INFORMING POLICY VIA DYNAMIC MODELS: ELIMINATING CHOLERA IN HAITI

BY JESSE WHEELER[1,a] , ANNAELAINE ROSENGART[1,b] ZHUOXUN JIANG[1,c],
KEVIN HAO EN TAN[1,d], NOAH TREUTLE[1,e] AND EDWARD IONIDES[1,f]

[1]*STATISTICS DEPARTMENT, UNIVERSITY OF MICHIGAN,* [a]*JESWHEEL@UMICH.EDU;* [b]*AELR@UMICH.EDU;*
[c]*ZHUOXUNJ@UMICH.EDU;* [d]*KEVTAN@UMICH.EDU;* [e]*NTREUTLE@UMICH.EDU;* [f]*IONIDES@UMICH.EDU*

Public health decisions must be made about when and how to
implement interventions to control an infectious disease epidemic.
These decisions should be informed by data on the epidemic as well
as current understanding about the transmission dynamics. Such de-
cisions can be posed as statistical questions about scientifically moti-
vated dynamic models. [Thus, we encounter the methodological task
of building credible, data-informed decisions for stochastic, partially
observed, nonlinear dynamic models. This necessitates addressing the
tradeoff between biological fidelity and model simplicity, and the re-
ality of misspecification for models at all levels of complexity. THE
CURRENT MS VERSION DOES NOT QUITE AIM AT THESE
ISSUES. BUT, THEY ARE HIGHLY RELEVANT , AND GIVE
A BROADER STATISTICAL CONTEXT FOR THE EPIDEMIO-
LOGICAL CASE STUDY. MAYBE WE CAN EDIT IN THAT DI-
RECTION?] As a case study, we consider a cholera epidemic in Haiti.
The 2010 introduction of cholera to Haiti led to an extensive outbreak
and sustained transmission, eliminated in 2019 with the help of vac-
cination and other public health measures. We study three models
developed by expert teams to advise on vaccination policies. We as-
sess methods used for fitting and evaluating these models, leading
to recommendations for future studies. Better statistical fit facili-
tates interpretation of the fitted model, but caution is nevertheless
required in drawing policy conclusions based on causal interpreta-
tions of the models.

1. Introduction. Quantitative models for dynamic systems offer potential for de-
signing effective control measures. Regulation of biological populations is a fundamental
topic in epidemiology, ecology, fisheries and agriculture. Quantitative models for these
population dynamics may be nonlinear and stochastic, with the resulting complexities
compounded by incomplete understanding of the underlying biological mechanisms and
by partial observability of the system variables. Developing and testing such models,
and assessing their fitness for guiding policy, is a challenging statistical task. Questions
of interest include: What indications should we look for in the data to assess whether
the model-based inferences are trustworthy? What diagnostic tests and model varia-
tions can and should be considered in the course of the data analysis? What are the
possible trade-offs of increasing model complexity, such as the inclusion of interactions
across spatial units?

This case study investigates the use of dynamic models and spatiotemporal data
to inform a policy decision in the context of the cholera outbreak in Haiti, which
started in 2010. We build on a multi-group modeling exercise by Lee et al. (2020a)
in which four expert modeling teams developed models to the same dataset with the

---

goal of comparing conclusions on the feasibility of eliminating cholera by a vaccination campaign. Model 1 is stochastic and describes cholera at the national level; Model 2 is deterministic with spatial structure, and includes transmission via contaminated water; Model 3 is stochastic with spatial structure, and accounts for measured rainfall. Model 4 has an agent-based construction, featuring considerable mechanistic detail but limited ability to calibrate these details to data. The strengths and weaknesses of the agent-based modeling approach (Tracy, Cerdá and Keyes, 2018) are outside the scope of this article, and we focus on Models 1–3.

The four independent teams were given the task of projecting the potential effect of various vaccination deployment strategies on the cholera epidemic in Haiti. [TODO: MAYBE DISCUSS MORE ABOUT THE VACCINES, LIKE WHAT EFFECTS THEY ARE KNOWN TO HAVE, NOTE HOW VACCINE SUPPLIES ARE LIMITED, ETC.] Certain data were shared between the groups, including demography and vaccination history; vaccine efficacy was also fixed at a shared value between groups. Beyond this, the groups made autonomous decisions on what to include and exclude from their models; this autonomy reduced the possible effect that assumptions about the dynamic system may have on the final conclusion of the study. Despite this autonomy, and largely adhering to existing guidelines on creating models to inform policy (Behrend et al., 2020; Saltelli et al., 2020), the consensus across the four models was that an extensive nationwide vaccination campaign would be necessary to eliminate cholera from Haiti. This conclusion is inconsistent with the fact that there have been no confirmed cases since February, 2019 without additional vaccination efforts (Ferguson, 2022). The failure of Lee et al. (2020a) to correctly predict the elimination of cholera has been the topic of some discussion (Francois, 2020; Rebaudet, Gaudart and Piarroux, 2020; Henrys et al., 2020; Lee et al., 2020b).

Here, we demonstrate that attention to certain statistical considerations could have improved the fit to previous data, as well as leading to forecasts that match more closely the observed outcome. Based on this retrospective analysis, we offer suggestions on fitting mechanistic models to dynamic systems for future studies. We proceed by introducing Models 1–3 in Sec. 2; in Sec. 3, we present a methodological approach to examining and refining these models. In Sec. 4, we use improved model fits to project cholera incidence in Haiti under various vaccination campaigns. We then conclude with a discussion on the use of mechanistic models to inform policy decisions in Sec. 5.

2. Models. We refer to models that focus on learning relationships between variables in a dataset as associative, and we refer to models that incorporate a known scientific property of the system as causal or mechanistic. The danger in using forecasting techniques which rely on associative models to predict the consequence of interventions is called the Lucas critique in an econometric context. Lucas et al. (1976) pointed out that it is naive to predict the effects of an intervention on a given system based entirely on historical associations. To successfully quantify the effect of an intervention, a model should therefore both provide a quantitative explanation of existing data and should additionally have a causal interpretation: a manipulation of the system should correspond quantitatively with the corresponding change to the model. We therefore focus our analysis on mechanistic statistical models, which provides a statistical fit to the available data while also supporting a causal intepretation.

