootics (Otvos et al., 1987).

In microbial-control data sets, even the sprayed populations are fundamentally observational, in the sense that post-spray conditions are uncontrolled. This is important because factors such as host density and host heterogeneity in infection risk affect the rate at which individuals contact each other and transmit the disease (Keeling and Rohani, 2008), but inferring individual-level mechanisms from observational data alone can be very difficult (King et al., 2008). Experiments in contrast can provide deep insights into individual-level mechanisms, but may not include mechanisms that occur in natural epizootics (Cory and Hoover, 2006). The problem of inferring how individuals affect population-level disease spread is especially difficult for baculoviruses, because transmission occurs when uninfected larvae, while feeding on foliage, consume infectious "occlusion bodies" released from dead, infected larvae (Cory and Myers, 2003). Baculovirus epizootics at the scale of forests are therefore widely assumed to be driven by transmission at the scale of branches (Cory and Myers, 2003), but it is not obvious how this assumption can be tested.

The problem of how to relate individual-level mechanisms to population-level phenomena is long-standing in ecology (Levin, 1992), however, and methods have therefore been proposed to solve the problem. In one such method, the parameters of a mechanistic model are first estimated from data collected at the individual level, the parameters are then used in the model to make predictions, and finally the model predictions are compared to data collected at the population level (Kareiva and Odell, 1987). Agreement between model predictions and population-level data is then taken as confirmation that individual-level mechanisms can help explain population-level

bioRxiv preprint doi: https://doi.org/10.1101/285080; this version posted March 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

phenomena, on the grounds that agreement between large-scale data and a model parameterized from small-scale data is unlikely to occur by chance.

Implementations of this approach, however, typically use only point estimates of the model parameters (Grimm et al., 2005; Keeling et al., 2003; Longini et al., 2005), neglecting the substantial uncertainty in the parameter values (Elderd et al., 2006). Here we instead follow a formal Bayesian statistical approach, in which experimental data are used to construct prior distributions on model parameter values, and observational epizootic data are used to calculate posteriors. To assess whether our experimental data do in fact help us to understand the observational data, we then use the WAIC model-selection criterion (Gelman et al., 2014), to compare the fit of models with priors constructed from experiments, to the fit of models with uninformative or "vague" priors. As we will show, models with priors constructed from experimental data can sometimes explain the observational data as well as models with uninformative priors, suggesting that transmission at the scale of branches helps explain epizootics at the scale of forests.

Irrespective of the priors, the epizootic data are best explained by models that allow for density-dependent transmission and host variation in infection risk. This result is important for microbial control, because naturally occurring epizootics have often decimated unsprayed populations in the same year and in the same general area as epizootics initiated by baculovirus sprays (Polivka et al., 2012; Scott and Spiegel, 2002). The occurrence of epizootics in low density, unsprayed populations has been a surprise to pest managers, but our results suggest that host densities in control plots have been high enough to drive intense epizootics even when initial infection rates are very low, confirming basic epidemiological theory (Keeling and Rohani, 2008). Using microbial control data to test disease models can thus produce useful insights into disease ecology, while providing guidance for pest control.

bioRxiv preprint doi: https://doi.org/10.1101/285080; this version posted March 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Methods

At roughly 10-year intervals (Mason, 1996), tussock moth densities rise from levels at which larvae are undetectable, to levels at which defoliation is severe (Shepherd et al., 1988). The economic costs of timber losses drove the USDA Forest Service to propagate naturally occurring virus in the lab, producing a formulation known as Tussock Moth Biocontrol-1, or "TMB-1" (Martignoni, 1999). The USDA Forest Service has used TMB-1 widely in the western USA (Polivka et al., 2012; Scott and Spiegel, 2002; Shepherd et al., 1984), while the Canadian Forest Service has deployed a similar formulation in British Columbia (Maclauchlan et al., 2009).

In spray formulations, baculovirus from lab-reared insects is mixed with water or oil and sprayed onto foliage from the air using helicopters, or more rarely from the ground using hoses. Unsprayed control populations are often established in areas that are of lesser concern, because of lower tussock moth densities or less valuable timber. The occurrence of epizootics in controls has been surprising not just because initial infection rates were consistently low, roughly 0.5% in one study (Polivka et al., 2012), but also because the controls were far enough from spray plots (1-50km) that virus drift between plots was unlikely (Otvos et al., 1987; Polivka et al., 2012; Scott and Spiegel, 2002).

Because higher initial densities of virus in sprayed plots have not always led to higher mortality relative to control plots, managers have hypothesized that naturally occurring virus may be more infectious than sprayed virus (Polivka et al., 2012), possibly because of the effects of prolonged low-temperature storage on the sprayed virus (Reed et al., 2003). Epidemiological theory, however, suggests that high virus mortality in control plots may instead have been due to the amplifying effects of multiple rounds of transmission. To test these competing hypotheses, and to test general theories of pathogen spread, we built competing epidemiological models, and we fit the models to data on epizootics in sprayed and unsprayed Douglas-fir tussock moth populations. Our competing models assume either that sprayed and naturally occurring baculoviruses have the same transmission rates, or that sprayed virus has a different (presumably

bioRxiv preprint doi: https://doi.org/10.1101/285080; this version posted March 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

lower) transmission rate. We fit our models using a non-linear model-fitting algorithm (Kennedy et al., 2014), and we chose the best model using the Watanabe-Akaike Information Criterion (WAIC), a Bayesian generalization of the more traditional Akaike Information Criterion (AIC) (Gelman et al., 2014; Hooten and Hobbs, 2015; Watanabe, 2010, 2013).

We also tested whether parameters based on our experiments produce accurate model predictions. To do this, we used our experimental data to construct prior distributions on the parameters, and we used our observational data to calculate the likelihood of the parameters. Under Bayes's law, priors and likelihoods together make up the posterior distribution of the parameters, which can be used to calculate WAIC scores. We then compared WAIC scores for the case in which we used uninformative priors, to the cases in which we used priors based on experimental data.

Priors that are constructed from data are usually referred to as "informative priors", because they constrain the prior distribution in a way that improves the model's ability to make predictions. Informative priors therefore add information to the model, much in the way that including additional data in a likelihood calculation adds information to the model (Bernardo and Smith, 2009). When we use ecological field experiments to construct priors, however, the priors may not improve the model's ability to predict observational data, because experiments are often carried out at smaller scales than the scales at which natural phenomena occur (Kareiva and Andersen, 1988). This is especially true in our case, because our experiments were carried out on single branches, whereas the epizootic data were collected in forest plots of up to 10 hectares.

Our experiment-based prior distributions might therefore be constrained in a way that biases the model predictions and leads to a poorer likelihood. In our case, testing whether priors based on our experiments lead to inaccurate or imprecise model predictions is equivalent to testing whether branch-scale transmission processes help determine the dynamics of forest-scale epizootics. By allowing us to compare models with vague priors to models with priors constructed from the experimental data, the WAIC therefore allowed us to test whether our experiment-based priors provided useful information about our observational data, and thus whether small-scale

bioRxiv preprint doi: https://doi.org/10.1101/285080; this version posted March 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

transmission helps drive large-scale epizootics.

Because our experiments were carried out at a much smaller scale than the scale of epizootics, it was almost inevitable that priors based on the experiments would produce larger, inferior WAIC scores than models with uninformative priors. As we will show, however, for models with priors based on experiments, WAIC scores were only moderately worse. Experiment-based priors therefore improved our understanding of the mechanisms driving epizootics, even though these priors slightly reduced the accuracy and precision of the model predictions.

Baculovirus Natural History and a Stochastic SEIR Model

In herbivorous insects like the tussock moth, baculovirus transmission occurs when larvae accidentally consume infectious virus particles known as "occlusion bodies", leading to death roughly two weeks later (Cory and Myers, 2003). Shortly after death, viral enzymes dissolve the insect integument, releasing occlusion bodies onto the foliage, where they are available to be consumed by additional larvae (Miller, 1997). Multiple rounds of infection can occur during the larval period, which is early June to mid-August in our study areas in Washington, Idaho, and Colorado, USA, and British Columbia, Canada. The severe epizootics that result can decimate the insect population, especially when larval densities are high (Otvos et al., 1987; Shepherd et al., 1984).

Epizootics are terminated by larval pupation, or by reductions in host density that are so severe that the probability of transmission is extremely low (Fuller et al., 2012). The virus then survives until the next spring by contaminating egg masses, although survival on the forest floor may also play a role (Thompson and Scott, 1979). So far as can be known, the Douglas-fir tussock moth is the only organism in its range that is susceptible to the baculovirus, although *Orgyia* spp. from other parts of North America have been successfully infected in the lab (Rohrmann, 2014).

