Supplement to "Informing policy via dynamic models: Cholera in Haiti"

Jesse Wheeler, AnnaElaine L. Rosengart, Zhuoxun Jiang,
Kevin Tan, Noah Treutle and Edward L. Ionides

Department of Statistics, University of Michigan

March 4, 2024

Supplementary Content

S1 Model Diagrams

Each of the dynamic models considered in this manuscript can be fully described using the model descriptions in the manuscript, coupled with the additional information described in Sections 2 and 3 of this supplement. Despite this, diagrams of dynamic systems are often helpful to understand the equations. In this section, we give three diagrams representing Models 1–3, respectively. Because the models are defined by their mathematical equations and numeric implementation, these diagrams are not unique visual representations of the model. Alternative representations that may be helpful in understanding the models explored in this paper were made by Lee et al. (2020b).

S1.1 Model 1

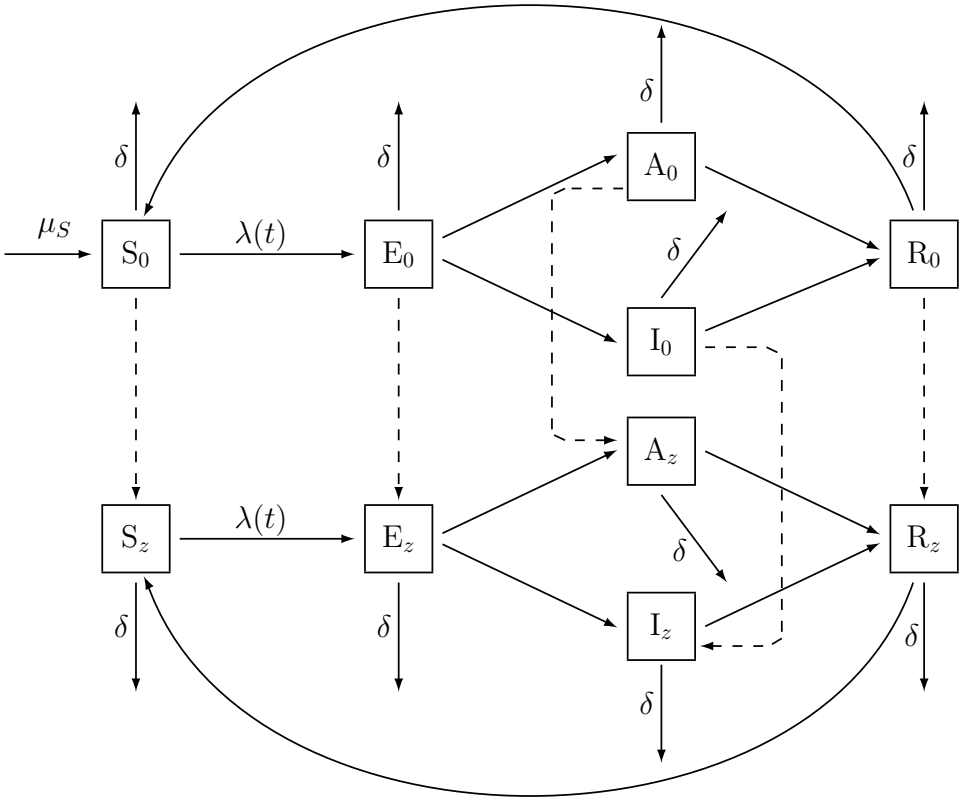

Fig S-1: A flow diagram for the SEAIR model.

S1.2 Model 2

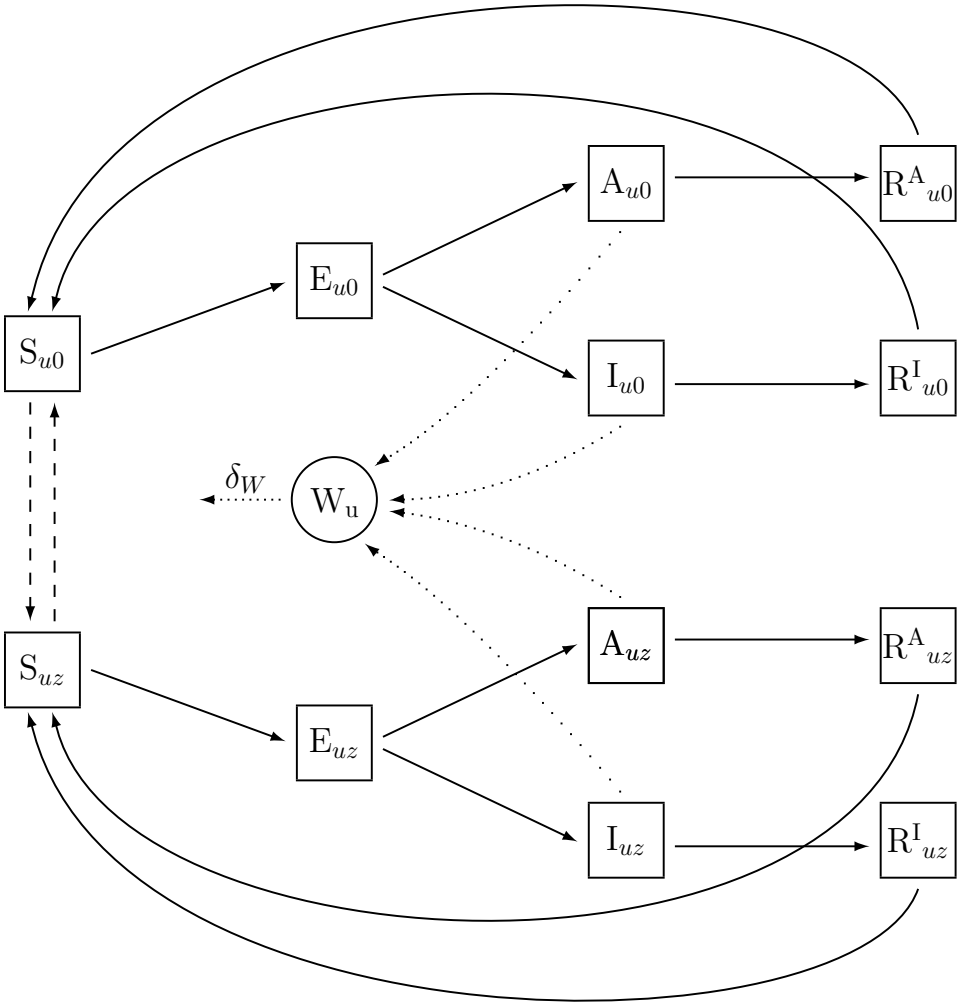

Fig S-2: A flow diagram for the SEAIR model 2. This is a constant population model, there are no births/deaths. Vaccinations are assumed to only be given to susceptible individuals, and vaccine immunity wanes only with susceptible vaccinated individuals.

S1.3 Model 3

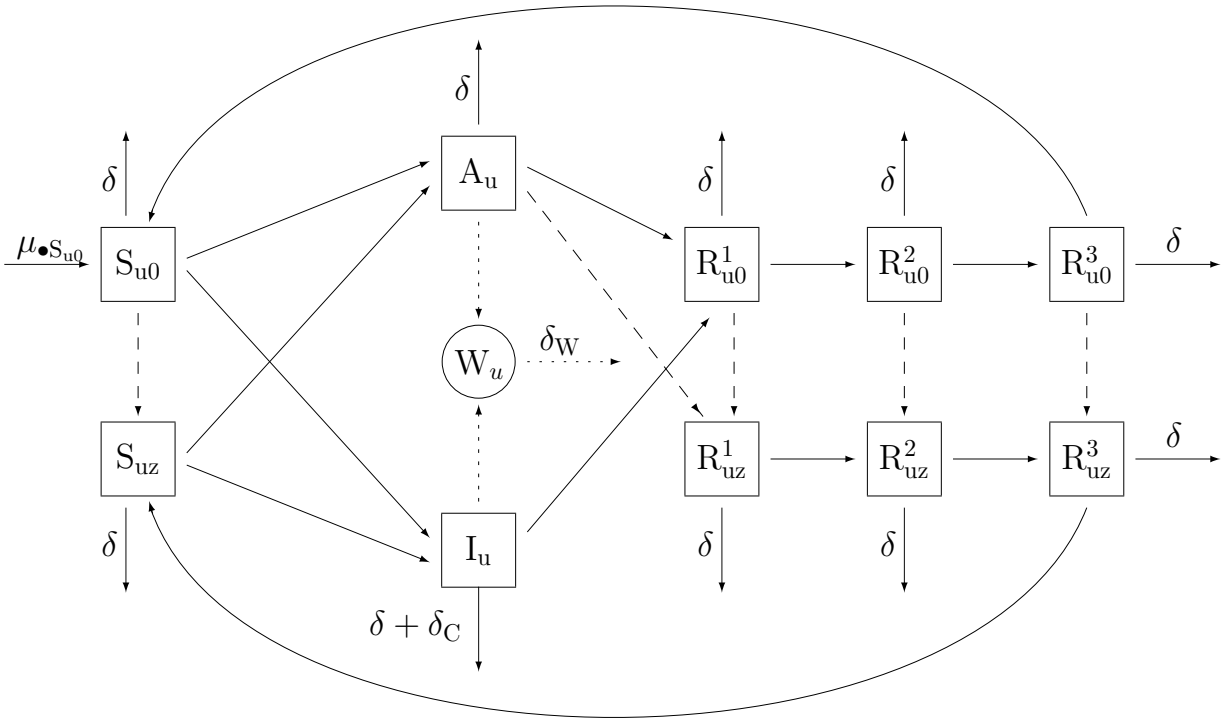

Fig S-3: A flow diagram for the SAIR model 3. This model assumes a constant population while also including a mechanism for births/deaths; all deaths are balanced by births into the unvaccinated susceptible compartment, so the birth rate $\mu_{\bullet S_{u0}}$ corresponds to the sum total deaths from the remaining compartments. The model assumes that symptomatic individuals will not be vaccinated, hence no vaccination arrow exiting the $I_u 0$ compartment.

S2 Markov chain and differential equation interpretations of compartment flow rates

In Sections 2.1, 2.2 and 2.3 of the main article, we define compartment models in terms of their flow rates. For a discrete population model, these rates define a Markov chain. For a continuous

and deterministic model, the rates define a system of ordinary differential equations. Here, we add additional details to clarify the mapping from a collection of rate functions to a fully specified process. Our treatment follows Bretó et al. (2009).

A general compartment model is a vector-valued process $X(t) = (X_1(t), \ldots, X_c(t))$ denoting the (integer or real-valued) counts in each of c compartments, where t is any continuous value in the interval $[t_0, \infty)$ for some real valued starting time t_0. The compartments may also have names, but to set up general notation we simply refer to them by their numerical index. The basic characteristic of a compartment model is that $X(t)$ can be written in terms of the flows $N_{ij}(t)$ from i to j. A flow into compartment i from outside the system is denoted by $N_{\bullet i}$, and a flow out of the system from compartment i is denoted by $N_{i\bullet}$. We call $\bullet$ a source/sink compartment, though it is an irregular compartment since $X_\bullet(t)$ is not defined. These flows are required to satisfy a "conservation of mass" identity:

$$X_i(t) = X_i(t_0) + N_{\bullet i}(t) - N_{i\bullet}(t) + \sum_{j \neq i} N_{ji}(t) - \sum_{j \neq i} N_{ij}(t). \tag{S1}$$

Each *flow* $N_{ij}(t)$ is associated with a *rate* function $\mu_{ij} = \mu_{ij}(t, X(t))$, where we include the possibility that i or j takes value $\bullet$.

There are different ways to use a collection of rate functions to build a fully specified model. We proceed to describe the ones we use in this paper: via a system of ordinary differential equations (Sec. S2.1), a simple Markov counting system (Sec. S2.2), and an overdispersed Markov counting system (Sec. S2.3). Other representations include stochastic differential equations driven by Gaussian noise or Gamma noise (Bhadra et al., 2011).

S2.1 Ordinary differential equation (ODE) interpretation

A basic deterministic specification is

$$dN_{ij}/dt = \mu_{ij}\big(t, X(t)\big) X_i(t), \quad i \in 1{:}c, \quad j \in 1{:}c \cup \{\bullet\}, \quad i \neq j, \tag{S2}$$

where $\mu_{ij}\big(t, X(t)\big)$ is called a per-capita rate or a unit rate. Flows into the system require special treatment since $X_i(t)$ in (S2) is not defined for $i = \bullet$. Instead, we specify

$$dN_{\bullet i}/dt = \mu_{\bullet i}\big(t, X(t)\big). \tag{S3}$$

This is the the interpretation and implementation used for Model 2 in our study.