The deliberate limitation of coordination between the groups of Lee et al. (2020a) allows us to treat the models as fairly independently developed expert approaches to understanding cholera transmission. However, it results in some differences in notation, and choices of data subsets for analysis, that hinder direct comparison. Here, we have
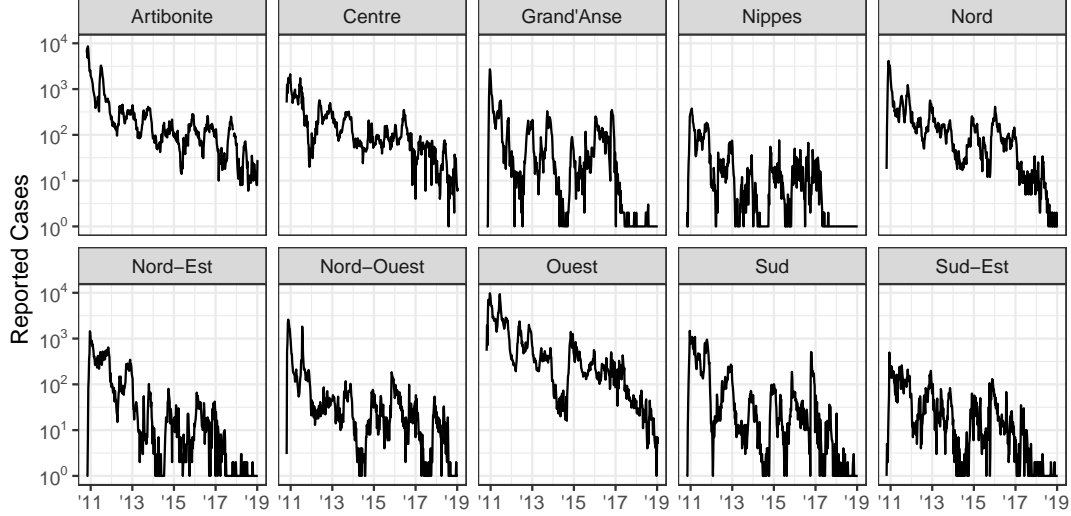
Fig 1. Reported Cholera cases in the outbreak in Haiti from 2010-2019.

put all three models into a common notational framework. Translations back to the original notation of Lee et al. (2020a) are given in Table S-1.

Each model describes the cholera dynamics using a latent state vector $\boldsymbol{X}^{(m)}(t)$ and an observable response vector $\boldsymbol{Y}^{(m)}(t)$ for each continuous time-point $t \in \mathcal{T}$, where $\mathcal{T}$ is the set of all time-points and $m \in \{1, 2, 3\}$ indexes the model. The latent state vector $\boldsymbol{X}^{(m)}$ consists of individuals labeled as susceptible (S), infected (I), asymptomatically infected (A), vaccinated (V), and recovered (R) for $m \in \{1, 2, 3\}$, with various sub-divisions sometimes considered across the three models under consideration. Models 2 and 3 have metapopulation structure, meaning that each individual is a member of a spatial unit, denoted by $u \in 1{:}U$, which corresponds here to membership of the $U = 10$ administrative departments.

In this study, we focus our attention on important statistical, rather than scientific, details that may improve inference on a dynamic system; we therefore refrain from commenting on the assumptions in each model or proposing our own alternative description of cholera dynamics, and instead rely on the scientific expertise of Lee et al. (2020a).

2.1. *Model 1.* $\boldsymbol{X}^{(1)}(t) = (S_z(t), E_z(t), I_z(t), A_z(t), R_z(t), z \in 0 : Z)$ describes suscep-tible, latent (exposed), infected (and symptomatic), asymptomatic, and recovered in-dividuals in vaccine cohort z. Here, $z = 0$ corresponds to unvaccinated individuals, and $z \in 1{:}Z$ describes hypothetical vaccination programs.

The force of infection is

$$(1) \qquad \lambda(t) = \Big(\sum_{z=0}^{Z} I_z(t) + \epsilon \sum_{z=0}^{Z} A_z(t) \Big)^{\nu} \frac{d\Gamma(t)}{dt} \beta(t)/N$$

where $\beta(t)$ is a periodic cubic spline representation of seasonality, given in terms of a B-spline basis $\{s_j(t), j \in 1{:}6\}$ and parameters $\beta_{1:6}$ as

$$(2) \qquad \log \beta(t) = \sum_{j=1}^{6} \beta_j s_j(t).$$

The process noise $d\Gamma(t)/dt$ is multiplicative Gamma-distributed white noise, with infinitesimal variance parameter σ_{proc}^2, which was held constant at $\sigma_{\text{proc}}^2 = 0$ in Lee et al. (2020a). The inclusion of gamma white noise in the process model permits the construction of an over-dispersed Markov chain (Bretó and Ionides, 2011), and has previously been found to increase a model's ability to fit data (Stocks, Britton and Höhle, 2020).

Per-capita transition rates are given in Equations 3-10:

$$\text{(3)} \qquad \mu_{S_z E_z} = \lambda(t)$$

$$\text{(4)} \qquad \mu_{E_z I_z} = \mu_{EI}(1 - f_z(t))$$

$$\text{(5)} \qquad \mu_{E_z A_z} = \mu_{EI} \, f_z(t)$$

$$\text{(6)} \qquad \mu_{I_z R_z} = \mu_{A_z R_z} = \mu_{IR}$$

$$\text{(7)} \qquad \mu_{R_z S_z} = \mu_{RS}$$

$$\text{(8)} \qquad \mu_{S_0 S_z} = \mu_{E_0 E_z} = \mu_{I_0 I_z} = \mu_{A_0 A_z} = \mu_{R_0 R_z} = \eta_z(t)$$

$$\text{(9)} \qquad \mu_{S_z D} = \mu_{E_z D} = \mu_{I_z D} = \mu_{A_z D} = \mu_{R_z D} = \delta$$

$$\text{(10)} \qquad \mu_{D S_0} = \mu_S$$

where $z \in 0\!:\!Z$. Here, μ_{AB} is a transition rate from compartment A to B. We have an additional demographic source and sink compartment D modeling entry into the study population due to birth or immigration, and exit from the study population due to death or immigration. Thus, μ_{AD} is a rate of exiting the study population from compartment A and μ_{DB} is a rate of entering the study population into compartment B.