Because pathogen transmission and host reproduction occur at different times of the year, we can describe baculovirus epizootics using a model of a single epizootic. We began with a standard

bioRxiv preprint doi: https://doi.org/10.1101/285080; this version posted March 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Susceptible-Exposed-Infectious-Recovered or "SEIR" model from human epidemiology (Keeling and Rohani, 2008), and we modified the model to allow for heterogeneity in host infection risk. Heterogeneity in infection risk includes a stochastic component that reflects the effects of weather, and a deterministic component that reflects inherent differences between hosts. The model is:

$$\frac{dS}{dt} = -\bar{\nu}e^{\epsilon_t} SP \left[\frac{S(t)}{S(0)}\right]^{C^2}, \tag{1a}$$

$$\frac{dE_1}{dt} = \bar{\nu}e^{\epsilon_t} SP \left[\frac{S(t)}{S(0)}\right]^{C^2} - m\delta E_1, \tag{1b}$$

$$\frac{dE_i}{dt} = m\delta E_{i-1} - m\delta E_i \quad (i = 2, \ldots, m), \tag{1c}$$

$$\frac{dP}{dt} = m\delta E_m - \mu P. \tag{1d}$$

As in standard SEIR models, susceptible hosts S that become infected proceed through m exposed classes, E_i. Exposed hosts eventually die of the infection, joining the infectious-cadaver class P, which is equivalent to the the I class of standard models. Cadaver infectiousness decays at rate μ, due mostly to UV radiation (Thompson and Scott, 1979). Hosts move between exposed classes at rate $m\delta$, so that the time spent in a single exposed class follows an exponential distribution with mean time $1/(m\delta)$. The total time in the m exposed classes is the sum of m such distributions, and a well-known theorem has shown that the sum follows a gamma distribution with mean, $1/\delta$, and variance, $1/(m\delta^2)$ (Keeling and Rohani, 2008). For the tussock moth virus, the variance in the speed of kill is quite low (Dwyer (1991), also Online Appendices), so in practice we fixed $m = 200$, rather than estimating m from the observational data.

The term $\left[\frac{S(t)}{S(0)}\right]^{C^2}$ allows for inherent host variation in infection risk, following models originally developed for the gypsy moth baculovirus (Dwyer et al., 1997). This term reflects an assumption that there is a distribution of infection risk in the host, with mean $\bar{\nu}$ and coefficient of variation C, where individual infection risk is represented by the transmission rate ν (Dwyer et al., 2000). This is an approximation to a more computationally intensive partial-differential equation model. The approximation is highly accurate if transmission rates follow a gamma distribution, but it is also reasonably accurate if transmission rates instead follow a log-normal

bioRxiv preprint doi: https://doi.org/10.1101/285080; this version posted March 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

distribution, which has a fatter tail than a gamma distribution (G. Dwyer, unpublished). In general, we assume that transmission occurs among larvae in the fourth instar (= larval stage), the instar that has the biggest impact on cumulative infection rates (Dwyer, 1991; Otvos et al., 1987). To allow for the smaller size of hatchlings, we multiply the initial virus density by the "ratio" parameter, ρ, which is equivalent to the ratio of the number of virus particles produced by a first-instar cadaver to the number produced by a fourth-instar cadaver.

By including the parameter ϵ_t, we allow for daily, stochastic fluctuations in average transmission $\bar{v}$, such that the value of ϵ_t for a given day is drawn from a normal distribution with mean 0 and standard deviation, σ. By exponentiating ϵ_t, we eliminate the possibility of negative transmission, which would be biologically meaningless. Allowing for stochasticity is important because baculovirus transmission varies due to changes in insect feeding behavior, which in turn may respond to stochastic fluctuations in weather (Eakin et al., 2015). Also, stochasticity may interact in complex ways with disease dynamics, such that the model predictions become more or less variable as an epizootic proceeds, or as initial host and pathogen densities vary between epizootics. By including stochasticity, we allow for this type of variation, ensuring that our estimation procedures are statistically robust. Because weather conditions likely also varied between populations, we estimated a value of the standard deviation σ for each population in our observational data set.

Although the model includes biological details that are not included in classical disease models (Kermack and McKendrick, 1927), the model's dynamics are qualitatively similar to the dynamics of classical models, as shown by fig. 1 for the deterministic case (stochasticity $\sigma = 0$). Notably, there is a threshold value of the uninfected-host density at $\frac{\mu}{\bar{v}}$, above which the introduction of a small density of infectious hosts will lead to an increase in the density of infected hosts, and below which the density of infected hosts will always decline. If the density of uninfected hosts is far enough above the threshold, severe epizootics may occur even if the initial density of infected individuals is very low. This behavior is a basic feature of density-dependent disease models (Keeling and Rohani, 2008), and it plays a key role in explaining baculovirus epizootics

bioRxiv preprint doi: https://doi.org/10.1101/285080; this version posted March 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

in unsprayed populations.

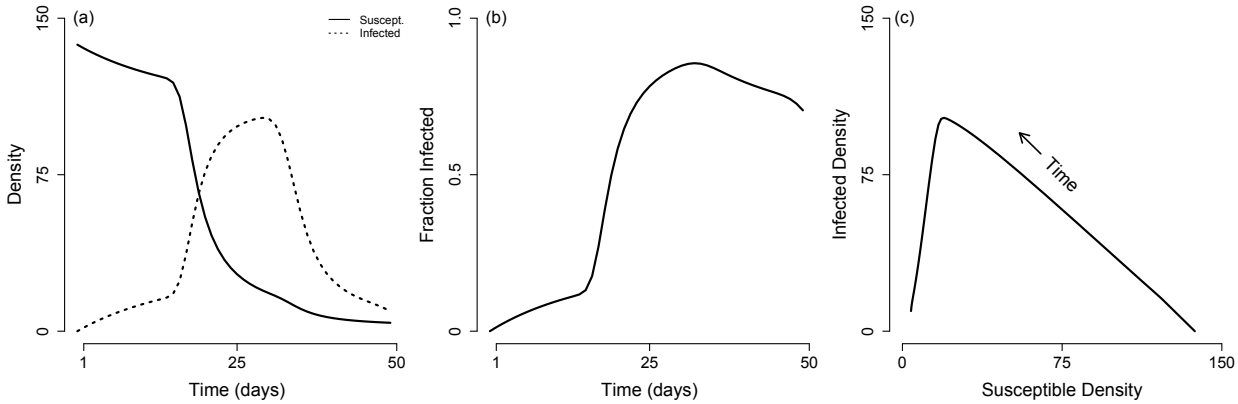

Figure 1: Dynamics of the deterministic version of the stochastic SEIR model, equations 1. (a) Simulated 50-day epizootic, showing only infected and susceptible host densities. In the model, the total infected host density is equivalent to the sum of the densities in the exposed classes, $\sum E_m$. (b) From the same simulation, the fraction of the host population that is infected, calculated as $\frac{\sum E_m}{\sum E_m + S}$. (c) Phase portrait of the same simulation.

To test whether naturally occurring virus and sprayed virus differ in infectiousness $\bar{v}$, we fit two classes of models (Polivka et al., 2017). In the "1-$\bar{v}$" class, we assumed that the average transmission rate was the same for naturally occurring virus and sprayed virus. The 1-$\bar{v}$ models therefore assume that differences in infection rates between epizootics were due only to differences in initial densities of hosts and pathogens. In the "2-$\bar{v}$" class of models, we instead allowed for the possibility that sprayed virus has a different infectiousness than natural virus, as reflected in differences in transmission parameters $\bar{v}$ between sprayed and unsprayed plots.

The epizootic data

Our epizootic data came from 7 unsprayed control plots and 5 sprayed treatment plots. At the beginning of the larval period at each site, initial host and pathogen population densities were estimated using standard methods, which provided initial conditions for our models (Online

bioRxiv preprint doi: https://doi.org/10.1101/285080; this version posted March 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Appendices). The data consist of approximately weekly observations of the fraction of larvae infected with the virus, for up to 50 days, approximately mid-June to mid-August. In sprayed plots, insects were collected within 7 days after application of the virus, while in control plots, the start of collections was more variable, particularly at sites where there was no concurrent spray project. Collected insects were reared in the lab in individual 2 oz (59.1 ml) cups with artificial diet at 25°C until pupation or death. Smears from dead insects were examined under a light microscope for the presence of occlusion bodies, which are easily visible at 400× power (Fleming-Davies et al., 2015; Polivka et al., 2017).

Experimental data

To estimate average transmission $\bar{v}$ and heterogeneity in transmission C, we used a field experiment in which we allowed for a single round of transmission on Douglas-fir branches in the field. Measuring transmission in the field ensures that larval feeding, and thus baculovirus transmission, occur naturally (Cory and Hoover, 2006). We then fit a simplified version of our SEIR model to the fraction of insects infected in the experiment (Online Appendices). We used three baculovirus isolates: TMB-1, the isolate used for biocontrol, and two naturally occurring isolates collected from Washington and New Mexico, which we refer to as "WA" and "NM". Each branch encompassed roughly 0.15 m^2 of foliage, whereas our epizootic data were collected in populations that covered roughly 10 hectares.

To estimate the average speed of kill, $1/\delta$, we infected larvae by allowing them to feed on Douglas-fir foliage that was contaminated with a sprayed virus solution (Online Appendices). Because there is no reason to believe that laboratory conditions lead to biased estimates of the speed of kill, this experiment was carried out in the laboratory. Because speed of kill is affected by temperature, we were careful to hold the larvae at temperatures typical of field conditions (Polivka et al., 2017).

To estimate the ratio parameter, ρ, we again infected larvae in the laboratory, but in this case we infected both hatchlings and fourth instars using virus-contaminated artificial diet (Online

bioRxiv preprint doi: https://doi.org/10.1101/285080; this version posted March 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Appendices). We held these larvae in the lab until death or pupation, and we counted the number of occlusion bodies per dead larva for both instars, using a light microscope and a hemocytometer.

