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

Jesse Wheeler, Anna Rosengart, Zhuoxun Jiang,
Kevin Hao En Tan, Noah Treutle, Edward L. Ionides

Department of Statistics, University of Michigan

December 19, 2022

Supplementary Content

S1 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) [TODO: Cite the short course].

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 $N_{\circ i}$, and a flow out of the system from compartment i is denoted by $N_{i\circ}$. We call $\circ$ a source/sink compartment, though it is an irregular compartment since $X_{circ}(t)$ 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. S1.1), a simple Markov counting system (Sec. S1.2), and an overdispersed Markov counting system (Sec. S1.3). Other representations include stochastic differential equations driven by Gaussian noise or Gamma noise (Bhadra et al., 2011).

S1.1 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.

S1.2 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

$$\mathbb{P}\big[N_{ij}(t+\delta) - N_{ij}(t) = 0 \,|\, X(t)\big] = 1 - \delta\mu_{ij}\big(t, X(t)\big) X_i(t) + o(\delta), \tag{S4}$$

$$\mathbb{P}\big[N_{ij}(t+\delta) - N_{ij}(t) = 1 \,|\, X(t)\big] = \delta\mu_{ij}\big(t, X(t)\big) X_i(t) + o(\delta), \tag{S5}$$

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

$$\mathbb{P}\big[N_{\bullet i}(t+\delta) - N_{\bullet i}(t) = 0 \,|\, X(t)\big] = 1 - \delta\mu_{\bullet i}\big(t, X(t)\big) + o(\delta), \tag{S6}$$

$$\mathbb{P}\big[N_{\bullet i}(t+\delta) - N_{\bullet i}(t) = 1 \,|\, X(t)\big] = \delta\mu_{\bullet i}\big(t, X(t)\big) + o(\delta). \tag{S7}$$

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).

S1.3 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 S2, below. Therefore, the numerical scheme in Sec. S2 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).

S2 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 our 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

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

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.

S3 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).

S3.1 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):

$$Y_n \mid \Delta N_{EI}(n) = z(n) \sim \text{NB}\left(\rho z(n), \psi\right), \tag{S15}$$

where Y_n is the reported cholera cases at time $n \in 1:N$ and $\Delta N_{EI}(n)$ is the sum total of individuals across vaccination compartment z who moved from compartment E_z to I_z since observation $n-1 \in 0:N-1$, and $\text{NB}(\rho z, \psi)$ is the negative binomial distribution with mean ρz and variance $\rho z\big(1+\frac{\rho z}{\psi}\big)$.

S3.2 Model 2

As mentioned in the main text, 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 as the loss function.

$$\log\big(Y_{u,n}+1\big) \mid \Delta N_{EI}(n) = z(n) \sim \text{N}\big(\log\big(\rho z(n)+1\big), \psi^2\big), \tag{S16}$$

where $\Delta N_{EI}(n)$ is the sum total of individuals across vaccination compartment z and unit u who moved from compartment E_{uz} to I_{uz} since observation $n-1 \in 0:N-1$. 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 $\mu = \log\big(\Delta N_{EI}(n)+1\big)$ 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. In the file `choleraEqsPublic.py`, Lee et al. (2020a) create several functions, where each function represents a set of coupled differential equations that could potentially be used to model cholera incidence data. Each function returns a vector (or `numpy` array) that represents the change in each state variable for a single time step, including the variable `dC`, which tracks the number of new infections and is used to obtain the reported case counts. Following their comments and code, it appears that the function `choleraEqs10WithoutVaccinationNetwork` was used to describe the dynamics of the epidemic phase, and `choleraEqs11WithoutVaccinationNetwork` was used for the endemic stage (see Sec S6.2 for more details). Because their models were fit using least squares, the code in these functions suggest that the measurement model for the epidemic phase is

$$Y_{u,n} \mid \Delta N_{EI}(n) = z(n) \sim \text{N}\big(\rho z(n), \psi^2\big), \tag{S17}$$

which is similar to our measurement model, the primary difference being that the measurement model is applied to raw case counts rather than log-transformed case counts. The measurement model for the endemic phase modifies the epidemic model by counting both asymptomatically infected (A) and symptomatically infected (I) individuals in the case counts:

$$Y_{u,n} \mid \Delta N_{EI}(n) = z_1(n), \Delta N_{EA}(n) = z_2(n) \sim \text{N}\big(\rho\big(z_1(n)+z_2(n)\big), \psi^2\big), \tag{S18}$$

where the notation for $\Delta N_{EA}(n)$ is similar to $\Delta N_{EI}(n)$, described above.

S3.3 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:

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

where $\Delta N_{S_{uz}I_{uz}}(n)$ is the number of individuals who moved from compartment S_{uz} to I_{uz} since observation $n-1 \in 0:N-1$.

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.

S4 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 could be known, or could be 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.

S4.1 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 ($I(0)$ and $E(0)$, respectively) using the IF2 algorithm, implemented as `mif2` in the `pomp` package. Finally, the initial proportion of Susceptible individuals S_0 is calculated as $S(0) = 1 - I(0) - E(0)$.

S4.2 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.