In this model, the advantage afforded to vaccinated individuals is an increased probability that an infection is asymptomatic. Conditional on infection status, vaccinated individuals are also less infectious than their non-vaccinated counterparts by a rate of $\epsilon = 0.05$ in Eq. (1). In (5) and (4) the asymptomatic ratio for non-vaccinated individuals is set $f_0(t) = 0$, so that the asymptomatic route is reserved for vaccinated individuals. For $z \in 1\!:\!Z$, the vaccination cohort z is assigned a time τ_z, and we take $f_z(t) = c\,\theta^*(t - \tau_z)$ where $\theta^*(t)$ is efficacy at time t since vaccination for adults, taken from Lee et al. (2020a), Table S4, and $c = \left(1 - (1 - 0.4688) \times 0.11\right)$ is a correction to allow for reduced efficacy in the 11% of the population aged under 5 years. Single and double vaccine doses were modeled by changing the waning of protection; protection was assumed to be equal between single and double dose until 52 weeks after vaccination, at which point the single dose becomes ineffective.

2.2. *Model 2.* Susceptible individuals are in compartments $S_{uz}(t)$, where $u \in 1\!:\!U$ corresponds to the $U = 10$ departments, and $z \in 0\!:\!4$ describes vaccination status:

$z = 0$: Unvaccinated or waned vaccination protection.
$z = 1$: One dose at age under five years.
$z = 2$: Two doses at age under five years.
$z = 3$: One dose at age over five years.
$z = 4$: Two doses at age over five years.

Individuals can progress to a latent infection E_{uz} followed by symptomatic infection I_{uz} with recovery to R_{uz} or asymptomatic infection A_{uz} with recovery to R_{uz}^A. The force of infection depends on both direct transmission and an aquatic reservoir, $W_u(t)$, and is given by

$$\text{(11)} \qquad \lambda_u(t) = 0.5\left(1 + a\cos(2\pi t)\right)\frac{\beta_W W_u(t)}{W_{\text{sat}} + W_u(t)} + \beta\left\{\sum_{z=0}^{4} I_{uz}(t) + \epsilon \sum_{z=0}^{4} A_{uz}(t)\right\}$$

The latent state is therefore described by the vector $\boldsymbol{X}^{(2)}(t) = \big(S_{uz}(t), E_{uz}(t), I_{uz}(t),$ $A_{uz}(t), R_{uz}(t), R_{uz}^A(t), W_u, u \in 1{:}U, z \in 0{:}4\big)$. A cosine term is included in Eq. (11) in order to account for annual seasonality. Here we note the addition of a phase parameter ϕ to allow for a possible shift in the seasonality component in the force of infection; the original implementation of Model 2 in Lee et al. (2020a) sets $\phi = 0$, not allowing for such a shift.

Individuals move from department u to v at rate T_{uv}, and aquatic cholera moves at rate T_{uv}^W. The nonzero transition rates are

$$(12) \qquad \mu_{S_{uz}E_{uz}} = \theta_z \, \lambda$$

$$(13) \qquad \mu_{E_{uz}I_{uz}} = f\mu_{EI}, \quad \mu_{E_{uz}A_{uz}} = (1-f)\mu_{EI}$$

$$(14) \qquad \mu_{I_{uz}R_{uz}} = \mu_{A_{uz}R_{uz}^A} = \mu_{IR}$$

$$(15) \qquad \mu_{R_{uz}S_{uz}} = \mu_{R_{uz}^A S_{uz}} = \mu_{RS}$$

$$(16) \qquad \mu_{S_{uz}S_{vz}} = \mu_{E_{uz}E_{vz}} = \mu_{I_{uz}I_{vz}} = \mu_{A_{uz}A_{vz}} = \mu_{R_{uz}R_{vz}} = \mu_{R_{uz}^A R_{vz}^A} = T_{uv}$$

$$(17) \qquad \mu_{S_{u1}S_{u0}} = \mu_{S_{u3}S_{u0}} = \omega_1$$

$$(18) \qquad \mu_{S_{u2}S_{u0}} = \mu_{S_{u4}S_{u0}} = \omega_2$$

$$(19) \qquad \mu_{DW_u} = \mu_W \left\{ \sum_{z=0}^{4} I_{uz}(t) + \epsilon_W \sum_{z=0}^{4} A_{uz}(t) \right\}$$

$$(20) \qquad \mu_{W_u D} = \delta_W$$

$$(21) \qquad \mu_{W_u W_v} = w_r T_{uv}^W$$

In (16), we suppose a gravity model:

$$(22) \qquad T_{uv} = v_{\text{rate}} \times \frac{\text{Pop}_u \text{Pop}_v}{D_{uv}^2},$$

where Pop_u is the mean population for department u, D_{uv} is a distance measure estimating average road distance between rancomly chosen members of each population, and $v_{\text{rate}} = 10^{-12}$ was treated as a fixed constant. In (21), T_{uv}^W is a measure of river flow between departments. The unit of $W_u(t)$ is cells per ml, with dose response modeled via a saturation constant of W_{sat} in (11).

2.3. *Model 3.* The latent state is described as $\boldsymbol{X}^{(3)}(t) = \big(S_{uz}(t), I_{uz}(t), A_{uz}(t),$ $R_{uzk}(t), W_u(t), u \in 0{:}U, z \in 0{:}4, k \in 1{:}3\big)$. Here, $z = 0$ corresponds to unvaccinated, $z = 2j-1$ corresponds to a single dose on the jth vaccination campaign in unit u and $z = 2j$ corresponds to receiving two doses on the jth vaccination campaign. $k \in 1{:}3$ models non-exponential duration in the recovered class before waning of immunity. The force of infection is

$$(23) \qquad \lambda_u(t) = \beta_{W_u} \frac{W_u(t)}{W_{\text{sat}} + 2W_u(t)} + \beta_u \sum_{v \neq u} \big(I_{v0}(t) + \epsilon A_{v0}(t)\big)$$

$$(24) \qquad \mu_{S_{uz}I_{uz}} = f \, \lambda_u \big(1 - \eta_{uz}(t)\big) \, d\Gamma/dt$$

$$(25) \qquad \mu_{S_{uz}A_{uz}} = (1-f) \, \lambda_u \big(1 - \eta_{uz}(t)\big) \, d\Gamma/dt$$

