

## RESEARCH ARTICLE

# When and why direct transmission models can be used for environmentally persistent pathogens

Lee Benson<sup>1,2,3,4\*</sup>, Ross S. Davidson<sup>2,4</sup>, Darren M. Green<sup>3</sup>, Andrew Hoyle<sup>1</sup>,  
Mike R. Hutchings<sup>2</sup>, Glenn Marion<sup>4</sup>

**1** Computing Science and Mathematics, University of Stirling, Stirling, United Kingdom, **2** Disease Systems, Animal And Veterinary Sciences, Scotland's Rural College, Edinburgh, United Kingdom, **3** Institute of Aquaculture, University of Stirling, Stirling, United Kingdom, **4** Biomathematics and Statistics Scotland, Edinburgh, United Kingdom

✉ These authors contributed equally to this work.

‡ AH and GM also contributed equally to this work.

\* [lee.benson@bioss.ac.uk](mailto:lee.benson@bioss.ac.uk)



## OPEN ACCESS

**Citation:** Benson L, Davidson RS, Green DM, Hoyle A, Hutchings MR, Marion G (2021) When and why direct transmission models can be used for environmentally persistent pathogens. *PLoS Comput Biol* 17(12): e1009652. <https://doi.org/10.1371/journal.pcbi.1009652>

**Editor:** Claudio José Struchiner, Fundação Getúlio Vargas: Fundacao Getulio Vargas, BRAZIL

**Received:** June 7, 2021

**Accepted:** November 16, 2021

**Published:** December 1, 2021

**Copyright:** © 2021 Benson et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** There are no primary data in the paper; all materials and code are available at [https://github.com/cenlb/DT\\_Models\\_Paper\\_Code](https://github.com/cenlb/DT_Models_Paper_Code).

**Funding:** G.M., R.S.D. and M.R.H. are supported by the Scottish Government's Rural and Environment Science and Analytical Services Division (RESAS), Innovation Project BIO/01/16. L. B. was supported by a studentship funded by Biomathematics and Statistics Scotland, SRUC and University of Stirling. The funders had no role in

## Abstract

Variants of the susceptible-infected-removed (SIR) model of Kermack & McKendrick (1927) enjoy wide application in epidemiology, offering simple yet powerful inferential and predictive tools in the study of diverse infectious diseases across human, animal and plant populations. Direct transmission models (DTM) are a subset of these that treat the processes of disease transmission as comprising a series of discrete instantaneous events. Infections transmitted indirectly by persistent environmental pathogens, however, are examples where a DTM description might fail and are perhaps better described by models that comprise explicit environmental transmission routes, so-called environmental transmission models (ETM). In this paper we discuss the stochastic susceptible-exposed-infected-removed (SEIR) DTM and susceptible-exposed-infected-removed-pathogen (SEIR-P) ETM and we show that the former is the timescale separation limit of the latter, with ETM host-disease dynamics increasingly resembling those of a DTM when the pathogen's characteristic timescale is shortened, relative to that of the host population. Using graphical posterior predictive checks (GPPC), we investigate the validity of the SEIR model when fitted to simulated SEIR-P host infection and removal times. Such analyses demonstrate how, in many cases, the SEIR model is robust to departure from direct transmission. Finally, we present a case study of white spot disease (WSD) in penaeid shrimp with rates of environmental transmission and pathogen decay (SEIR-P model parameters) estimated using published results of experiments. Using SEIR and SEIR-P simulations of a hypothetical WSD outbreak management scenario, we demonstrate how relative shortening of the pathogen timescale comes about in practice. With attempts to remove diseased shrimp from the population every 24h, we see SEIR and SEIR-P model outputs closely coincide. However, when removals are 6-hourly, the two models' mean outputs diverge, with distinct predictions of outbreak size and duration.

study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

## Author summary

Mathematical models of the spread and progression of communicable disease in populations are important tools in efforts to prevent and control outbreaks. A common class of disease models assume that infection is transmitted directly from infectious to susceptible individuals when they are in close proximity—so called direct transmission models. These are used widely and have proven invaluable as simplified descriptions of a wide array of infectious diseases in diverse populations. However, many pathogens spread through indirect, environmental routes of transmission, for example via contact with contaminated water sources in the case of cholera, or inhalation of infectious airborne droplets for respiratory infections, such as Covid-19.

We show that direct transmission models work well for such pathogens with short environmental lifetimes and where hosts shed pathogens into the environment at high rates. This means that we do not require information about environmental pathogen levels to understand the behaviour of outbreaks caused by these pathogens. When shedding rates are also low, e.g., with macroparasitic infections, or when variable environmental factors play a role in transmissibility, then explicit modelling of both the pathogen and environmental transmission will provide a more accurate picture than a direct transmission approximation.

## Introduction

The famed Susceptible-Infectious-Recovered (SIR) compartmental model framework of Kermack and McKendrick [1], and its many subsequent extensions (see [2], for example), stand as prominent examples of what can be gained from simple models of complex systems. In addition to the assumption that the host population can be divided into a finite number of discrete states, transmission of infection within such models is characterised by a force of infection that depends linearly upon the numbers of infectious individuals. Throughout this paper, we call such a model a direct transmission model (DTM) (as in [3]). The proportionality constant, known as the transmission rate, is often interpreted as the rate at which individuals within the population come into contact with each other, times the probability that such a contact leads to the transmission of infection (termed an infectious contact in [4]), times the probability of successful transmission. However, this simplified representation has been extended, for example, to account for non-uniform frequency of contact among the population and levels of infectiousness varying across individuals and over time.

Such approaches have been pivotal in gaining valuable insights into the dynamics and patterns of the spread of disease throughout many varied populations. Recently, for example, DTMs have served as the basis for understanding drivers of spatial spread of Ebola virus [5], the likely effectiveness of scaling up certain vaccination, treatment and testing regimes in the fight to control hepatitis B [6] and the importance of targeting household transmission of MRSA as a preventative strategy [7]. Incorporation of a spatial element into the DTM framework enables the observed spatial-temporal trajectory of the 2001 foot and mouth outbreak in the UK to be closely replicated and provides insight for control [8, 9]. DTMs have also been recently drafted into the effort to understand and predict the dynamics of SARS-CoV-2 [10, 11].

Despite these successes, DTMs may not always be appropriate, e.g., when members of the host population are in contact with environmental sources of infection, such as pools of

pathogens residing on surfaces or in water bodies. The focus of this paper is to critique and support the use of DTMs in describing the spread of disease in the presence of such environmental pools of persistent pathogens. Examples of relevant disease systems include cholera [12, 13], avian influenza [14, 15] and even respiratory infections, such as SARS-CoV-2 [16]. Our case study (in Case study: White spot disease in penaeid shrimp) focusses on infectious disease spread in aquaculture systems which are likely to feature a large degree of environmental transmission. In scenarios such as these it is prolonged exposure to these sources, in addition to direct infectious contact between individuals, that gives rise to new infections, in the most general case. Environmental transmission models (ETM, introduced below) are a second family of candidate models for such infections. As well as describing direct transmission of disease between hosts, ETMs also allow indirect, environmental routes of transmission due to interaction between the hosts and external pathogen, whose dynamics are in turn described in terms of shedding from infectious hosts as well as pathogen decay (or loss of viability). Epidemiological models, including DTMs and ETMs, are approximations of complex biological processes driving disease transmission and progression. The main goal of this paper is to show that the relative timescale separation between the host and pathogen populations determines whether environmental transmission due to contact with external pathogens shed by infectious hosts can be more parsimoniously described as direct transmission within the DTM framework. A key advantage of this approach is that we do not need information about environmental pathogen levels in order to fit DTMs.

A critical step in applications is the fitting of DTMs to data on disease outbreaks. Using Bayesian model fitting methods and graphical posterior predictive checks (GPPCs) that target observable characteristics of an outbreak, such as its final size and when it peaks, we show that DTMs fit very well to simulated host-disease event times that would occur when the infection is transmitted environmentally but the rates of pathogen emission and removal are high (Results). We show that this is explained by the force of infection within an ETM behaving increasingly like that of a DTM when the timescale of the pathogen population is shorter than that of the host population (Results). When fitted to simulated outbreak data, we find that DTMs still make accurate predictions of outbreak size and duration even when the underlying data generating process comprises a low rate of pathogen emission and long pathogen lifetimes. It is only the rate at which outbreaks grow towards their peak that is poorly predicted in such cases.

These issues are further highlighted in a case study illustrating the use of DTMs as approximate descriptions of outbreaks of disease due to an environmentally persistent pathogen (Case study: White spot disease in penaeid shrimp). Using parameter values estimated from published data on WSD infection in penaeid shrimp, we explore how imperfect interventions that aim to remove dead and diseased hosts at regular intervals impact outbreak control in closed populations of this aquaculture disease system. With removal attempts spaced at 24 hourly intervals, average outbreak trajectories, final outbreak size, outbreak duration, are accurately captured using a DTM, without need to model the pathogen load. When the frequency of the removal events are increased to every 6 h, we begin to see divergence between the two models, so that, e.g., the DTM predicts slightly larger outbreaks of shorter duration than the ETM. This case study illustrates the potential practical consequences of ignoring issues of timescale separation when applying DTMs to environmentally transmitted pathogens. Control and other processes in such disease systems may be accurately captured by DTMs at one timescale but are poorly represented at others; in this case underestimating the benefit of high frequency removal.

Tables 1 and 2 contain summaries of symbols and abbreviations used throughout this paper.

**Table 1. Summary of symbols and their units used throughout this paper.**

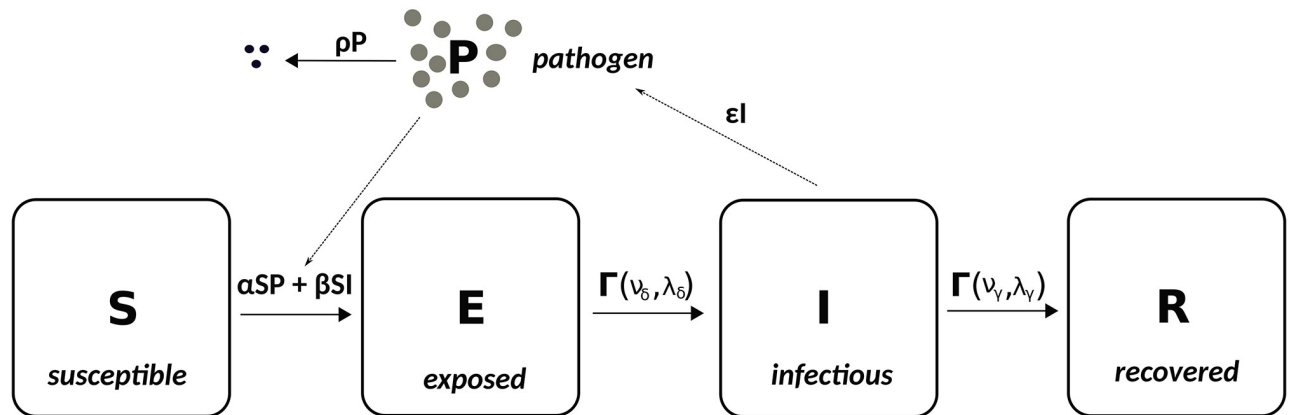
Symbol	description	units
$S_t, E_t, I_t, R_t$	numbers of susc., exp'd, infectious and rem'd hosts at time $t$	hosts
$N$	total host population size	hosts
$P_t$	size of environmental pathogen pool at time $t$	virions
$\alpha$	environmental transmission rate	virion <sup>-1</sup> h <sup>-1</sup>
$\beta$	direct transmission rate	host <sup>-1</sup> h <sup>-1</sup>
$\nu_\delta, \nu_\gamma$	(gamma-distributed) exposed and infectious lifetime shape parameters	-
$\lambda_\delta, \lambda_\gamma$	exposed and infectious lifetime rate parameters	h <sup>-1</sup>
$\delta, \gamma$	(exponentially distributed) exposed and infectious lifetime rate parameters	h <sup>-1</sup>
$\epsilon$	pathogen emission rate	host <sup>-1</sup> h <sup>-1</sup>
$\rho$	pathogen decay rate	h <sup>-1</sup>
$f(t)$	force of infection at time $t$	h <sup>-1</sup>
$\mathbf{t}^E, \mathbf{t}^I, \mathbf{t}^R$	times of exposure, onsets of infectivity, removals	h
$m$	final outbreak size	hosts
$\omega_\beta, \omega_\gamma, \omega_\delta$	prior exponential rate parameters for $\beta, \gamma, \delta$	h <sup>-1</sup>
$\kappa$	host index corresponding to index exposure	-
$t_\kappa^E$	time of index exposure	h
$t_{\text{peak}}$	time of outbreak peak	h
$I_{\text{max}}$	size of outbreak at peak	hosts
$c$	relates to interpolated value for $t_{\text{peak}}$	-
$\hat{\beta} \left( = \beta + \alpha \frac{\epsilon}{\rho} \right)$	effective direct transmission rate	host <sup>-1</sup> h <sup>-1</sup>
$\hat{\delta}, \hat{\gamma}$	reciprocals of mean simulated exposed and infectious lifetimes	h <sup>-1</sup>
$R_0, \hat{R}_0$	basic reproduction ratio for SEIR / SEIR-P models	-
$\beta_{\text{ingest}}, \tilde{\beta}_{\text{ingest}}$	rates of transmission due to ingestion in resp. Tuyen, et al. and Lotz & Soto	host <sup>-1</sup> h <sup>-1</sup>
$\beta_{\text{cohab}}, \tilde{\beta}_{\text{cohab}}$	similar to above for cohabitation	host <sup>-1</sup> h <sup>-1</sup>
$\alpha_L$	lower estimate of $\alpha$ (Case Study)	ml virion <sup>-1</sup> h <sup>-1</sup>
$\rho_U$	upper estimate of $\rho$ (Case Study)	h <sup>-1</sup>
$\pi_E, \pi_I$	probability of removal of exposed, infectious hosts (Case Study)	-