S4.3 Model 3

As described in the main test, the latent states are initialized by estimating of the number of infected, asymptomatic, and recovered individuals at time t_0 by using observed case counts at time t_1. The initial value for the bacteria compartment is then determined using the rainfall dynamics of Eq. 23. For the departments of Grande'Anse, Nord-Ouest, and Sud-Est, estimating the latent states at time t_0 using the observed data at time t_1 is problematic (see Sec. S5), so we instead estimate initial value of the latent state $I_{u0}(t_0)$ for the integers $u \in \{3, 7, 9\}$, which correspond to the aforementioned departments. The remaining hidden states are estimated by enforcing model dynamics using this result.

S5 Calibrating Model 3 to observed cases

In this section, we provide more detail on the process that was used to estimate the coefficients used in Model 3. In particular, we discuss why we decided to include additional model parameters that were not considered by Lee et al. (2020a) (parameters related to initial conditions and Hurricane Matthew). 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 does not exist published examples the IBPF algorithm used for data analysis outside of the papers in which the algorithm was introduced (Ning and Ionides, 2021; Ionides et al., 2022), which at least partially motivates the creation of this section.

Lee et al. (2020a) were only able to estimate 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 all available data. Our preliminary results (not shown) suggest that a single search for the MLE using the IBPF algorithm is insufficient; this result is consistent with the findings of Ionides et al. (2022), who used a sequence of successive searches for the MLE in order to properly maximize the model likelihood. We therefore use a similar approach by performing three successive searches for the MLE. The first search is performed by obtaining initial values for the parameters by uniformly sampling values from a predefined hypercube. Subsequent searches use the resulting parameter values that correspond the largest model likelihoods.

We use the technique described above to fit the fully coupled version of Model 3 proposed by Lee et al. (2020a) to all observed cholera data. The maximum likelihood we obtained with the model was -18046, which is higher than the log-ARMA benchmark (-18062). For meta-population models, it is worth considering how well the model fits the data to each spatial unit. This can be done by looking at conditional log-likelihoods, which is part of the output of the `bpfilter` algorithm in the `spatPomp` package. The likelihoods for each department, compared to the corresponding log-ARMA benchmark, are displayed in Fig. S-1. The figure demonstrates that while the log-likelihood of the fitted model outperforms the log-ARMA benchmark, the model has lower likelihoods than the benchmark for some departments. This result warrants further investigation of the model fit.

In an attempt to determine why the model is underperforming in some departments, we plot the conditional log likelihoods of each observation given the previous observations for each department in Fig. S-2. This figure reveals that the fitted model has difficult time explaining 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

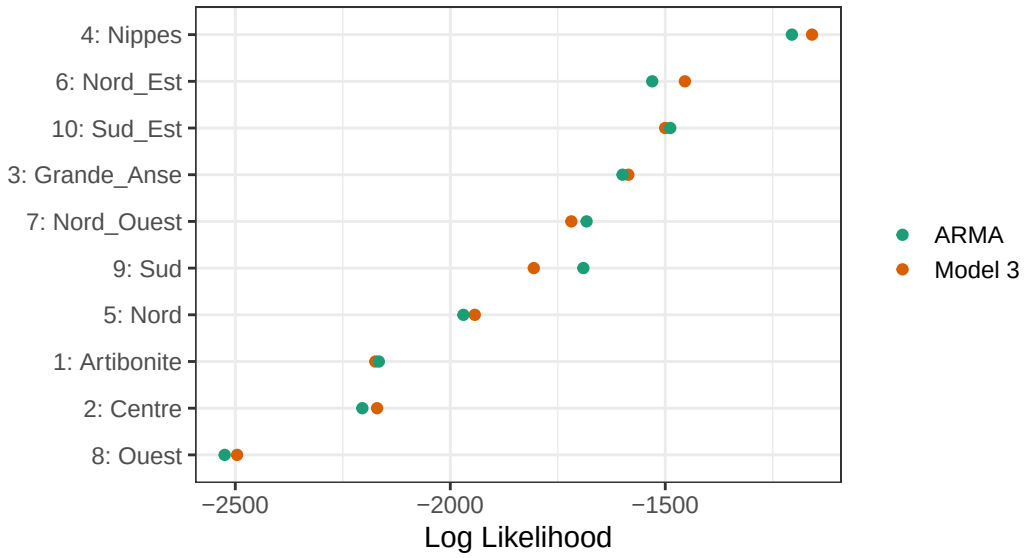

Figure S-1: Log-likelihoods of Model 3 for each department compared to the corresponding log-ARMA benchmark prior to the inclusion of parameters for initial values and Hurricane Matthew.

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 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}.