$$(26) \qquad \mu_{I_{uz}R_{uz1}} = \mu_{A_{uz}R_{uz1}} = \mu_{IR}$$

$$(27) \qquad \mu_{I_{uz}S_{u0}} = \delta + \delta_C$$

$$(28) \qquad \mu_{A_{uz}S_{u0}} = \delta$$

$$(29) \qquad \mu_{R_{uz1}R_{uz2}} = \mu_{R_{uz2}R_{uz3}} = 3\mu_{RS}$$

$$(30) \qquad \mu_{R_{uzk}S_{u0}} = \delta + 3\mu_{RS} \, \mathbf{1}_{\{k=3\}}$$

$$(31) \qquad \mu_{DW_u} = \left[1 + a\big(J(t)\big)^r\right] D_i \, \mu_W \left[I_{u0}(t) + \epsilon_W A_{u0}(t)\right]$$

$$(32) \qquad \mu_{W_u D} = \delta_W$$

As with Model 1, $d\Gamma_u(t)/dt$ is multiplicative Gamma-distributed white noise in (24) and (25). In (31), $J_u(t)$ is a unit-less measurement of precipitation that has been standardized by dividing the observed rainfall at time t by the maximum recorded rainfall in departement u during the epidemic, and D_u is the average population density. Demographic stochasticity is accounted for by modeling non-cholera related death rate δ in each compartment, along with an additional death rate δ_C in (27) to account for cholera induced deaths among infected individuals. We note that all deaths are balanced by births into the susceptible compartment in (28) and (30), thereby maintaining constant population in each department.

3. Methods.

3.1. *Model Fitting.* Many dynamic systems of scientific and public interest are highly complex by nature; therefore when developing a mathematical description of a dynamic system, it is often necessary to impose assumptions on the system in order to make inference possible. One common assumption in epidemiology is a homogeneous mixing population [REF???]. Other common considerations include whether the proposed model should be stochastic or deterministic; whether the model should have change points in parameter values or should otherwise make adjustments for changes in the dynamic system; and whether the proposed model should include any spatial heterogeneity at a scale permissible by the observed data. Ultimately, these decisions will result in a different approach to fitting the model.

The models considered in this study represent a diverse selection of possible modelling decisions. Each of the models independently describe the cholera dynamics as a Partially Observed Markov Process (POMP), since each model is structured in such a way that, given the partially-observable state of the system at time n, the future state of the system is independent of its past. Because of the spatial structure of Models 2 and 3, we refer to these models as SpatPOMPs (Asfaw et al., 2021).

In the following subsections, we describe approaches to fitting mechanistic models and discuss potential trade-offs of the selected model complexity. We begin by fitting the parameters of Model 2, as the deterministic data generating process makes the task of parameter estimation the easiest of the three considered models.

3.1.1. *Model 2.* Model 2 is a deterministic model, which can be viewed as a special case of a POMP (or a and SpatPOMP) with no randomness in the dynamic model. By combining a deterministic process model with a simple Gaussian measurement model, Model 2 reduces model fitting to a least squares calculation over parameters in a set of differential equations. Deterministic compartment models have a long history in the field of infectious disease epidemiology (Kermack and McKendrick, 1927; Brauer, 2017; Varghese et al., 2021), and can be justified by asymptotic considerations in a large-population limit (Dadlani et al., 2020; Ndii and Supriatna, 2017).

Lee et al. (2020a) originally fit two versions of model 2 based on a presupposed change in cholera transmission from a epidemic phase to endemic phase that occurred in March, 2014. We follow their decision to re-estimate model parameters at this break-point, but we include a requirement that the latent state $\boldsymbol{X}^{(2)}(t)$ at the start of the endemic period must be the same as the state at the end of the epidemic period. This additional constraint is sensible for a mechanistic interpretation of the latent state,

though it adds difficulty to the task of obtaining a model fit that closely resembles the observed data. To combat this added difficulty, and to allow for a possible shift in the seasonality component in the force of infection, we introduce an additional phase parameter ϕ in Eq. 11.

We further increase model flexibility by fitting the parameter μ_W rather than treating it as fixed. We implemented this model using the `spatPomp R` package (Asfaw, Ionides and King, 2021). The model was then fit using the subplex algorithm, implemented in the `subplex` package (King and Rowan, 2020). A comparison of the trajectory of the fitted model to the data is given in Fig 8.

3.1.2. *Model 1.* Another approach to modelling a dynamic system is through probabilistic models. With a probabilistic model, we suppose the existence of a joint density $f_{\boldsymbol{X}_{0:N}^{(m)}, \boldsymbol{Y}_{1:N}^{(m)}}$, with $\boldsymbol{X}_{0:N}^{(m)}$ denoting the unobservable Markov process of model m at times $0\!:\!N = \{0, 1, \ldots, N\}$, and $\boldsymbol{Y}_{1:N}^{(m)}$ denoting the observable process of the system at times $1\!:\!N$. Under this framework, the observed data y_{1n}^* are assumed to be a single realization of the model $y_{1n}^* \sim f_{\boldsymbol{X}_{0:N}^{(m)}, \boldsymbol{Y}_{1:N}^{(m)}}(x_{0:N}, y_{1:N}; \theta)$, where θ is a parameter vector that indexes the model. Using a probabilistic model results in several advantages, including the ability to estimate the variability of the system. Furthermore, because each draw from the joint distribution represents a potential outcome of the dynamic system, best/worst case scenarios under the assumptions of the model can be easily obtained via simulation.

Once a model for the system has been proposed, the parameter vector θ needs to be estimated using the observed data. There exist several modelling techniques that can be used to obtain maximum likelihood estimates (MLE) of the parameters in stochastic dynamic models, including the EM algorithm, Kalman Filter (and extensions) (Evensen, 2009), and iterated filtering algorithms (Ionides et al., 2015). In order to use models that are scientifically meaningful rather than only those that are statistically convenient, we restrict ourselves to parameter estimation techniques for POMP models that have the plug-and-play property, which is that the fitting procedure only requires a way to simulate the latent process instead of the ability to evaluate transition densities (Bretó et al., 2009; He, Ionides and King, 2010). Plug-and-play algorithms include ABC and PMCMC (Toni et al., 2009; Andrieu, Doucet and Holenstein, 2010), but we focus on likelihood maximization via iterated filtering algorithms, which modify the well-known particle filter (Arulampalam et al., 2002) by performing a random walk for each parameter and particle. These perturbations are carried out iteratively over multiple filtering operations, using the collection of parameters from the previous filtering pass as the parameter initialization for the next, and decreasing the random walk variance at each iteration.