<https://doi.org/10.1371/journal.pcbi.1009652.t001>

**Table 2. Summary of abbreviations employed in the text.**

Abbreviation	meaning
DTM	direct transmission model
ETM	environmental transmission model
DT (ET)	direct (environmental) transmission
SIR	susceptible-infectious-removed DTM
SEIR	susceptible-exposed-infectious-removed DTM
SIR-P	suseptible-exposed-infectious-removed-pathogen ETM
SEIR-P	similar to SIR-P
MCMC	Markov chain Monte Carlo
GPPC	graphical posterior-predictive check
DTA	direct transmission approximation
IDP	immigration-death process
SIWR	susceptible-infectious-waterborne reservoir-removed
WSD	white spot disease
WSSV	white spot syndrome virus

<https://doi.org/10.1371/journal.pcbi.1009652.t002>



**Fig 1. Susceptible-exposed-infectious-removed-pathogen (SEIR-P) environmental transmission model diagram illustrating the four host compartments ((S)usceptible, (E)xposed, (I)nfectious and (R)emoved, or recovered) and single pathogen compartment (P) of the model.** The solid arrows indicate the movement of hosts between host compartments and the loss of viable pathogen from the system. The dotted arrows indicate how the host and pathogen parts of the model influence each other. The parameters  $\alpha$  and  $\beta$  are the *environmental* and *direct transmission rates* while  $\epsilon$  and  $\rho$  are the rates of pathogen emission and decay. The time that hosts spend in the E and I compartments is determined by the chosen *exposed* and *infectious* lifetime distributions, e.g., gamma distributions with shape parameters  $v_E$  and  $v_I$  and rate parameters  $\lambda_E$  and  $\lambda_I$ , as shown here. When exponential distributions are assumed, then the rate parameters are denoted respectively by  $\delta$  and  $\gamma$  (see Table 1 for a summary of all symbols used throughout the paper). When  $\alpha = \epsilon = \rho = 0$  we obtain the SEIR direct transmission model as a submodel.

<https://doi.org/10.1371/journal.pcbi.1009652.g001>

## Models for direct and environmental transmission of disease

**A direct transmission model: Susceptible-Exposed-Infectious-Removed (SEIR).** The stochastic, compartmental SEIR model (see Fig 1) referred to throughout this paper is a DTM that treats the focal population (the *hosts*) as divided into four sub-populations: hosts that are susceptible to disease (S), have been exposed to the disease but not yet infectious (E), infectious (I) and recovered or removed from the population (R). Hosts in the R compartment play no further role in the spread of disease. Here we consider outbreaks started by the introduction of a single host to a wholly susceptible population of size  $N$  (this is known as the *index exposure*), however, our results are generalisable to greater numbers of initial exposures. In addition a closed population is assumed, so that there is no immigration, births or non disease-induced mortality. Hosts remain in the E and I compartments for periods of time determined by the *exposed* and *infectious lifetime distributions*, i.e., random variables with continuous, positive distributions. In the case of exponentially-distributed lifetime distributions, with rates  $\delta$  and  $\gamma$  respectively for the exposed and infectious states, the resulting process is a continuous time Markov chain and the current state of the system is fully determined by the number of hosts in the four compartments. The SEIR model may be specialised by stipulating that hosts spend a period of zero duration in the E compartment, resulting in a SIR model.

In Case study: White spot disease in penaeid shrimp we modify this model by additionally allowing, at regular intervals, each host in the E and the I compartments to go directly to the R compartment with probabilities  $\pi_E$  and  $\pi_I$ . This represents regular attempts, with error, to remove all exposed and infectious hosts from the system.

The direct transmission assumption means that secondary cases are generated at a rate dependent on the number of infected individuals. Here we adopt the standard approach to modelling this. The probability that each susceptible host at time  $t$  becomes exposed to disease over the short time interval  $(t, t + h]$ , is  $\beta I_t h$  to first order, where  $\beta$  is the *direct transmission rate* and  $I_t$  is the number of infectious hosts present at time  $t$ . The force of infection, i.e., the rate of secondary infections, per susceptible host, or the imminent risk of infection that is

faced by each susceptible host, at time  $t$  is therefore expressed as

$$f(t) = \beta I_t. \quad (1)$$

This means that the force of infection is proportional (by the factor  $\beta$ ) to the number of infected hosts currently present. When the number of infected hosts increases or decreases, there is a corresponding instantaneous change in the force of infection. This is a key feature of this and other DTMs and represents an important modelling assumption.

**An environmental transmission model: Susceptible-Exposed-Infectious-Removed-Pathogen (SEIR-P).** Here we define a class of models that represent both direct (DT) and environmental transmission (ET) of disease (see, for example, [17]) where the latter occurs via interaction of susceptible hosts with environmental pools of infectious pathogen. SEIR-P describes the time evolution of two populations: the hosts, divided into S, E, I and R sub-populations, measured in hosts (as in the DTM SEIR, described above) and the pathogen population (P—measured here in virions) in the environment, external to the hosts. When hosts enter state I they begin to emit pathogen at the fixed rate  $\epsilon$ , i.e., they begin to contribute to an increase in the environmental pathogen load. The pathogen population decays exponentially at rate  $\rho$ .

Each susceptible host at time  $t$  now becomes exposed to disease over the short time interval  $(t, t + h]$  with probability  $(\alpha P_t + \beta I_t)h$  to first order, where the summand  $\beta I_t$  represents that part of the force of infection contributed by direct transmission, as in the SEIR model, and  $\alpha$  and  $P_t$  represent the *indirect* or *environmental transmission rate* and size of the pathogen population (in virions) at time  $t$ , respectively. The force of infection is now

$$f(t) = \alpha P_t + \beta I_t, \quad (2)$$

which depends linearly on both the size of the environmental pathogen load and the number of infectious hosts. A change in the number of infected hosts now produces a delayed response in the force of infection due to the pathogen load taking time to either build up or decay.

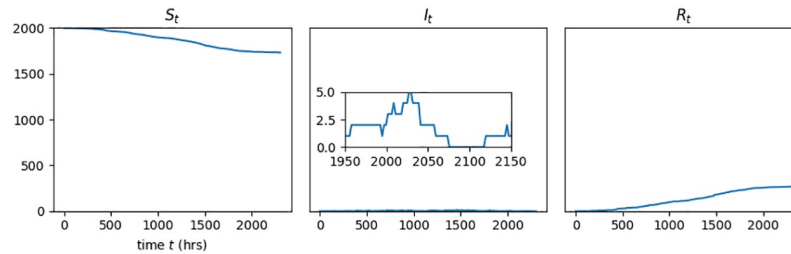
Setting  $\alpha$  to zero and restricting attention to the four host population compartments recovers the SEIR model described above and setting  $\beta = 0$  produces a pure ET model. Finally, stipulating that hosts pass instantaneously from the E to the I compartment yields a SIR-P model with both DT and ET.

### The direct transmission approximation as timescale limit of SEIR-P process

The two populations described by an ETM, the hosts and environmental pathogen, each have a timescale characterising their evolution, and the extent to which these differ has a qualitative effect upon the behaviour of the model. For example, ETMs with shed pathogen retaining infectivity for durations comparable with the typical host infectious lifetime can exhibit outbreaks that appear to have died out, in terms of infected individuals, only to restart (see Fig 2). Similar behaviour is prohibited under the SIR model, where zero infectious hosts implies that there is no more force of infection to drive the outbreak forward. On the other hand, ETMs with short-lived pathogens produce host-disease dynamics that are reproducible with a host-only DTM, as we now demonstrate.

The probability that a susceptible host at time  $t$  becomes exposed within the short interval of time  $(t, t + h]$  is  $(\alpha P_t + \beta I_t)h$ , to first order. The relationship between the qualitative behaviour of the SEIR-P model and pathogen timescale comes down to the appearance of  $P_t$  in this expression. How does  $P_t$  behave? At time  $t$ , each infectious host is emitting pathogen at the rate  $\epsilon$  so, overall, new pathogen is entering the population at the rate  $\epsilon I_t$ . At the same time, the





**Fig 2. Host S, I and R outputs from a single realisation of SIR-P model with environmental transmission only (i.e.,  $\beta = 0 \text{ host}^{-1} \text{ h}^{-1}$ ) among a host population of size  $N = 2000$ .** The rates of environmental transmission and pathogen emission are  $\alpha = 1.325 \times 10^{-10} \text{ virion}^{-1} \text{ h}^{-1}$  and  $\epsilon = 3462.5 \text{ virion host}^{-1} \text{ h}^{-1}$  and the rates of pathogen decay,  $\rho$ , and host recovery,  $\gamma$ , are both  $0.029 \text{ h}^{-1}$  (equivalent to a half-life of 24 h). The inset in the central panel is the key feature of interest and shows how the number of infectious hosts goes to zero roughly between the times  $t = 2075 \text{ h}$  and  $2120 \text{ h}$ . Under the SIR model, as soon as the number of infectious hosts reaches zero no further secondary infections are possible. However, within the SIR-P model with comparable pathogen and host infectious lifetimes, outbreaks can appear to die off but later continue, due to the force of infection from long-living pathogens.

<https://doi.org/10.1371/journal.pcbi.1009652.g002>

pathogen decays exponentially at rate  $\rho$ , or equivalently, each pathogen element remains viable on average for a duration  $1/\rho$  (in appropriate units). This amounts to  $P_t$  being an immigration-death process (IDP) with inhomogenous immigration rate  $\epsilon I_t$  and death rate  $\rho$ . As a consequence of elementary theory of Markov chains (see, e.g., [18]), if we momentarily regard the number of infectives as fixed,  $I_t = i$  for  $t \geq 0$ , then the distribution of  $P_t$  tends to Poisson  $\left(\frac{\epsilon i}{\rho}\right)$  and

$$\mathbf{E}(P_t) \rightarrow \frac{\epsilon i}{\rho} \text{ as } t \rightarrow \infty. \tag{3}$$

Increasing the values of the parameters  $\epsilon$  and  $\rho$ , while keeping their ratio constant, increases the rate of convergence but has no other effect upon this limiting behaviour, with the limiting distribution unchanged. Returning now to variable  $I_t$ , in the limiting case,  $P_t$ 's behaviour can be characterised approximately as being in equilibrium, i.e.,  $P_t \sim \text{Poisson}\left(\frac{\epsilon i}{\rho}\right)$  during intervals of constant  $I_t = i$ , and jumping without transition to a new equilibrium when  $I_t$  changes.

Consequently, for large  $\epsilon$  and  $\rho$ , we may approximate the probability that a susceptible host becomes exposed over the interval  $(t, t + h]$  as

$$(\alpha \mathbf{E}(P_t) + \beta I_t)h = \left(\alpha \frac{\epsilon}{\rho} + \beta\right) I_t h = \hat{\beta} I_t h. \tag{4}$$

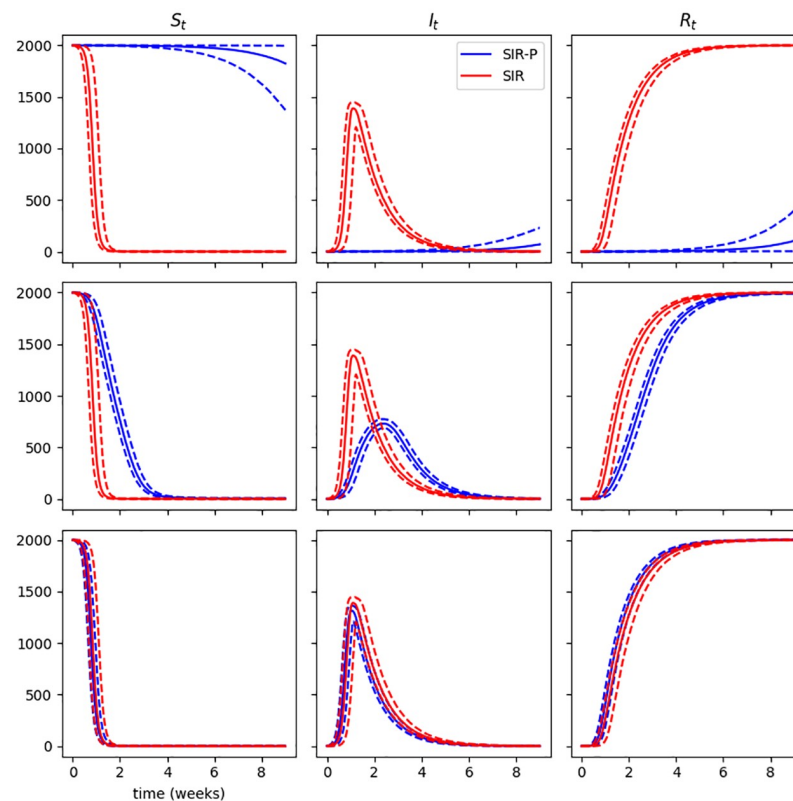
The part of the SEIR-P model that describes the host disease dynamics is approximately a direct transmission SEIR model, with an *effective* transmission rate  $\hat{\beta}$  and E and I lifetime distributions unchanged. This SEIR model is the *direct transmission approximation* (DTA) of the ETM.