Other points of concern include the observations at or near the start of the recorded cases, for example, in Nord-Ouest and Sud-Est. These low conditional log-likelihoods are a result of having zero or near zero confirmed cases in the first week of observed data, followed by a rapid growth of cases. The model cannot explain the rapid growth of cholera in the absence of starting cases because the initialization scheme forces departments with no recorded cases in the first week to have hidden states $I_{u0}(0) = A_{u0}(0) = B_u(0) = 0$. In reality, however, it is possible to have some infected or asymptomatic individuals in a department when no cases are recorded due to under-reporting. This may be particularly true at the onset of an infectious disease outbreak, as lower public awareness and inadequate testing infrastructure may result in many cases going undetected. We therefore include and estimate I_{0u} that determines the initial number of Infected individuals at the start of the pandemic for the departments Grande'Anse, Nord-Ouest, and Sud-Est; each of these departments have both low starting case counts followed by rapid growth, and low initial values of the conditinoal log-likelihood.

We refit Model 3 after introducing these seven parameters—three initial value parameters and four parameters related to Hurricane Matthew—using the same successive search scheme described above. The resulting model has a log-likelihood value of −17850.4. We also provide figures for

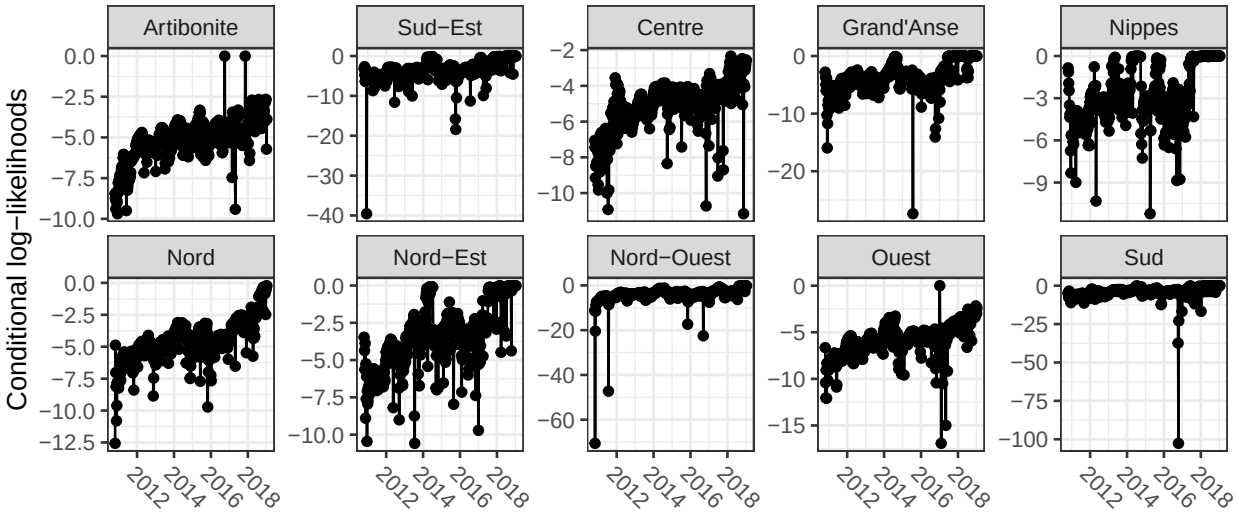

Figure S-2: Conditional log-likelihoods of Model 3 prior to the inclusion of parameters for initial values and Hurricane Matthew.

department specific log-likelihoods and conditional log-likelihoods in Figs. S-3 and S-4.

While the addition of the Hurricane parameters seems to have resulted in a dramatic increase of conditional likelihoods around and after October 2016, the inclusion of the initial value parameters does not appear to have fixed the low conditional log-likelihood values in Sud-Est and Nord-Ouest. This could be due to many factors, including the possibility that the coupling between units forces low initial value parameters in a given department in order for the model to fit the data in the remaining departments. Whatever the case may be, the inclusion of the seven parameters resulted in a 195.2 increase in log-likelihood, which is a highly statistically significant difference. One could similarly test whether the inclusion of the initial value parameters resulted in a significantly better fit to the observed data, but we do not do this here due to the large computational effort that would be required.

S5.1 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, as this compartment drives transmission dynamics in SIR models for infectious disease outbreaks.

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-5 shows the percentage of individuals that are in the susceptible compartment from various simulations