The use of likelihood-based inference allows for the statistical testing of potential improvements to the model. We demonstrate this capability by proposing a linear trend ζ in transmission in Eq. (2):

$$(33) \qquad \log \beta(t) = \sum_{j=1}^{6} \beta_s s_j(t) + \zeta \bar{t}$$

Where $\bar{t}$ is the linear mapping $\bar{t} : [0, N] \to [-1, 1]$ of the time t. The proposal of a linear trend in transmission is a result of observing an apparent decrease in reported cholera infections from 2012-2019 in Fig. 1. While several factors may contribute to this decrease, one explanation is that case-area targeted interventions (CATIs), which included education sessions, increased monitoring, household decontamination, soap

distribution, and water chlorination in infected areas (Rebaudet et al., 2019), may have greatly reduced cholera transmission (Rebaudet et al., 2021).

We perform a statistical test to determine whether or not the data indicate the presence of a linear trend in transmissibility. To do this, we perform a profile-likelihood search on the parameter ζ and obtain a confidence interval via a Monte Carlo Adjusted Profile (MCAP) (Ionides et al., 2017). Similar to Model 2, this model was originally implemented in two parts: an epidemic phase from October 2010 through March 2015, and an endemic phase from March 2015 onward. As before, we continue to allow the re-estimation of some model parameters at the start of the endemic phase ($\rho, \sigma_{\text{proc}}^2$ and ψ) but require that the latent Markov process $X(t)$ carry over from one phase into the next. The resulting confidence interval for ζ is $(-0.109, -0.014)$, with the full results displayed in Fig. 2. These results indicate that the inclusion of a trend improves the model fit to the data. The reported results for Model 1 in the remainder of this article were obtained with the inclusion of the parameter ζ.
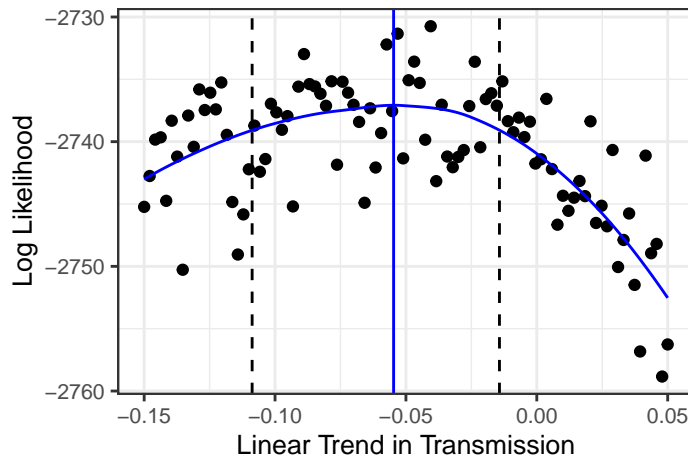

FIG 2. *Monte Carlo adjusted profile of ζ. The blue curve is the profile, the blue line indicates the MLE, and the dashed lines indicate the confidence interval.*

Model 1 is implemented using the `pomp` package King et al. (2009), which contains an implementation of the IF2 algorithm that was used to compute the likelihood profile. Simulations from the fitted model compared to the observed data are given in Fig 3.

3.1.3. *Model 3.* Model 3 is also of a probabilistic model. In this model, both the latent and observable processes can be factored into department specific processes which interact with each other. We denote the latent and measurement processes for Model 3 as $\boldsymbol{X}^{(3)}(t_{0:N}) = X_{1:U,0:N}^{(3)}$, and $\boldsymbol{Y}^{(3)}(t_{1:N}) = Y_{1:U,1:N}$. The decision to model the system via metapopulation models versus a model aggregated to a larger spatial scale is one of great scientific interest, and evidence for the former approach has been provided in previous studies (King et al., 2015). Note that this evidence alone does not automatically discredit conclusions drawn via nationally aggregated models, as an argument can easily be made that one should prefer a simple model over a complex one (Saltelli et al., 2020; Green and Armstrong, 2015). Still, researchers that intend to use a mechanistic model to describe a dynamic system should design the model to be as realistic as their scientific understanding of the system and their computational abilities permit. Fitting
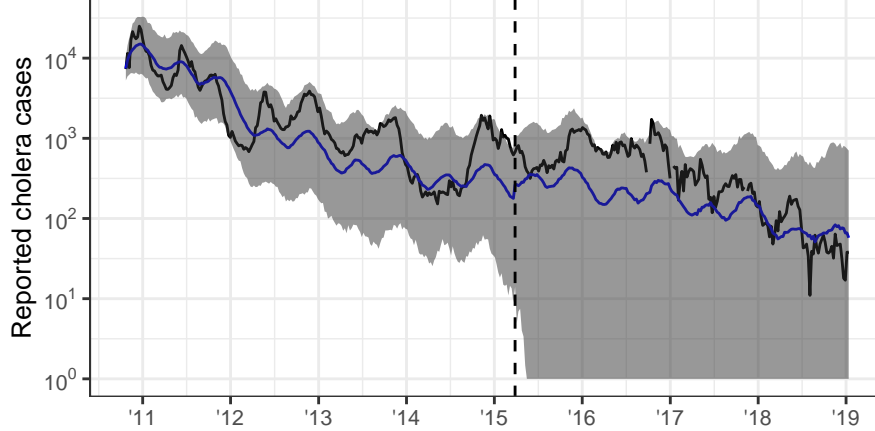
FIG 3. *Simulations from Model 1 compared to reported cholera cases. The black curve is observed data, the blue curve is median of 500 simulations from the fitted model, and the vertical dashed line represents break-point when parameters are refit.*

scientifically flexible metapopulation models, however, remains a challenging statistical problem; this is due to the fact that the approximation error of particle filters grows exponentially in the dimension of the model (Rebeschini and van Handel, 2015; Park and Ionides, 2020). Algorithms that are based on the particle filter therefore become computationally intractable as the number of spatial units increase.