This approximation of the sub-process,  $P_t$ , is the stochastic analogue of the “quasi-steady state approximation” of the pathogen concentration discussed by Tien and Earn in their susceptible-infected-waterborne reservoir-removed (SIWR) ordinary differential equation model of cholera outbreaks among humans [19], in which the pathogen concentration in water sources is restricted to the “critical manifold” of fixed points of the flow. This is related to the concept of “timescale separation” [20] within complex systems, whereby one component of

the system is undergoing changes rapidly enough to be considered as jumping instantaneously between equilibria.

The vanishing lag between changes in the number of shedding, infectious hosts, and the resulting response in the force of infection is the reason why ET via short-lived pathogen can be approximated as DT. Systems with ET that results from the accumulation of relatively long-lived pathogen retain a *memory* of the size of the infectious host sub-population since hosts that have since been removed, or have ceased to shed, may still be the cause of new exposures via the pathogen that they had previously emitted. This delay between changes in the number of infectious hosts and their effect on the system dynamics violates the implicit assumption of DTMs that the force of infection is directly related to the number of infectious hosts, or the number of hosts who are shedding pathogen. As the pathogen lifetime decreases, so does this memory effect, and we see an increasing similarity with the dynamics produced by direct transmission.

Fig 3 demonstrates this behaviour by showing susceptible, infectious and removed sub-population sizes averaged over a number of independent simulations from three SIR-P models



**Fig 3. Susceptible, infectious and removed host sub-population sizes of SIR-P (blue) process averaged over 5000 simulations for fixed environmental transmission rate  $\alpha = 5.95 \times 10^{-7}$  virion $^{-1}$  h $^{-1}$  and host mortality rate  $\gamma = 5.95 \times 10^{-3}$  h $^{-1}$ . The rates of pathogen emission,  $\epsilon$ , and pathogen removal,  $\rho$ , are increased while keeping their ratio fixed,  $\frac{\epsilon}{\rho} = 50.0$ . **Top row:**  $\epsilon = 2.98 \times 10^{-2}$  host $^{-1}$  h $^{-1}$ ,  $\rho = 5.95 \times 10^{-4}$  h $^{-1}$ , **middle row:**  $\epsilon = 2.98$  host $^{-1}$  h $^{-1}$ ,  $\rho = 5.95 \times 10^{-2}$  h $^{-1}$ , **bottom row:**  $\epsilon = 2.98 \times 10^2$  host $^{-1}$  h $^{-1}$ ,  $\rho = 5.95$  h $^{-1}$ . For comparison, the same sub-population sizes for the DTA SIR process with fixed direct transmission rate  $\beta = \alpha \frac{\epsilon}{\rho} = 2.98 \times 10^{-1}$  host $^{-1}$  h $^{-1}$  and  $\gamma = 5.95 \times 10^{-3}$  h $^{-1}$  are plotted in (red). Median population sizes indicated by bold lines, dashed lines indicate 5<sup>th</sup> and 95<sup>th</sup> percentiles. The top row of panels show two processes that are visibly distinct in their outputs, but with a hundred-fold increase in the pathogen decay rate, a closer alignment between the two sets of trajectories can be seen in the middle row. In the last case (bottom row) there is no difference on the scale of the plots between the SIR-P and SIR model outputs.**

<https://doi.org/10.1371/journal.pcbi.1009652.g003>



with fixed  $\alpha$  and  $\gamma$  and increasing  $\epsilon$  and  $\delta$ , with  $\frac{\epsilon}{\rho}$  held constant. These are plotted in each case against similar outputs from a SIR model with the same  $\gamma$  and  $\beta$  chosen to equal  $\alpha \frac{\epsilon}{\rho}$ , so that the SIR model is the DTA of the SIR-P model. As shown in the figure, increasing rates of pathogen emission and decay rate lead to increasing similarity between the SIR-P and SIR model outputs, and in the case of short-lived pathogen, the SIR-P and DTA SIR outputs are indistinguishable on the scale of the plots.

## Results

### Estimating DTA parameters from outbreak data

The presence of environmental transmission violates an assumption of DTMs: that the force of infection is directly related to the number of infectious hosts currently in the system. Here we fitted the DT SEIR model to data from simulated SEIR-P outbreaks with varied rates of pathogen emission and decay and assessed the goodness of fit of the model using GPPCs (Materials and methods). The simulated data consists of times of onset of infectivity and host removal, i.e., the times of entry of hosts into the I and the R states, the observation of which is feasible in some cases (see Case study: White spot disease in penaeid shrimp).

The data set specifications are summarised in Table 3 and the MCMC methods of model fitting and a description of the GPPCs are found in Materials and methods and Section A in S1 Appendix. The first three data sets were generated using a SEIR-P process in which there was both environmental transmission due to pathogen as well as direct transmission from host to host. The rates of pathogen emission,  $\epsilon$ , and decay,  $\rho$ , were increased with  $\frac{\epsilon}{\rho}$  kept constant. The fourth data set has no environmental transmission rate, and so is the output of a DTM.

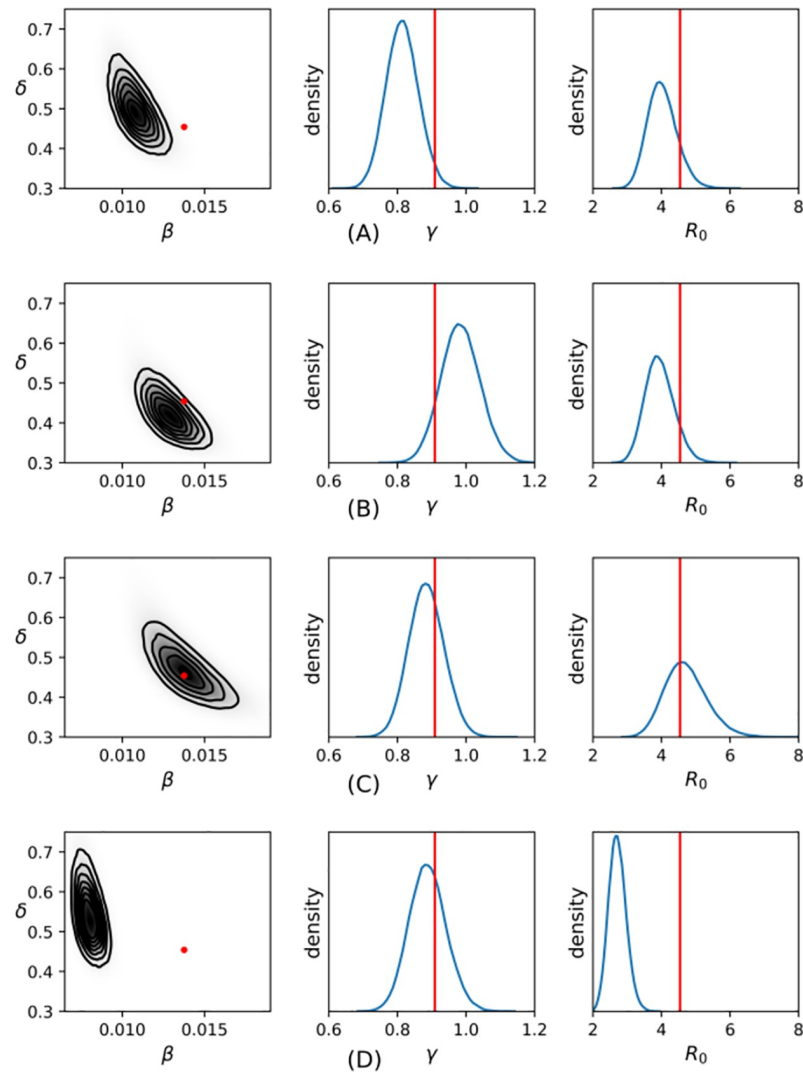
**MCMC convergence and posterior coverage.** Convergence of the MCMC sampling chains was checked by running two separate chains for each fitted model with separated initial values and observing that they converge to a common stationary distribution. Trace plots for the two parameters that were MCMC sampled,  $\beta$  and  $\delta$  can be found in Fig A in S1 Appendix.

**Table 3. Scenarios for simulation study.**

Data generating process	parameter values	N	m
A. long-lived pathogen	$\alpha = 0.001 \text{ virion}^{-1} \text{ d}^{-1}, \beta = 0.007 \text{ host}^{-1} \text{ d}^{-1}$	300	295
	$I_j - E_j \sim \text{Gamma}(1.10, 0.5 \text{ d}^{-1})$		
	$R_j - I_j \sim \text{Gamma}(1.10, 1.0 \text{ d}^{-1})$		
	$\epsilon = 5.4 \text{ virion host}^{-1} \text{ d}^{-1}, \rho = 0.8 \text{ d}^{-1}$		
B. intermediate pathogen	$\alpha = 0.001 \text{ virion}^{-1} \text{ d}^{-1}, \beta = 0.007 \text{ host}^{-1} \text{ d}^{-1}$	300	294
	$I_j - E_j \sim \text{Gamma}(1.10, 0.5 \text{ d}^{-1})$		
	$R_j - I_j \sim \text{Gamma}(1.10, 1.0 \text{ d}^{-1})$		
	$\epsilon = 54.0 \text{ virion host}^{-1} \text{ d}^{-1}, \rho = 8.0 \text{ d}^{-1}$		
C. short-lived pathogen	$\alpha = 0.001 \text{ virion}^{-1} \text{ d}^{-1}, \beta = 0.007 \text{ host}^{-1} \text{ d}^{-1}$	300	297
	$I_j - E_j \sim \text{Gamma}(1.10, 0.5 \text{ d}^{-1})$		
	$R_j - I_j \sim \text{Gamma}(1.10, 1.0 \text{ d}^{-1})$		
	$\epsilon = 5.4 \times 10^4 \text{ virion host}^{-1} \text{ d}^{-1}, \rho = 0.8 \times 10^4 \text{ d}^{-1}$		
D. direct transmission only	$\alpha = 0.0 \text{ virion}^{-1} \text{ d}^{-1}, \beta = 0.0075 \text{ host}^{-1} \text{ d}^{-1}$	300	272
	$I_j - E_j \sim \text{Gamma}(1.10, 0.5 \text{ d}^{-1})$		
	$R_j - I_j \sim \text{Gamma}(1.10, 1.0 \text{ d}^{-1})$		

Total host population size =  $N$ . Final outbreak size =  $m$ .

<https://doi.org/10.1371/journal.pcbi.1009652.t003>



**Fig 4. Density estimates of SEIR parameter posterior distributions and  $R_0$  for long-lived pathogen (A), intermediate pathogen (B), short-lived pathogen (C) and direct-transmission only (D) data sets.** The red dot in the leftmost panels indicates  $(\hat{\beta}, \hat{\delta})$ , where  $\hat{\beta} = \alpha\epsilon/\rho + \beta$  and  $\hat{\delta} = (\mathbb{E}(I_j - E_j))^{-1}$ . The red vertical lines in the central and rightmost panels indicate  $\hat{\gamma} = (\mathbb{E}(R_j - I_j))^{-1}$  and  $\hat{R}_0 = \frac{\hat{\beta}N}{\hat{\gamma}}$ , respectively. The marginal posterior distributions  $\gamma$  and  $(\beta, \delta)$  are conditionally independent and so are plotted separately.

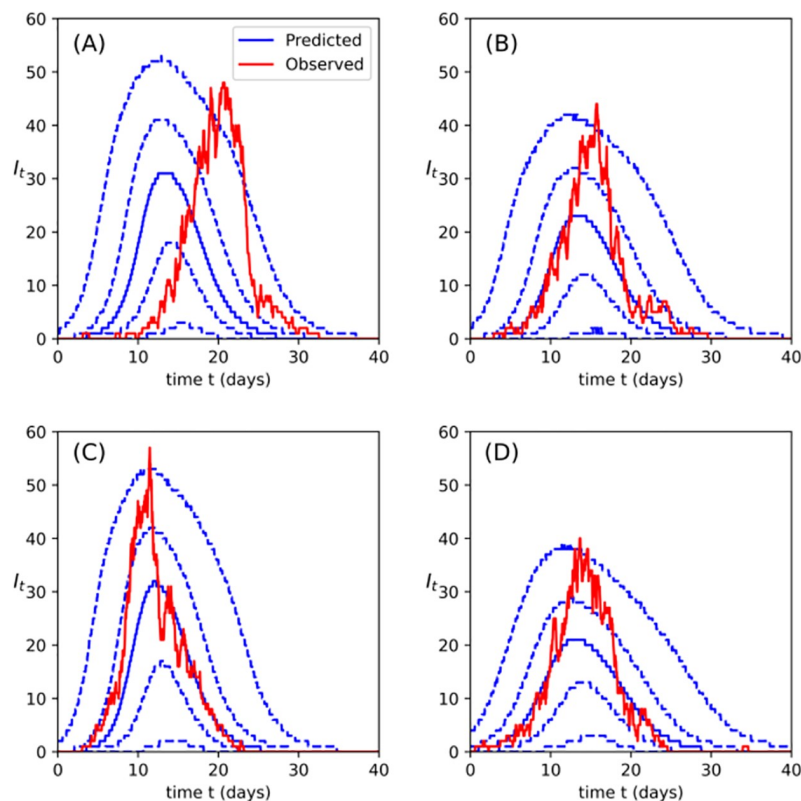
<https://doi.org/10.1371/journal.pcbi.1009652.g004>

Fig 4 is a graphical summary of the samples obtained from the posterior distribution of the parameters and of the SEIR basic reproduction ratio (the expected number of new cases of infection that result from addition of a single infected host into a large, wholly susceptible population),  $R_0 = \frac{\beta N}{\gamma}$ , for each of the fitted models. For comparison, these are plotted against the reference values  $\hat{\beta}, \hat{\delta}, \hat{\gamma}$  and  $\hat{R}_0$ , where  $\hat{\beta} = \alpha\epsilon/\rho + \beta$  is the effective transmission rate defined in The direct transmission approximation as timescale limit of SEIR-P process. The quantity  $\hat{\delta} = \mathbb{E}(t^I - t^E)^{-1}$  and the quantity  $\hat{\gamma} = \mathbb{E}(t^R - t^I)^{-1}$  are, respectively, the reciprocals of the mean E and I lifetimes in the underlying process and  $\hat{R}_0 = \frac{\hat{\beta}N}{\hat{\gamma}}$  is the basic reproductive ratio for the SEIR-P process, according to the survival function formulation (see, e.g., [21, 22]).