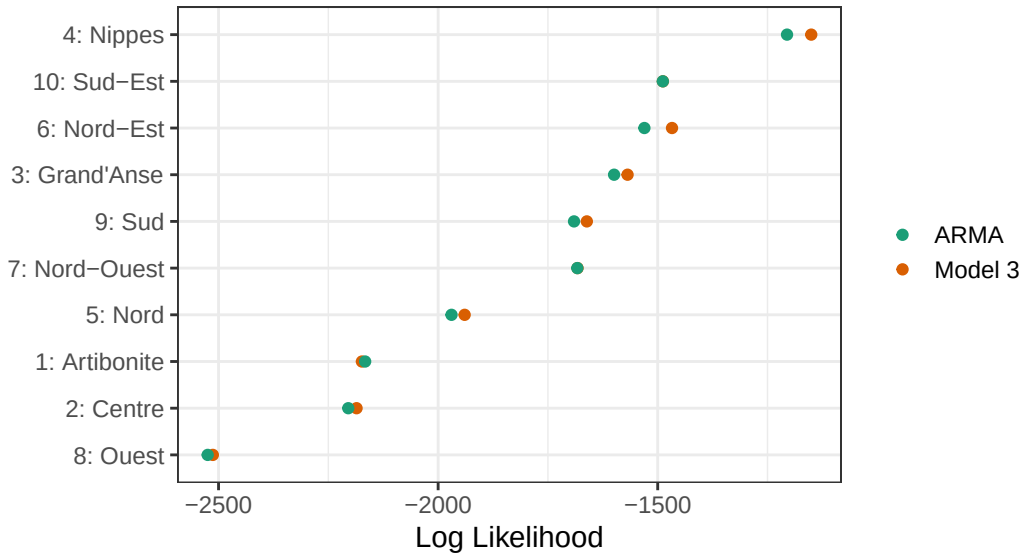

Figure S-3: Log-likelihoods of Model 3 for each department compared to the corresponding log-ARMA benchmark after adding and estimating parameters related to initial values and Hurricane Matthew.

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 also appears true in later stages of the time period of observed data, where the data suggests that there is a replenishment of the susceptible compartment, but the Model simulations typically retain a smaller proportion of susceptible individuals. This result demonstrates that the calibrated model favors a more rapid growth in cholera cases than what is typically seen in the observed data, which explains why the model fails to predict the absence of cases between February, 2019 and October, 2022.

S6 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 1. 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 reproducibility.

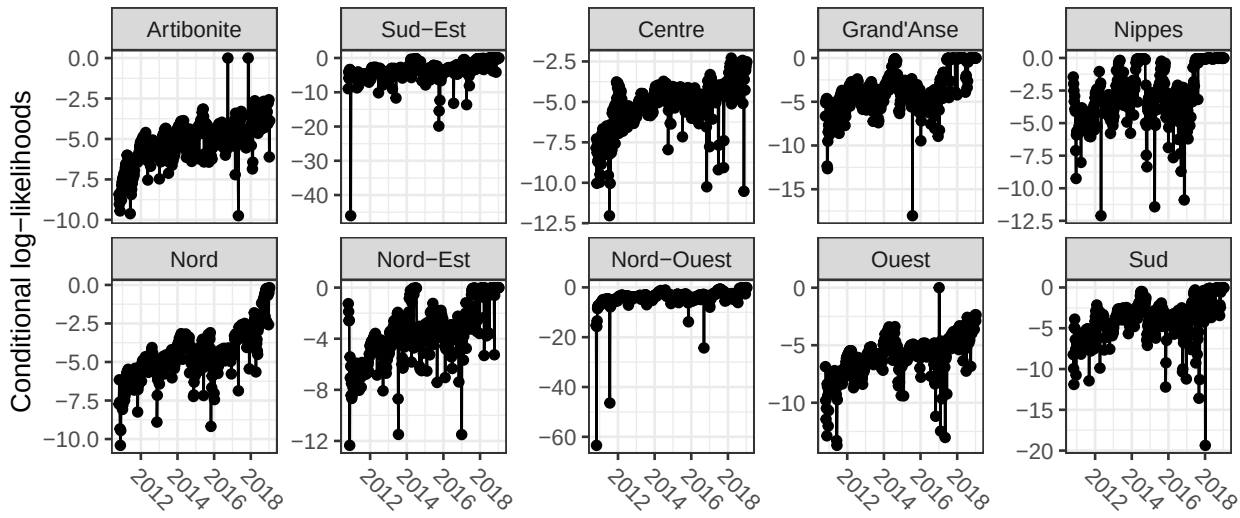

Figure S-4: Conditional log-likelihoods of Model 3 after adding and estimating the parameters for initial values and Hurricane Matthew.

S6.1 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 (as mentioned in our 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 were several hundred log-likelihood units lower than other parameter sets. Such a large difference in log-likelihoods is well beyond the threshold of statistical uncertainty determined by Wilks' theorem.

To fully reproduce the results of the Model 1 authors, it is necessary to use the exact same set of model parameters. Because the parameters used to obtain the result given by Lee et al. (2020a) are not 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. Still, running the publicly available source code resulted in a set of parameters that are visually similar to those used by the Model 1 authors (See Figures S-6 and S-7). 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-8).

Because the model forecasts provided by Lee et al. (2020a) come from various sets of parameters—