S2.2 Simple Markov counting system interpretation

A continuous time Markov chain can be specified via its infinitesimal transition probabilities. A basic approach to this is to define

$$\begin{aligned} \mathbb{P}\big[N_{ij}(t+\delta) - N_{ij}(t) = 0 \,|\, X(t)\big] &= 1 - \delta\mu_{ij}\big(t, X(t)\big) X_i(t) + o(\delta), & \text{(S4)} \\ \mathbb{P}\big[N_{ij}(t+\delta) - N_{ij}(t) = 1 \,|\, X(t)\big] &= \delta\mu_{ij}\big(t, X(t)\big) X_i(t) + o(\delta), & \text{(S5)} \end{aligned}$$

for $i \in 1{:}c$ and $j \in 1{:}c \cup \{\bullet\}$ with $i \neq j$. As with the ODE case, we need special attention for flows into the system, and we define

$$\begin{aligned} \mathbb{P}\big[N_{\bullet i}(t+\delta) - N_{\bullet i}(t) = 0 \,|\, X(t)\big] &= 1 - \delta\mu_{\bullet i}\big(t, X(t)\big) + o(\delta), & \text{(S6)} \\ \mathbb{P}\big[N_{\bullet i}(t+\delta) - N_{\bullet i}(t) = 1 \,|\, X(t)\big] &= \delta\mu_{\bullet i}\big(t, X(t)\big) + o(\delta). & \text{(S7)} \end{aligned}$$

Together with the initial conditions $X(0)$, equations (S4)–(S7) define a Markov chain. Each flow is a simple counting process, meaning a non-decreasing integer-valued process that only has jumps of size one. We therefore call the Markov chain a simple Markov counting system (SMCS). The infinitesimal mean of every flow is equal to its infinitesimal variance (Bretó and Ionides, 2011) and so an SMCS is called equidispersed. We note that the special case of Model 1 used by Lee et al. (2020a) (with $\sigma_{\text{proc}} = 0$) is an SMCS. To permit more general mean-variance relationships for a Markov counting system, we must permit jumps of size greater than one. The utility of overdispersed models, where the infinitesimal variance of the flow exceeds the infinitesimal mean, has become widely recognized (Stocks et al., 2020; He et al., 2010).

S2.3 Overdispersed Markov counting system interpretation

Including white noise in the rate function enables the possibility of an overdispersed Markov counting system (Bretó and Ionides, 2011; Bretó et al., 2009; He et al., 2010). Since rates should be non-negative, Gaussian noise is not appropriate and gamma noise is a convenient option that has found various applications (Romero-Severson et al., 2015; Subramanian et al., 2020). Specifically, we consider a model given by

$$\mu_{ij}\big(t, X(t)\big) = \bar{\mu}_{ij}\big(t, X(t)\big)\, d\Gamma_{ij}(t)/dt, \tag{S8}$$

where $\Gamma_{ij}(t)$ is a stochastic process having independent gamma distributed increments, with

$$\mathbb{E}\big[\Gamma_{ij}(t)\big] = t, \quad \mathrm{Var}\big[\Gamma_{ij}(t)\big] = \sigma_{ij}^2 t. \tag{S9}$$

Formally interpreting the meaning of (S8) is not trivial, and we do so by constructing a Markov process $X(t)$ as the limit of the Euler scheme descrived in Section S3, below. Therefore, the numerical scheme in Sec. S3 can be taken as a definition of the meaning of (S8). The Markov chain defined by the limit of this Euler scheme as the step size decreases is an overdispersed Markov counting system, with the possibility of instantaneous jumps of size greater than one (Bretó and Ionides, 2011).

S3 Numerical solutions to compartment models

Models may be fitted and their implications assessed via numerical solutions (i.e., simulations) from the model equations. All the analyses we consider have this simulation-based property, known as plug-and-play or equation-free or likelihood-free. The numerical solutions to the model are arguably of more direct scientific interest than the exact solutions to the postulated equations. For ODE models, numerical methods are well studied and a standard numerical solution package such as `deSolve` in `R` is adequate for many purposes. For SMCS and ODMCS models, exact schemes are feasible when the number of events is small, which may be the case for small populations. However, for applicability to larger populations, we use instead the following Euler scheme. Write δ for an Euler time step, and ΔN_{ij} for the numerical approximation to $N_{ij}(t+\delta) - N_{ij}(t)$ given $X(t)$. For each i and j in $1:c \cup \{\bullet\}$ with $i \neq j$, we draw independent Gamma distributed noise increments with mean δ and variance $\sigma_{ij}^2\delta$, denoted using a mean-variance parameterization of the gamma distribution as

$$\Delta\Gamma_{ij} \sim \mathrm{gamma}(\delta, \sigma_{ij}^2\delta). \tag{S10}$$

In the case of an SMCS model, $\sigma_{ij} = 0$ for all i and j, so we have $\Delta\Gamma_{ij} = \delta$. Then, for $i \neq \bullet$ and $j \neq i$, and writing

$$\mu_{ij} = \bar{\mu}_{ij}\big(t, X(t)\big)\Delta\Gamma_{ij}/\delta, \tag{S11}$$

we calculate transition probabilities

$$p_{ij} = \exp\left\{-\sum_{k\in 1:c\cup\{\bullet\}} \mu_{ik}\,\delta\right\} \frac{\mu_{ij}}{\sum_{k\in 1:c\cup\{\bullet\}}\mu_{ik}}, \tag{S12}$$

$$p_{ii} = 1 - \sum_{j\neq i} p_{ij}. \tag{S13}$$

These probabilities correspond to competing hazards for every individual in compartment i to transition to some compartment j, interpreting $j = i$ to mean that the individual remains in i. Then, $\big(\Delta N_{i1}, \ldots, \Delta N_{ic}, \Delta N_{i\bullet}\big)$ has the multinomial distribution where $X_i(t)$ individuals are allocated independently to $1 : c \cup \{\bullet\}$ with probabilities given by (S12) and (S13). We use the `reulermultinom` function in the `pomp` package to draw from this multionomial distribution.

Different treatments of demographic flows—such as birth, death, immigration and emigration—are possible. For the case $i = \bullet$, the treatment used by Model 1 is to set

$$\Delta N_{\bullet j} \sim \text{poisson}(\mu_{\bullet j}\delta), \tag{S14}$$

an independent Poisson random variable with mean $\mu_{\bullet j}\delta$.

Models 2 and 3 used an alternative approach, balancing the total number of flows in and out of the compartment, i.e., $\sum_i N_{\bullet i}(t) = \sum_i N_{i\bullet}(t)$, in order to make the model consistent with the known total population. In this case, we formally model the death rate as a rate of returning to the susceptible class S, and use external transitions from $\bullet$ into S to describe only net population increase.

S4 Measurement Models

Each POMP requires specification of a measurement model, which is a statistical description of how observations on the system are obtained. In general, we used the same measurement models that were reported by Lee et al. (2020a) unless specifically noted in the following subsections.

S4.1 Model 1

In this model, the advantage afforded by vaccination is an increased probability that an infection is asymptomatic. Therefore, under the assumptions of this model, all reported cases are assumed to be a fraction of individuals that transition from the exposed to the infected compartment, as noted in Eq. (S15):

$$Y_n \mid \Delta N_{E.I.}(n) = \delta \sim \text{NB}\left(\rho\delta, \psi\right), \tag{S15}$$

where Y_n is the reported cholera cases at time $t_n \in t_1 : t_N$ and $\Delta N_{E.I.}(n)$ is the sum total of individuals across each vaccination compartment $z \in 1 : Z$ who moved from compartment E_z to I_z since observation t_{n-1}. Here, $\text{NB}\left(\rho\delta, \psi\right)$ denotes a negative binomial distribution with mean $\rho\delta$ and variance $\rho\delta\big(1 + \frac{\rho\delta}{\psi}\big)$.

S4.2 Model 2

Model 2 was fit using reported case counts that were transformed using the natural logarithm. We fit Model 2 using the subplex algorithm in the `subplex` package, using a Gaussian measurement model (Eq. (S16)) on the log transformed cases within each unit. The final loss function that is maximized is the product of the likelihoods of the individual units, or the sum of the log-likelihood of the individual units. The measurement model for individual units is given in Eq. (S16).

$$\log\left(Y_{u,n}+1\right) \mid \Delta N_{E_{u\cdot}I_{u\cdot}}(n) = \delta_u \sim \mathrm{N}\left(\log\left(\rho\delta_u+1\right), \psi^2\right), \tag{S16}$$

where $\Delta N_{E_{u\cdot}I_{u\cdot}}(n)$ is the sum total of individuals across vaccination compartment $z \in 0:4$ within unit u who moved from compartment E_{uz} to I_{uz} since observation t_{n-1}. Therefore, because the natural logarithm of observed case counts (plus one, to avoid taking the logarithm of zero) has a normal distribution, $Y_{u,n}+1$ is assumed to follow a log-normal distribution with log-mean parameter $\log\left(\rho\Delta N_{E_{u\cdot}I_{u\cdot}}(n)+1\right)$ and log-variance ψ^2. We note that fitting a model with this measurement model is equivalent to fitting using least squares, with $\log(Y_{u,n}+1)$ as the response variable.

This measurement model differs from that used by Lee et al. (2020a), who fit the model in two stages: epidemic and endemic phases. Although their text and supplement material do not explicitly describe the measurement model used, inspection of the code provided with their submission suggests a change in measurement model between the epidemic and endemic phases. The measurement model they used for the epidemic phase is

$$Y_{u,n} \mid \Delta N_{E_{u\cdot}I_{u\cdot}}(n) = \delta_u \sim \mathrm{N}\left(\rho\delta_u, \psi^2\right), \tag{S17}$$

The measurement model they used for the endemic phase modifies the epidemic model by counting both asymptomatically infected (A) and symptomatically infected (I) individuals in the case counts:

$$Y_{u,n} \mid \Delta N_{E_{u\cdot}I_{u\cdot}}(n) = \delta_{u1}, \Delta N_{E_{u\cdot}A_{u\cdot}}(n) = \delta_{u2} \sim \mathrm{N}\left(\rho\big(\delta_{u1}+\delta_{u2}\big), \psi^2\right), \tag{S18}$$

where the notation for $\Delta N_{E_{u\cdot}A_{u\cdot}}(n)$ is similar to $\Delta N_{E_{u\cdot}I_{u\cdot}}(n)$, described above.

S4.3 Model 3

In this model, reported cholera cases are assumed to stem from individuals who develop symptoms and seek healthcare. Therefore reported cases are assumed to come from an over-dispersed negative binomial model, given the increase in infected individuals:

$$Y_{u,n} \mid \Delta N_{S_{u\cdot}I_u}(t) = \delta_u \sim \mathrm{NB}\left(\rho\delta_u, \psi\right), \tag{S19}$$

where $\Delta N_{S_{u\cdot}I_{uz}}(n)$ is the number of individuals who moved from compartment S_{uz} to I_u since observation t_{n-1}.

This measurement model is a minor change from that used by Lee et al. (2020a), which allowed for a change in the reporting rate on January 1st, 2018. The fitted values of the reporting rate—before and after January 2018—were 0.97 and 0.097, respectively. An instantaneous change from near perfect to almost non-existent reporting can be problematic, as it forces the model to explain the observed reduction in reported cases as a decrease in the reporting of cases, rather than a decrease in the prevalence of cholera. This shift was justified by a "change of the case definition that occurred on January 1st, 2018"; this claim was not cited, and we could find no evidence that such a drastic change in the reporting rate would be warranted. We therefore do not allow a change in reporting rate when fitting Model 3.

S5 Initial Values

To perform inference on POMP models, it is necessary to propose an initial probability density for the latent process $f_{X_0}(x_0;\theta)$, including the possibility that the initial values of the latent states are known, or are a non-random function of the unknown parameter vector, θ. This density is used to obtain initial values of the latent state when fitting and evaluating the model. For each of the models considered in this analysis, the initial conditions are derived by enforcing the model dynamics on reported cholera cases. It is also sometimes necessary to fit some initial value parameters in order to help determine initial values for weakly identifiable compartments. In the following subsections, we mention initial value parameters that were fit for each model.

S5.1 Model 1

For this model, the number of individuals in the Recovered and Asymptomatic compartments are set to zero, but the initial proportion of Infected and Exposed individuals is estimated as initial value parameters ($I_{0,0}$ and $E_{0,0}$, respectively) using the IF2 algorithm, implemented as `mif2` in the `pomp` package. Finally, the initial proportion of Susceptible individuals $S_{0,0}$ is calculated as $S_{0,0} = 1 - I_{0,0} - E_{0,0}$. This model for the initial values of the latent states matches that which was used by Lee et al. (2020a).