The position of  $\hat{\beta}$ ,  $\hat{\delta}$  and  $\hat{\gamma}$  close to the centre of the parameter posterior distribution in the short-lived pathogen case in Fig 4C means that a SEIR process with direct transmission rate  $\hat{\beta} = \alpha_p^\varepsilon + \beta$  and (exponentially-distributed) E and I lifetimes with means that match those of the underlying process is the most likely model from among the class of SEIR models with exponential lifetimes. As a result, the fitted model produces a very good estimate of  $R_0$  in relation to the underlying SEIR-P process.

As the pathogen lifetime increases in Fig 4A and 4B, leading to a lesser degree of timescale separation, the fitted models both underestimate  $R_0$  and  $\hat{\beta}$ ,  $\hat{\delta}$  and  $\hat{\gamma}$  lie further from the centre of the parameter posterior distribution (in the tails in the long-lived pathogen case).

**Assessing DTA model fit against ET outbreak.** First, we compare the outbreak size trajectories predicted by the fitted model with the trajectory obtained from the data. These were obtained by simulating SEIR event times with parameter values drawn from the posterior samples, as discussed in Materials and methods. If the model fits well then we should expect these posterior-predicted trajectories to look similar to the observed trajectory [23, 24]. Fig 5

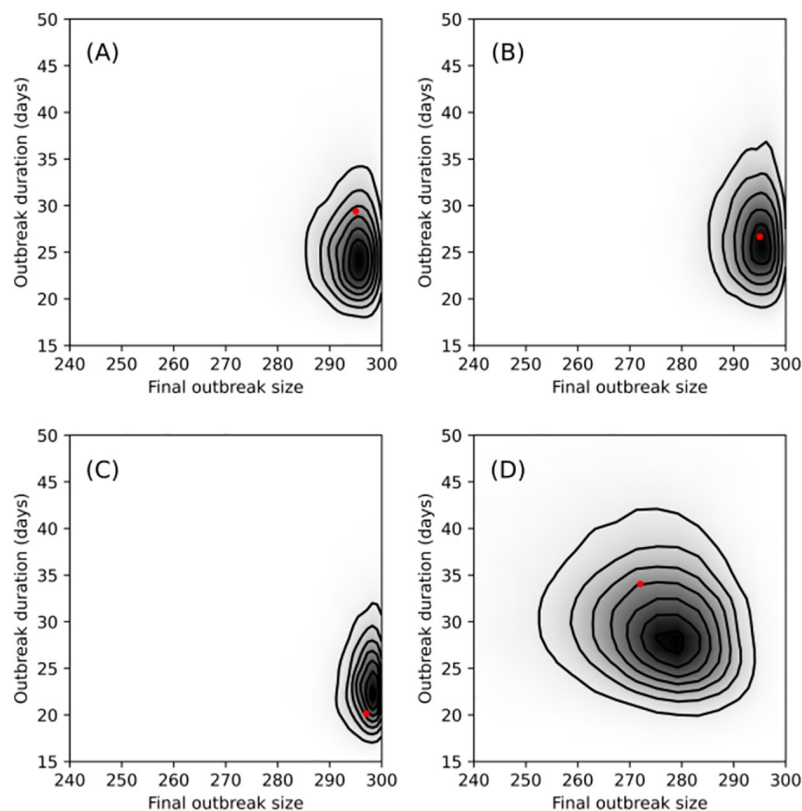


**Fig 5. Observed outbreak size trajectories,  $I_t$ , over course of a single simulated outbreak (solid red line): long-lived (A), intermediate (B) and short-lived pathogen (C) and direct transmission only (D).** These are compared with the trajectories obtained from SEIR model with MCMC-sampled parameters values, with *small* outbreaks ( $\leq 50$ ) discarded. The time axis was discretised (400 points) and 5<sup>th</sup>, 25<sup>th</sup>, 50<sup>th</sup> (median), 75<sup>th</sup> and 95<sup>th</sup> percentiles of the SEIR-predicted outbreak size were estimated at each discrete time point. The solid blue line indicates the median outbreak size while the dashed blue lines are the other percentiles. In the short-lived and direct transmission only cases, the shape of the predicted outbreak size trajectories (as indicated by the blue lines) mirrors that of the observed outbreak size, with the solid red and blue lines aligning at the initial exponential growth phase, as well as at the end of the outbreak when the number of infective hosts dies out. This is not the case for the long-lived pathogen case, for which the model predicts earlier onset of growth of the outbreak, peaking somewhat earlier than was observed.

<https://doi.org/10.1371/journal.pcbi.1009652.g005>

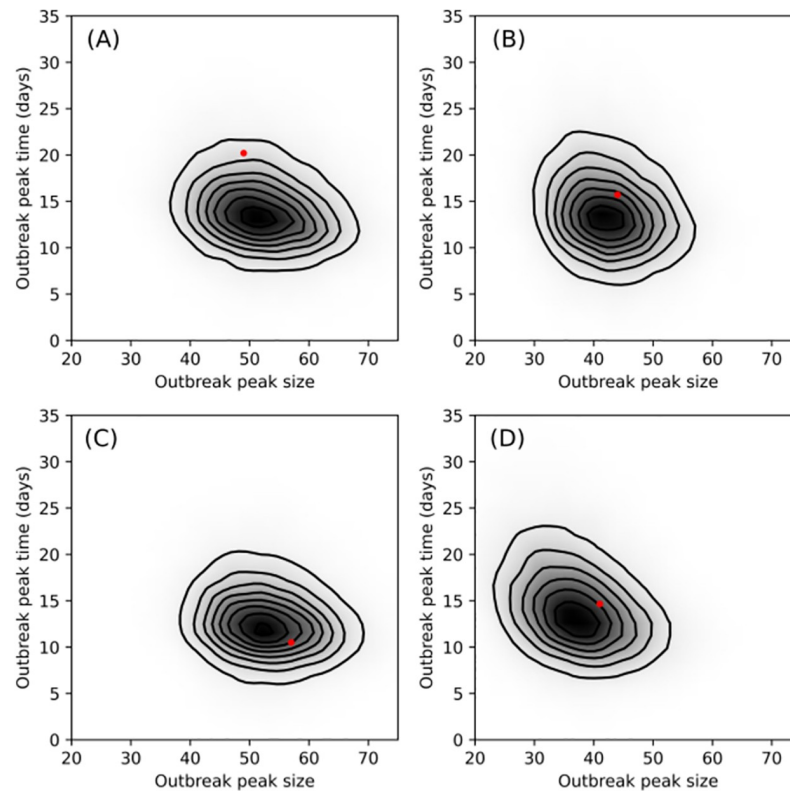
graphically compares the posterior-predictive outbreak size trajectories for the four cases described above with the underlying observed outbreak trajectory. The outbreak size trajectory from the data is indicated by the solid red line and superposed on this is a graphical summary of posterior predictive outbreak size trajectories. The solid blue line indicates the median predicted number of infectious hosts, while the blue dashed lines are the 5<sup>th</sup>, 25<sup>th</sup>, 75<sup>th</sup> and 95<sup>th</sup> percentiles. In the cases of intermediate and short-lived pathogen, the predicted model output appears to agree well with the data, as is the case with DT only. However, for the long-lived pathogen, the model appears to predict outbreaks that reach their peak and begin to recede sooner than was observed in the data. Nonetheless, the fitted model does agree with the data in terms of peak outbreak size.

Fig 6 compares graphically the final outbreak size and the outbreak duration associated with each of the four sets of onset of infectivity and removal times with the same quantities drawn from their respective posterior-predictive distributions. Fig 7 is similar, comparing size and time of outbreak peak (i.e., the outbreak at its largest) (see [Materials and methods](#) for details). As is indicated clearly in the plots, for the long-lived pathogen case, the model makes



**Fig 6. Graphical comparisons of final outbreak size (total hosts infected during outbreak) vs. outbreak duration (latest removal time minus time of first onset of infectivity) with their posterior predictive distributions.** The red dot indicates observed value of statistics from one outbreak: long-lived (A), intermediate (B) and short-lived pathogen (C) and direct transmission only (D). The shading and contours were obtained from a kernel density estimate after simulating 15000 SEIR outbreak trajectories with parameter values taken from the MCMC samples obtained while fitting the SEIR model, with *small* outbreaks ( $\leq 50$ ) discarded. In the case of long-lived pathogen, the fitted model tends to predict shorter duration outbreaks but otherwise agrees with the data in terms of final outbreak size. This is indicated by the red dot aligning horizontally with the darkest part of the density estimate but being shifted vertically. Better agreement between the data and fitted model is evident in the short-lived and intermediate pathogen and DT-only cases.

<https://doi.org/10.1371/journal.pcbi.1009652.g006>



**Fig 7. Graphical comparisons of size of outbreak peak, i.e., the size of  $I_t$  at its largest, and time of outbreak peak, as defined in main body of text with their posterior predictive distributions.** The red dot indicates observed value of statistics from one outbreak: long-lived (A), intermediate (B) and short-lived pathogen (C) and direct transmission only (D). The shading and contours were obtained from a kernel density estimate after simulating 15000 SEIR outbreak trajectories with parameter values taken from the MCMC samples obtained while fitting the SEIR model, with *small* outbreaks ( $\leq 50$ ) discarded. For long-lived pathogen, the fitted SEIR model predicts that outbreaks peak, on average, at the size observed in the data. However, the model predicts outbreaks that peak *earlier*. This is evident in panel (A), where the predicted outbreak size trajectories clearly peak earlier than the observed outbreak size trajectory indicated by the solid red line. Better agreement between data and model predictions are visible in panels (C) and (D).

<https://doi.org/10.1371/journal.pcbi.1009652.g007>

a poor prediction of when the outbreak peaks and the how long the outbreak persists before finally dying out. Better agreement is evident for the intermediate and short-lived pathogen, more so, in fact, than is evident in the DT only case.

### Case study: White spot disease in penaeid shrimp

In this section we focus on white spot disease (WSD) in penaeid shrimp, since this is a key example of an infectious disease transmitted via both DT and ET due to a pathogen known to be long-lived in the environment, under the right conditions. We formulate a SEIR-P model of the WSD-penaeid system in which attempts are made at regular intervals to remove dead and diseased hosts from the system. We estimate its parameters from published data and then compare the effect upon the outbreak trajectories resulting from stepping up the frequency of removals from every 24 h to every 6 h.

This increase in the frequency of removals is an example of how relative shortening of the SEIR-P timescales might come about in practice. We see that with removals every 24 h the SEIR-P host-disease dynamics are closely replicated by its direct transmission approximation

(DTA), whereas SEIR-P and DTA visibly diverge in their averaged outputs with removals every 6 h.

## Background: WSD and its impact

WSD is a devastating viral disease caused by the white spot syndrome virus (WSSV) that affects a wide host range, including penaeid shrimp, such as the Asian tiger shrimp, *Penaeus monodon* and the whiteleg shrimp, *Litopenaeus vannamei*. Disease onset and mortality occur in shrimp quickly after exposure, with 90% of farmed stocks typically lost to disease within 2 d to 7 d [25] and nearly 100% mortality of experimentally infected shrimps observed after 5 d to 7 d [26]. Under laboratory conditions, WSSV has been shown to retain its infectivity in sea water for up to 12 days and in sun dried and water-logged pond sediment for up to 19 and 35 days, respectively [27]. Through periodic sampling of seawater from abandoned shrimp culture ponds and surrounding canals in Vietnam, where previously an outbreak of WSD had led to 100% mortality of cultivated shrimp, the authors of [28] found that WSSV remained detectable for up to 20 months. The detection rates declined throughout the duration of the study with the steepest declines observed between July and December of both 2001 and 2002. The authors suggest this is linked to decreased plankton biomass during that period, which in turn suggests that WSSV is able to replicate within certain plankton species. Esparza-Leal, et al. [29] suggest that free-floating WSSV virions have the potential to infect shrimp in pond water at around 27°C, whereas pond water temperatures in range of 30–33°C prohibit infection. In [29] it was noted that detectability of WSSV varied among pond water samples taken simultaneously from the same pond, leading the authors to note a degree of stochasticity in relation to the waterborne pathogen load.

## Modelling WSD and estimation of SEIR-P parameters

Estimates of rates of WSD transmission via ingestion and cohabitation have been made by Lotz and Soto [30] for *Litopenaeus vannamei* and by Tuyen, et al [31] for both *Litopenaeus vannamei* and *Penaeus monodon*. Each of these estimates rely on the assumption that the force of infection is proportional to the number of infected shrimp currently present in the tank ( $I_t$ ) and therefore responds immediately to changes in this number.