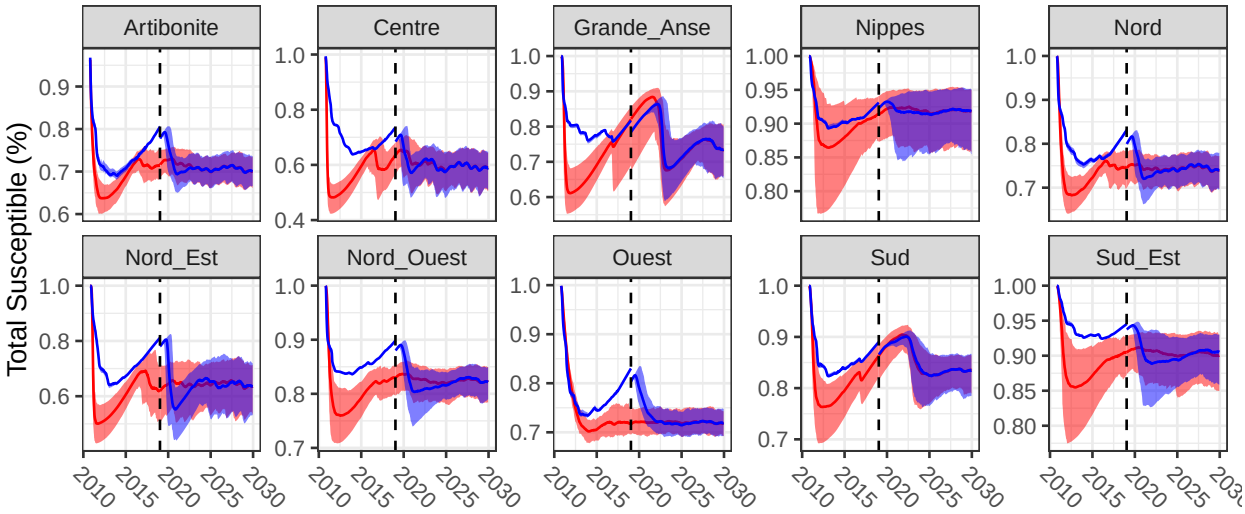

Figure S-5: Percentage of individuals that are in the susceptible compartment. Simulations from Model 3 under initial conditions are displayed in red; simulations from the filtering distribution of model are displayed in blue. The dashed line represents the end of the observed data.

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 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.

S6.2 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., 2021), which facilitates the creation of meta-population models. We note that the travel and water matrices

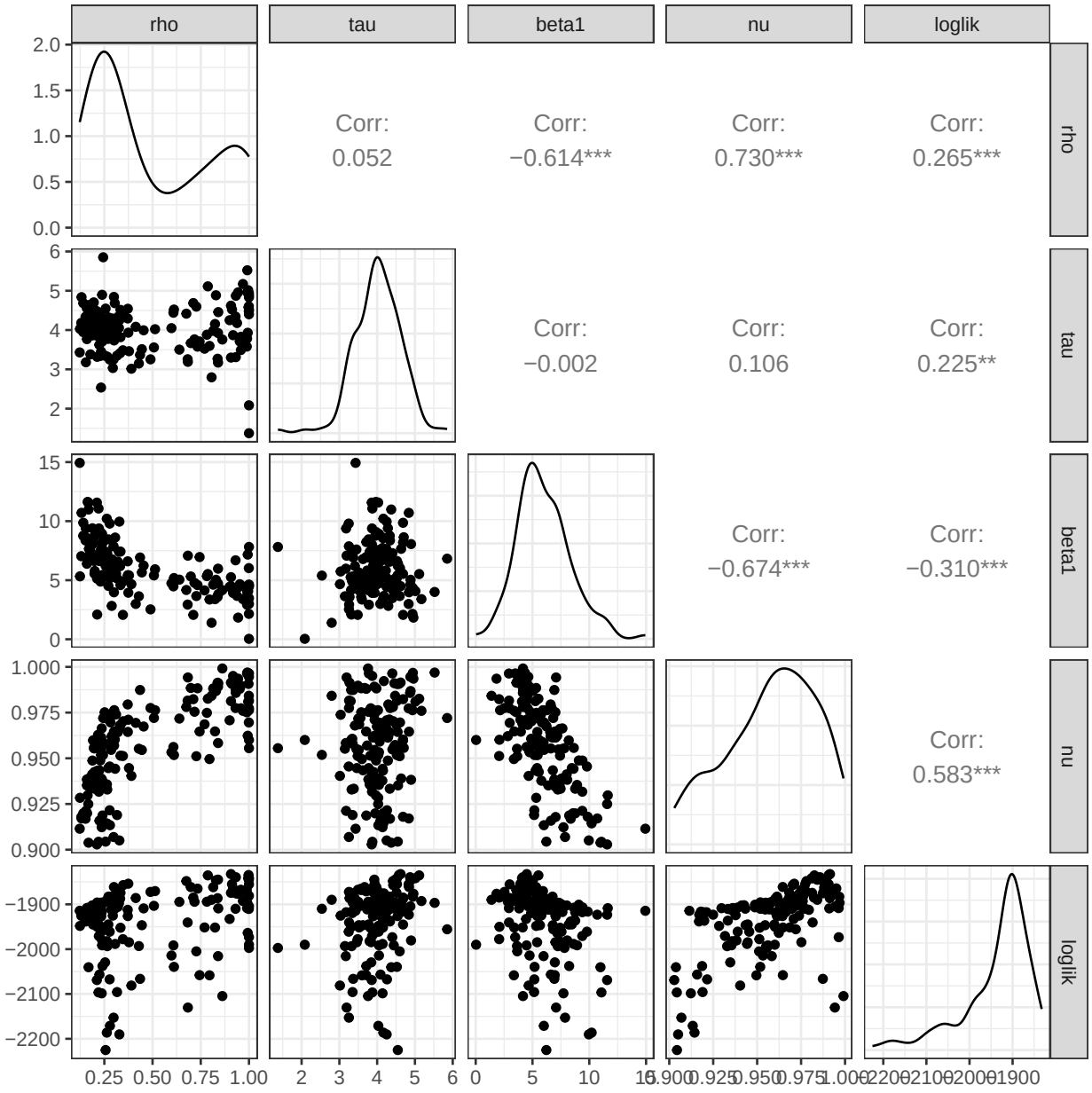

Figure S-6: Compare to Figure S8 of Lee et al. (2020) supplement.