S5.2 Model 2

Model 2 assumes that the initial values are a deterministic function of the reporting rate and the initial case reports, and so no initial value parameters need to be estimated. Initial values for latent state compartments are chosen so as to enforce the model dynamics on the observed number of cases. Specifically, the model sets $I_{u0}(0) = y^*_{1u}/\rho$ for each unit $u \in 1:10$, where $I_{u0}(t)$ is the number of infected individuals in unit u at time t, vaccination scenario $z = 0$, y^*_{tu} is the reported number of cases, and ρ is the reporting rate. It is futher assumed that there are no individuals in the recovered compartment, as the epidemic has just begun. This model for the initial values of the latent states matches that which was used by Lee et al. (2020a).

The decision to fix initial values so that they satisfy the dynamics of the model has the benefit of reducing the number of estimated parameters and enforcing latent states at time t_0 to be consistent with the calibrated model. The risk of fixing initial values rather than estimating them is doing so may have substantial effects on the model dynamics, and hence on the consequences of the analysis. To consider the impact of the chosen model approach, we consider an alternative initialization model that enables flexible estimation of certain latent states. Specifically, we initialize $I_{u0}(0) = \tilde{I}_{u0}$ for $u \in 1:10$, where $\{\tilde{I}_{u0}\}_{u=1}^{10}$ is a set of additional model parameters. We then fix $S_{u0}(0) = \text{pop}_u - I_{u0}(0)$, and all other starting values are set to zero, as with the fixed value approach.

The AIC of this alternative approach is 43854.0, compared to the fixed approach with an AIC of 43926.5. This alternative initialization approach results in quantitative improvement to the model-fit, but does not result in qualitative differences in the conclusions made using this model. Figure S-4 displays the trajectory of the model with this alternate initialization model. Table S-1 gives the estimated initial values. The estimated value of the latent state is similar to the fixed value in Artibonite and Centre, where the largest number of cases are present at the start of the epidemic. Because of this, the qualitative dynamics do not differ by much when the parameters are estimated versus held constant, despite the improvement in model fit measured by AIC.

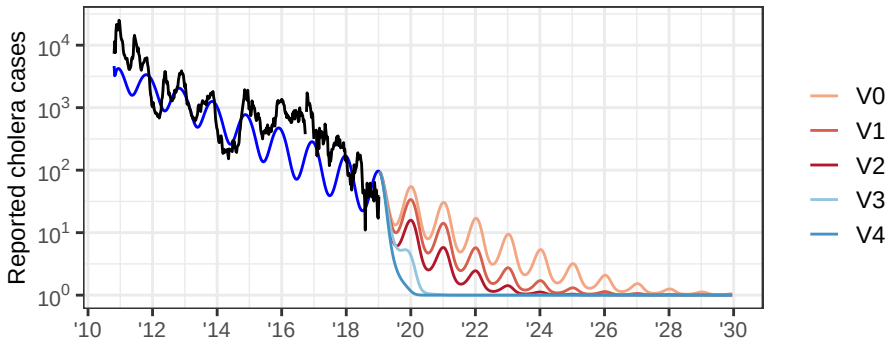

Fig S-4: **Simulated trajectory of alternate initialization of Model 2.** The black line shows the nationally aggregated weekly cholera incidence data. The blue curve from 2012-2019 is the trajectory of the calibrated version of Model 2. Compare to Fig. 4 of the article.

Table S-1: Initial values estimated for alternative initialization model for Model 2 compared to the fixed-value initialization model.

Latent State	Department	Calibrated Model Value	Fixed Value
$I_{1,0}(0)$	Artibonite	27125	31170
$I_{2,0}(0)$	Centre	2095	2535
$I_{3,0}(0)$	Grand'Anse	64	0
$I_{4,0}(0)$	Nippes	35	0
$I_{5,0}(0)$	Nord	0	85
$I_{6,0}(0)$	Nord-Est	0	0
$I_{7,0}(0)$	Nord-Ouest	3233	10
$I_{8,0}(0)$	Ouest	878	2700
$I_{9,0}(0)$	Sud	50	0
$I_{10,0}(0)$	Sud-Est	0	0

S5.3 Model 3

The latent states of this model are initialized by enforcing the model dynamics on the incidence data from the start of the recorded cases until time t_0, requiring that some of the available data be used to determine the initial values of the latent states. This is the same approach that was taken by Lee et al. (2020a), who used the value t_0 = 2014-02-22; this choice of t_0 results in modeling roughly only 60% of the available data, some of which is later discarded for alternative reasons (Lee et al., 2020b).

We do not see any immediate reason that this model could not be extended to cover a larger range of the data, and chose the value t_0 = 2010-11-13. This choice of t_0 corresponds to using approximately 1% of the available data to determine initial values of the latent states. In addition to modeling a larger portion of the available data, this choice of t_0 corresponds to an important real-world event, as daily reports from each of the departments were not being sent to the health ministry until November 10, 2010 (Barzilay et al., 2013); this choice of t_0 therefore makes $\boldsymbol{Y}(t_1)$ the first week of data once daily reports are being sent to the health ministry. The few observation

times that exist before t_0 are used to initialize the model by enforcing model dynamics on these preliminary observations. For convenience, we denote these observations as t_{-3}, t_{-2} and t_{-1}; as before, we let $y_{u,-k} = Y_u\big(t_{-k}\big)$ denote the observed case count for unit u at time point t_{-k}, where $k \in 1:3$. Equations for the initial values of non-zero latent states are provided in Eqs. (S20)–(S24); these equations match those that were used by Lee et al. (2020a), the primary change being a change to the value of t_0.

$$I_u\big(t_0\big) = \frac{y^*_{u,-1}}{7\rho\left(\mu_{IR} + (\delta + \delta_C)/365\right)}, \tag{S20}$$

$$A_u\big(t_0\big) = \frac{I_{u0}\big(t_0\big)\,(1-f)}{f}, \tag{S21}$$

$$R_{u01}\big(t_0\big) = R_{u02}\big(t_0\big) = R_{u03}\big(t_0\big) = \left(\frac{\sum_{k=-3}^{0} y^*_{u,k}}{\rho f} - \big(I_{u0}\big(t_0\big) + A_{u0}\big(t_0\big)\big)\right)/3 \tag{S22}$$

$$S_{u0}\big(t_0\big) = \mathrm{Pop}_u - I_u\big(t_0\big) - A_u\big(t_0\big) - \sum_{k=1}^{3} R_{u0k}\big(t_0\big) \tag{S23}$$

$$B_u\big(t_0\big) = \big[1 + a\tilde{J}^r\big]\mathrm{Den}_u\,\mu_W\big[I_u(t_0) + \epsilon_W A_{u_0}(t)\big]/\mu_W. \tag{S24}$$

In Eq. (S24), $\tilde{J} = 0.002376$ is the median adjusted rainfall over the observation period. One important consideration to make with this parameter initialization model is when $y^*_{u,-1} = 0$, which occurs for units $u \in \{3, 4\}$, which correspond to the Grand'Anse and Nippes departments, respectively. When this is the case, each of the infectious $I_u(t_0)$, asymptomatic $A_u(t_0)$, and bacterial reservoir $W_u(t_0)$ compartments have a value of zero. Because Model 3 models cholera transmission primarily by means of the bacterial reservoir, this makes it nearly impossible for an outbreak to occur. Therefore for units $u \in \{3, 4\}$, we introduce initial value parameters $I^3_{0,0}$ and $I^4_{0,0}$, and calibrate these parameter values using the data. The resulting parameter estimates are used to obtain the remaining non-zero initial values of the latent states using Eqs. (S21)–(S24).

S6 Calibrating Model 3 to observed cases

In this section, we provide more detail on the process that was used to estimate the coefficients of Model 3. In particular, we discuss why we decided to include additional model parameters—those that are associated with the behavior of the system during Hurricane Matthew—that were not considered by Lee et al. (2020a). To calibrate this model, we used the iterated block particle filter (IBPF) method of Ionides et al. (2022). Due to the novelty of this algorithm, there exists only a few examples where the IBPF algorithm is used for data analysis (Li et al., 2023; Ionides et al., 2022), which is one motivation of the inclusion additional details related to fitting and diagnosing the model fit provided here.

Lee et al. (2020a) only estimated model parameters to a simplified version of Model 3 on a subset of the available data, as no method existed at the time of their publication to fit a fully coupled meta-population model to disease incidence data. Building on their results, we fit the fully coupled version of Model 3 to (nearly) all available data, reserving only a few observations to use to calibrate the initial conditions of the model (see the supplement for initialization models for more details). To maximize model likelihoods, we relied on parameter estimates obtained while

calculating profile-likelihood confidence intervals, as this calculation requires many replicated IBPF searches. In our preliminary investigations that were done prior to conducting a profile likelihood search, we found that it was necessary to use multiple searches for the MLE, periodically pruning away less successful searches. To do this, the first collection of searches is performed by obtaining initial values for the parameters by uniformly sampling values from a predefined hypercube. A subsequent refinement search used parameter values corresponding to the largest model likelihoods as starting parameter values. The need for multiple searches does not appear to be uncommon, as a similar approach was used by Ionides et al. (2022). While computationally intensive, profile likelihoods proved to be an effective alternative to maximizing model likelihoods without the need to apply this multistage heuristic.

We use the iterative fitting / pruning technique described above to fit the fully coupled version of Model 3 proposed by Lee et al. (2020a). The maximum likelihood we obtained after two rounds of searching was -17549, which is higher than the benchmark model (-17933). While beating a simple associative benchmark is promising, this does not immediately imply that the model is a good description of the system. Additional investigation of parameters estimates and their corresponding implications on model based conclusions should always be conducted. For meta-population models, it is worth considering how well the model fits the data to each spatial unit. The likelihoods for each department, compared to the corresponding benchmark model, are displayed in Fig. S-5. The figure demonstrates that while the log-likelihood of the fitted model outperforms the auto-regressive negative binomial benchmark model at the aggregate level, Model 3 has lower likelihoods for some departments.

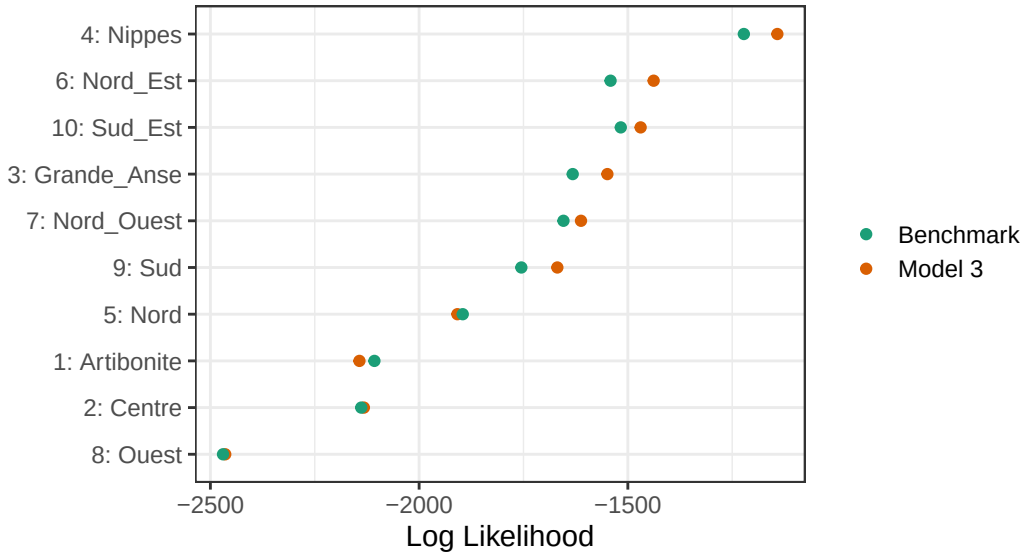

Fig S-5: Log-likelihoods of Model 3 for each department compared to the corresponding benchmark model prior to the inclusion of parameters that account for Hurricane Matthew.

In addition to considering the conditional log-likelihoods of each unit, one can consider conditional log-likelihoods of each observation. When compared to a benchmark, this level of detail can provide useful information about which observations are well described by the model and which are not. In Fig. S-6, we plot the conditional log-likelihoods of Model 3 for each observation. Typically

it is most useful to compare the conditional log-likelihoods of the model under consideration to a benchmark, as plotting only conditional log-likelihoods without a comparison may not be helpful. In this case, however, the same insight can be drawn using a figure without a benchmark comparison, so we do not include the benchmark in order to avoid the issue of over-plotting.