Tuyen, et al, decompose the rate of direct transmission into two parts arising from ingestion and cohabitation,  $\beta = \beta_{\text{ingest}} + \beta_{\text{cohab}}$  (our notation), so that the force of infection at time  $t$  is  $\frac{\beta I_t}{N}$  (where they assume that transmission is frequency dependent). They estimate the two components of  $\beta$  using regression analysis of data obtained via an immersion challenge experiment where the relative amounts of exposure via the two routes are controlled. This is done for *Litopenaeus vannamei* and *Penaeus monodon* independently, and for both species combined.

Lotz and Soto expose *Litopenaeus vannamei* to WSSV exclusively via either ingestion or cohabitation for a set duration and estimate  $\tilde{\beta}_{\text{ingest}}$  and  $\tilde{\beta}_{\text{cohab}}$  from the numbers of shrimp that later developed the disease. Using a Reed-Frost model of epidemics (e.g., see [32]), the latter two quantities are probabilities of disease transmission per distinct susceptible-infected shrimp pair during a time interval of duration  $\Delta t$ . The force of infection here is approximately  $\frac{\tilde{\beta} I_t}{\Delta t}$  (see Section B in S1 Appendix), where  $\tilde{\beta}$  is either  $\tilde{\beta}_{\text{ingest}}$  or  $\tilde{\beta}_{\text{cohab}}$ . Lotz and Soto found  $\tilde{\beta}_{\text{cohab}}$  to be not significantly different from zero in a first experiment and to be over an order of magnitude smaller than  $\tilde{\beta}_{\text{ingest}}$  when the experiment was repeated ( $\tilde{\beta}_{\text{ingest}} = 0.56$ ,  $\tilde{\beta}_{\text{cohab}} = 0.02$ ). Such a relatively low rate of transmission due to cohabitation led the authors to omit this from their model of WSD in *Litopenaeus vannamei*, described in [33]. Tuyen, et al, found a similar result in the case of *Penaeus monodon* ( $\beta_{\text{ingest}} = 0.22 \text{ h}^{-1}$ ,  $\beta_{\text{cohab}} = 0.0026 \text{ h}^{-1}$ ) but for *Litopenaeus*



*vannamei* they in fact found that the reverse was true, in that the rate of transmission via cohabitation was greater than via ingestion ( $\beta_{\text{ingest}} = 0.0038 \text{ h}^{-1}$ ,  $\beta_{\text{cohab}} = 0.018 \text{ h}^{-1}$ ).

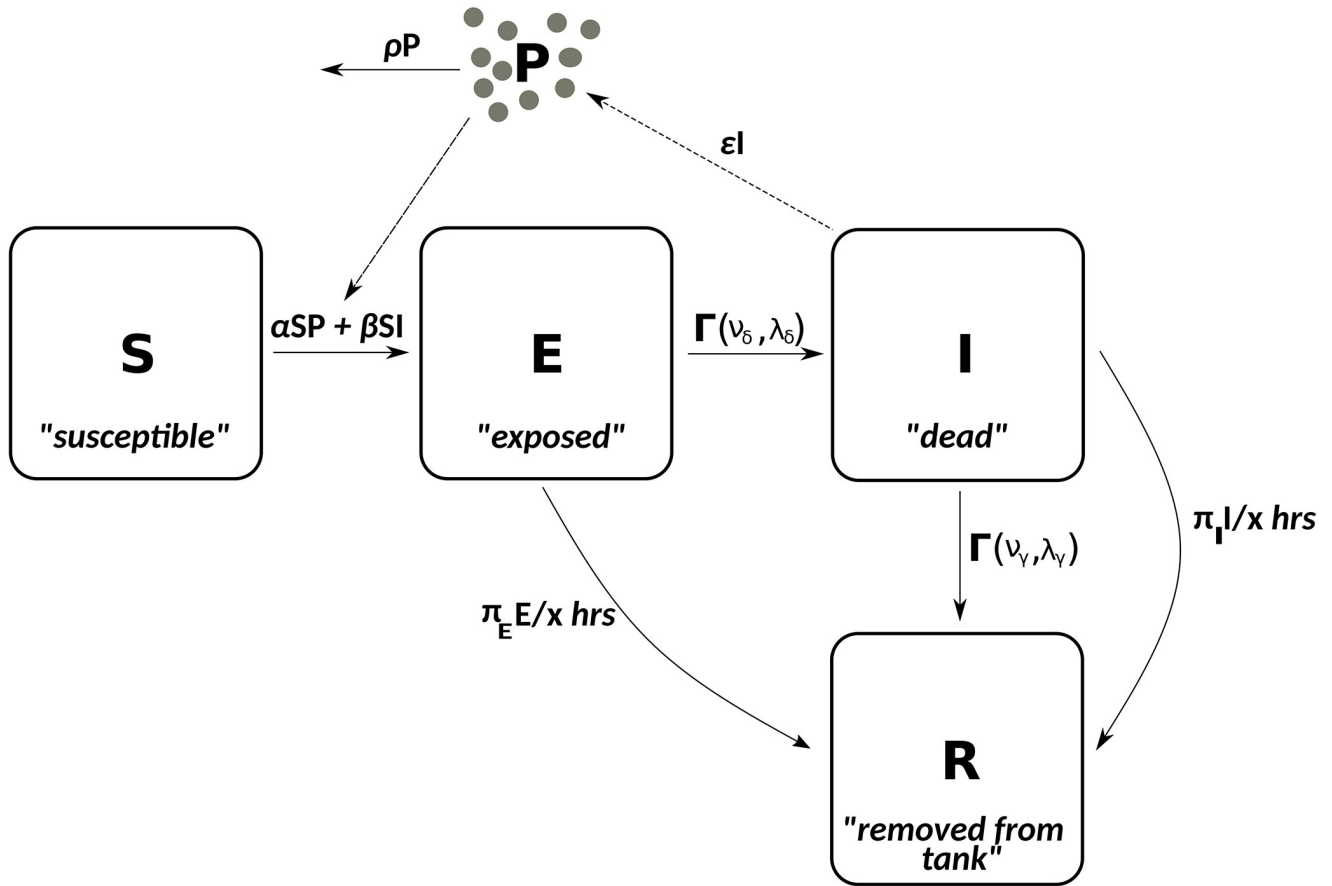
Underlying the estimates of  $\beta_{\text{cohab}}$  and  $\tilde{\beta}_{\text{cohab}}$  above is the assumption that the force of infection responds without delay to a change in the number of infected shrimp,  $I_t$ . This assumption is indeed valid across a wide variety of cases in which the size of the environmental pathogen load responds more or less rapidly to changes in  $I_t$ , as discussed in the previous section. However, given the slow rate of decay of infectivity of WSSV and its persistence in water bodies long after outbreaks have occurred, it is perhaps fruitful to consider the relationship between environmental pathogen load and the rate of environmental transmission as described, for example, by the SEIR-P model described in Models for direct and environmental transmission of disease. For example, Wang, et al. (2012) [34] find that an environmental transmission model similar to SEIR-P of avian flu among duck populations was able to account for the complex periodic outbreak patterns of the disease over long time periods.

Whereas Lotz and Soto and Tuyen, et al. characterise the route of transmission due to cohabitation as being implicitly direct, with rate  $\beta_{\text{cohab}}$ , (since its rate is directly proportional to the number of infectious hosts), we aim here instead to characterise this as environmental transmission with rate  $\alpha$ , as described in The direct transmission approximation as timescale limit of SEIR-P process. As far as we know, there is no published estimate of this quantity for WSD among penaeids. We obtain a lower estimate of  $\alpha_L = 10^{-4} \text{ ml virion}^{-1} \text{ h}^{-1}$  for *Penaeus monodon*, along with an upper estimate of the pathogen decay rate,  $\rho_U = 0.005 \text{ h}^{-1}$ , from the results of the WSSV viability in seawater experiment by Kumar, et al., ([27], details in Section C in S1 Appendix). Lotz and Soto use a shrimp density of 12 animals per square metre of water surface in their experiment in order to mimic densities of wild populations [30]. In the simulation study, described below, we adopt a nominal water volume of  $46.2 \text{ m}^3$  and water surface area of  $77 \text{ m}^2$  to obtain a similar host density with 1000 shrimps initially in the system. Since the estimates of  $\alpha$  and  $\beta$  both have dimensions  $\text{volume} \times \text{virion}^{-1} \times \text{time}^{-1}$  and  $\text{volume} \times \text{host}^{-1} \times \text{time}^{-1}$ , respectively, we scale these by this nominal volume before carrying out the simulations.

The pathogen emission rate,  $\epsilon$ , is chosen from within a range of known shedding rates for waterborne viruses (e.g., see [35, 36]). Since each dead shrimp contributes  $\frac{\epsilon}{\rho}$  to the environmental pathogen load, at equilibrium, and therefore  $\alpha \frac{\epsilon}{\rho}$  to the force of infection via environmental transmission, we choose a direct transmission rate  $\beta = 10 \times \alpha \frac{\epsilon}{\rho}$ . This is in accord with the relative sizes of Lotz and Soto's estimates of  $\tilde{\beta}_{\text{ingest}} \approx 10 \times \tilde{\beta}_{\text{cohab}}$ .

The S, E, I and R compartments of the SEIR-P model (summarised in Fig 8) are, respectively, shrimp that are susceptible (S), have been exposed, but still alive (E), dead, and now causing new infections either via shedding virus into the water body due to decay or being scavenged upon (I) and physically removed from the system (R). We assume for simplicity that there is no viral shedding or infectivity during the E stage and that times from exposure to mortality are gamma-distributed with shape and rate parameters  $\nu_\delta$  and  $\lambda_\delta$  such that the mean time from exposure to mortality ( $\frac{\nu_\delta}{\lambda_\delta}$ ) agrees with the estimate given by [31].

There are two processes by which shrimp are removed from the system. Firstly, there is the long process of decay characterised by gamma-distributed times in the I compartment, with shape and rate parameters  $\nu_\gamma$  and  $\lambda_\gamma$  with mean  $\frac{\nu_\gamma}{\lambda_\gamma} = 333.3 \text{ h} \approx 14 \text{ d}$ . Secondly, removals occur probabilistically from both the E and I compartments at regularly spaced time points with probabilities of success of  $\pi_E = 0.05$  and  $\pi_I = 0.95$ , respectively, so that  $\text{bin}(E_t, \pi_E)$  and  $\text{bin}(I_t, \pi_I)$  shrimp are removed at each removal point,  $t$ . Shrimp that are dead are therefore removed at the first removal time, post-mortem, with probability 0.95 and at the second with probability



**Fig 8. Summary of SEIR-P model of WSD among penaeids.** Parameter values are listed in Table 4. The arrow from I to R, labelled  $\Gamma(v_\gamma, \lambda_\gamma)$ , represents removal of dead hosts after a gamma-distributed time to full natural decay. The curved arrow from I to R represents removal at one of the x-hourly removal attempts, with probability  $\pi_i$ , similarly for the curved arrow from E to I. The *direct transmission approximation* (DTA) of the SEIR-P model is obtained by replacing  $\alpha S_i P_t + \beta S_t I_t$  above the leftmost arrow with  $\hat{\beta} = \alpha \frac{\epsilon}{\rho} + \beta$  and setting  $\epsilon = \rho = 0$ .

<https://doi.org/10.1371/journal.pcbi.1009652.g008>

0.9975. This means that it is highly unlikely that shrimp are removed from the system due to natural decay in this scenario. We assume that no living, disease-free shrimp are accidentally removed in this process. All of these model quantities are summarised together in Table 4.

### Impact of removal frequency on WSD outbreaks among penaeids

Using simulations, we study outbreak patterns under 24 and 6-hourly removals under the SEIR-P model described above. Alongside these we also look at those of the DTA of this model, where  $\hat{\beta} = \alpha \frac{\epsilon}{\rho} + \beta$ , for comparison. Density plots for the final outbreak size and outbreak duration,  $\max(t^R)$ , of the SEIR-P model and the DTA are displayed in Fig 9 for both 24 and 6 hourly removals. Figs 10 and 11 show the simulated outbreak trajectories for the four host compartments of the SEIR-P model and DTA. The top row in these two figures are typical individual outbreak trajectories while the bottom row are trajectories averaged over  $3 \times 10^4$  independent simulations, with “small” outbreaks of fewer than 10 infections removed.

Figs 9A and 10 show that attempting to remove the dead shrimp from the system at 24 h intervals, even with a success rate of 95%, is not sufficient to prevent large outbreaks of WSD, with outbreaks overwhelmingly affecting more than 90% of the shrimp population and lasting

Table 4. Parameter estimates and sources for SEIR-P model of WSSV in penaeids.