Model fitting and model selection

Modern methods of fitting models to data generally rely on calculations of likelihood scores, which are measures of the goodness of fit of a model. On a log scale, and with a normal likelihood function, maximizing the log likelihood is equivalent to minimizing the sum of squares. In our case, however, the data are fractional infection rates, and so we used a beta-binomial likelihood function, which allows for the binomial nature of the data, while also allowing for the possibility of extra-binomial variation ((McCullagh and Nelder, 1989), also Online Appendices). Because our experimental and observational data sets are independent, we could have calculated likelihoods by simply adding the log likelihoods of the two data sets, following a maximum-likelihood approach (Bolker, 2008). If we had added experimental and observational likelihoods, however, the total likelihood score would have been very different than if we had only calculated likelihoods based on the observational data. A maximum likelihood approach would thus make it impossible to formally compare model predictions with and without the experimental data.

We therefore followed a Bayesian approach. Bayes' theorem can be written as:

$$P(\theta|D) \quad \propto \quad \pi(\theta)\mathcal{L}(\theta|D), \tag{2}$$

where $P(\theta|D)$ is the posterior probability of the model fit to data D. The symbol $\pi(\theta)$ is the prior probability of the parameter values, and $\mathcal{L}(\theta|D)$ is the likelihood of the parameters. Because in general posterior probabilities are only used for comparison purposes, we only need to calculate the posterior probability up to a constant of proportionality, hence our use of the proportion symbol $\propto$.

The posterior distribution of the parameters for each model can then be used to calculate a WAIC score for that model. Like AIC and other model-selection criteria (Konishi and Kitagawa,

bioRxiv preprint doi: https://doi.org/10.1101/285080; this version posted March 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

2008), WAIC identifies the best model by balancing the poorer fit of simpler models against the more uncertain predictions of more complex models (Gelman et al., 2014). Unlike AIC, however, WAIC is fully Bayesian, and so in calculating the ability of the model to predict future or out-of-sample data, WAIC takes into account the effects of the model's priors on the model's predictive performance (Online Appendices). Although it is possible for a model with priors constructed from experimental data to have a better (lower) WAIC score than a model with uninformative priors, given the difference in scales between our experimental data and our epizootic data, such an outcome seems unlikely. For models with priors based on experiments, we therefore treat WAIC scores that are only modestly worse (larger) than the best score as an indication that mechanisms at a small scale help to explain disease spread at a large scale.

Results

Epizootics

The number of larvae collected varied between studies due to differences in overall design, initial host density, and the resources available for collections. Also, near the end of epizootics, sample sizes usually dropped because high baculovirus mortality made it hard to find live larvae (cadavers disintegrate too rapidly to allow for meaningful collections). The number of larvae collected per week therefore ranged from 7 to 197, with a mean of 95.6, and with a total across sites of 5830 (Otvos et al., 1987; Polivka et al., 2012; Scott and Spiegel, 2002). In fitting our models, we then had 61 observations of virus infection rates, 30 at control sites and 31 at sprayed sites. Initial host densities spanned almost two orders of magnitude, ranging from 7.20 to 232.0 larvae/m^2, with a mean of 103.2. The control and sprayed sites had similar ranges and means of initial larval host density (Online Appendices).

Initial infection rates in controls were low, but increased slowly over the larval period (fig. 2), as expected, such that it took weeks for the infection rate in populations with lower initial host densities to reach high levels. Sprayed sites in contrast received an initial inundation of the

bioRxiv preprint doi: https://doi.org/10.1101/285080; this version posted March 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

pathogen, and so infection rates in those sites increased within a week or two after the spray application. Host population collapse was therefore rapid in the sprayed sites, leading to weaker effects of initial host density.

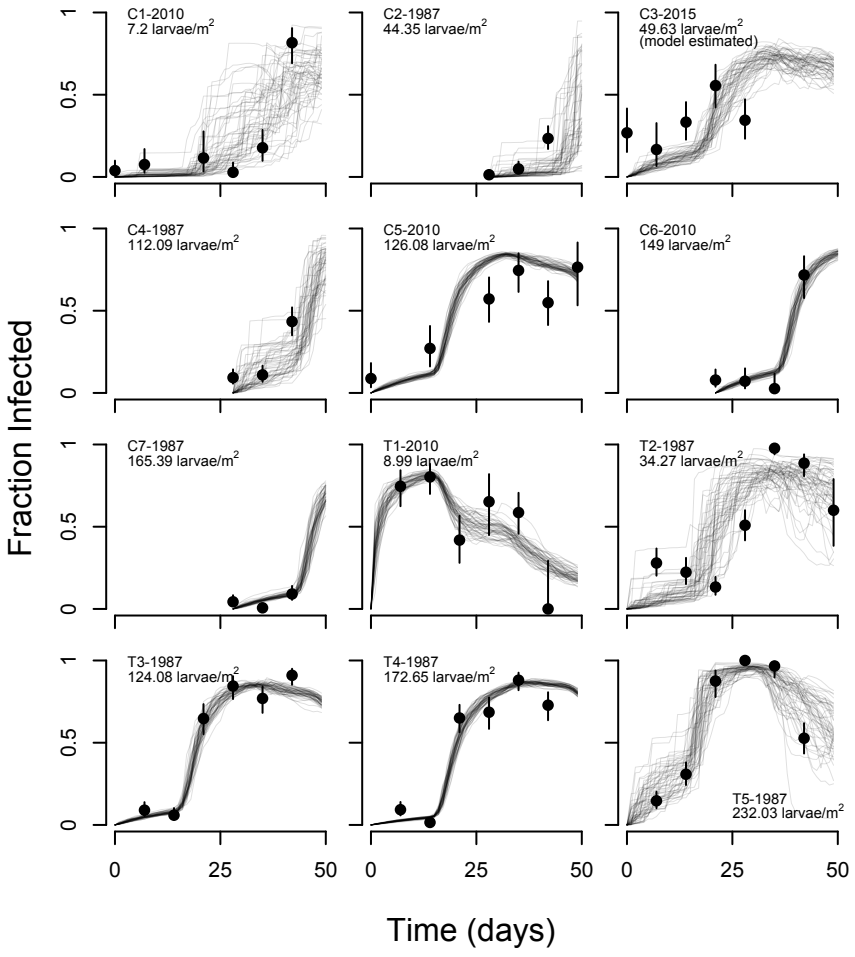

Figure 2: Stochastic realizations of the best fitting model (gray lines), which uses uninformative priors on all parameters, versus the data (black points with 95% binomial confidence intervals as error bars). The labels C and T stand for Control and Treatment, respectively, and are followed by the year of observation. The initial larval host density is also shown. Note that for the Colorado site (C3-2015), the initial larval host density was estimated from the data. Here and in subsequent figures, for two of the populations (C3-2015, C4-1987), we show the model's predictions after the last data point was collected, to illustrate the overall dynamics of the pathogen.

bioRxiv preprint doi: https://doi.org/10.1101/285080; this version posted March 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Experiments

Compared to the case in which hosts are identical, variation in risk leads to higher infection rates at low virus density, because some hosts are highly susceptible, and lower infection rates at high virus density, because some hosts are highly resistant (Dwyer et al., 1997). On a log scale, the fraction uninfected is then a nonlinear (concave) function of baculovirus density (Online Appendices). For two of the three strains that we tested, the best-fit model has a strongly nonlinear shape (fig. 3), and so the model that provides the best overall explanation for our experimental data includes host heterogeneity in infection risk.

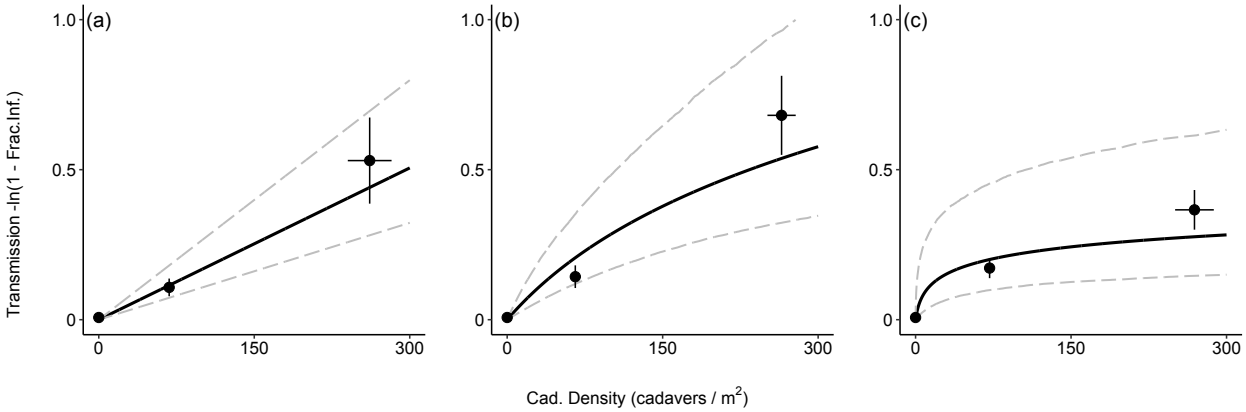

Figure 3: Results of the field transmission experiment. Best-fit transmission parameters (average transmission, $\bar{v}$, and heterogeneity, C) for three viral isolates, in terms of medians with interquartile ranges. (a) WA ($\bar{v}$: 0.006 (0.004, 0.007); C: 0.60 (0.34, 0.93)), (b) TMB-1 ($\bar{v}$: 0.012 (0.009, 0.017); C: 1.52 (1.17, 1.88)), and (c) NM ($\bar{v}$: 0.076 (0.027, 0.309); C: 4.11 (3.31, 4.96)). In the figure, the solid lines represent the median fitted expectations assuming heterogeneity. The gray dashed lines represent boot-strapped 95% credible intervals. Error bars are one standard error of the mean.