Parameters that must be fit in Model 3 are primarily shared between each departement, the exception to this being the parameters β_{W_u}, and β_u, which are unique for each departement $u \in 1\!:\!10$. To fit this model, Lee et al. (2020a) simplified the parameter estimation problem by fitting independent departement-level models to the data. The shared parameters were calibrated using the cholera incidence data from Artibonite, and the departement-specific parameters (β_{W_u} and β_u) were fit using the data from their respective departement. Reducing a spatially-heterogeneous model to individual units in this fashion requires special treatment of any interactive mechanisms between spatial units, such as found in Eq. (23). In particular, when considering a model for departement u, the values $I_\nu(t)$ and $A_\nu(t)$, are unknown for $u \neq v \in 1\!:\!10$. A first order approximation of Eq. (23) for each departement $v \in 1\!:\!10$ can be obtained using the weekly number of observed cholera cases in each departement:

$$(34) \qquad I_v(t) + A_v(t) \approx \frac{365}{7\rho} y_v^*(t) \left(\frac{1}{\delta + \delta_C + \mu_{IR}} + \frac{1 - f}{f(\delta + \mu_{IR})} \right)$$

This approximation leads to department-specific models that are conditionally independent given the reported number of cholera infections in the remaining departments. Here, we refer to a collection of POMP models that are independent across units as a PanelPOMP.

In the case of a PanelPOMP, an extension of the IF2 algorithm, known as Panel Iterated Filtering (PIF) (Bretó, Ionides and King, 2020a), can be used to obtain the MLE, which solves the curse of dimensionality for this class of models. A major advantage of this algorithm over fitting each unit-specific model separately is that PIF can be used to fit both unit-specific parameters and shared parameters; in this way, the calibration of shared parameters involves all of the available data instead of just an arbitrarily chosen subset.

We then fit the panel version of Model 3 using a slight modification of the PIF algorithm, which we call the Block Panel Iterated Filter (BPIF). The BPIF algorithm is implemented in the `panelPomp` package (Bretó, Ionides and King, 2020b) in the R programming language, and can be used by adding the argument `block = TRUE` in the `mif2` function. pseudo-code for this algorithm is provided in Algorithm S1 in the supplementary material. Therefore the reported parameters were obtained using the PanelPOMP version of Model 3, and simulations for the various vaccination campaigns (Sec. 4) were obtained using these same parameters in the SpatPOMP version of the model.
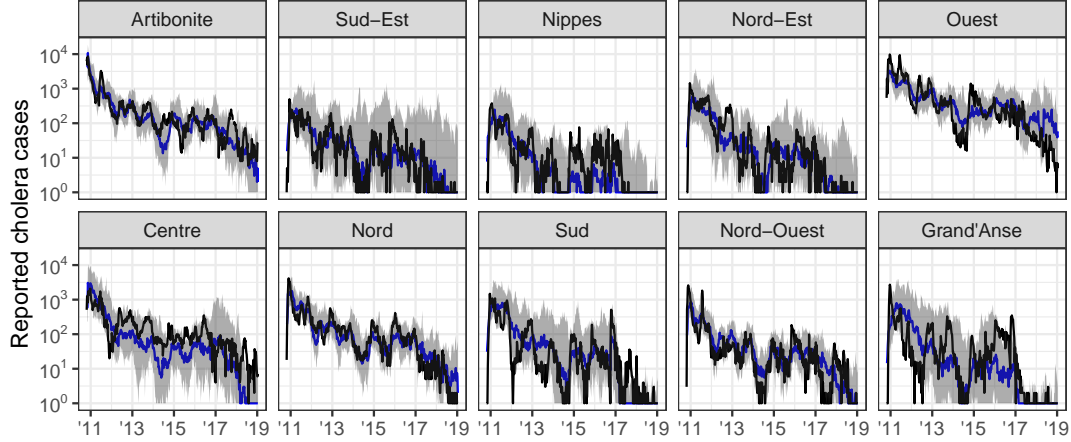

FIG 4. *Simulations from initial conditions using panelPOMP version of Model 3. The black curve represents true case count, the blue curve the median of 500 simulations from the model, and the grey ribbons representing* 95% *confidence interval.*

Fitting Model 3 simplifies the problem of parameter estimation, but it also introduces additional technicalities that must be addressed. One concern is that of obtaining model projections, which was the primary goal of the original modeling exercise. The simplified PanelPOMP version of Model 3 relies on the observed cholera cases as a covariate, which are unavailable for use in projections. To address this, we use the MLE obtained for the PanelPOMP approximation of Model 3 as an estimate for the parameters in fully coupled version of the model, which was implemented using the `spatPomp` package. Simulations from the SpatPOMP version of the model, using the parameters that were fit with the PanelPOMP version of the model, are displayed in Fig. 5.

While simulations from the SpatPOMP version of the model seem similar to the observed data, we note in Table 1 that using the MLE obtained using the approximation in Eq. (34) results in a lower likelihood value for the SpatPOMP version of the model than the PanelPOMP version. A natural question to ask is if the MLE obtained using the approximation in Eq. (34) could result in a drastically different set of parameters than those that would be obtained by estimating the MLE in the fully coupled model. Unfortunately, the answer to this question would require the ability to accurately and reliably estimate the MLE of the fully coupled model, which is beyond the scope of this article. This highlights the need for advancements in statistical methodology that permit inference on models with coupled metapopulation dynamics.
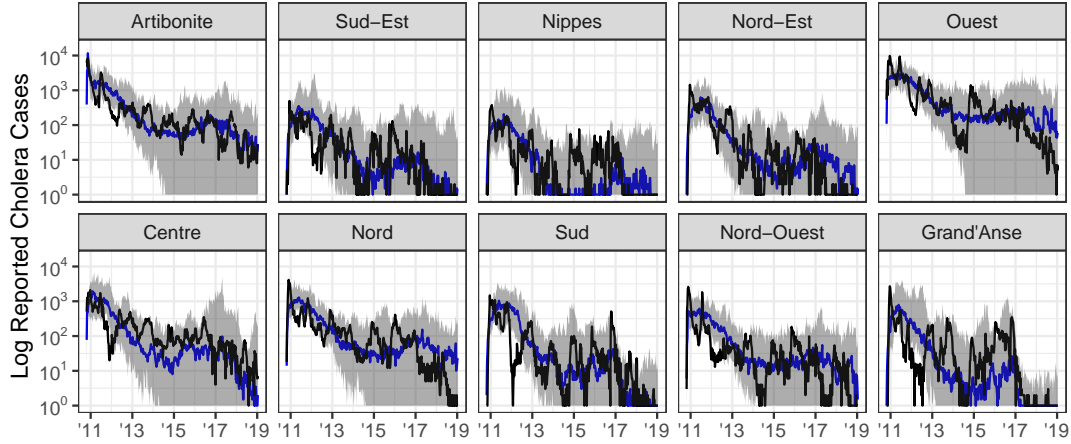
FIG 5. *Simulations from initial conditions using SpatPOMP version of Model 3. The black curve represents true case count, the blue line the median of 500 simulations from the model, and the grey ribbons representing* 95% *confidence interval.*