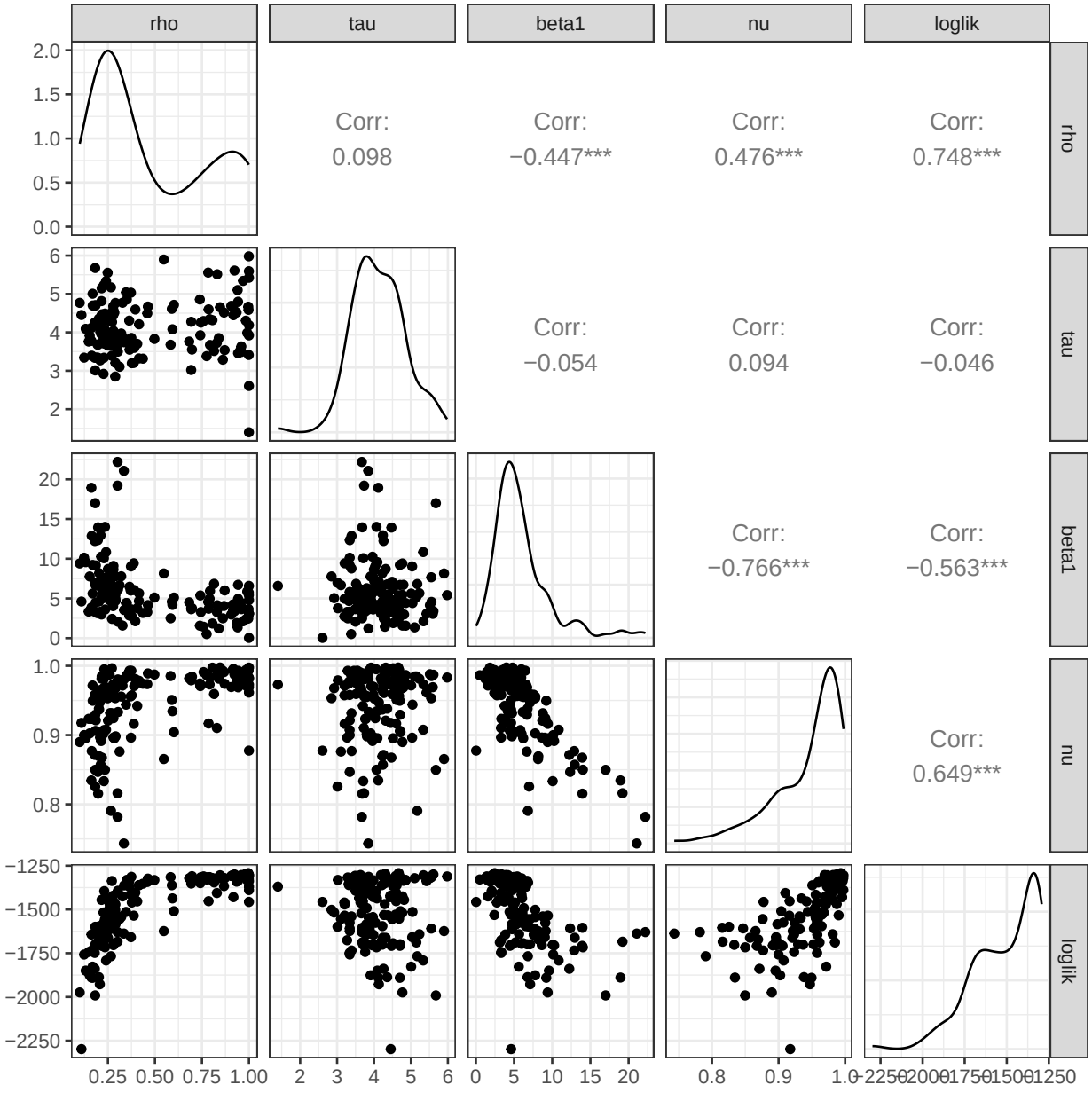

Figure S-7: Bivariate relationships between variables after fitting endemic period. Compare to S9 of Lee et al. (2020) supplement.