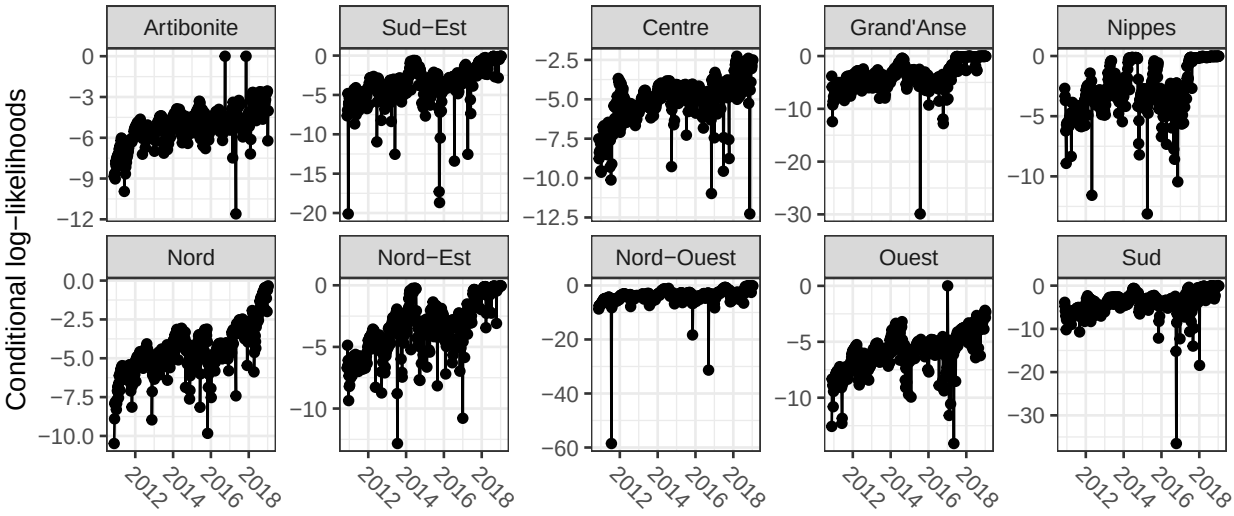

Fig S-6: Conditional log-likelihoods of Model 3 prior to the inclusion of the Hurricane Matthew related parameters.

Fig. S-6 reveals that the fitted model poorly describes certain features of the data. For example, many departments (in particular Sud) have observations with lower conditional log-likelihoods near October 2016 than at other time points. Further investigation of the model output reveals that the model is struggling to explain the sudden surge in cholera cases that occurred at this time, which coincides with the time that Hurricane Matthew struck Haiti. While the model does include a mechanism to account for increased cholera transmission due to large rainfall events, the mechanism does not appear to be sufficient to capture the damaging effects of the hurricane, which had the greatest impact in the the Sud and Grand'Anse departments (Ferreira, 2016). This result led us to include parameters β_{Wu}^{hm} and h_u^{hm} in Eq. 23 of the main text, which allows for an increase in the transmission rate between environmental cholera and humans for in Sud and Grand'Anse during and after the hurricane. The effect of the hurricane on cholera transmission is assumed to have an exponential decay, where the magnitude is determined by β_{Wu}^{hm} and the duration of the effect determined by h_u^{hm}.

We refit Model 3 after introducing these hurricane-related parameters. The resulting model has a log-likelihood value of -17332.9. The inclusion of these parameters resulted in an overall increase of 216.4 log-likelihood units. Such a large difference in log-likelihoods is well beyond the threshold of statistical uncertainty determined by Wilks' theorem, suggesting that the data highly favor the inclusion of the additional parameters. The addition of the Hurricane parameters also increases in conditional likelihoods for each observation, particularly around October 2016 (Fig. S-7).

Now that the model with additional parameters has been calibrated to the incidence data, we plot the conditional log-likelihood of each department compared to a benchmark model in Fig. S-8. The difference in log-likelihoods between Model 3 and the benchmark model is smallest in

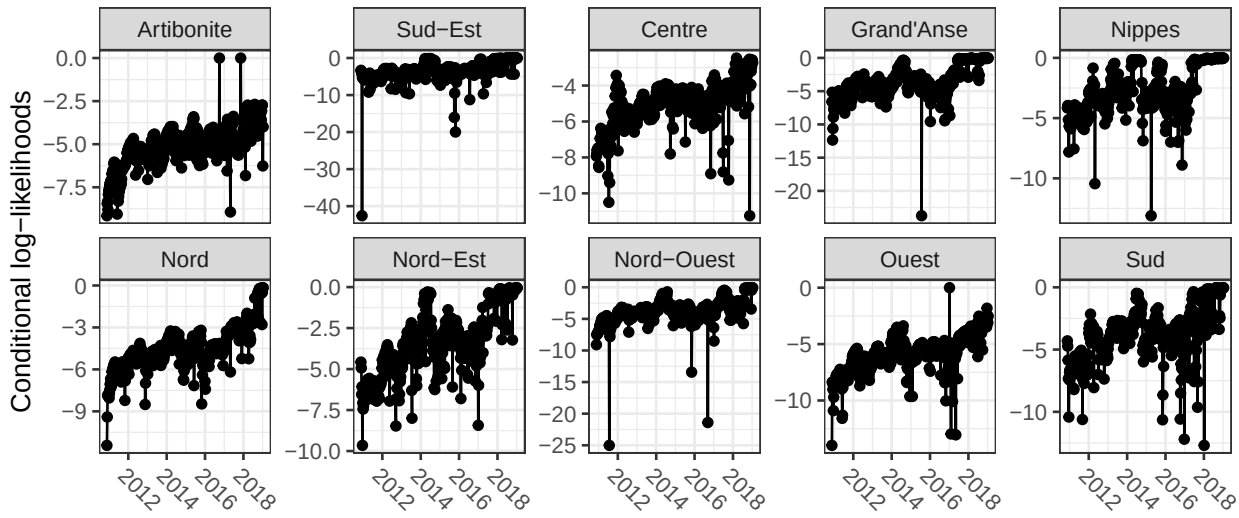

Fig S-7: Conditional log-likelihoods of Model 3 after adding and estimating the parameters related to Hurricane Matthew.

the departments Artibonite, Nord, Ouest and Centre. Each of these departments also exhibited the most sustained cholera transmission, defined by having the fewest number of weeks with no recorded cholera cases. Specifically, these four departments have zero cholera cases recorded in less than 4% of the available data, and all remaining departments—except for Nord-Ouest, which has approximately 9.5% of cases that are zeros and also exhibits the next smallest difference in log-likelihoods—have zero cases recorded in at least 14% of the available weekly data. This result suggests that the quantitative advantage Model 3 has over its respective benchmark is primarily due to the model's ability to describe a resurgence of cases after a department records a week of zero cholera cases. This result may be unsurprising in the context of the models that we are comparing. Because the cholera transmission in individual departments likely depends on the national prevalence of cholera and the Vibrio cholerae bacteria, our spatially-independnet benchmark model that relies exclusively on the previous number of case within any given unit has a difficult time predicting a resurgence of cases.

The difference in log-likelihoods between Model 3 and its benchmark model for each individual units suggests that Model 3 has a relatively poor fit for the four units with the most sustained cholera transmission. The simple four parameter benchmark has a higher likelihood than Model 3 for the Artibonite and Nord departments, and also has log-likelihoods only a few units smaller than Model 3 for the departments Ouest and Centre. This is particularly worrisome given that these four departments account for more than 77% of the total number of reported cholera cases.

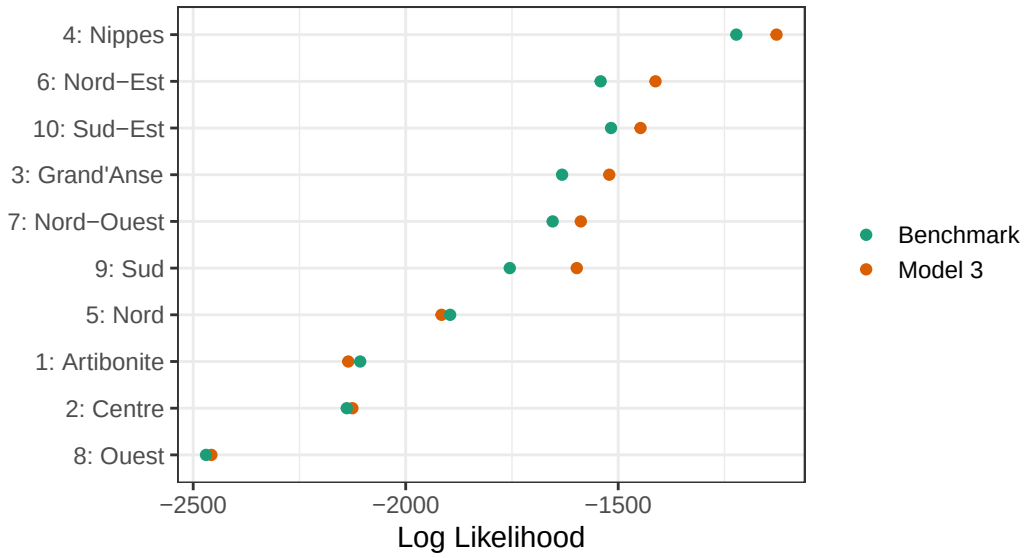

Fig S-8: Log-likelihoods of Model 3 for each department compared to the corresponding benchmark model after adding and estimating parameters related to Hurricane Matthew.

S6.1 Examining the Hidden States of the Calibrated Model

For mechanistic models, beating a suitable statistical benchmark does not alone guarantee that the model provides an accurate description of a dynamic process. Indeed, a good statistical fit does not require the model to assert a causal explanation. For example, reconstructed latent variables should make sense in the context of alternative measurements of these variables (Grad et al., 2012). We demonstrate this principle by examining the latent state of the calibrated model. In particular, we examine the compartment of susceptible individuals under various scenarios. This analysis can also provide insight into why the calibrated model fails to outperform the benchmark model on the four departments with the most sustained cholera transmission.

Recall that the filtering distribution for the calibrated version of Model 3 at time t_k is defined as the distribution of the latent state at time t_k given the data from times $t_1 : t_k$, i.e. $f^{(3)}_{\boldsymbol{X}_k|\boldsymbol{Y}_{1:k}}(\boldsymbol{x}_k|\boldsymbol{y}^*_{1:k};\hat{\theta})$. In general, one may expect simulations from the filtering distribution of a model with a good statistical fit to result in hidden states that are highly consistent with the observed data because the filtering distribution is conditioned on the observed data. Fig. S-9 shows the percentage of individuals that are in the susceptible compartment from various simulations of the model: simulations from Model 3 under initial conditions are displayed in red; simulations from the filtering distribution of model are displayed in blue. This figure shows that simulations from initial conditions tends to result in a much more rapid depletion of susceptible individuals at the start of the epidemic than simulations from the filtering distribution, suggesting the calibrated model has a propensity to predict larger outbreaks than what is typically seen in the data. This result demonstrates that the calibrated model favors a more rapid growth in cholera cases than what is typically seen in the observed data, providing a possible explanation as to why the model fails to beat the simple benchmark for each spatial unit. This results hints at the possibility of model mispecification, and warrants a degree of caution in interpreting the model's outputs.

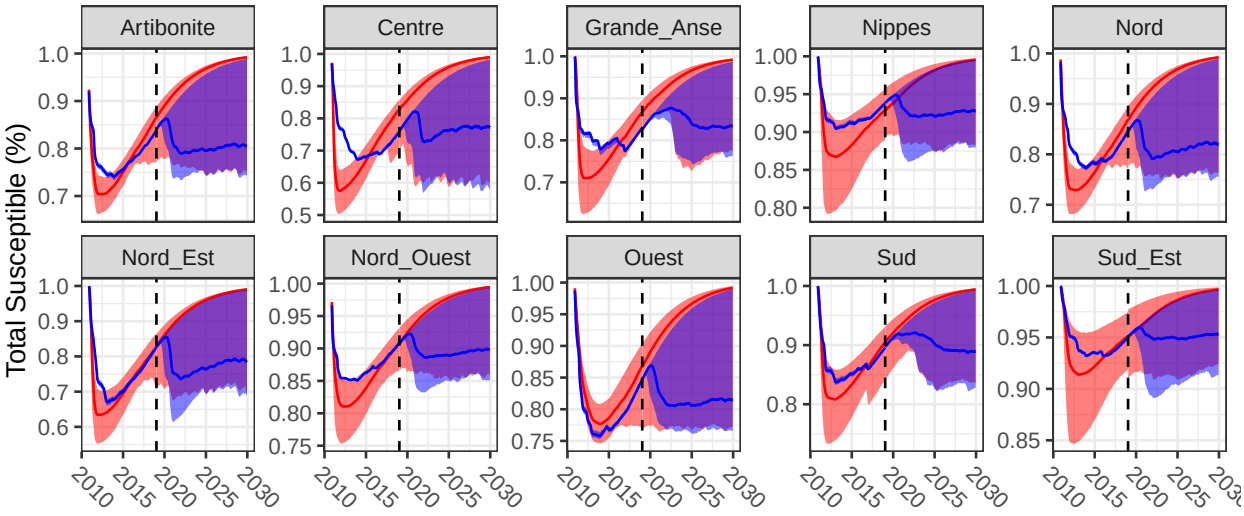

Fig S-9: Percentage of individuals that are in the susceptible compartment. Simulations from Model 3 under initial conditions are displayed in red; simulations from the filtering distribution of model are displayed in blue. The dashed line represents the end of the observed data.