3.2. *Model Diagnostics.* [How about this edit?] Researchers may use fitted models to make inference and draw conclusions about the dynamic system under investigation. Because of the casual nature of mechanistic models, the validity of these conclusions is dependent on the ability of the proposed model to approximate the system in question. It is often impossible to know, however, the extent to which a proposed model provides a faithful representation of the system. In such cases, we rely on the quantitative and statistical ability of the model to match the observed data, which we call the model's goodness-of-fit. Goodness-of-fit provides evidence supporting the causal interpretation of a model, but cannot by itself rule out the possibility of alternative explanations.

One common approach to assess a mechanistic model's goodness-of-fit is to compare several simulations from the fitted model to the observed data; this is done to determine whether or not the observed data are a plausible realization of the fitted model. While these comparisons are useful, relying only on these qualitative comparisons may lead to over confidence on the goodness-of-fit. A relevant example is the original parameter estimates provided by Lee et al. (2020a): these parameters resulted in simulations that closely resembled the observed data, yet resulted in model likelihoods that were—in some cases—drastically smaller than likelihoods that can be achieved via the likelihood based optimization techniques that were used (see Table 1). Alternative forms of model validation should therefore be used in conjunction with visual comparisons of simulations to observed data.

The goodness-of-fit of a model can be assessed by comparing a quantitative description of the model fit (such as MSE, predictive accuracy, or model likelihood) to that of other proposed models. This type of comparison provides insight into how each model performs relative to the others, but does not provide insight on the performance of each model on a given dataset unless the performance of one of the models is widely understood. We refer to such a model as a benchmark, and examples of these models include logistic regression and ordinary least squares.

Comparing associative models using an objective function seems natural, as these models are designed to optimize objective functions given the available data. In the case of mechanistic modeling, however, comparing models based on an objective measure is less natural because each mechanistic model is designed to provide a scientific

description of the dynamic system. Any comparison between mechanistic models would therefore ideally be based on the scientific validity of each proposed model, which is often impossible to know. Still, parameters in mechanistic models are fit using techniques that are designed to maximize an objective such as likelihood, and it is expected that a good scientific description of the dynamic system should result in more optimal values of an objective evaluation than would a poor description of the system. In this sense, one could justify comparing mechanistic models based on some objective measurement such as model likelihoods. We do not suggest that the model with the highest likelihood given the data automatically provides a better scientific description of the dynamic system; still, a comparison based on model likelihoods may help guide researchers to know the plausibility of a proposed model compared to another given their relative ability to describe the observed data.

This comparison still only provides a relative measure of model fit compared to other mechanistic models. The resulting challenge is that no general benchmark exists for mechanistic models that would allow researchers to assess the goodness-of-fit by comparing to a known standard. In some cases, a possible benchmark model could be a generally accepted truth, but often no such truth is available. Because of this, we propose the use of an ARMA(2, 1) model as a potential benchmark. Despite being an associative model, the reason that an ARMA(2, 1) model is an appealing benchmark is because the performance of this class of models is widely understood, and often provides a good description of the data with very limited effort. One question that follows is whether it is useful to assess mechanistic models by their statistical fit, if this is not their primary purpose? To what extent can we say the data support a scientific model (A) over another model (B) based on improved fit to data, if neither fit better than an associative model (C)?

When comparing mechanistic models to non-mechanistic benchmarks, we do not require the benchmark model to optimize a measure of statistical fit, nor do we suggest that a mechanistic model must have superior fit to the benchmark in order to be of scientific interest. It is possible that a successful mechanistic model may have better statistical fit than simple benchmarks such as log-linear models; still, the specific threshold of beating, or failing to beat, an arbitrary benchmark is similarly arbitrary. In this case, however, a large shortfall in the fit of a mechanistic model to that of a log-linear model may demonstrate room for substantial improvement. The use of a common benchmark may also be beneficial when developing models with varying spatial scale, as direct comparisons between models fit to data with different levels of spatial aggregation are meaningless (e.g. comparing Model 1 to Model 3). Likelihoods of Models 1–3 and their respective ARMA(2, 1) benchmark models are provided in Table 1. [WE SAY THAT WE HAVE AN IN-DEPTH DESCRIPTION OFMODEL VALIDATION, BUT ONLY PROPOSE/DEMONSTRATE TWO WAYS TO DO THIS. SHOULD WE MENTION OTHER THINGS, LIKE CONDITIONAL LIKELIHOODS / ESS, CHECKING PLAUSABILITY OF FIT MODEL PARAMETERS, OR COMPARING AUTO-CORRELATIONS BETWEEN SIMULATIONS AND OBSERVATIONS (LIKE AARON KING HAS DONE)? I DON'T THINK WE NECESSARILY HAVE TO SHOW ALL OF THESE, SINCE THE ARTICLE IS ALREADY TOO LONG, BUT MAYBE MENTIONING THEM? WOULD MENTIONING THESE WITHOUT DEMONSTRATING THEM RAISE CONCERNS FROM REVIEWERS?]