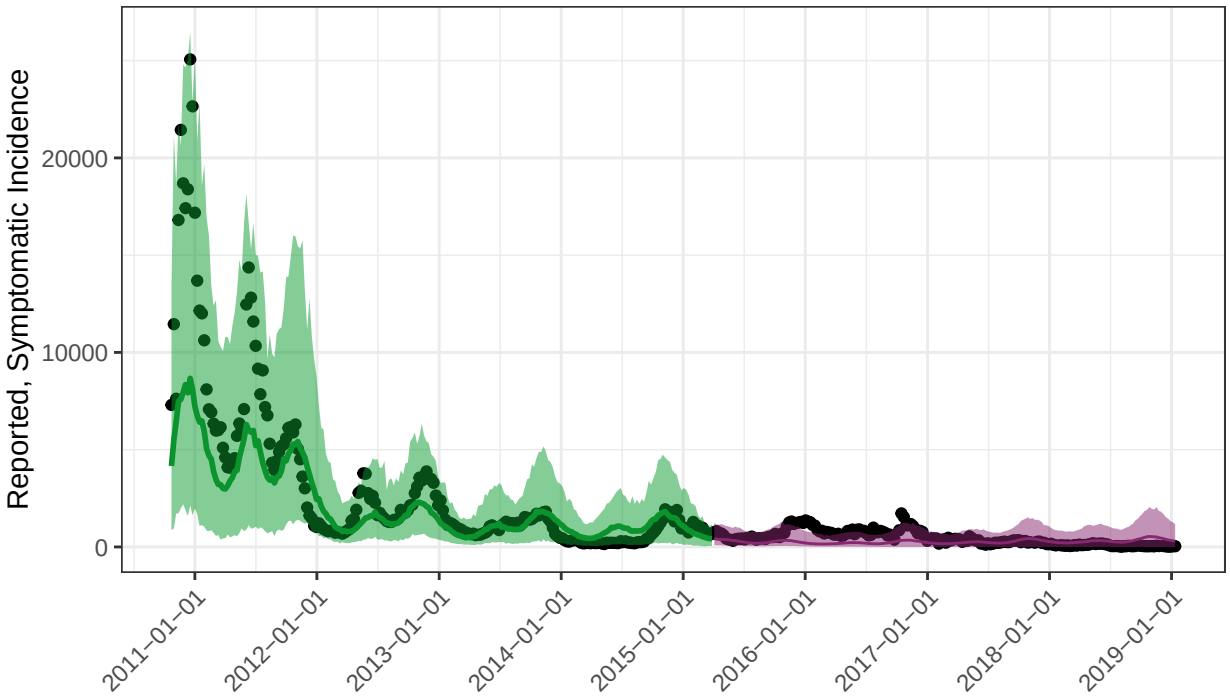

Figure S-8: Simulations using parameter sets that were generated by running source code provided by Lee et al (2020). Compare to Figure S7 of Lee et al. (2020b). The upper bound for the likelihood of this model is -3031.

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-9, are very similar to the trajectories shown in Figure S15 of Lee et al. (2020b).