	Description		source / comment
$\alpha$	transmission (cohabitation)	$10^{-4}$ ml virion <sup>-1</sup> h <sup>-1</sup>	estimated from [27] (Section C in S1 Appendix)
		$2.16 \times 10^{-12}$ virion <sup>-1</sup> h <sup>-1</sup>	scaled by 46.2 m <sup>3</sup>
$\beta$	transmission (ingestion)	$8.64 \times 10^{-4}$ shrimp <sup>-1</sup> h <sup>-1</sup>	$10 \times \alpha_{\rho}^{\epsilon}$ (see [30] and above discussion)
$\hat{\beta}$	direct transmission (DTA)	$9.5 \times 10^{-4}$ shrimp <sup>-1</sup> h <sup>-1</sup>	$\alpha_{\rho}^{\epsilon} + \beta$
$\nu_{\delta}$	mortality (shape)	1.5	
$\lambda_{\delta}$	mortality (rate)	0.0112 h <sup>-1</sup>	[31]
$\nu_{\gamma}$	removal (decay) (shape)	2.0	
$\lambda_{\gamma}$	removal (decay) (rate)	0.006 h <sup>-1</sup>	
$\pi_E$	success of removal (from E)	0.05	
$\pi_I$	success of removal (from I)	0.95	
$\epsilon$	WSSV shedding	$2 \times 10^5$ virion shrimp <sup>-1</sup> h <sup>-1</sup>	(see e.g. [35, 36])
$\rho$	loss of WSSV infectivity	0.005 h <sup>-1</sup>	estimated from [27] (Section C in S1 Appendix)

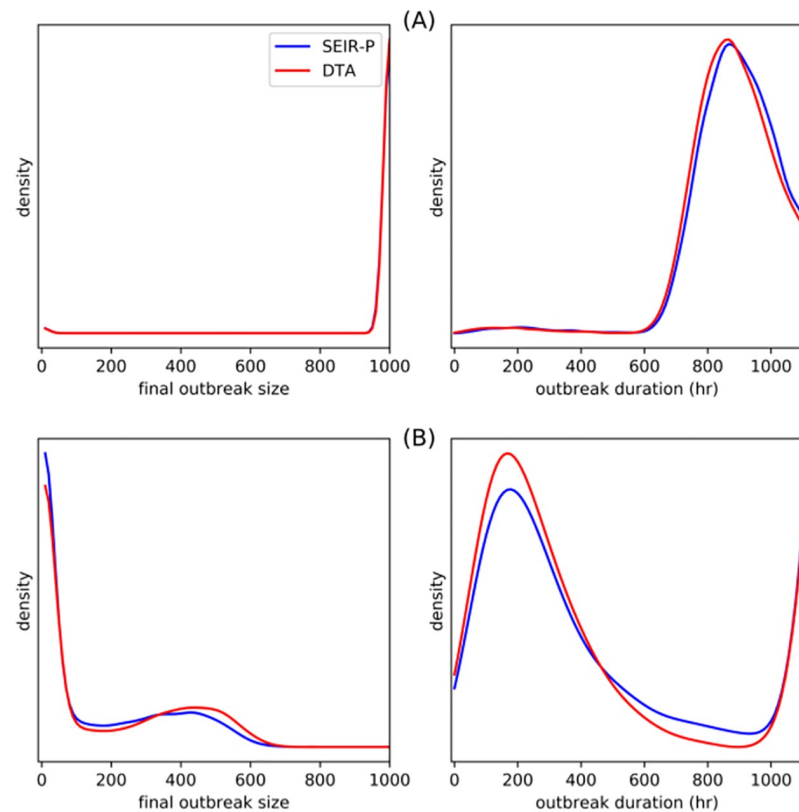
<https://doi.org/10.1371/journal.pcbi.1009652.t004>

more than 600 h from index exposure to final removal. Increasing the intensity of surveillance, however, by removing dead and diseased shrimp every 6 h, eliminates many such large outbreaks, limiting the final outbreak size to about 60% of the population. Additionally, outbreaks tend to be eradicated sooner, at around 200 h, although a sizeable proportion continue for longer (see Fig 9B).

It is interesting to compare the SEIR-P and DTA trajectories when going from 24 to 6-hourly removals, since the time that infectious shrimp are in the system is reduced by about a quarter, on average. This is an example of how two distinct degrees of host and pathogen timescale separation may be observed for the same host-disease system. The plots in Figs 9A and 10 suggest very close alignment between the SEIR-P and DTA models in their host compartment dynamics, final outbreak sizes and outbreak durations, meaning that we can faithfully reproduce the environmental transmission without needing to model the pathogen load. Under such a scenario, most hosts remain infectious for less than 24 h. Nonetheless, we see that the resulting outbreak patterns are captured equally well by the DTM as by the full SEIR-P model. The shortening of the host timescale by removing every 6 h is sufficient, however, to begin to observe divergence between the SEIR-P and the DTA, most noticeably, perhaps, in the distributions of the final outbreak size and outbreak duration (Fig 9B). Indeed, the DTA underestimates, on average, the reduction in final outbreak size and overestimates the reduction in outbreak duration. The DTA outbreaks at 6-hourly removals tend to grow slightly faster than the SEIR-P outbreaks (see Fig 11).

## Discussion

We have seen in Results that the SIR and SEIR models approximate the host-disease dynamics arising from a combination of direct and environmental transmission, as modelled by SIR-P or SEIR-P, that this approximation improves with increasing rates of pathogen shedding and decay and that when fitting these models to data, using Bayesian inference and data augmentation, they are highly robust to violations of the assumption of direct transmission. For example, these results suggest that the direct transmission approximation will be suitable for modelling transmission of SARS-CoV-2 within a closed environment, such as a hospital, since it has a half life of about 1 h in aerosols and 1 h, 3.5 h, 5.75 h and 7 h on copper, cardboard, stainless steel and plastic surfaces [37] but the mean infectious period is considerably longer: 5 d to 11 d



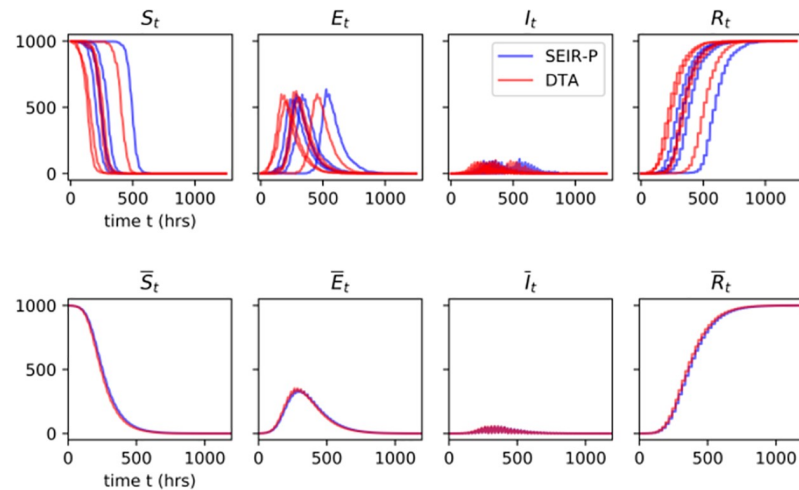
**Fig 9. Estimated density plots of final outbreak size (left panels) and outbreak duration (right) for the SEIR-P (blue) and DTA (red) models of WSD under (A) 24-hourly removals, where both quantities are distributed very similarly under the two models, and (B) 6-hourly removals.** Increasing the removal frequency tends to reduce the size and duration of outbreaks, although some larger outbreaks still occur. The benefit of increasing the removal frequency, in terms of reduction in mean final outbreak size, is underestimated slightly by the DTA and the reduction in outbreak duration is over-estimated.

<https://doi.org/10.1371/journal.pcbi.1009652.g009>

for asymptomatic cases, up to 4 d for presymptomatic cases [38] and about 7 d for symptomatic cases [39].

Tien & Earn, in their investigation of multiple transmission routes of cholera among humans [19], cite the rate of pathogen decay in the waterbody as the important factor in determining whether one should consider modelling the environmental and direct routes of transmission separately, or combined as one direct route. As suggested by the simulation study in Case study: White spot disease in penaeid shrimp, a viral lifetime of 200 h versus a much shorter host mean infectious lifetime of around 24 h also results in disease dynamics closely reproducible with a DTM, in spite of the low rate of pathogen decay. In this case the high rate of pathogen shedding produces sufficient host-pathogen timescale separation in order that DT provides a good, approximate description of the transmission via both direct and environmental routes. When the rates of shedding and pathogen decay are both low then we do not expect the DT approximation to work. Macro-parasite infections are one class of disease system within this grouping and our results indicate why models describing the complex host-parasite interaction, similar to those of similar to that of Anderson & May [40, 41], are often used for these systems (e.g., see [42]).

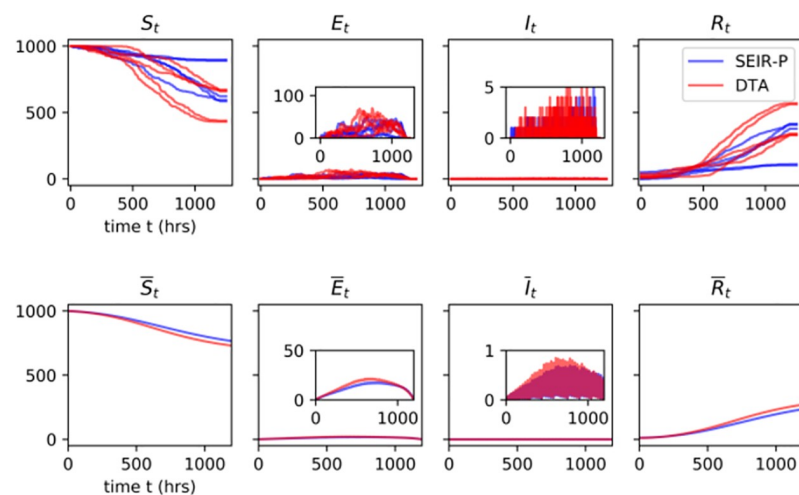
While individual outbreak trajectories appear very similar, statistical comparison over many runs reveals a strong and practically important divergence between environmental and



**Fig 10. Simulations of the SEIR-P (blue) and DTA (red) models of WSD in penaeid shrimp with removals of exposed (E) and dead (I) hosts at 24-hourly intervals, with probabilities of success 0.05 and 0.95, respectively.** Single outbreak trajectories (top row) and averages over 30 000 independent simulations with small outbreaks (fewer than 10 infections) excluded (bottom row). The zig-zag pattern in the 3<sup>rd</sup> panel on the bottom row is due to the periodic removals. The averaged model outputs show a high degree of similarity between SEIR-P and DTA, meaning that at these timescales the environmental transmission of WSD can be well approximated with direct transmission among the hosts.

<https://doi.org/10.1371/journal.pcbi.1009652.g010>

direct transmission models of WSD among penaeids under more effective disease control (i.e., more frequent removals). The model fitting and checking in [Results](#) were done under the rather special scenario that both the times of onset of infectivity and host removal are known so as to construct the outbreak size trajectories displayed in [Fig 5](#), which provides the clearest indication of lack of model fit in the long-lived pathogen case. However, DTMs can and have been fitted to a wide range of partial epidemic data [[43](#), [44](#)] and our conclusions will hold in such scenarios. Nonetheless GPPCs in general may not be the sharpest way to detect departure



**Fig 11. Simulations of the SEIR-P (blue) and DTA (red), as in [Fig 10](#), with removals at 6-hourly intervals.** Although the outbreaks of single SEIR-P and DTA trajectories appear similar, a small but definite divergence between the two models appears when studying their averaged outputs.

<https://doi.org/10.1371/journal.pcbi.1009652.g011>

from direct transmission when the data from an outbreak is less complete, as is often the case. *Exposure time residuals* (ETRs) (Lau, et al [45]) could potentially yield a numerical measure of model fit. ETRs are defined, relative to some putative model, as joint functions of the data, latent variables and parameters and their joint posterior predictive density should approximate an independent uniform sample when the parameters are close to the mode of their posterior. However, in the case of there being latent variables, such as when the host event times are not fully observed, analysis of their high-dimensional posterior distribution is not straightforward. Common methods of model comparison, such as model evidence [46] and the Bayesian and the deviance information criteria (BIC & DIC) would require an alternative ETM fitted to the same data in order to make a comparison. It is an open question whether ETMs can be fitted to host-disease events without measurement of the environmental pathogen load or strong prior information about the pathogen shedding and decay rates. Tien and Earn [19] comment that even when pathogen dynamics are slow, parameters quantifying rates of environmental and direct transmission ( $\alpha$  and  $\beta$ ) are still unidentifiable from disease incidence data alone. Methods that measure pathogen density in the waterbody, such as polymerase chain reaction [47, 48] are therefore required in order to quantify environmental transmission from data.

Among the key assumptions of the SEIR-P model of WSD among penaeids is that the rate of pathogen shedding,  $\epsilon$ , is constant, both across the host population and temporally within a single host over the course of their infection. There are, indeed, a few studies in which the rate of pathogen shedding has been measured in the same host at multiple time points, and these suggest that rates of pathogen emission do indeed vary over time (e.g., the investigation by Wargo, et al., of infectious hematopoietic necrosis virus shedding in juvenile rainbow trout [49] and that by Jones, et al., regarding repeated measures of respiratory viral load of SARS-CoV-2 [50]). What is gained by fitting a model with fixed shedding rate to epidemiological data is an idea of the “average” rate of shedding, both across the population and over time, even though the model’s estimates of risks at particular time points are either under or over-estimated. Another assumption is that the environmental pathogen is uniformly mixed throughout the water body. In the context of small to mid-sized tanks, this is reasonable, but when, e.g., considering the spread of infection between a local network of shrimp farming ponds then we would perhaps consider additionally incorporating a contact structure between multiple, uniformly mixed sites, as in [44]. Further complications may arise, however, in larger bodies of water where pathogen density perhaps varies spatially due to factors including water temperature or the existence of eddy currents, and temporally, due to effects of diffusion that cannot be neglected. Computational fluid dynamics, which numerically models flows of air and water, can be coupled with epidemiological models in order to incorporate uneven spatial pathogen densities and predict flows of pathogen carrying air and water currents (see, e.g., the study of the spatial pattern of affected households in the 2001 Amoy Gardens outbreak of SARS [51, 52]).

Nonetheless, even simple models, as long as we can adequately quantify them from data, offer approximations of useful quantities, such as the likely size and duration of outbreaks. This work should offer reassurance to readers that direct transmission models, with their simple picture of disease transmission, remain powerful tools as their field of application grows.