Fig. 3 also shows that the three isolates differed strongly, such that the NM isolate had much higher heterogeneity in transmission, and in general higher average transmission rates $\bar{v}$ were associated with higher variation in transmission C. With only 3 isolates, we cannot reach general

bioRxiv preprint doi: https://doi.org/10.1101/285080; this version posted March 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

conclusions, but in the gypsy moth baculovirus, more extensive data revealed a negative correlation between the average and the CV of transmission, consistent with this pattern (Fleming-Davies et al., 2015). A similar correlation may therefore occur in the tussock moth baculovirus. More immediately, the variation across isolates is important when we use our experimental data to explain the epizootic data, because in the epizootic data, we have no information about the isolates that were present. In constructing informative priors from our experimental data, we therefore allowed for uncertainty due to isolate variation by pooling the marginal posterior distributions for the three isolates and inflating the pooled variance slightly (Online Appendices).

When we measured the speed of kill of infected larvae in the laboratory, the average was $14.9d$, with very low variance. The mean value of the death rate is then $\delta = 1/14.9 = 0.067$. Because our sample size was modest ($n = 38$ larvae), we inflated the variance on the prior for δ (Online Appendices). This precautionary variance inflation, however, had no effect on our results.

Model selection

In general, our MCMC routines converged rapidly, except for models with no heterogeneity in infection risk ($C = 0$, table 1). Relative to the no-heterogeneity case, heterogeneity in infection risk modulates the effect of host density, such that, at high host density, heterogeneity reduces infection rates, due to the dominating effects of resistant individuals, while at low host density, heterogeneity increases infection rates, due to the presence of at least a few highly susceptible individuals (Dwyer et al., 1997). This effect is roughly apparent in fig. 2, in that infection rates were high across a very broad range of densities. Because of this effect, models that do not account for heterogeneity provide poor fits to data from populations at either very low or very high density.

This lack of fit led to a biased and highly uncertain value of the death rate parameter δ. In this obligately lethal pathogen, a fast death rate of infected-but-not-yet-infectious hosts causes more rapid disease spread and thus higher infection rates. When fitting the no-heterogeneity model

bioRxiv preprint doi: https://doi.org/10.1101/285080; this version posted March 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

to the data from low-density populations, the fitting routine therefore attempted to improve the fit by increasing δ, so that the best (median) estimate was much higher than our experimental estimate (median speed of kill $(1/\delta) = 5.17d$, compared to an experimental median of $14.9d$). The necessity of also fitting the high-density populations, however, meant that δ was not well-identified, leading to a large upper bound (95% credible interval (Bayesian confidence bound) on the speed of kill: 2.45 - $16.10d$). Moreover, as we will show, the lower bounds on the credible intervals for the heterogeneity parameters for all the 1-$\bar{v}$ models are well above zero (lowest 95% credible interval for $C = 0.61$, Online Appendices). Heterogeneity in infection risk therefore appears to be substantial in tussock moth populations. We therefore conclude that heterogeneity in infection risk plays a key role in epizootics of the tussock moth virus, and we do not consider the no-heterogeneity models further.

WAIC scores (table 1) then showed that the best explanation for the epizootic data comes from models that assume that there is no difference between sprayed virus and naturally occurring virus. As table 1 shows, two of the 2-$\bar{v}$ models had ΔWAIC< 3, and were thus indistinguishable from the best 1-$\bar{v}$ models (Bolker, 2008). For the best 2-$\bar{v}$ model, however, the credible interval on the difference between the two values of $\bar{v}$ for the model included zero (CI: -0.13, 2.56, Online Appendices), indicating that the best 1-$\bar{v}$ model and the best 2-$\bar{v}$ model are effectively identical. Because the 1-$\bar{v}$ model is more parsimonious, we conclude that differences in infectiousness between sprayed virus (TMB-1) and naturally occurring virus were negligible. The epizootic data are thus better explained by models that allow only for differences in initial host and pathogen densities across populations.

Table 1 also shows that, within the 1-$\bar{v}$ class of models, the model with uninformative priors on all parameters is the best model, as we expected. Nevertheless, for three 1-$\bar{v}$ models with priors based on experimental data, WAIC scores were less than about 3, and thus were effectively indistinguishable from the score for the best model. These three models were the model with an informative prior on the death rate δ, a model with an informative prior on the heterogeneity parameter C, and a model with informative priors on δ, C, and the ratio parameter ρ. In the Online

bioRxiv preprint doi: https://doi.org/10.1101/285080; this version posted March 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Model	Informative Priors	Average Likelihood $\sum -\log(\hat{\hat{L}})$	Penalty Score $p\text{WAIC}_2$	WAIC	ΔWAIC
1 $\bar{v}$	**None**	**-197.33**	**5.95**	**406.58**	**0**
	δ	**-197.05**	**6.34**	**406.79**	**0.21**
	$C\ \rho\ \delta$	**-197.77**	**7.00**	**409.52**	**2.94**
	C	**-197.09**	**7.71**	**409.61**	**3.03**
	ρ	-198.38	7.44	411.64	5.06
	$\bar{v}$	-199.20	7.08	412.57	5.99
	$C\ \rho\ \delta\ \bar{v}$	-198.13	8.55	413.37	6.79
	None, $C = 0^a$	-208.62	No Convergence		
2 $\bar{v}$, Spray & Control	δ	**-196.63**	**6.74**	**406.74**	**0.16**
	None	**-196.89**	**7.35**	**408.47**	**1.89**
	$C\ \rho\ \delta$	-196.46	8.81	410.54	3.96
	$\bar{v}$, TMB-1^b	-198.24	9.10	414.68	8.10

Table 1: WAIC model selection for observational data. Models for which ΔWAIC$<\approx 3$ are considered to be indistinguishable from the best model, and are therefore shown in bold face. ^{a}Because the model with no heterogeneity in transmission ($C = 0$) did not converge, the likelihood score for that model is a rough estimate based on non-converged MCMC samples. ^{b}For this model, the prior on the $\bar{v}$ value in the sprayed plots was based on our experimental estimate for the sprayed virus, TMB-1.

Appendices, we show that the visual fit of these models to the data is almost indistinguishable from the fit of the best model, and here we discuss what this result means for the relevance of our experiments for understanding epizootics.

First, our laboratory measure of the speed of kill, the inverse of δ, apparently provides a reasonably accurate representation of speeds of kill in nature. This is at least moderately surprising, given that temperatures control death rates, and given that temperatures in nature likely

bioRxiv preprint doi: https://doi.org/10.1101/285080; this version posted March 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

fluctuate in a complex way, while temperatures in our lab experiment were tightly constrained. Temperature variation in nature may therefore play only a minor role in epizootic dynamics. Our measure of the ratio parameter ρ may similarly be not that far from typical values of ρ in nature. This result is perhaps less surprising, given that ρ depends only on counts of occlusion bodies, which are likely to be the same in the lab as in the field. The success of models with experiment-based priors on heterogeneity C is more surprising, given that C was measured in a field experiment. Mechanisms that affect variation in infection risk across individuals may therefore act in similar ways at the scale of single branches, and at the scale of entire epizootics.

In contrast, models that had informative priors on $\bar{v}$ provided meaningfully worse explanations than the best model, with ΔWAIC values of 6 or more (Table 1). There may therefore be mechanisms operating at large spatial scales and long temporal scales that are not accounted for in our experiments. For example, Douglas-fir tussock moth larvae tend to occur at slightly higher densities near the crowns of trees (C. Mehmel, unpublished), and this aggregation may increase infection rates relative to our experiments (Dwyer and Elkinton, 1993).

Evidence for effects of such large-scale mechanisms on the transmission of this or other baculoviruses, however, is almost entirely lacking (D'Amico and Elkinton, 1995; Dwyer, 1991). Moreover, failed efforts to estimate the decay rate μ experimentally suggest that μ may be very low for this virus (Dwyer (1991), Online Appendices). This is important because the median values of μ estimated for models with experiment-based priors on transmission $\bar{v}$ are much lower than the median values of μ estimated for the other models. Better experimental estimates of μ might thus support the models with experiment-based priors on $\bar{v}$.