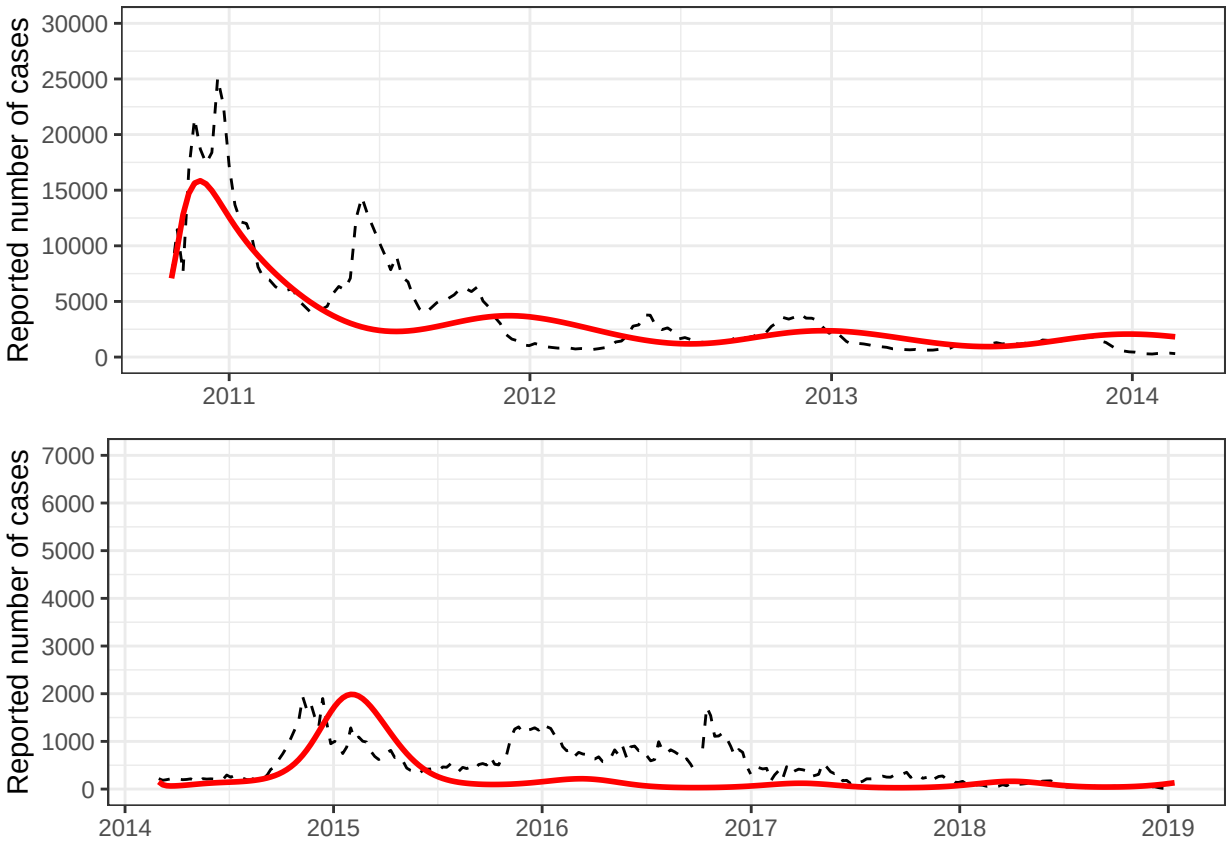

Figure S-9: Model 2 trajectories using the `haitipkg`. Compare to Figure S15 of Lee et al. (2020b).

There are minor differences between Figure S-9 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 be attributed to implementing the model in two different programing 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 (See Table S-1) 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 had values that were far more precise. 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 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 described by Eq. S17 for both phases of the epidemic, as this seemed to visually match the results of Lee et al. (2020a) most closely.

S6.3 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 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-10, we plot simulations from this model. This figure should 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.

S7 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

[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`.

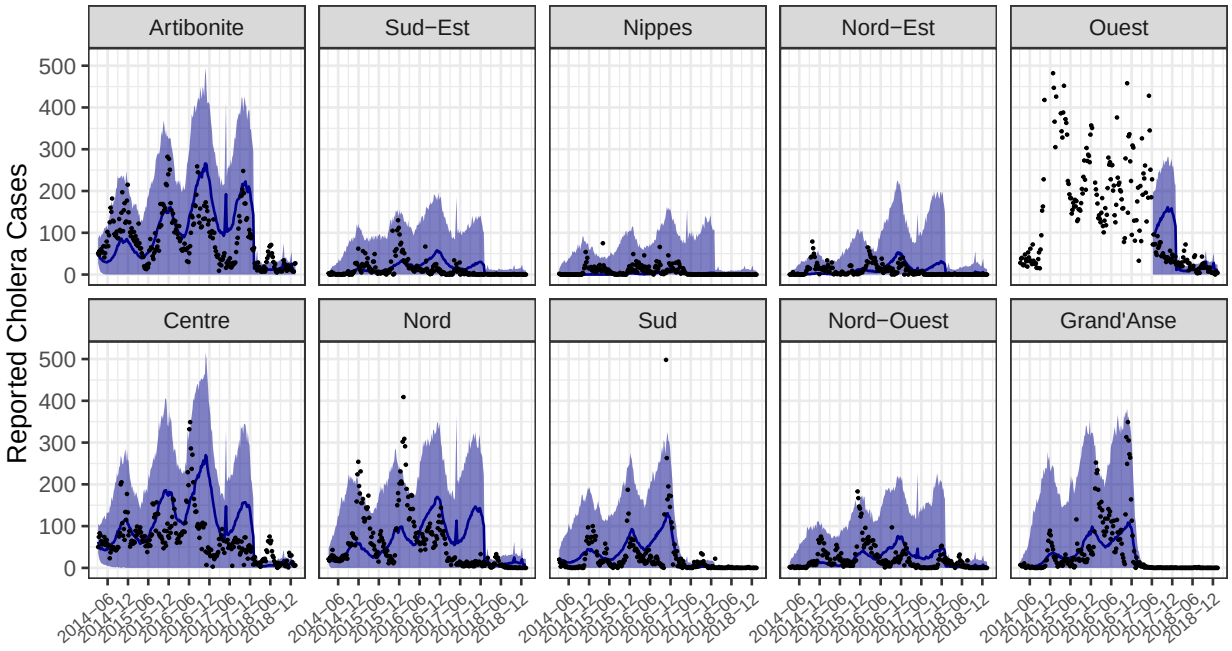

Figure S-10: Simulations from Model 3. Compare to Figure S18 of Lee et al. (2020a).

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 $i \in 1:k$, let Θ_i be a random vector of model parameters. Then, letting Θ denote the true model parameters, we endow the set $\{\Theta_1, \Theta_2, \ldots, \Theta_k\}$ with a discrete uniform distribution, such that $P\big(\Theta = \Theta_i\big) = \frac{1}{K}$ for all values $i \in 1:k$. 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. Instead, we choose Θ_i to be the output of a stochastic routine applied to the observed data by setting Θ_i 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 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 in in the discrete prior distribution.

S8 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-1 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	κ	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-1: Translations between our common notation and notation used by Lee et al. (2020a)

Supplementary References

Asfaw, K., Ionides, E. L., and King, A. A. (2021). `spatPomp`: R package for statistical inference for spatiotemporal partially observed Markov processes. `https://github.com/kidusasfaw/spatPomp`.

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.

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., Ning, N., and Wheeler, J. (2022). An iterated block particle filter for inference on coupled dynamic systems with shared and unit-specific parameters. arXiv:2206.03837.

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.

Ning, N. and Ionides, E. L. (2021). Iterated block particle filter for high-dimensional parameter learning: Beating the curse of dimensionality.

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.