S7 Replication of Lee et al. (2020a)

In this article we claimed that we were able to obtain better fits to the observed data using the same models that were proposed by Lee et al. (2020a). Along with visual comparisons to the data, this claim was supported by comparing likelihoods and AIC values in Table 2 in the manuscript. Because model likelihoods were not provided by Lee et al. (2020a), it is necessary to replicate these models in order to obtain likelihood estimates. Here we would like to thank the authors of Lee et al. (2020a), who provided detailed descriptions of their models, which enabled us to build on their work. In the following subsections, we use our `R` package `haitipkg` to reproduce some of the results of Lee et al. (2020a). This reproduction allows us to estimate the likelihoods of the Lee et al. (2020a) version of Models 1–3, and also provides a demonstration of the importance and usefulness of reproducible research.

S7.1 Model 1 Replication

The model was implemented by a team at Johns Hopkins Bloomberg School of Public Health (hereafter referred to as the Model 1 authors) in the `R` programming language using the `pomp` package (King et al., 2016). Original source code is publicly available with DOI: 10.5281/zenodo.3360991. The final results reported by the Model 1 authors were obtained by using several different parameter sets rather than a single point estimate. According to the supplement materials, this was because model realizations from a single parameter set retained substantial variability, but multiple realizations from a collection of parameter sets resulted in a reasonable visual fit to the data. We are also inclined to believe that the use of multiple parameter values was in part intended to account for parameter uncertainty—the importance of which was discussed in the main text—an effort by the Model 1 authors that we applaud. Simulations from each of the parameter sets, however, were treated with equal importance when being used to diagnose the model fit and make inference on the system. This is problematic given Figures S8 and S9 of the supplement material, which suggest that some parameter sets that were used for inference may have been several hundred units of log-likelihood lower than other parameter sets that were simultaneously used to make forecasts. Such a large difference in log-likelihoods is well beyond the threshold of statistical uncertainty determined by Wilks' theorem, resulting in the equal use of statistically inferior parameter sets in order to make forecasts and conduct inference on the system.

To fully reproduce the results of the Model 1 authors, it is necessary to use the exact same set of model parameters that were originally used to obtain the results presented by Lee et al. (2020a). Because these parameter sets were not made publicly available, we relied on the source code provided by the Model 1 authors to approximately recreate the parameter set. Due to software updates since the publication of the source code, we were unable to produce the exact same set of parameters. Running the publicly available source code, however, resulted in a set of parameters that are visually similar to those used by the Model 1 authors (See Figures S-10 and S-11). Furthermore, simulations using the set of parameters produced by the source code appear practically equivalent to those displayed by Lee et al. (2020a) (See Figure S-12).

Because the model forecasts provided by Lee et al. (2020a) come from various sets of parameters—which each correspond to a unique log-likelihood value—it is not obvious how one would obtain an estimate for the log-likelihood of the model that was used for simulations by the Model 1 authors. One approach could be to calculate the logarithm of the weighted mean of the likelihoods for each parameter sets used to obtain the forecasts, where the weights are proportional to the number of

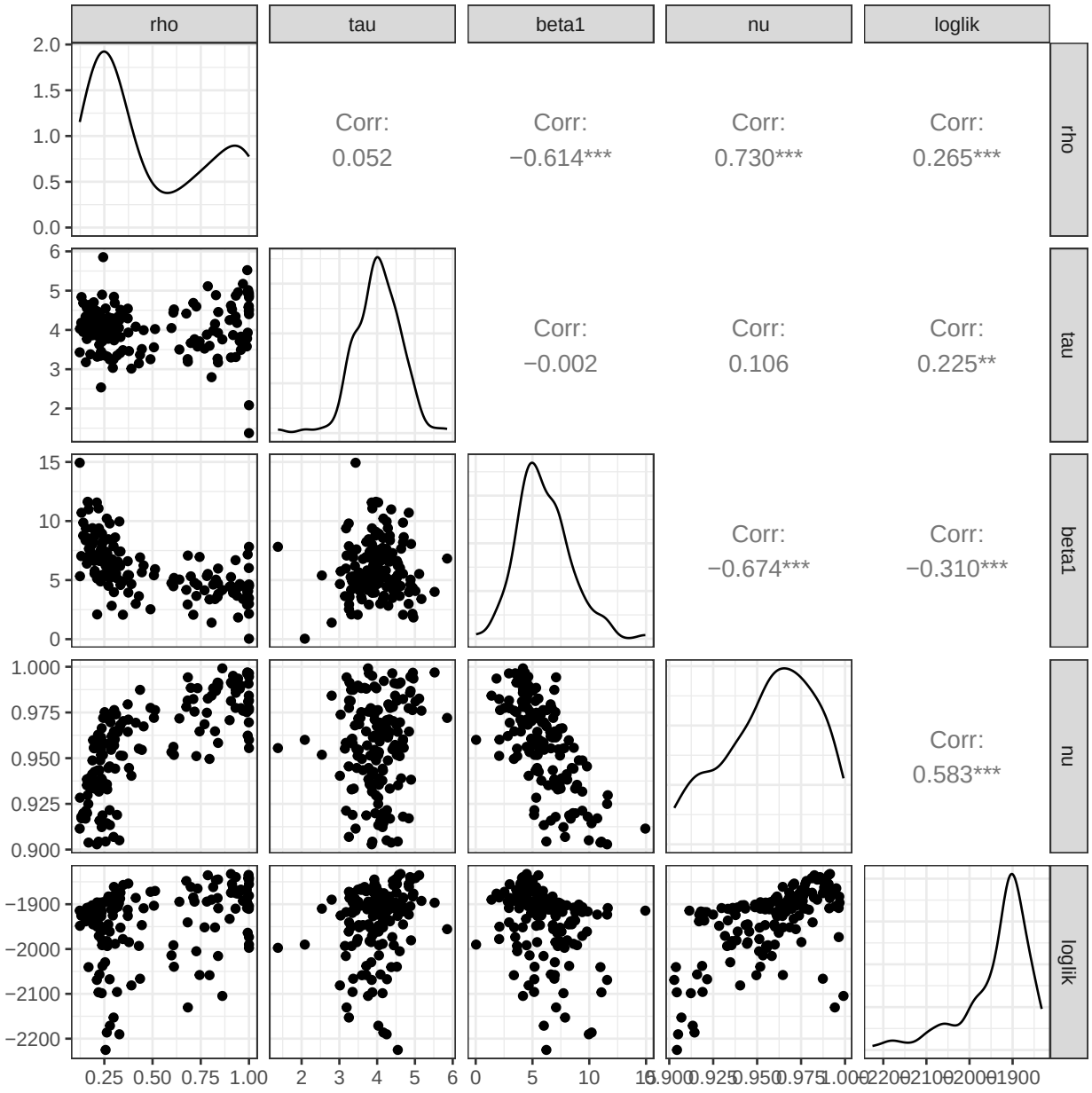

Fig S-10: Bivariate distributions of parameter estimates after fitting epidemic phase of the Model 1 following the procedure described by Lee et al. (2020). Compare to Figure S8 in the supplement of Lee et al. (2020).

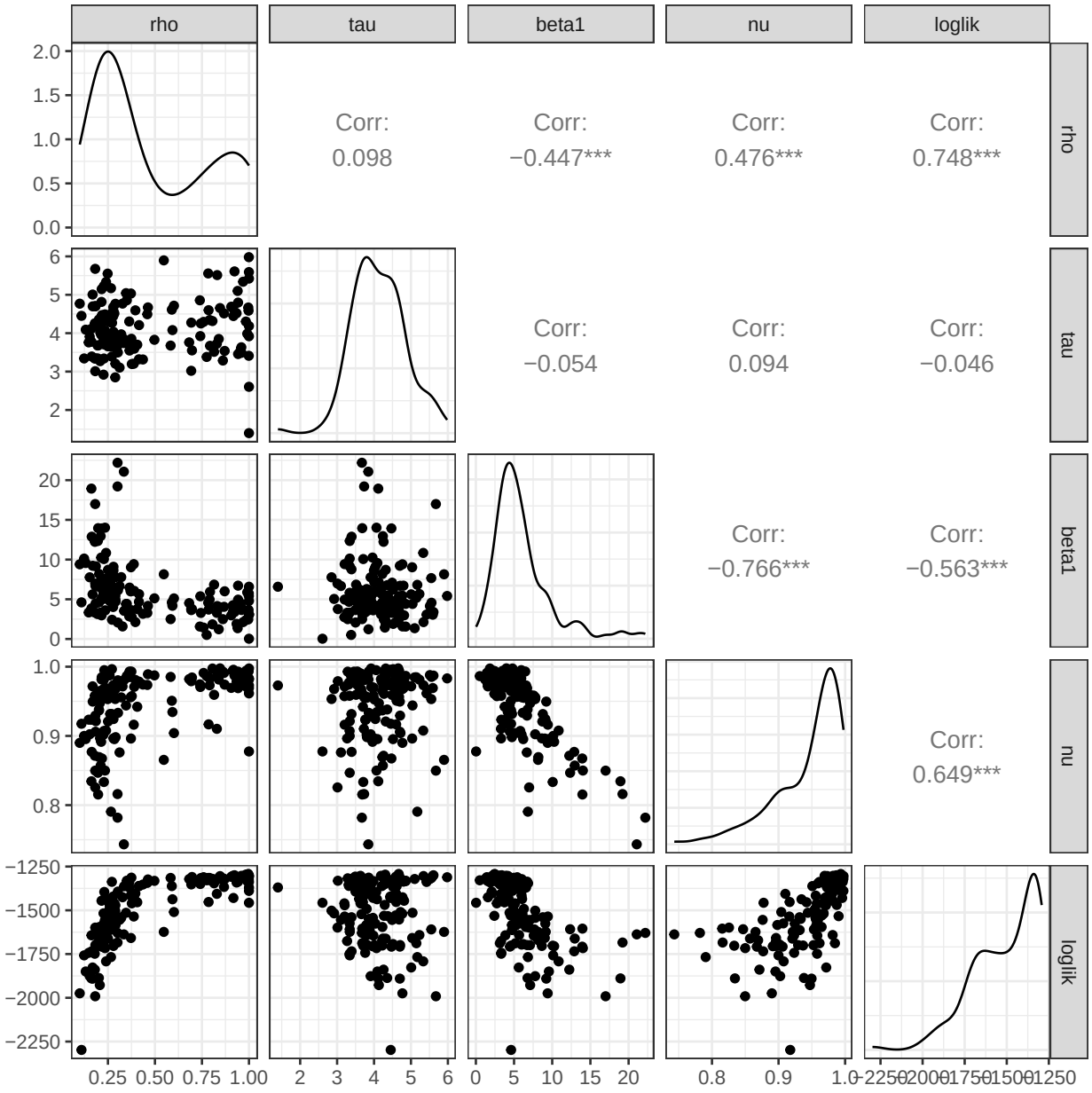

Fig S-11: Bivariate distributions of parameter estimates after fitting endemic phase of the Model 1 following the procedure described by Lee et al. (2020). Compare to Figure S9 in the supplement of Lee et al. (2020).