The worse fit provided by the models with informative priors on $\bar{v}$ therefore may not have a direct biological interpretation. Instead, it may arise from the statistical effects of compounded error in the observational and experimental data sets. Notably, our experimental estimate of $\bar{v}$ depends on our estimate of the ratio parameter ρ, such that high values of ρ lead to low estimates of $\bar{v}$ (Online Appendices). Estimating ρ is difficult enough that such underestimation may have been a significant problem (Online Appendices), and this may be why our experimental estimate

bioRxiv preprint doi: https://doi.org/10.1101/285080; this version posted March 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

of $\bar{\nu}$ is substantially lower than the posterior median values of $\bar{\nu}$ from the cases in which we used uninformative priors (fig. 4). For the other parameters the posterior medians when we used experiment-based priors were remarkably similar to the posterior medians when we used uninformative priors.

Also, the predictions of the SEIR model, equations (1a)-(1d), are quite sensitive to small changes in the transmission parameter $\bar{\nu}$, but are only moderately sensitive to changes in the heterogeneity parameter C (Dwyer et al., 1997). The models with informative priors on $\bar{\nu}$ may therefore have performed poorly because small errors in our estimate of $\bar{\nu}$ have bigger effects than small errors in our estimate of C. This problem may have been made worse by the strong variation across pathogen isolates that is apparent in our experimental data (fig. 3), a point that we consider further in the Discussion. Moreover, the fit to data of the model with experiment-based priors on C, ρ, δ, and $\bar{\nu}$ is only slightly worse (fig. 5) than the fit of the model with all uninformative priors (fig. 2), in spite of the former model's substantially worse WAIC score.

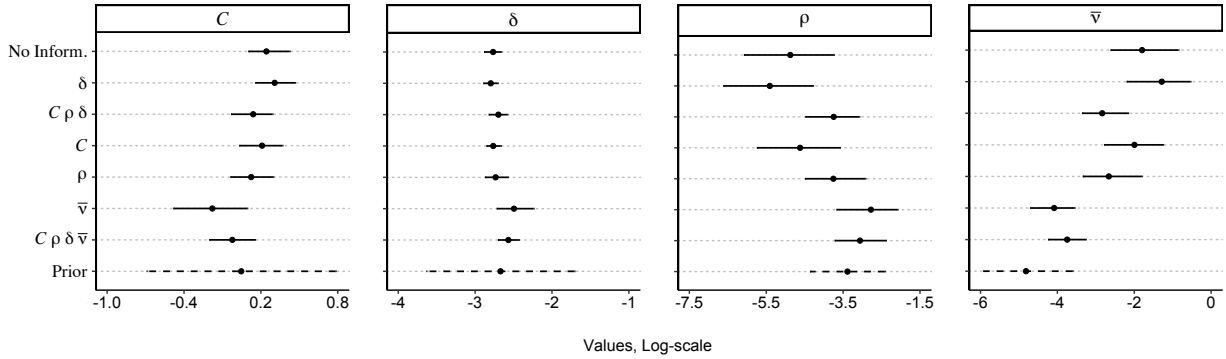

Figure 4: Marginal posterior parameter estimates (columns) from the 1-$\bar{\nu}$ class of models (rows), compared to the informative prior distributions derived from experiments (dashed lines). Dots are median values, while bars are 95% credible intervals. Note that each panel has a different scale on its horizontal axis.

Another reason why the models with informative priors on the transmission parameter have larger (poorer) WAIC scores is that the best way for these models to fit the low-density plots is by assuming that stochasticity was higher in those plots, an effect that is clearly apparent

bioRxiv preprint doi: https://doi.org/10.1101/285080; this version posted March 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

in fig. 5. Increased stochasticity produces poorer WAIC scores because the model-complexity term increases with the variance in the likelihood, which of course increases with stochasticity. Surprisingly, however, increased stochasticity also increases infection rates. This is because increased stochasticity raises the rate at which both very low and very high transmission rates occur in our simulations, but higher transmission rates have disproportionately strong effects (in the Online Appendices, we prove this assertion). When fitting models with an informative prior on the transmission rate, the fitting algorithm therefore compensates for low transmission rates by increasing the stochasticity parameter σ.

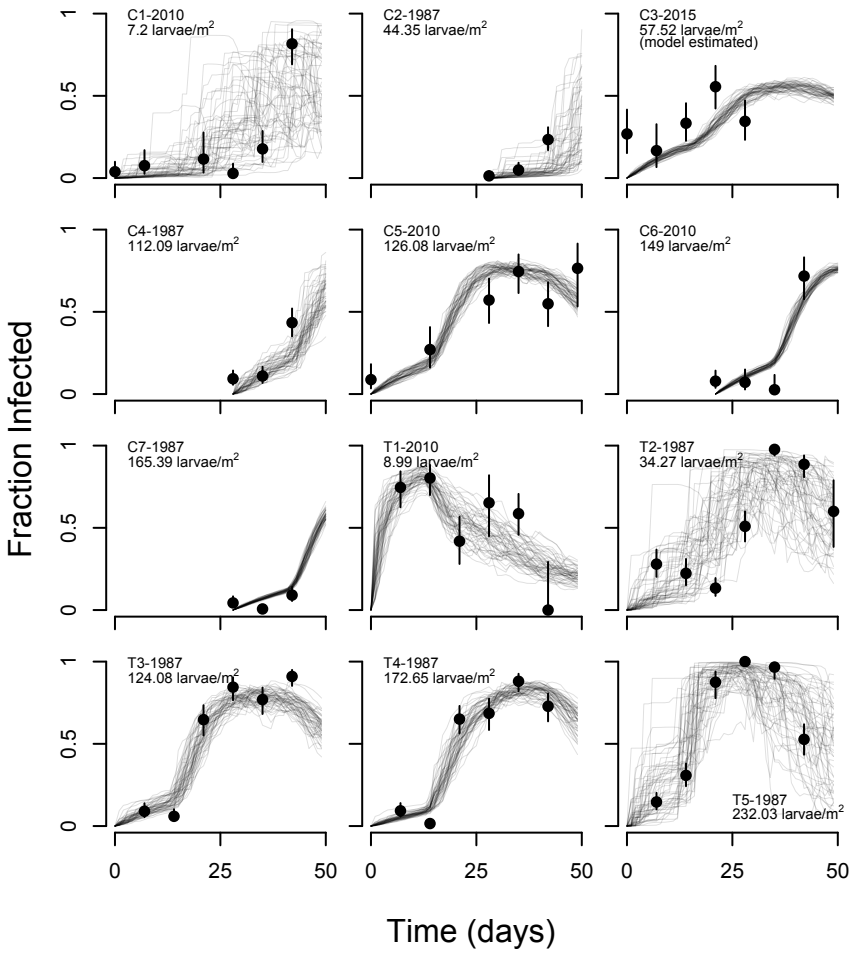

Figure 5: Stochastic realizations of the model with informative priors on $\bar{v}$, C, ρ, and δ, with symbols and labels as in fig. 2

bioRxiv preprint doi: https://doi.org/10.1101/285080; this version posted March 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

The Disease-density Threshold

As table 2 shows, our estimates of the disease-density threshold $\mu/\bar{v}$ are remarkably consistent, even though our estimates of μ and $\bar{v}$ differ depending on whether we take into account our experimental data. This is a consequence of a strong correlation in the posterior samples of $\bar{v}$ and μ (Online Appendices), which in turn reflects the similar effects that transmission rates and removal rates have on disease dynamics. From a statistical perspective, the correlation means that separate estimation of $\bar{v}$ and μ requires more data than if the two parameters were not correlated, supporting our use of experimental data.

Model (Informative Priors)	Disease-Density Threshold ($\mu/\bar{v}$)
None	0.29 (0.01, 1.04)
δ	0.16 (0.003, 0.60)
$C\ \rho\ \delta$	0.29 (0.001, 1.66)
C	0.40 (0.03, 1.19)
$C\ \rho\ \delta\ \bar{v}$	0.29 (0.01, 1.57)
ρ	0.47 (0.01, 2.18)
$\bar{v}$	1.05 (0.03, 4.38)

Table 2: Estimates of the disease-density threshold for each model, presented in order of increasing WAIC scores. Estimates are shown as medians with 95% credible intervals in parentheses. Thresholds are in units of larvae/m^2.