## Materials and methods

### Bayesian model fitting and inference

**SEIR posterior, likelihood and prior densities.** In [Results](#) we fit the SEIR model with exponential exposed and infectious lifetimes to data generated by both the SEIR and SEIR-P processes. Since in practice, the times of host exposure,  $\mathbf{t}^E = \{t_1^E, \dots, t_m^E\}$ , are very rarely



observed directly, these are treated as *missing data*. It is often the case that the times of onset of infectivity,  $\mathbf{t}^I = \{t_1^I, \dots, t_m^I\}$ , are also unobservable. However, we will be using the SEIR and SEIR-P models to describe WSSV in penaeid shrimp (see Case study: White spot disease in penaeid shrimp), for which onset of infectiousness coincides roughly with the death of the shrimp. Therefore, in this particular case, the times of entry into the I state, corresponding to death (and onset of infectiousness), and the R state, corresponding to removal from the system (either due to complete decay or physical removal from the waterbody) are feasibly observable.

Bayesian inference regarding the parameters of the SEIR model is based entirely on the *posterior density*, which in the case of exponentially distributed exposed and infectious lifetimes is

$$p(\beta, \delta, \gamma, \mathbf{t}^E | \mathbf{t}^I, \mathbf{t}^R) \propto p(\mathbf{t}^E, \mathbf{t}^I, \mathbf{t}^R | \beta, \delta, \gamma) p(\beta, \delta, \gamma) \quad (5)$$

where  $\mathbf{t}^I$  and  $\mathbf{t}^R = \{t_1^R < \dots < t_m^R\}$  are the observed times of onset of infectivity and removal and  $\mathbf{t}^E$  are the unobserved times of exposure (with indices corresponding to hosts, so that the host that is exposed at time  $t_i^E$  becomes infectious at  $t_i^I$  and is removed at  $t_i^R$ ). The symbol  $m$  denotes the final outbreak size. The two factors on the right hand side are respectively the *likelihood* and the *prior densities*. The prior density summarises our knowledge and uncertainty about the parameters *prior* to observing the data. Throughout, we will assume that the three parameters are *a priori* independent, i.e.,

$$p(\beta, \delta, \gamma) = p(\beta)p(\delta)p(\gamma) \quad (6)$$

and that

$$\begin{aligned} p(\beta) &\sim \text{Exp}(\omega_\beta) \\ p(\gamma) &\sim \text{Exp}(\omega_\gamma) \\ p(\delta) &\sim \text{Exp}(\omega_\delta) \text{ or } \text{U}(0, 10) \end{aligned} \quad (7)$$

We follow [43] in choosing exponentially-distributed priors since their functional form leads to convenient expressions for the full conditional distributions of the parameters, making it easier to sample from the posterior distribution. We choose the values  $\omega_\beta$ ,  $\omega_\delta$  and  $\omega_\gamma = 0.001$  for each marginal prior distribution's rate parameter so that each has a mean  $1 \times 10^3$  and variance  $1 \times 10^6$  (in appropriate units) and the resulting prior density is approximately flat over a large area of parameter space. This means that almost all of the information expressed by the posterior distribution comes from the likelihood. This kind of prior is described as *uninformative* since it expresses, in probabilistic language, that we have minimal knowledge about the values of the parameters prior to observing the data. In circumstances where we are fitting models to real data, without prior knowledge about the parameters of the model, it is necessary to ensure that a specific choice of uninformative prior is not unduly influential on posterior inferences. We perform such a *sensitivity analysis* by refitting the same model with several distinct, uninformative priors, comparing the resulting posterior distributions and checking that these are largely unaffected by choice of prior. However, since in what follows, we are fitting models to simulated data with known parameter values, we select priors in advance that we know to be sufficiently uniform on the scale of the likelihood. We will see, in what follows, that even with no prior knowledge about the direct transmission rate,  $\beta$ , (as expressed by its uninformative marginal prior density) it is possible to estimate this quantity in the absence of observed exposure times (see, e.g., [43, 44, 53]).

In the case of long-lived pathogen, SEIR is far from the "correct" model for the data (e.g., see [54, 55] for accounts of fitting mis-specified models) and this can present issues with the convergence of the MCMC chains since there are no strong candidates among parameter values that simultaneously explain the data. In order to aid convergence, therefore, the alternative,

uniformly-distributed prior for  $\delta$  is used in the long-lived pathogen case only (Estimating DTA parameters from outbreak data). By placing an upper bound (in this case, 10.0) on the support of  $\delta$ , we stipulate that we seek a model with non-negligible latent infectious periods, of duration no shorter than one tenth of a day.

The likelihood,  $p(\mathbf{t}^E, \mathbf{t}^I, \mathbf{t}^R | \beta, \delta, \gamma)$ , describes how the data depend on the parameters of the model, which, in the case of exponentially distributed times in the E and I states is

$$\begin{aligned}
 p(\mathbf{t}^E, \mathbf{t}^I, \mathbf{t}^R | \beta, \gamma, \delta) &\propto \prod_{i \neq \kappa} \{\beta I_{i, E-}\} \times \exp\left\{-\beta \int_{t_{\kappa}^E}^{\max(\mathbf{t}^R)} S_t I_t dt\right\} \times \dots \\
 &\dots \times \prod_{j=1}^m \delta e^{-\delta(t_j^I - t_j^E)} \times \prod_{j=1}^m \gamma e^{-\gamma(t_j^R - t_j^I)}.
 \end{aligned}
 \tag{8}$$

where  $I_{i, E-}$  is the number of infectious hosts *immediately before* the  $i^{\text{th}}$  exposure time. See Section A in [S1 Appendix](#).

The products over  $j = 1, \dots, m$  in (8) are the contributions to the likelihood from each of the exponentially-distributed times spent in states E and I. The terms in the product over  $i \neq \kappa$  are the contributions from the exposure times, excluding the index exposure, which each have associated hazard  $\beta S_t I_t$ .

Similarly to [43],  $t_{\kappa}^E$  represents the (perhaps unobserved) first exposure time. In [Results](#) we fit these models in scenarios where the exposure times have not been observed and so the time of the index exposure is unknown.

For details of model fitting, see Section A in [S1 Appendix](#).

**Model checking using graphical posterior-predictive checks (GPPC).** Graphical posterior predictive checks (GPPC) [23, 24] are used here to test for departure from DT model assumptions. These are a standard model checking tool, offering a visual comparison of quantities derived from the observed data,  $h(\mathbf{t}^I, \mathbf{t}^R)$ , with  $h(\mathbf{t}^I, \mathbf{t}^R)$ , where  $\mathbf{t}^I, \mathbf{t}^R$  are simulated from the posterior-predictive distribution of the fitted model, with density

$$\begin{aligned}
 p(\mathbf{t}^I, \mathbf{t}^R | \mathbf{t}^E, \mathbf{t}^R) &= \int p(\mathbf{t}^I, \mathbf{t}^R | \beta, \delta, \gamma, \mathbf{t}^E, \mathbf{t}^I, \mathbf{t}^R) p(\beta, \delta, \gamma, \mathbf{t}^E | \mathbf{t}^I, \mathbf{t}^R) d\beta d\delta d\gamma d\mathbf{t}^E \\
 &= \int p(\mathbf{t}^I, \mathbf{t}^R | \beta, \delta, \gamma) p(\beta, \delta, \gamma, \mathbf{t}^E | \mathbf{t}^I, \mathbf{t}^R) d\beta d\delta d\gamma d\mathbf{t}^E
 \end{aligned}
 \tag{9}$$

Uncertainty regarding parameter values is expressed by drawing from their posterior distribution. The idea is to check that the fitted model replicates the original data with reasonable probability, with no large, systematic disagreements between the data and model predictions.

Among the salient features of a disease outbreak are its size, at both peak and completion, and characteristic timescales, e.g., time from index exposure to peak of outbreak, and total duration. Such statistics are of interest in their own right and there are known formulas in the deterministic case for peak and final outbreak sizes and initial exponential growth rates for SIR, SEIR and similar models [56–58]. For the GPPCs here we obtain a probabilistic picture of similar quantities with the timing of the outbreak’s peak standing as proxy for the initial growth rate. The following four quantities are considered.

1. final outbreak size,  $m$ , i.e., total hosts who become exposed during outbreak
2. outbreak duration,  $\max(\mathbf{t}^R) - \min(\mathbf{t}^I)$
3. time of outbreak peak,  $t_{\text{peak}}$
4. size of outbreak at peak,  $I_{\text{max}} = \max\{I_t, t \geq 0\}$ .

Since the simulated data contains times of onset of infectivity and host removal, the above quantities are indeed directly calculable. Additionally, the outbreak size trajectory,  $I_t$ , over the course of an outbreak will be examined. Due to likely correlations between  $m$  and  $\max(t^R) - \min(t^I)$  and between  $I_{\max}$  and  $t_{\text{peak}}$ , these are plotted bivariately.

The time of outbreak peak,  $t_{\text{peak}}$ , is interpolated between the first and last times that the outbreak size is within the range  $cI_{\max} \leq I_t \leq I_{\max}$ , where  $0 < c < 1$ , i.e.,

$$t_{\text{peak}} = \frac{\max\{t : I_t \geq cI_{\max}\} - \min\{t : I_t \geq cI_{\max}\}}{2}. \quad (10)$$

Calculating  $t_{\text{peak}}$  this way, rather than simply taking the time that the outbreak size reaches its maximum, avoids the complication of the outbreak hitting its maximum size more than once and, more importantly, aims to reduce the variance of its posterior-predictive distribution, and therefore produce a sharper test for model departure. Here  $c$  is chosen to be 0.3, since the GPPC outputs were not found to be sensitive to the particular value chosen.

## Supporting information

**S1 Appendix. Supplementary text.** Detailed description of Metropolis-cooled MCMC routine, derivation of force of infection for Reed-Frost epidemic models, parameter estimation for WSSV SEIR-P model and diagnostic MCMC trace plots. (PDF)

## Author Contributions

**Conceptualization:** Lee Benson, Ross S. Davidson, Darren M. Green, Andrew Hoyle, Mike R. Hutchings, Glenn Marion.

**Formal analysis:** Lee Benson.

**Funding acquisition:** Ross S. Davidson, Andrew Hoyle, Mike R. Hutchings, Glenn Marion.

**Investigation:** Lee Benson.

**Methodology:** Lee Benson, Ross S. Davidson, Andrew Hoyle, Glenn Marion.

**Project administration:** Lee Benson, Ross S. Davidson, Andrew Hoyle, Glenn Marion.

**Supervision:** Ross S. Davidson, Darren M. Green, Andrew Hoyle, Mike R. Hutchings, Glenn Marion.

**Writing – original draft:** Lee Benson.

**Writing – review & editing:** Darren M. Green, Andrew Hoyle, Mike R. Hutchings, Glenn Marion.

## References

1. Kermack WO, McKendrick AG. A contribution to the mathematical theory of epidemics. Proceedings of the Royal Society of London Series A, Containing Papers of a Mathematical and Physical Character. 1927; 115(772):700–721.
2. Kleczkowski A, Hoyle A, McMenemy P. One model to rule them all? Modelling approaches across One-Health for human, animal and plant epidemics. Philosophical Transactions of the Royal Society B: Biological Sciences. 2019; 374(1775):20180255. <https://doi.org/10.1098/rstb.2018.0255> PMID: 31056049
3. McCallum H. How should pathogen transmission be modelled? Trends in Ecology & Evolution. 2001; 16(6):295–300.
4. Diekmann O. Mathematical epidemiology of infectious diseases: model building, analysis, and interpretation. Wiley series in mathematical and computational biology. Chichester: Wiley; 2000.