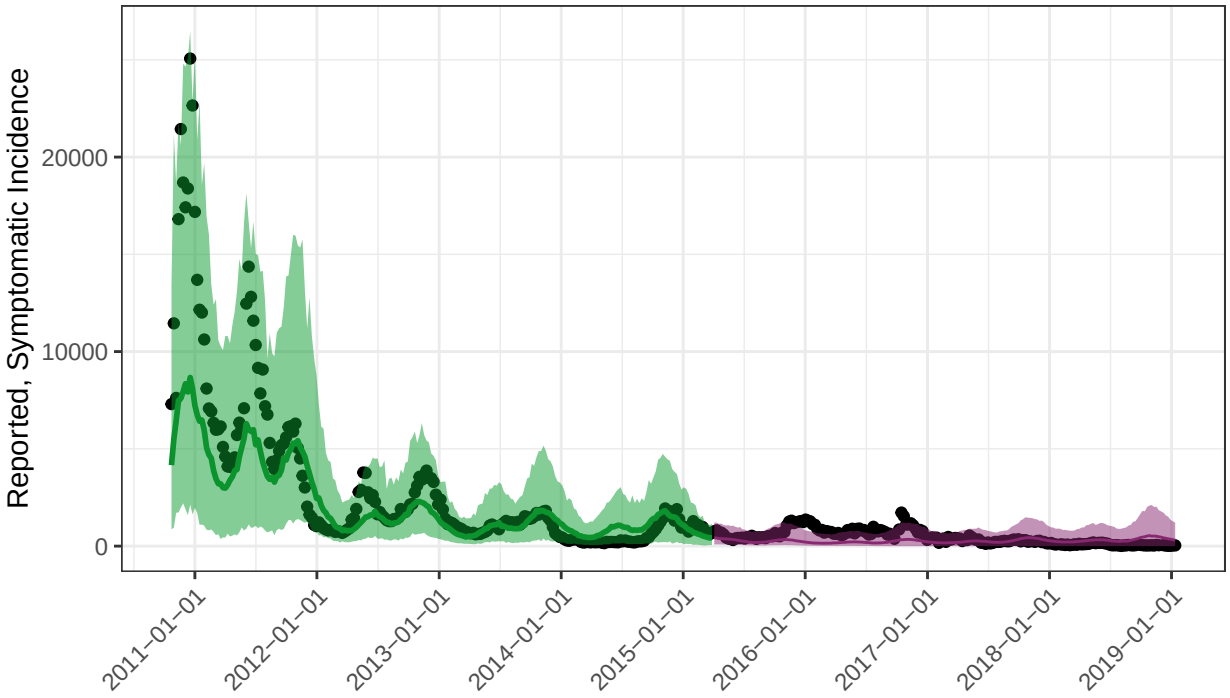

Fig S-12: Simulations from Model 1 using parameter sets that were generated by running source code provided by Lee et al (2020). Compare to Figure S7 in the supplement of Lee et al. (2020). The upper bound for the likelihood of this model is -3031.

times the parameter set was used. However, in an effort to not underestimate the likelihood of the model of the Model 1 authors, we report the estimated log-likelihood as the log-likelihood value corresponding to the parameter set with the largest likelihood value, even though the majority of simulations were obtained using parameter sets with lower likelihood values. In this sense, we consider the log-likelihood reported in Table 1 of the main text to be an upper-bound of the log-likelihood of the model used by Lee et al. (2020a). For each parameter set, the log-likelihood was estimated using a particle filter, implemented as the `pfilter` function in the `pomp` package.

S7.2 Model 2 Replication

Model 2 was developed by a team that consisted of members from the Fred Hutchinson Cancer Research Center and the University of Florida (hereafter referred to as the Model 2 authors). While Model 2 is the only deterministic model we considered in our analysis, it contains perhaps the most complex descriptions of cholera in Haiti: Model 2 accounts for movement between spatial units; human-to-human and environment-to-human cholera infections; and transfer of water between spatial units based on elevation charts and river flows.

The source code that the Model 2 authors used to generate their results was written in the `Python` programming language, and is publicly available at `10.5281/zenodo.3360857` and its accompanying GitHub repository `https://github.com/lulelita/HaitiCholeraMultiModelingProject`. In order to perform our analysis in a unified framework, we re-implemented this model in the `R` programming language using the `spatPomp` package (Asfaw et al., 2023), which facilitates the creation of meta-population models. We note that the travel and water matrices used to implement the complex dynamics in Model 2 (Lee et al., 2020b) are not available in either the Zenodo archive or the GitHub repository; instead, we obtained these matrices via personal correspondence with the Model 2 authors. Using these matrices, and the point estimates for model parameters provided by (Lee et al., 2020b), we created trajectories of the cholera dynamics and compared this to available data. These trajectories, shown in Figure S-13, are very similar to the trajectories shown in Figure S15 of Lee et al. (2020b).

There are minor differences between Figure S-13 and Figure S15 of Lee et al. (2020b). While the discrepancy appears minor, the deterministic nature of Model 2 implies that an exact replication of model trajectories should be possible. In this case, these discrepancies may possibly be attributed to implementing the model and plotting the model trajectory in two different programming languages. Another potential explanation for the discrepancy is that the parameters that we used are only approximately the same as those used by Lee et al. (2020b). For example, the parameters β, β_W had reported values of 9.9×10^{-7} and 4.03×10^{-2}, respectively (Table S13 of Lee et al. (2020b)), but were actually fit to data and therefore likely these values have been rounded. Additionally, our implementation of Model 2 used a time scale of years and many of the parameters were reported on a weekly scale, so small differences may result due to unit conversions. The collective effect of these small differences in model parameters likely will result in small differences in model trajectories.

Some additional concerns about being able to accurately replicate the results of Lee et al. (2020a) are valid. Details about the measurement models and how latent states were initialized for the epidemic model were not provided by Lee et al. (2020b) and therefore these details must be inferred by looking at the provided source code. According to repository comments, the files `fitInPieces3paramsCleanMay2019Public.py` and `fitInPiecesMuWithFracSusFixedAllInfectionsPublic.py` were used to fit the epidemic and endemic phases of the model respectively, although it is apparent that these exact files were not used to obtain the reported results since

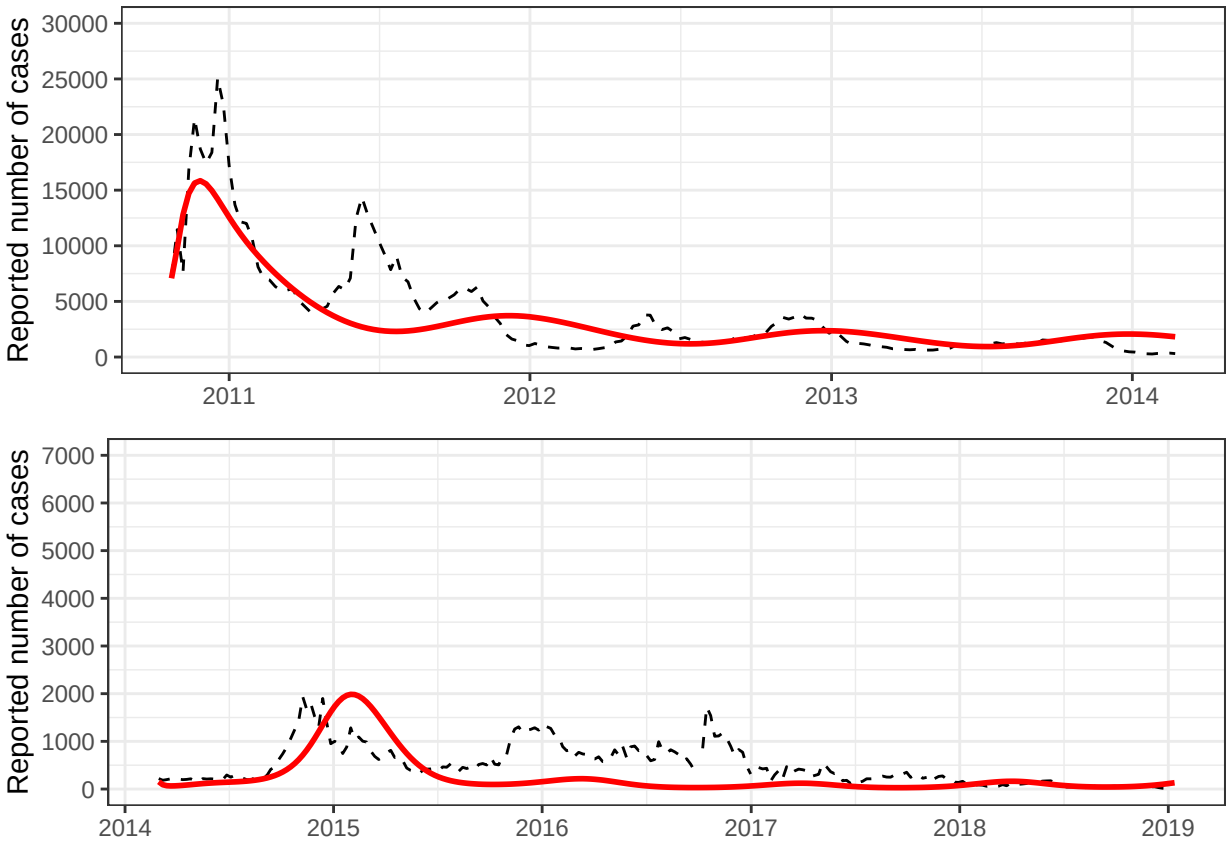

Fig S-13: Model 2 trajectories using the `haitipkg`. Compare to Figure S15 in the supplement of Lee et al. (2020b).

the files contain some variable-naming errors that make it impossible to run the files without making modifications [1]. The inability to replicate the results by Lee et al. (2020a) by running the provided source code makes checking whether or not a our numeric implementation faithfully represents their results very difficult. Additionally, the script that was said to been used to obtain the results reported by Lee et al. (2020a) appears to use a different measurement model than what was described in the supplemental material, again making it difficult to fully replicate the result of Lee et al. (2020a) without being able to easily run the provided source code. In this case, we chose to use measurement model that considers only symptomatic individuals for both phases of the epidemic, as this seemed to visually match the results of Lee et al. (2020a) most closely.

S7.3 Model 3 Replication

Model 3 was developed by a team of researchers at the Laboratory of the Swiss Federal Institute of Technology in Lausanne, hereafter referred to as the Model 3 authors. The code that was originally used to implement Model 3 is archived with the DOI: `10.5281/zenodo.3360723`, and

[1] One example of why the code cannot be run that the file loads functions from a non-extant file named `choleraEqs.py` in line 13 rather than `choleraEqsPublic.py`.

also available in the public GitHub repository: `jcblemai/haiti-mass-ocv-campaign`. Because the code was made publicly available, and final model parameters were reported in the supplementary material of Lee et al. (2020a), we were able to reproduce Model 3 by directly using the source code. In Fig. S-14, we plot simulations from this model. This figure can be compared to Figure S18 of Lee et al. (2020a). We note that slight differences may be accounted for due to variance in the model simulations and the difference in programming language used to produce the figure. Overall, the high standard of reproducibility that was achieved by the Model 3 authors facilitated the ability to readily replicate their model and results.

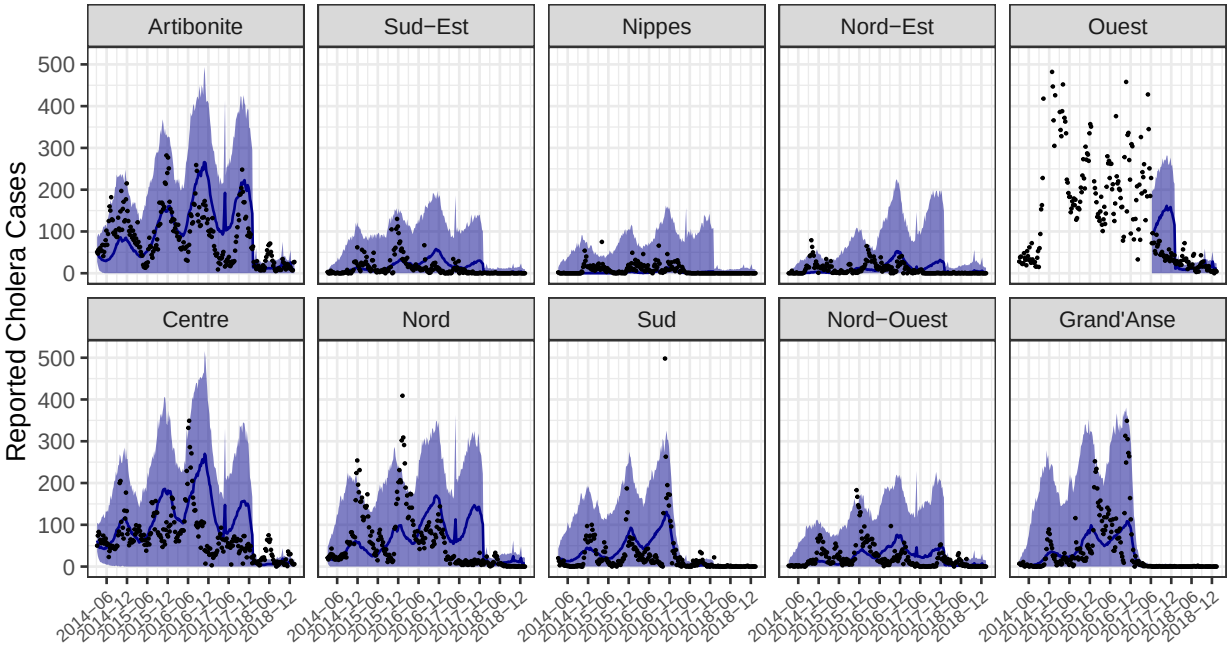

Fig S-14: Simulations from Model 3. Compare to Figure S18 in the supplement of Lee et al. (2020a).