Our estimates of the disease-density threshold suggest that it may be effective to spray tussock moth populations that are well below densities at which significant defoliation occurs. Densities slightly above the threshold are too low to produce defoliation, but are apparently high enough to support at least modest epizootics. The reason why epizootics do not normally occur at such densities is probably that there is a delay between when the insect population begins to increase,

bioRxiv preprint doi: https://doi.org/10.1101/285080; this version posted March 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

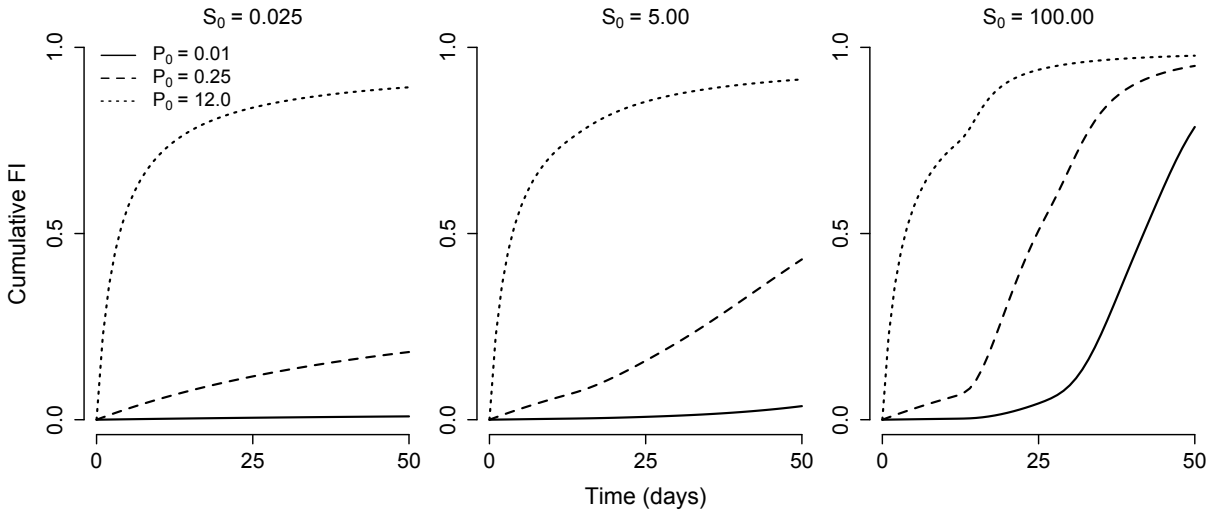

Figure 6: Simulated trajectories of cumulative mortality (i.e. cumulative fraction infected (FI)), where S_0 and P_0 are the initial host densities (larvae m^{-2}) and virus densities (infectious cadaver equivalents m^{-2}), respectively.

and when the baculovirus population begins to increase. This effect is known from both long-term insect outbreak models (Dwyer et al., 2000), and classical predator-prey models (Kot, 2001). In the case of the tussock moth baculovirus, the delay may be so severe that it may make economic sense to spray tussock moth populations that would otherwise require an additional 2 or 3 years of increases before natural baculovirus epizootics would occur.

This point can be made more explicit by considering how the model behavior varies with initial host and pathogen densities. The middle panel in fig. 6 shows that, for an initial density of 5 larvae/m^2, the input of a modest amount of virus can lead to a reasonably severe epizootic. This density is lower than all but one of the spray plots in our data set (fig. 2), emphasizing that spray programs could be successful even at low densities. In the data for the treatment plot in 2010, for example, spraying a population with an initial density of 9 larvae/m^2 produced high mortality.

Meanwhile, at high host densities, infection rates are likely to be high even for very low ini-

bioRxiv preprint doi: https://doi.org/10.1101/285080; this version posted March 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

tial pathogen densities, an effect that is well known from epidemiological theory (Keeling and Rohani, 2008). This effect is apparent in the rightmost panel in fig. 6, for which the initial host density of 100 larvae/m^2 is lower than in 3 of our 5 spray plots. Virus mortality may therefore have been high in those 3 plots even if they had not been sprayed. Previous researchers recommended that "a natural virus level greater than 25% indicates a [tussock moth] population that is already about to collapse from an epizootic, so direct control is not necessary" (Brookes et al., 1978; Stelzer, 1979). Our work suggests that this rule of thumb is too conservative, especially in high-density populations.

Discussion

The success of our models (fig. 2) strongly supports the hypothesis that density-dependent transmission drives epizootics of the tussock moth baculovirus. For many vertebrate diseases in contrast, the logistical hurdles of data collection often make it difficult to test for effects of density on transmission (McCallum, 2016; McCallum et al., 2001). In the insect pathology literature, the importance of density is widely acknowledged (Moreau and Lucarotti, 2007), but there is also a long tradition of explaining epizootics by invoking complex mechanisms other than density (Fuxa and Tanada, 1987). Perhaps for this reason, disease-density thresholds are invoked less commonly in microbial control than in wildlife conservation, yet the unexpected natural epizootics discussed here are not unusual in microbial control (Moreau and Lucarotti, 2007; Podgwaite, 1985; Wollam et al., 1978). A clearer consideration of the effects of host and pathogen densities could therefore improve control strategies (Brown, 1987; Moscardi, 1999; Podgwaite, 1985).

This is especially true for efforts to control tussock moth populations, which rarely consider host density as a factor influencing the effectiveness of control. Instead, host density is used to predict how much defoliation damage will occur, and thus whether control measures should be implemented (Maclauchlan et al., 2009; Mason and Wickman, 1991). Previous protocols in this system have nevertheless suggested that host densities are important and that costs could

bioRxiv preprint doi: https://doi.org/10.1101/285080; this version posted March 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

be reduced by spraying earlier, lowering the dosage of sprayed virus, or reducing the area of application (Maclauchlan et al., 2009; Otvos et al., 1998). By forecasting the severity of natural epizootics relative to microbial control, parameterized epidemiological model like ours could serve as useful tools in such efforts (Otvos et al., 1998), leading to decisions that incorporate the relative costs of spraying versus defoliation.

Indeed, our models confirm that in some cases the best way to reduce the costs of tussock moth control is to forgo a spray program altogether (fig. 6, table 2). Previous efforts to estimate initial virus incidence in tussock moth populations, however, have often relied on small sample sizes (Stelzer, 1979). At small sample sizes, the virus may be undetected even though it is present at a high enough density to cause a severe epizootic in a dense host population. Inaccurate estimation of the initial fraction infected can therefore lead to the unnecessary application of virus spray, suggesting that more intensive sampling may be worth the effort. This is especially true given that the model predictions are more strongly affected by such inaccurate estimation when host densities are intermediate. It is nevertheless important to recognize that low initial infection rates may mean that it takes weeks or months for substantial mortality to occur. During this delay, defoliation may be economically significant, and so microbial control may be necessary despite a high observed population density.

Our estimates of the heterogeneity parameter C are mostly greater than 1, and thus high enough to prevent outbreaks in insect-pathogen models unless there is fluctuating selection for disease resistance (Elderd et al., 2008). Natural selection may therefore play a key role in tussock moth population cycles, as in the gypsy moth (Páez et al., 2017). More broadly, genetic variation in susceptibility to disease is omnipresent (Keeling and Rohani, 2008), and so it seems likely that the lack of evidence for host-variation effects in other pathogens is due to a lack of consideration of host-variation effects, rather than to an actual lack of host variation. From a biological-control perspective, high heterogeneity means that a substantial fraction of the population will survive even a severe epizootic. This does not mean, however, that defoliation will occur in the following year, because over-wintering virus may decimate the host population in the following year.

bioRxiv preprint doi: https://doi.org/10.1101/285080; this version posted March 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

It is also important to point out the ways in which our models do not explain the data. First, the success of models with experiment-based priors on the heterogeneity parameter suggests that transmission at a small scale helps drive epizootics at a large scale, but the poor performance of models that use priors on the transmission rate suggests that larger scale mechanisms also play a role. Second, our models require at least moderate stochasticity in transmission to explain the epizootic data, but they do not identify the source of this stochasticity. An important future direction is therefore to consider aspects of the biology of the disease that are not in the model. For example, there are two morphotypes of the baculovirus that may differ in infectiousness (Martignoni and Iwai, 1977), and the frequency of the morphotypes varies with latitude (Williams et al., 2011). Allowing for this variation in our models was impractical, because the data do not include information about which morphotype was present in each epizootic, but pathogen variation may help explain why models with informative priors on transmission were less successful. More concretely, the latitudinal variation in the virus means that microbial control of the Douglas-fir tussock moth may be improved if different morphotypes are used at different latitudes. Understanding the consequences of morphotype variation for epizootic severity is therefore a key priority in future work. Long-term host-pathogen models have shown that insect pathogens can drive host population cycles (Anderson and May, 1980; Hochberg, 1989), and our work suggests that more mechanistic versions of such models (Dwyer et al., 2000) could be useful in guiding microbial control. Previous efforts to model Douglas-fir tussock moth population dynamics in contrast have focused on insect population growth and forest biology (Bousfield et al., 1984; Colbert et al., 1979; Mason and Wickman, 1991; Monserud and Crookston, 1982; Moore and Hatch, 1981), while providing only highly simplified descriptions of baculovirus dynamics (Brookes et al., 1978). Another future direction is therefore to extend the SEIR-type models that we have presented here to allow for long-term dynamics.

Parameterized models of baculovirus dynamics that could be used in microbial control programs are exceedingly rare (Brown, 1987). This is partly because baculovirus sprays are not widely used, due to the high cost of *in vivo* propagation, and the failure of *in vitro* propagation

bioRxiv preprint doi: https://doi.org/10.1101/285080; this version posted March 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

methods, which lead to the emergence of strains that do not produce occlusion bodies (Lacey et al., 2015). As is often the case in environmental management, however, a major obstacle to the development of useful models has been the difficulty of estimating model parameters (Grenfell et al., 2002). As our works shows, high-performance computing and advanced statistical algorithms are now sufficient to provide reasonably accurate parameter values. The high levels of computing resources that we used are nevertheless often unavailable to environmental managers, emphasizing the usefulness of collaborations between managers and university researchers (Joseph et al., 2013).