5. Merler S, Ajelli M, Fumanelli L, Gomes MFC, Piontti APy, Rossi L, et al. Spatiotemporal spread of the 2014 outbreak of Ebola virus disease in Liberia and the effectiveness of non-pharmaceutical interventions: a computational modelling analysis. *The Lancet Infectious Diseases*. 2015; 15(2):204–211. [https://doi.org/10.1016/S1473-3099\(14\)71074-6](https://doi.org/10.1016/S1473-3099(14)71074-6) PMID: 25575618
6. Nayagam S, Thursz M, Sicuri E, Conteh L, Wiktor S, Low-Beer D, et al. Requirements for global elimination of hepatitis B: a modelling study. *The Lancet Infectious Diseases*. 2016; 16(12):1399–1408. [https://doi.org/10.1016/S1473-3099\(16\)30204-3](https://doi.org/10.1016/S1473-3099(16)30204-3) PMID: 27638356
7. Di Ruscio F, Guzzetta G, Bjørnholt JV, Leegaard TM, Moen AEF, Merler S, et al. Quantifying the transmission dynamics of MRSA in the community and healthcare settings in a low-prevalence country. *Proceedings of the National Academy of Sciences*. 2019; 116(29):14599–14605. <https://doi.org/10.1073/pnas.1900959116>
8. Keeling MJ. Dynamics of the 2001 UK Foot and Mouth Epidemic: Stochastic Dispersal in a Heterogeneous Landscape. *Science*. 2001; 294(5543):813–817. <https://doi.org/10.1126/science.1065973> PMID: 11679661
9. Keeling MJ. Models of foot-and-mouth disease. *Proceedings of the Royal Society B: Biological Sciences*. 2005; 272(1569):1195–1202. <https://doi.org/10.1098/rspb.2004.3046> PMID: 16024382
10. Chen TM, Rui J, Wang QP, Zhao ZY, Cui JA, Yin L. A mathematical model for simulating the phase-based transmissibility of a novel coronavirus. *Infectious Diseases of Poverty*. 2020; 9(1):1–8. <https://doi.org/10.1186/s40249-020-00640-3> PMID: 32111262
11. Monteiro LHA. Short Communication. *Ecological Complexity*. 2020; p. 100836. <https://doi.org/10.1016/j.ecocom.2020.100836> PMID: 32921471
12. Vezzulli L, Pruzzo C, Huq A, Colwell RR. Environmental reservoirs of *Vibrio cholerae* and their role in cholera. *Environmental Microbiology Reports*. 2010; 2(1):27–33. <https://doi.org/10.1111/j.1758-2229.2009.00128.x> PMID: 23765995
13. Almagro-Moreno S, Taylor RK. Cholera: Environmental Reservoirs and Impact on Disease Transmission. In: *One Health*. vol. 1. Washington, DC, USA: ASM Press; 2014. p. 149–165. Available from: <http://doi.wiley.com/10.1128/9781555818432.ch10>.
14. Stallknecht DE, Shane SM, Kearney MT, Zwank PJ. Persistence of Avian Influenza Viruses in Water. *Avian Diseases*. 1990; 34(2):406. <https://doi.org/10.2307/1591428> PMID: 2142420
15. Hauck R, Crossley B, Rejmanek D, Zhou H, Gallardo RA. Persistence of Highly Pathogenic and Low Pathogenic Avian Influenza Viruses in Footbaths and Poultry Manure. *Avian Diseases*. 2017; 61(1):64–69. <https://doi.org/10.1637/11495-091916-Reg> PMID: 28301246
16. Zhang R, Li Y, Zhang AL, Wang Y, Molina MJ. Identifying airborne transmission as the dominant route for the spread of COVID-19. *Proceedings of the National Academy of Sciences*. 2020; 117(26):14857–14863. <https://doi.org/10.1073/pnas.2009637117> PMID: 32527856
17. Salama N, Rabe B. Developing models for investigating the environmental transmission of disease-causing agents within open-cage salmon aquaculture. *Aquaculture Environment Interactions*. 2013; 4(2):91–115. <https://doi.org/10.3354/aei00077>
18. Anderson WJ. *Continuous-Time Markov Chains*. Springer Series in Statistics. New York, NY: Springer New York; 1991. Available from: <http://link.springer.com/10.1007/978-1-4612-3038-0>.
19. Tien JH, Earn DJD. Multiple Transmission Pathways and Disease Dynamics in a Waterborne Pathogen Model. *Bulletin of Mathematical Biology*. 2010; 72(6):1506–1533. <https://doi.org/10.1007/s11538-010-9507-6> PMID: 20143271
20. Gunawardena J. Time-scale separation—Michaelis and Menten’s old idea, still bearing fruit. *FEBS Journal*. 2014; 281(2):473–488. <https://doi.org/10.1111/febs.12532> PMID: 24103070
21. Heesterbeek JAP, Dietz K. The concept of  $R_0$  in epidemic theory. *Statistica Neerlandica*. 1996; 50(1):89–110. <https://doi.org/10.1111/j.1467-9574.1996.tb01482.x>
22. Heffernan JM, Smith RJ, Wahl LM. Perspectives on the basic reproductive ratio. *Journal of The Royal Society Interface*. 2005; 2(4):281–293. <https://doi.org/10.1098/rsif.2005.0042> PMID: 16849186
23. Gelman A, Carlin JB, Stern HS, Rubin DB. *Bayesian Data Analysis*. Chapman and Hall/CRC; 1995. Available from: <https://www.taylorfrancis.com/books/9781135439415>.
24. Gelman A, Shalizi CR. Philosophy and the practice of Bayesian statistics. *British Journal of Mathematical and Statistical Psychology*. 2013; 66(1):8–38. <https://doi.org/10.1111/j.2044-8317.2011.02037.x> PMID: 22364575
25. Lo CF, Kou GH. Virus-associated White Spot Syndrome of Shrimp in Taiwan: A Review. *Fish Pathology*. 1998; 33(4):365–371. <https://doi.org/10.3147/jsfp.33.365>
26. Chou H, Huang C, Wang C, Chiang H, Lo C. Pathogenicity of a baculovirus infection causing white spot syndrome in cultured penaeid shrimp in Taiwan. *Diseases of Aquatic Organisms*. 1995; 23(3):165–173. <https://doi.org/10.3354/dao023165>

27. Satheesh Kumar S, Ananda Bharathi R, Rajan JJS, Alavandi SV, Poornima M, Balasubramanian CP, et al. Viability of white spot syndrome virus (WSSV) in sediment during sun-drying (drainable pond) and under non-drainable pond conditions indicated by infectivity to shrimp. *Aquaculture*. 2013; 402-403:119–126. <https://doi.org/10.1016/j.aquaculture.2013.04.001>
28. Quang ND, Hoa PTP, Da TT, Anh PH. Persistence of white spot syndrome virus in shrimp ponds and surrounding areas after an outbreak. *Environmental Monitoring and Assessment*. 2009; 156(1-4):69–72. <https://doi.org/10.1007/s10661-008-0463-7> PMID: 18696239
29. Esparza-Leal HM, Escobedo-Bonilla CM, Casillas-Hernández R, Álvarez-Ruiz P, Portillo-Clark G, Valerio-García RC, et al. Detection of white spot syndrome virus in filtered shrimp-farm water fractions and experimental evaluation of its infectivity in *Penaeus* (*Litopenaeus*) *vannamei*. *Aquaculture*. 2009; 292(1-2):16–22. <https://doi.org/10.1016/j.aquaculture.2009.03.021>
30. Soto MA, Lotz JM. Epidemiological Parameters of White Spot Syndrome Virus Infections in *Litopenaeus vannamei* and *L. setiferus*. *Journal of Invertebrate Pathology*. 2001; 78(1):9–15. <https://doi.org/10.1006/jipa.2001.5035> PMID: 11500088
31. Tuyen NX, Verreth J, Vlak JM, de Jong MCM. Horizontal transmission dynamics of White spot syndrome virus by cohabitation trials in juvenile *Penaeus monodon* and *P. vannamei*. *Preventive Veterinary Medicine*. 2014; 117(1):286–294. <https://doi.org/10.1016/j.prevetmed.2014.08.007> PMID: 25189688
32. Abbey H. An Examination of the Reed-Frost Theory of Epidemics. *Human Biology*. 1952; 24(3):201. PMID: 12990130
33. Lotz JM, Soto MA. Model of white spot syndrome virus (WSSV) epidemics in *Litopenaeus vannamei*. *Diseases of Aquatic Organisms*. 2002; 50(3):199–209. <https://doi.org/10.3354/dao050199> PMID: 12219976
34. Wang RH, Jin Z, Liu QX, van de Koppel J, Alonso D. A Simple Stochastic Model with Environmental Transmission Explains Multi-Year Periodicity in Outbreaks of Avian Flu. *PLoS ONE*. 2012; 7(2):e28873. <https://doi.org/10.1371/journal.pone.0028873> PMID: 22363397
35. Wargo AR, Scott RJ, Kerr B, Kurath G. Replication and shedding kinetics of infectious hematopoietic necrosis virus in juvenile rainbow trout. *Virus Research*. 2017; 227:200–211. <https://doi.org/10.1016/j.virusres.2016.10.011> PMID: 27771253
36. Kim RK, Faisal M. Shedding of viral hemorrhagic septicemia virus (Genotype IVb) by experimentally infected muskellunge (*Esox masquinongy*). *The Journal of Microbiology*. 2012; 50(2):278–284. <https://doi.org/10.1007/s12275-012-1145-2> PMID: 22538657
37. van Doremalen N, Bushmaker T, Morris DH, Holbrook MG, Gamble A, Williamson BN, et al. Aerosol and Surface Stability of SARS-CoV-2 as Compared with SARS-CoV-1. *New England Journal of Medicine*. 2020; 382(16):1564–1567. <https://doi.org/10.1056/NEJMc2004973>
38. Byrne AW, McEvoy D, Collins AB, Hunt K, Casey M, Barber A, et al. Inferred duration of infectious period of SARS-CoV-2: rapid scoping review and analysis of available evidence for asymptomatic and symptomatic COVID-19 cases. *BMJ open*. 2020; 10(8):e039856. <https://doi.org/10.1136/bmjopen-2020-039856> PMID: 32759252
39. Wölfel R, Corman VM, Guggemos W, Seilmaier M, Zange S, Müller MA, et al. Virological assessment of hospitalized patients with COVID-2019. *Nature*. 2020; 581(7809):465–469. <https://doi.org/10.1038/s41586-020-2196-x> PMID: 32235945
40. Anderson RM, May RM. Regulation and Stability of Host-Parasite Population Interactions: I. Regulatory Processes. *The Journal of Animal Ecology*. 1978; 47(1):219. <https://doi.org/10.2307/3933>
41. May RM, Anderson RM. Regulation and Stability of Host-Parasite Population Interactions: II. Destabilizing Processes. *The Journal of Animal Ecology*. 1978; 47(1):249. <https://doi.org/10.2307/3934>
42. McPherson NJ, Norman RA, Hoyle AS, Bron JE, Taylor NGH. Stocking methods and parasite-induced reductions in capture: Modelling *Argulus foliaceus* in trout fisheries. *Journal of Theoretical Biology*. 2012; 312:22–33. <https://doi.org/10.1016/j.jtbi.2012.07.017> PMID: 22842010
43. Neal P, Roberts G. A case study in non-centering for data augmentation: Stochastic epidemics. *Statistics and Computing*. 2005; 15(4):315–327. <https://doi.org/10.1007/s11222-005-4074-7>
44. Gamado K, Marion G, Porphyre T. Data-Driven Risk Assessment from Small Scale Epidemics: Estimation and Model Choice for Spatio-Temporal Data with Application to a Classical Swine Fever Outbreak. *Frontiers in Veterinary Science*. 2017; 4(February):1–14. <https://doi.org/10.3389/fvets.2017.00016> PMID: 28293559
45. Lau MSY, Marion G, Streftaris G, Gibson GJ. New model diagnostics for spatio-temporal systems in epidemiology and ecology. *Journal of The Royal Society Interface*. 2014; 11(93):20131093–20131093. <https://doi.org/10.1098/rsif.2013.1093> PMID: 24522782

46. Pooley CM, Marion G. Bayesian model evidence as a practical alternative to deviance information criterion. *Royal Society Open Science*. 2018; 5(3):171519. <https://doi.org/10.1098/rsos.171519> PMID: 29657762
47. Ramírez-Castillo F, Loera-Muro A, Jacques M, Garneau P, Avelar-González F, Harel J, et al. Waterborne Pathogens: Detection Methods and Challenges. *Pathogens*. 2015; 4(2):307–334. <https://doi.org/10.3390/pathogens4020307> PMID: 26011827
48. Aintablian N, Walpita P, Sawyer MH. Detection of *Bordetella pertussis* and Respiratory Syncytial Virus in Air Samples from Hospital Rooms. *Infection Control and Hospital Epidemiology*. 1998; 19(12):918–923. <https://doi.org/10.2307/30142018> PMID: 9872529
49. Wargo AR, Scott RJ, Kerr B, Kurath G. Replication and shedding kinetics of infectious hematopoietic necrosis virus in juvenile rainbow trout. *Virus Research*. 2017; 227:200–211. <https://doi.org/10.1016/j.virusres.2016.10.011> PMID: 27771253
50. Jones TC, Biele G, Mühlemann B, Veith T, Schneider J, Beheim-schwarzbach J, et al. Estimating infectiousness throughout SARS-CoV-2 infection course. *Science*. 2021; 5273(May):eabi5273. <https://doi.org/10.1126/science.abi5273> PMID: 34035154
51. Yu ITS, Li Y, Wong TW, Tam W, Chan AT, Lee JHW, et al. Evidence of Airborne Transmission of the Severe Acute Respiratory Syndrome Virus. *New England Journal of Medicine*. 2004; 350(17):1731–1739. <https://doi.org/10.1056/NEJMoa032867> PMID: 15102999
52. Yu ITS, Qiu H, Tse LA, Wong TW. Severe Acute Respiratory Syndrome Beyond Amoy Gardens: Completing the Incomplete Legacy. *Clinical Infectious Diseases*. 2014; 58(5):683–686. <https://doi.org/10.1093/cid/cit797> PMID: 24319085
53. Streftaris G, Gibson GJ. Statistical Inference for Stochastic Epidemic Models. *Annals of Statistics*. 2002; 609616:1–8.
54. Berk RH. Limiting Behavior of Posterior Distributions when the Model is Incorrect. *The Annals of Mathematical Statistics*. 1966; 37(1):51–58. <https://doi.org/10.1214/aoms/1177699597>
55. Walker SG. Bayesian inference with misspecified models. *Journal of Statistical Planning and Inference*. 2013; 143(10):1621–1633.
56. Feng Z. Final and peak epidemic sizes for SEIR models with quarantine and isolation. *Mathematical Biosciences and Engineering*. 2007; 4(4):675–686. <https://doi.org/10.3934/mbe.2007.4.675> PMID: 17924718
57. Miller JC. A Note on the Derivation of Epidemic Final Sizes. *Bulletin of Mathematical Biology*. 2012; 74(9):2125–2141. <https://doi.org/10.1007/s11538-012-9749-6> PMID: 22829179
58. Ma J. Estimating epidemic exponential growth rate and basic reproduction number. *Infectious Disease Modelling*. 2020; 5:129–141. <https://doi.org/10.1016/j.idm.2019.12.009> PMID: 31956741