S8 Confidence Intervals for Model Parameters

In this section we provide confidence intervals for all model parameters, excluding those that take unique values for each spatial unit. For each model and parameter, we use principles of profile likelihood to obtain confidence intervals Pawitan (2001). Due to the non-linear and stochastic nature of Models 1 and 3, exact evaluations of the profile log-likelihood are difficult to obtain. Instead, the log-likelihood at each point of the profile is estimated using via Monte-Carlo based particle filter methods. We therefore obtain confidence intervals for the parameters of Model 1 and Model 3 using the Monte Carlo adjust profile (MCAP) algorithm Ionides et al. (2017).

In each subsection, we provide figures that show the curvature of the profile log-likelihood near the MLE (Figures S-15–S-17). In these figures, the parameter values are shown on the transformed scale in which the profile was calculated.

S8.1 Model 1 parameters

Parameter estimates for Model 1, along with the MCAP confidence intervals for the estimate, are given in Table S-2. Figure S-15 displays the Monte Carlo evaluations of the profile likelihood values, obtained using a particle filter.

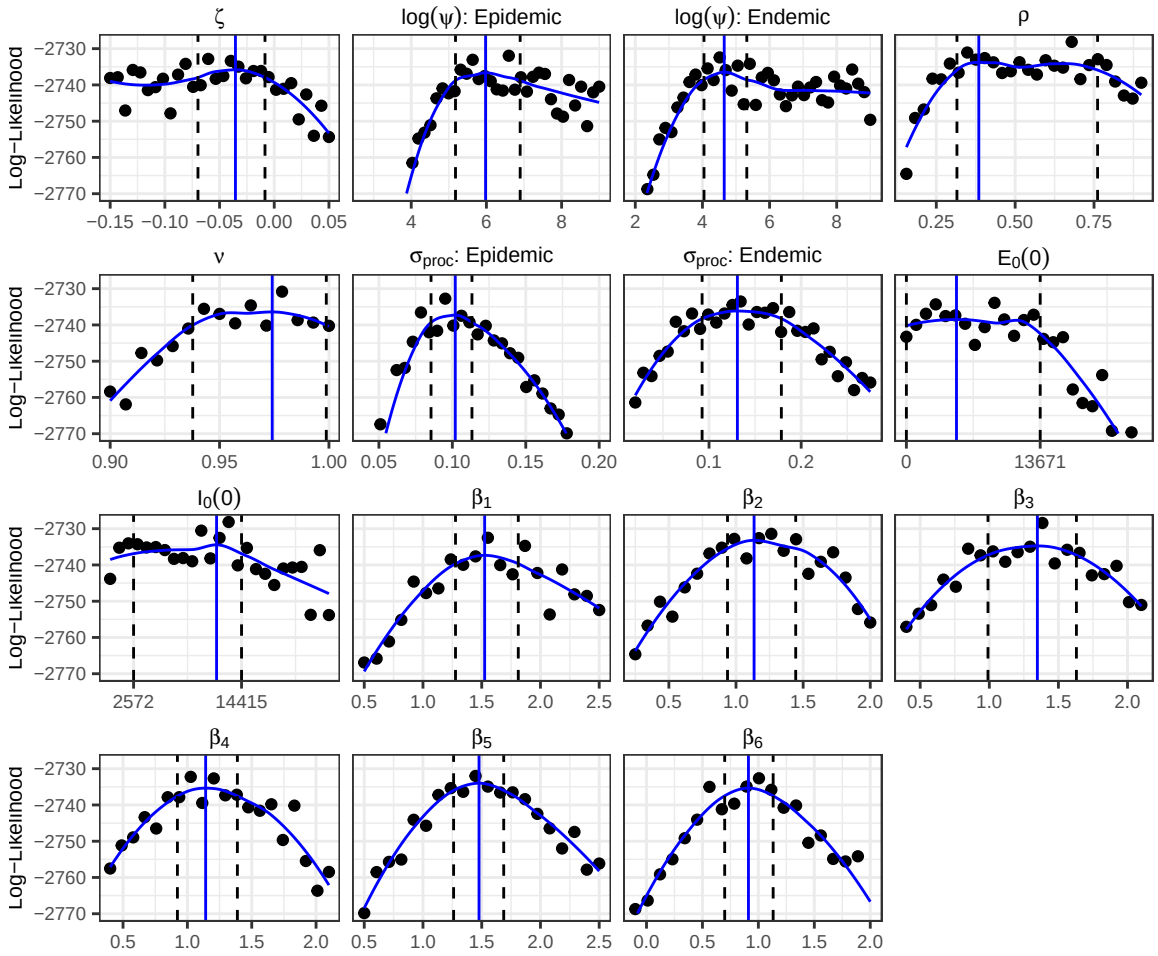

Fig S-15: MCAP confidence intervals for Model 1 parameters. The vertical blue line indicates the smoothed MLE.

Table S-2: Model 1 parameter estimates and their corresponding confidence intervals, obtained via the MCAP algorithm.

Mechanism	**Parameter**	**MLE**	**95% Confidence Interval**
Seasonality	ζ	-0.036	$(-0.070, -0.008)$
Seasonality	β_1	1.417	$(1.277, 1.811)$
Seasonality	β_2	1.169	$(0.937, 1.445)$
Seasonality	β_3	1.136	$(0.990, 1.630)$
Seasonality	β_4	1.140	$(0.922, 1.389)$
Seasonality	β_5	1.401	$(1.261, 1.687)$
Seasonality	β_6	0.988	$(0.699, 1.132)$
Observation Variance	ψ : Epi	279.147	$(177.226, 990.191)$
Observation Variance	ψ : End	78.326	$(57.171, 204.654)$
Reporting Rate	ρ	0.679	$(0.315, 0.761)$
Mixing Exponent	ν	0.978	$(0.938, 0.999)$
Process noise ($\text{wk}^{1/2}$)	σ_{proc} : Epi	0.092	$(0.085, 0.113)$
Process noise ($\text{wk}^{1/2}$)	σ_{proc} : End	0.118	$(0.092, 0.179)$
Initial Values	$I_0(0)$	7298	$(2572, 1.4415 \times 10^4)$
Initial Values	$E_0(0)$	350	$(1, 1.3671 \times 10^4)$

S8.2 Model 2 parameters

Parameter estimates for Model 2, along with the profile likelihood confidence intervals for each estimate, are given in Table S-3. Figure S-16 displays the profile log-likelihood curve near the MLE. In Table S-3, the confidence interval for μ_{RS}^{-1}, the duration of natural immunity due to cholera infection, is arbitrarily large (going to infinity). This is possible because the parameter that was estimated was μ_{RS}, and the true MLE for this parameter is zero (see Figure S-16). This suggests that the fitted model favors a regime where reinfection events are not possible. Similarly, the MLE for the parameter β, which controls the amount of cholera transmission from human to human, is zero. Because Model 2 fails to describe the incidence data as well as a simple statistical benchmark, we must be careful to not interpret these results as evidence that reinfections and human-to-human infection events do not occur. Instead, we may consider this as additional evidence of model mispecification.

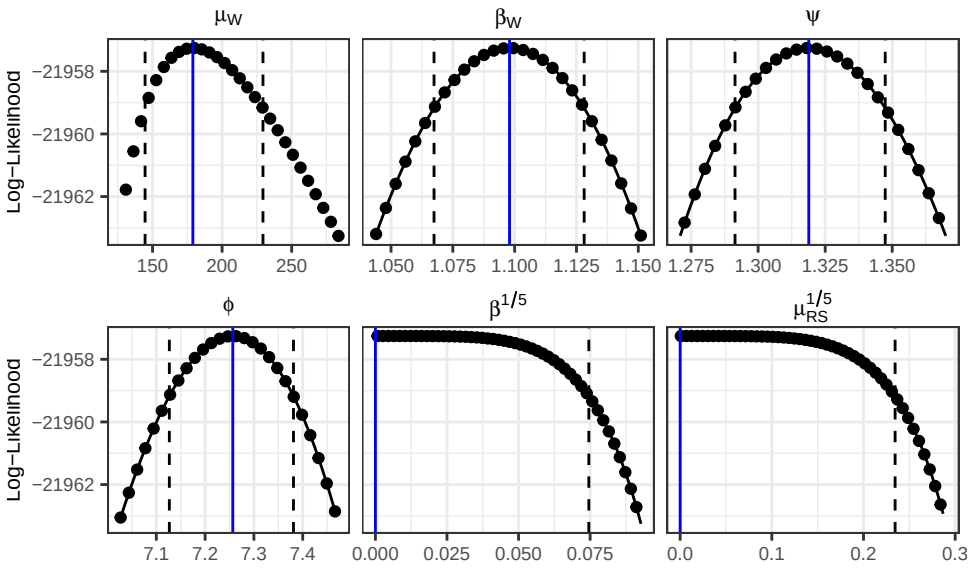

Fig S-16: MCAP confidence intervals for Model 2 parameters. The vertical blue line indicates the MLE.

Table S-3: Model 2 parameter estimates and their corresponding confidence intervals, obtained via profile likelihood.

Mechanism	**Parameter**	**MLE**	**95% Confidence Interval**
Human to water shedding (wk^{-1})	μ_W	179.2	$(144.6, 229.4)$
Water to Human Infection (yr^{-1})	β_W	1.098	$(1.067, 1.128)$
Observation Variance	ψ	1.319	$(1.291, 1.347)$
Seasonality	ϕ	0.974	$(7.127, 7.381)$
Human to Human Infection (yr^{-1})	β	$5.97 \times 10^{-15*}$	$[0, 2.3 \times 10^{-6})$
Immunity (yr)	μ_{RS}^{-1}	$1.4 \times 10^{11*}$	$(1410, \text{inf})$

*As evident in Figure S-16, the true MLE for these parameters is 0 and ∞, respectively; this value could not be obtained numerically due to the parameter transformation applied to the parameter for the model fitting processes.

S8.3 Model 3 parameters

Parameter estimates for Model 3, along with the MCAP confidence intervals for the estimate, are given in Table S-4. Figure S-17 displays the Monte Carlo evaluations of the profile likelihood values, obtained using a particle filter.

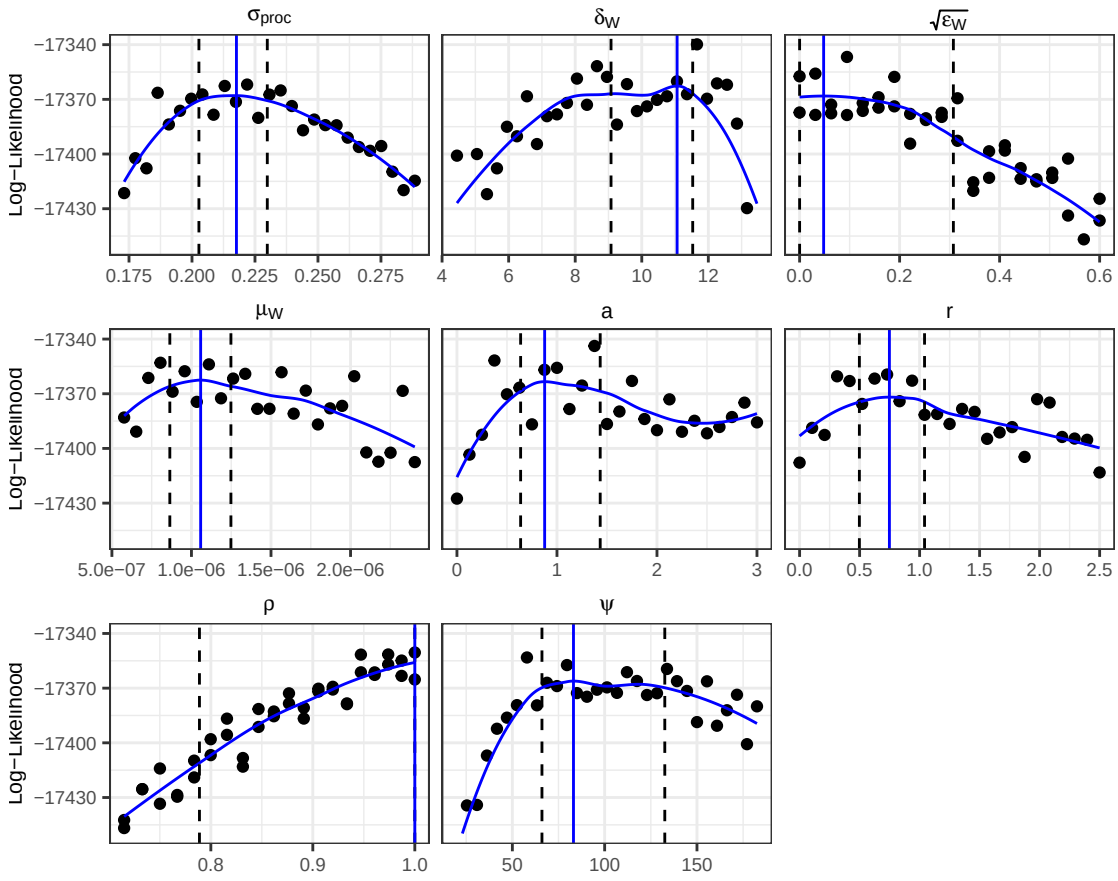

Fig S-17: MCAP confidence intervals for Model 3 parameters. The vertical blue line indicates the smoothed MLE.