Our Bayesian approach has the advantage that it allowed us to test the extent to which our small-scale, experimental data help us to understand large-scale epizootic data. More conventional maximum-likelihood methods have driven the development of flexible "plug-and-play" software, which has been used to fit complex models like ours to data (Ionides et al., 2006, 2015; King et al., 2016), leading to important insights into disease dynamics (King et al., 2016; Martinez et al., 2016; Ranjeva et al., 2017). Maximum likelihood methods, however, do not allow for comparison of models that do and do not take into account prior information, thus preventing comparison of models that do and do not take into account experimental data. A Bayesian version of particle filtering shows great promise, but sometimes fails to converge on the best parameters (Ionides et al., 2015).

bioRxiv preprint doi: https://doi.org/10.1101/285080; this version posted March 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Acknowledgments

We are grateful for the support of dedicated and talented field technicians: Ruby An, Kate Lynne Logan, Rachel Hosman, Hannah Koch, Jenni Novak, Katherine Sirianni, Jeffrey Thorburn, Alison Hunter, Cara Skalisky, Alyssa Taylor, Jason Sims, and Rita Bennett. Roy Magelssen provided information on spray projects in general and some field support for the 2010 project on the Methow Ranger District. Natural epizootic data were provided with the assistance of Tom Eckberg, Idaho Department of Lands and by Rebecca Powell, Forest Health Protection, Rocky Mountain Region. JRM was funded by a US Department of Agriculture (USDA) National Institute of Food and Agriculture (NIFA) Postdoctoral Fellowship (2014-67012-22272). Additional work was funded by the Okanogan-Wenatchee National Forest (Methow Ranger District), and a grant to KMP and GD from the USDA Forest Service Pesticide Impacts Assessment Program. Helpful feedback was provided by Iral Ragenovich, USDA Forest Service, Region 6 Forest Health Protection. Imre Otvos kindly provided the epizootic data from 1987. M.E. Martignoni provided important encouragement to G.D. at a crucial career juncture decades ago.

bioRxiv preprint doi: https://doi.org/10.1101/285080; this version posted March 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Literature Cited

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

Anderson, R. M., and R. M. May. 1979. Population biology of infectious diseases: Part i. Nature 280:361–367.

———. 1980. Infectious-diseases and population-cycles of forest insects. Science 210:658–661.

———. 1992. Infectious diseases of humans: dynamics and control. Oxford University Press, Oxford.

Bernardo, J., and A. Smith. 2009. Bayesian Theory. Wiley Series in Probability and Statistics. Wiley.

Bolker, B. M. 2008. Ecological models and data in R. Princeton University Press.

Bousfield, W., J. Brickell, J. Cleeves, R. Dezellem, T. Gregg, T. Hofacker, R. Johnson, J. Lewis, G. Lightner, T. Quinn, et al. 1984. Economics of douglas-fir tussock moth control. Washington, DC: US Department of Agriculture, Forest Service .

Brookes, M. H., R. W. Stark, and R. W. Campbell. 1978. The Douglas-fir tussock moth: a synthesis. 1585. Forest Service, US Department of Agriculture.

Brown, G. 1987. Modeling. Pages 43–71 *in* J. Fuxa and Y. Tanada, eds. Epizootiology of Insect Diseases. John Wiley & Sons, New York, NY.

Colbert, J., W. Overton, and C. White. 1979. Documentation of the Douglas-fir Tussock Moth Outbreak-Population Model. Tech. rep., Pacific Northwest Forest and Range Experiment Station, Portland, Oregon.

Cory, J. S., and K. Hoover. 2006. Plant-mediated effects in insect-pathogen interactions. Trends in Ecology and Evolution 21:278–286.

bioRxiv preprint doi: https://doi.org/10.1101/285080; this version posted March 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Cory, J. S., and J. H. Myers. 2003. The ecology and evolution of insect baculoviruses. Annual Reviews of Ecology and Systematics 34:239–272.

D'Amico, V., and J. S. Elkinton. 1995. Rainfall effects on transmission of gypsy-moth (Lepidoptera: Lymantiidae) nuclear polyhedrosis-virus. Environmental Entomology 24:1144–1149.

Dwyer, G. 1991. The effects of density, stage and spatial heterogeneity on the transmission of an insect virus. Ecology 72:559–574.

Dwyer, G., J. Dushoff, J. S. Elkinton, and S. A. Levin. 2000. Pathogen-driven outbreaks in forest defoliators revisited: Building models from experimental data. American Naturalist 156:105–120.

Dwyer, G., and J. S. Elkinton. 1993. Using simple-models to predict virus epizootics in gypsy-moth populations. Journal Of Animal Ecology 62:1–11.

Dwyer, G., J. S. Elkinton, and J. P. Buonaccorsi. 1997. Host heterogeneity in susceptibility and disease dynamics: Tests of a mathematical model. American Naturalist 150:685–707.

Eakin, L., M. Wang, and G. Dwyer. 2015. The effects of the avoidance of infectious hosts on infection risk in an insect-pathogen interaction. Am. Nat. 185:pp. 100–112.

Elderd, B., V. Dukic, and G. Dwyer. 2006. Uncertainty in predictions of disease spread and public-health responses to bioterrorism and emerging diseases. Proceedings of the National Academy of Sciences 103:15693–15697.

Elderd, B. D., J. Dushoff, and G. Dwyer. 2008. Host-pathogen interactions, insect outbreaks, and natural selection for disease resistance. The American Naturalist 172:829–842.

Fleming-Davies, A. E., V. Dukic, V. Andreasen, and G. Dwyer. 2015. Effects of host heterogeneity on pathogen diversity and evolution. Ecology Letters 18:1252–1261.

bioRxiv preprint doi: https://doi.org/10.1101/285080; this version posted March 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Fuller, E., B. D. Elderd, and G. Dwyer. 2012. Pathogen persistence in the environment and insect-baculovirus interactions: Disease-density thresholds, epidemic burnout, and insect outbreaks. Am. Nat. 179:pp. E70–E96.

Fuxa, J. R., and Y. Tanada. 1987. Epizootiology of insect diseases. Wiley.

Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin. 2014. Bayesian Data Analysis, Third Edition. Chapman & Hall/CRC Press. New York, NY.

Grenfell, B., W. Amos, P. Arneberg, O. Bjornstad, J. Greenman, J. Harwood, P. Lanfranchi, A. McLean, R. Norman, A. Read, et al. 2002. Visions for future research in wildlife epidemiology. Ecology of Wildlife Diseases pages 151–164.

Grimm, V., E. Revilla, U. Berger, F. Jeltsch, W. M. Mooij, S. F. Railsback, H.-H. Thulke, J. Weiner, T. Wiegand, and D. L. DeAngelis. 2005. Pattern-oriented modeling of agent-based complex systems: lessons from ecology. science 310:987–991.

Hochberg, M. E. 1989. The potential role of pathogens in biological control. Nature 337:262–265.

Hooten, M. B., and N. Hobbs. 2015. A guide to bayesian model selection for ecologists. Ecological Monographs 85:3–28.

Hunter-Fujita, F. R., P. F. Entwistle, H. F. Evans, and N. E. Crook. 1998. Insect viruses and pest management. John Wiley and Sons: Somerset, New Jersey.

Ionides, E. L., C. Bretó, and A. A. King. 2006. Inference for nonlinear dynamical systems. Proceedings of the National Academy of Sciences of the United States of America 103:18438–43.

Ionides, E. L., D. Nguyen, Y. Atchadé, S. Stoev, and A. A. King. 2015. Inference for dynamic and latent variable models via iterated, perturbed Bayes maps. Proceedings of the National Academy of Sciences 112:719–724.

bioRxiv preprint doi: https://doi.org/10.1101/285080; this version posted March 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Joseph, M. B., J. R. Mihaljevic, A. L. Arellano, J. G. Kueneman, D. L. Preston, P. C. Cross, and P. T. J. Johnson. 2013. Taming wildlife disease: bridging the gap between science and management. Journal of Applied Ecology 50:702–712.

Kareiva, P., and M. Andersen. 1988. Spatial aspects of species interactions: the wedding of models and experiments. Pages 35–50 *in* Community ecology. Springer.

Kareiva, P., and G. Odell. 1987. Swarms of predators exhibit" preytaxis" if individual predators use area-restricted search. The American Naturalist 130:233–270.

Keeling, M., M. Woolhouse, R. May, G. Davies, and B. Grenfell. 2003. Modelling vaccination strategies against foot-and-mouth disease. Nature 421:136–142.

Keeling, M. J., and P. Rohani. 2008. Modeling Infectious Diseases in Humans and Animals. Princeton University Press.

Kennedy, D. A., V. Dukic, and G. Dwyer. 2014. Pathogen Growth in Insect Hosts: Inferring the Importance of Different Mechanisms Using Stochastic Models and Response-Time Data. American Naturalist 184:407–423.

Kermack, W., and A. McKendrick. 1927. A contribution to the mathematical theory of epidemics. Proceedings of the Royal Society of London, Series A 115:700–721.

King, A. A., E. L. Ionides, M. Pascual, and M. J. Bouma. 2008. Inapparent infections and cholera dynamics. Nature 454:877–U29.

King, A. A., D. Nguyen, and E. L. Ionides. 2016. Statistical Inference for Partially Observed Markov Processes via the R Package pomp. Journal of Statistical Software 69.

Konishi, S., and G. Kitagawa. 2008. Information criteria and statistical modeling. Springer Science & Business Media.

Kot, M. 2001. Elements of mathematical ecology. Cambridge University Press.

bioRxiv preprint doi: https://doi.org/10.1101/285080; this version posted March 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Lacey, L., D. Grzywacz, D. Shapiro-Ilan, R. Frutos, M. Brownbridge, and M. Goettel. 2015. Insect pathogens as biological control agents: back to the future. Journal of invertebrate pathology 132:1–41.

Levin, S. A. 1992. The problem of pattern and scale in ecology: the robert h. macarthur award lecture. Ecology 73:1943–1967.

Longini, I. M., A. Nizam, S. Xu, K. Ungchusak, W. Hanshaoworakul, D. A. Cummings, and M. E. Halloran. 2005. Containing pandemic influenza at the source. Science 309:1083–1087.

Maclauchlan, L., P. Hall, I. Otovs, and J. Brooks. 2009. An integrated management system for the douglas-fir tussock moth in southern british columbia. BC Journal of Ecosystems and Management 10:22–36.

Martignoni, M., and P. Iwai. 1977. Thermal inactivation characteristics of 2 strains of nucleopolyhedrosis virus (baculovirus subgroup A) pathogenic for *Orgyia pseudotsugata*. Journal of Invertebrate Pathology 30:255–262.

Martignoni, M. E. 1999. History of tm biocontrol-1, the first registered virus-based produced for control of a forest insect. The American Entomologist 45:30–37.

Martinez, P. P., A. A. King, M. Yunus, A. Faruque, and M. Pascual. 2016. Differential and enhanced response to climate forcing in diarrheal disease due to rotavirus across a megacity of the developing world. Proceedings of the National Academy of Sciences 113:4092–4097.

Mason, R. 1996. Dynamic behavior of Douglas-fir tussock moth populations in the Pacific northwest. Forest Science 42:182–191.

Mason, R., and B. Wickman. 1991. Integrated pest-management of the Douglas-fir tussock moth. Forest Ecology and Managements 39:119–130. Special Symposium at the 18th International Congress of Entomology, Vancouver, Canada, 1988.

bioRxiv preprint doi: https://doi.org/10.1101/285080; this version posted March 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

May, R. M., and R. M. Anderson. 1979. Population biology of infectious diseases: Part ii. Nature 280:455–461.

McCallum, H. 2016. Models for managing wildlife disease. Parasitology 143:805–820.

McCallum, H., N. Barlow, and J. Hone. 2001. How should pathogen transmission be modelled? Trends in ecology & evolution 16:295–300.

McCullagh, P., and J. Nelder. 1989. Generalized Linear Models. Chapman & Hall, Boca Raton, FL.

Mihaljevic, J., B. Hoye, and P. Johnson. 2017. Parasite metacommunities: Evaluating the roles of host community composition and environmental gradients in structuring symbiont communities within amphibians. The Journal of animal ecology .

Miller, L. K. 1997. The baculoviruses. Plenum Press.

Monserud, R. A., and N. L. Crookston. 1982. A user's guide to the combined stand prognosis and douglas-fir tussock moth outbreak model. United States. Intermountain Forest and Range Experiment Station. USDA Forest Service general technical report INT (USA) .

Moore, J., and C. Hatch. 1981. A Simulation Approach For Predicting The Effect Of Douglas-Fir Tussock Moth Defoliation On Juvenile Tree Growth And Stand Dynamics. Forest Science 27:685–700.

Moreau, G., and C. J. Lucarotti. 2007. A brief review of the past use of baculoviruses for the management of eruptive forest defoliators and recent developments on a sawfly virus in canada. Forestry Chronicle 83:105–112.

Moscardi, F. 1999. Assessment of the Application of Baculoviruses for Control of Lepidoptera. Annual Review of Entomology 44:257–289.

bioRxiv preprint doi: https://doi.org/10.1101/285080; this version posted March 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Otvos, I., L. Maclauchlan, P. Hall, and N. Conder. 1998. A management system to control douglas-fir tussock moth *Orgyia pseudotsugata* using OpNPV. Technology Transfer Note 11, Pacific Forestry Centre, Forestry Canada.

Otvos, I. S., J. C. Cunningham, and L. M. Friskie. 1987. Aerial application of nuclear polyhedrosis virus against douglas-fir tussock moth, *Orgyia pseudostugata* (Mcdunnough) (Lepidoptera: Lymantriidae). 1. impact in the year of application. Canadian Entomologist 119:697–706.

O'Dea, E. B., K. M. Pepin, B. A. Lopman, and C. O. Wilke. 2014. Fitting outbreak models to data from many small norovirus outbreaks. Epidemics 6:18–29.

Páez, D., V. Dukic, J. Dushoff, A. Fleming-Davies, and G. Dwyer. 2017. Effects of pathogen exposure on life-history variation in the gypsy moth (lymantria dispar). The American Naturalist accepted pending minor revision.

Podgwaite, J. 1985. Strategies for field use of baculoviruses. Pages 775–799 *in* K. Maramorosch and K. Sherman, eds. Viral Insecticides for Biological Control. Academic Press, Orlando, Florida.

Polivka, K., G. Dwyer, and C. Mehmel. 2017. Environmental persistence of a pathogen used in microbial insect control. Research Note PNW-RN-573, Pacific Northwest Research Station, USDA Forest Service.

Polivka, K., G. Dwyer, K. C. Skalisky, C. J. Mehmel, and J. L. Novak. 2012. Analysis of NPV epizootics during a Douglas-fir tussock moth management project in the Methow Ranger District (Okanogan-Wenatchee National Forest), 2010. Report to Okanogan-Wenatchee National Forest, Forest Health Protection .

Ranjeva, S. L., E. B. Baskerville, V. Dukic, L. L. Villa, E. Lazcano-Ponce, A. R. Giuliano, G. Dwyer, and S. Cobey. 2017. Recurring infection with ecologically distinct hpv types can explain high prevalence and diversity. Proceedings of the National Academy of Sciences page 201714712.

bioRxiv preprint doi: https://doi.org/10.1101/285080; this version posted March 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Reed, C., I. Otvos, R. Reardon, I. Ragenovich, and H. Williams. 2003. Effects of long-term storage on the stability of opmnpv dna contained in tm biocontrol-1. Journal of invertebrate pathology 84:104–113.

Rohrmann, G. F. 2014. Baculovirus nucleocapsid aggregation (MNPV vs SNPV): an evolutionary strategy, or a product of replication conditions? Virus Genes 49:351–357.

Ruder, M. G., T. J. Lysyk, D. E. Stallknecht, L. D. Foil, D. J. Johnson, C. C. Chase, D. A. Dargatz, and E. P. J. Gibbs. 2015. Transmission and epidemiology of bluetongue and epizootic hemorrhagic disease in north america: Current perspectives, research gaps, and future directions. Vector-Borne and Zoonotic Diseases 15:348–363.

Scott, D., and L. Spiegel. 2002. One and two year follow-up evaluation of TM Biocontrol-1 treatments to suppress Douglas-fir tussock moth in the Blue Mountains of northeastern Oregon and southeastern Washington. Technical Report BMPMSC-02-02, USDA Forest Service, Pacific Northwest Region.

Shepherd, R., D. Bennett, J. Dale, S. Tunnock, R. Dolph, and R. Thier. 1988. Evidence of synchronized cycles in outbreak patterns of Douglas-fir tussock moth, *Orgyia pseudotsugata* (McDunnough) (Lepidoptera:Lymantriidae). Memoirs of the Entomological Society of Canada pages 107–121.

Shepherd, R. F., I. S. Otvos, R. J. Chorney, and J. C. Cunningham. 1984. Pest-management of Douglas-fir tussock moth (Lepidoptera: Lymantriidae) - prevention of an outbreak through early treatment with a nuclear polyhedrosis-virus by ground and aerial applications. Canadian Entomologist 116:1533–1542.

Stelzer, R. 1979. How to determine the incidence of virus in egg masses. Douglas-fir tussock moth handbook. USDA agriculture handbook .

Thompson, C., and D. Scott. 1979. Production And Persistence Of The Nuclear Polyhedrosis-

bioRxiv preprint doi: https://doi.org/10.1101/285080; this version posted March 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Virus Of The Douglas-Fir Tussock Moth, Orgyia-Pseudotsugata (Lepidoptera, Lymantriidae), In The Forest Ecosystem. Journal Of Invertebrate Pathology 33:57–65.

Watanabe, S. 2010. Asymptotic Equivalence of Bayes Cross Validation and Widely Applicable Information Criterion in Singular Learning Theory. Journal of Machine Learning Research 11:3571–3594.

———. 2013. A widely applicable Bayesian information criterion. The Journal of Machine Learning Research 14:867–897.

Williams, H. L., K. S. Monge-Monge, I. S. Otvos, R. Reardon, and I. Ragenovich. 2011. Genotypic variation among Douglas-fir tussock moth nucleopolyhedrovirus (OpNPV) isolates in the western United States. Journal of Invertebrate Pathology 108:13–21.

Wollam, J. D., W. G. Yendol, and F. B. Lewis. 1978. Evaluation of aerially-applied nuclear polyhedrosis virus for suppression of the gypsy moth, Lymantria dispar L. Forest Service Research Paper NE-396:8 p.