Table S-4: Model 3 parameter estimates and their corresponding confidence intervals, obtained via the MCAP algorithm.

Mechanism	**Parameter**	**MLE**	**95% Confidence Interval**
Process Noise ($\text{wk}^{1/2}$)	σ_{proc}	0.218	$(0.203, 0.230)$
Water Survival (wk)	δ_W^{-1}	0.108	$(0.087, 0.110)$
Human to Water Shedding $\frac{\text{km}^2}{\text{wk}}$	μ_W	9.77×10^{-7}	$(8.64 \times 10^{-7}, 1.25 \times 10^{-6})$
Asymptomatic Shedding	ϵ_W	0.008	$(0.0, 0.095)$
Seasonality	a	1.000	$(0.637, 1.432)$
Seasonality	r	0.780	$(0.498, 1.041)$
Reporting Rate	ρ	0.983	$(0.789, 1.000)$
Observation Variance	ψ	88.578	$(66.034, 132.563)$

S9 Forecasting with parameter uncertainty

Let $f_{Y_{1:N}}(y_{1:N}|\theta)$ denote the pdf of the model under consideration, were θ is a parameter vector that indexes the model. Furthermore, denote the observed data as $y^*_{1:N}$. Because the uncertainty

in just a single parameter can lead to drastically different forecasts (Saltelli et al., 2020), parameter uncertainty should be considered when obtaining model forecasts when the goal is to influencing policy. In a Bayesian modelling paradigm, the most natural way to account for parameter uncertainty in model forecasts is to suppose that θ comes from a distribution f_Θ, and then to obtain J forecasts from the model where each forecast is obtained using parameters drawn from the posterior distribution $\theta_{1:J} \mid Y_{1:N} = y^*_{1:N} \sim f_\Theta\big(\theta | Y_{1:N} = y^*_{1:N}\big)$.

When frequentist methods are used, however, there does not exist a posterior distribution from which one could sample. A common approach could be to obtain a weighted average of the simulations from various models (Hoeting et al., 1999), but this can be problematic when forecasts from each model are very different from each other (Grueber et al., 2011). A similar approach that has been taken (King et al., 2015) is to obtain model forecasts using multiple sets of parameter values and then sample from the resulting forecasts using weights proportional to the corresponding likelihoods of the parameter values. This approach could be considered as empirical Bayes, as it is equivalent to using a discrete uniform prior where the set of values in the prior distribution is determined by a stochastic routine applied to the observed data, as discussed below.

For each $k \in 1:K$, let θ_k be a unique set of model parameters. Letting Θ denote a random vector of model parameters, we endow Θ with a discrete uniform distribution on the set $\{\theta_1, \theta_2, \ldots, \theta_K\}$, such that $P\big(\Theta = \theta_k\big) = \frac{1}{K}$ for all values $k \in 1:K$. Using this as a prior distribution, the posterior distribution of $\Theta | Y_{1:N} = y^*_{1:N}$ can be expressed as: $P\big(\Theta = \theta_k | Y_{1:N} = y^*_{1:N}\big) = \frac{f_{Y_{1:N}}(y^*_{1:N}|\theta_k)}{\sum_{l=1}^{K} f_{Y_{1:N}}(y^*_{1:N}|\theta_l)}$. In a standard empirical Bayes analysis, the values $\theta_1, \ldots, \theta_k$ of the prior distribution would be chosen using the observed data, resulting in a posterior distribution that weighs the prior parameter vectors proportional to their corresponding likelihoods. We choose θ_k to be the output of a stochastic routine applied to the observed data by setting θ_k to be the output of an iterated filtering algorithm. In practice, because the likelihood maximization routines of iterated filtering methods are stochastic, it is common to run the iterated filtering method multiple times (K) for each model in order to obtain a maximum likelihood estimate for model parameters. This results in a natural set of parameters near the MLE that could be used as the discrete prior distribution. Alternatively, the set $\{\theta_1, \theta_2, \ldots, \theta_K\}$ could be determined by first obtaining marginal confidence intervals for each element of the parameter vector Θ, and then creating a hypercube using the combination of marginal confidence intervals. The set $\{\theta_1, \theta_2, \ldots, \theta_K\}$ is then obtained by sampling uniformly K values from the resulting hypercube, as was done by (King et al., 2015).

S10 Translating to Lee et al. (2020a) notation

Since the models of Lee et al. (2020a) were developed independently, the choice of notation varies inconsistently between models. For our reanalysis, we rename parameters to provide a unified notation facilitating comparison between models. Table S-5 maps this notation back to the original notations, for reference.

Parameter	Our Notation	Lee et al. (2020a)		
		1	2	3
Reporting Rate	ρ	ρ	ρ	ϵ_1, ϵ_2
Mixing Coefficient	ν	ν	—	—
Measurement Over-Dispersion	ψ	τ	—	p
Birth Rate	μ_S	μ	—	—
Natural Mortality Rate	δ	δ	—	μ
Cholera Mortality Rate	δ_C	—	—	α
Latent Period	$1/\mu_{EI}$	$1/\sigma$	$1/\gamma_E$	—
Recovery Rate	μ_{IR}	γ	γ	γ
Loss of Immunity	μ_{RS}	α	σ	ρ
Symptomatic Ratio	f	$1-\theta_0$	k	σ
Asymptomatic Relative Infectiousness	ϵ	$1-\kappa$	red_β	—
Human-to-Water Shedding	μ_W	—	μ	θ_I
Asymptomatic Relative Shedding	ϵ_W	—	red_μ	θ_A/θ_I
Seasonal Amplitude	a	—	α_s	λ
Transmission	β	β	β	c
Water-to-Human	β_W	—	β_W	β
Bacteria Mortality	δ_W	—	δ	μ_β
Vaccination Efficacy	θ	θ_{vk}	$\theta_1, \theta_2, \theta_{1_5}, \theta_{2_5}$	η_{1d}, η_{2d}
Process Over-dispersion	σ_{proc}	—	—	σ_w

Table S-5: Translations between our common notation and notation used by Lee et al (2020).

Supplementary References

Asfaw, K., Park, J., King, A. A., and Ionides, E. L. (2023). Partially observed Markov processes with spatial structure via the R package spatPomp. *arXiv:2101.01157v3*.

Barzilay, E. J., Schaad, N., Magloire, R., Mung, K. S., Boncy, J., Dahourou, G. A., Mintz, E. D., Steenland, M. W., Vertefeuille, J. F., and Tappero, J. W. (2013). Cholera surveillance during the Haiti epidemic—the first 2 years. *New England Journal of Medicine*, 368(7):599–609. Includes case definition to 2013: The NCSS used a modified World Health Organization case definition for cholera that included acute watery diarrhea, with or without vomiting, in persons of all ages residing in an area in which at least one case of Vibrio cholerae O1 infection had been confirmed by culture.

Bhadra, A., Ionides, E. L., Laneri, K., Pascual, M., Bouma, M., and Dhiman, R. C. (2011). Malaria in Northwest India: Data analysis via partially observed stochastic differential equation models driven by Lévy noise. *Journal of the American Statistical Association*, 106:440–451.

Bretó, C., He, D., Ionides, E. L., and King, A. A. (2009). Time series analysis via mechanistic models. *Annals of Applied Statistics*, 3:319–348.

Bretó, C. and Ionides, E. L. (2011). Compound Markov counting processes and their applications to modeling infinitesimally over-dispersed systems. *Stochastic Processes and their Applications*, 121:2571–2591.

Ferreira, S. (2016). Cholera threatens Haiti after Hurricane Matthew. *BMJ*, 355:i5516.

Grad, Y. H., Miller, J. C., and Lipsitch, M. (2012). Cholera modeling: Challenges to quantitative analysis and predicting the impact of interventions. *Epidemiology*, 23(4):523.

Grueber, C. E., Nakagawa, S., Laws, R. J., and Jamieson, I. G. (2011). Multimodel inference in ecology and evolution: challenges and solutions. *Journal of Evolutionary Biology*, 24(4):699–711.

He, D., Ionides, E. L., and King, A. A. (2010). Plug-and-play inference for disease dynamics: Measles in large and small towns as a case study. *Journal of the Royal Society Interface*, 7:271–283.

Hoeting, J. A., Madigan, D., Raftery, A. E., and Volinsky, C. T. (1999). Bayesian model averaging: a tutorial (with comments by M. Clyde, David Draper and E. I. George, and a rejoinder by the authors. *Statistical Science*, 14(4):382 – 417.

Ionides, E. L., Breto, C., Park, J., Smith, R. A., and King, A. A. (2017). Monte Carlo profile confidence intervals for dynamic systems. *Journal of the Royal Society Interface*, 14:1–10.

Ionides, E. L., Ning, N., and Wheeler, J. (2022). An iterated block particle filter for inference on coupled dynamic systems with shared and unit-specific parameters. *Statistica Sinica,* pre-published online.

King, A. A., Domenech de Cellès, M., Magpantay, F. M., and Rohani, P. (2015). Avoidable errors in the modelling of outbreaks of emerging pathogens, with special reference to ebola. *Proceedings of the Royal Society B: Biological Sciences*, 282(1806):20150347.

King, A. A., Nguyen, D., and Ionides, E. L. (2016). Statistical inference for partially observed Markov processes via the R package pomp. *Journal of Statistical Software*, 69:1–43.

Lee, E. C., Chao, D. L., Lemaitre, J. C., Matrajt, L., Pasetto, D., Perez-Saez, J., Finger, F., Rinaldo, A., Sugimoto, J. D., Halloran, M. E., Longini, I. M., Ternier, R., Vissieres, K., Azman, A. S., Lessler, J., and Ivers, L. C. (2020a). Achieving coordinated national immunity and cholera elimination in Haiti through vaccination: A modelling study. *The Lancet Global Health*, 8(8):e1081–e1089.

Lee, E. C., Chao, D. L., Lemaitre, J. C., Matrajt, L., Pasetto, D., Perez-Saez, J., Finger, F., Rinaldo, A., Sugimoto, J. D., Halloran, M. E., Longini, I. M., Ternier, R., Vissieres, K., Azman, A. S., Lessler, J., and Ivers, L. C. (2020b). Supplement to: Achieving coordinated national immunity and cholera elimination in Haiti through vaccination: A modelling study. *The Lancet Global Health*, 8(8):e1081–e1089.

Li, J., Ionides, E. L., King, A. A., Pascual, M., and Ning, N. (2023). Machine learning for mechanistic models of metapopulation dynamics. *arxiv:2311.06702*.

Pawitan, Y. (2001). *In all likelihood: statistical modelling and inference using likelihood*. Oxford University Press.

Romero-Severson, E., Volz, E., Koopman, J., Leitner, T., and Ionides, E. (2015). Dynamic variation in sexual contact rates in a cohort of HIV-negative gay men. *American Journal of Epidemiology*, 182:255–262.

Saltelli, A., Bammer, G., Bruno, I., Charters, E., Di Fiore, M., Didier, E., Nelson Espeland, W., Kay, J., Lo Piano, S., Mayo, D., Pielke, R., Portaluri, T., Porter, T. M., Puy, A., Rafols, I., Ravetz, J. R., Reinert, E., Sarewitz, D., Stark, P. B., Stirling, A., van der Sluijs, J., and Vineis, P. (2020). Five ways to ensure that models serve society: a manifesto. *Nature*, 582:428–484.

Stocks, T., Britton, T., and Höhle, M. (2020). Model selection and parameter estimation for dynamic epidemic models via iterated filtering: Application to rotavirus in Germany. *Biostatistics*, 21(3):400–416.

Subramanian, R., Romeo-Aznar, V., Ionides, E., Codeço, C. T., and Pascual, M. (2020). Predicting re-emergence times of dengue epidemics at low reproductive numbers: DENV1 in Rio de Janeiro, 1986–1990. *Journal of the Royal Society Interface*, 17(167):20200273.