	Model 1	Model 2	Model 3 (panelPOMP)	Model 3 (SpatPOMP)
Log Likelihood	-2731.3 (-2966.8)[1]	-27805.3 (-28844.3)	-17601.3[2]	-19746.6 (-60060.4)[3]
Number of Fit Parameters	13 (20)	21 (26)	39 (29)	39 (29)
AIC	5488.7 (5973.6)[1]	55652.5 (57740.7)	35280.7[2]	39571.1 (120178.9)[3]
ARMA(2, 1) Log Likelihood	-2802.6	-18061.9	-18061.9	-18061.9

TABLE 1

Log-likelihood values for each models compared to their ARMA benchmarks. Values in parenthesis are corresponding values using Lee et al. (2020a) parameter estimates.

3.3. *Model Projections.* As previously noted, simulations from a well-fit mechanistic model may closely resemble the observed data $y_{1:N}^*$. In such comparisons, these simulations are random draws from the complete estimated joint distribution $f_{\boldsymbol{X}_{0:N}^{(m)}, \boldsymbol{Y}_{1:N}^{(m)}}\left(x_{0:N}, y_{1:N} \mid \hat{\theta}\right)$, where m identifies the model used for simulations. It can therefore be tempting to use simulations from this model up to time $N + s$, $f_{\boldsymbol{X}_{0:N+s}^{(m)}, \boldsymbol{Y}_{1:N+s}^{(m)}}\left(x_{0:N+s}, y_{1:N+s} \mid \hat{\theta}\right)$, with $s \geq 0$ to project the dynamic system up to a future time $N + s$, as it has been done in previous studies (Lee et al., 2020a; **?**). This approach, however, does not take advantage of the information about the state of the system that is contained in the observed data. Note that each of the models in question (and, more generally, all models that are described as POMPs) are Markovian, that is, the history of the process $\{\boldsymbol{X}_s^{(m)}, s \leq t\}$ for model m is uninformative about the future of the process $\{\boldsymbol{X}_s^{(m)}, s \geq t\}$, given the current state $\boldsymbol{X}_t^{(m)}$. In other words, observing the data $\boldsymbol{Y}_N^{(m)} = y_N^*$ at time N provides more information about the future state of the system than does the initial conditions. A more sound projection of the dynamic system into the future could therefore be obtained by simulating using the most recently known state of the system, that is, obtaining draws from the conditional density $f_{\boldsymbol{X}_{N:s}^{(m)}, \boldsymbol{Y}_{N:s}^m}\left(x_{N:s}, y_{N:s} \mid \hat{\theta}, \boldsymbol{X}_N^{(m)} = \boldsymbol{x}_N^{(m)}\right)$.

$\boldsymbol{X}_N^{(m)}$, however, is unobservable and therefore draws from the desired conditional density are unobtainable. Possible values of $\boldsymbol{X}_N^{(m)}$ given the observed data $\boldsymbol{Y}_{1:N}^{(m)} = y_{1:N}^*$ and the proposed model can easily be obtained, however, via the filtering distribution. Let $\hat{\boldsymbol{X}}_N^{(m),i}$, $i \in 1, 2, \ldots, J$ be iid draws from the filtering distribution at time N, with density $\hat{\boldsymbol{X}}_N^{(m),i} \sim f_{\boldsymbol{X}_N^{(m)} \mid \boldsymbol{Y}_{1:N}^{(m)}}(x_N \mid \hat{\theta}, y_{1:N}^*)$. A single model projection can then be obtained by simulating from the model $f_{\boldsymbol{X}_{N:s}^{(m)}, \boldsymbol{Y}_{N:s}^{(m)}}\left(x_{N:s}, y_{N:s} \mid \hat{\theta}, \boldsymbol{X}_N^{(m)} = \hat{\boldsymbol{X}}_N^{(m),i}\right)$. Intuitively, simulating the model starting at the filtering distribution of the most recently available

---

[1]The reported likelihood is an upper bound of the likelihood of the Lee et al. (2020a) model as it is the largest likelihood obtained using their parameter calibration regime.

[2]Parameters to department-specific models not provided by Lee et al. (2020a), as department fits were only used as an intermediary step to obtain parameter estimates of the coupled model.

[3]Model 3 was originally fit to only a subset of the data starting from March 2014 and did not include a large portion of data from Ouest in 2015-2016. On this subset, the parameters provided by Lee et al. (2020a) achieved a likelihood of -13202.1. On this same subset of data, our model achieved a likelihood of -8583.7.

time point is a more appropriate way to project a stochastic dynamic system into the future, as it is not expected that each simulation from initial conditions will result in a latent state at time N consistent with the model and the observation $\boldsymbol{Y}_N^{(m)} = y_N^*$. In this particular case study, projecting the future state of the cholera epidemic in Haiti starting from the draws from the filtering distribution allows the proposed models to benefit from the fact that very few cholera cases had been observed between 2018 and 2019, and that cases appear to be decreasing (i.e., the number of susceptible individuals may be small).

An additional consideration to make when obtaining model forecasts is that of parameter uncertainty. It has been noted that the uncertainty in just a single parameter can lead to drastically different projections (Saltelli et al., 2020). One possible approach to account for parameter uncertainty in model projections is by obtaining confidence intervals for each parameter, sampling parameters from the confidence intervals, simulating the model with the resulting parameter set, and then weighing the resulting model projections based on the likelihood of the given set of parameters, as was done by King et al. (2015).

Note that in order to obtain projections that are consistent with the observed data, one must first be able to sample from the filtering distribution given each set of parameters. This approach can be done for both deterministic and stochastic models, but requires a large number of computations, especially as the number of observations and model parameters increase. Because the focus of this study is on model fitting and evaluation, we do not provide model projections accounting for parameter uncertainty. Instead, we use the projections from point estimates to highlight a major deficiency of deterministic models, which is that the only variability in model projections is a result of parameter uncertainty, which leads to over-confidence in the projections. This observation is consistent with those made in King et al. (2015), and suggests that stochastic models should be preferred over deterministic models. [TO THE CREDIT OF MODEL 2, ITS PREDICTIONS OF POLICY IMPLICATIONS WERE AS GOOD AS, OR BETTER THAN, THE OTHER MODELS. PERHAPS THE WEAKNESSES OF DETERMINISTIC MODELS ONLY SHOW UP AS THE DATA ANALYSIS BECOMES MORE REFINED - FOR EXAMPLE, DETERMINISTIC MODELS MAY HAVE SUCH LOW STATISTICAL FIT THAT WE CONCLUDE THEY ARE NOT RELIABLE AS ASSOCIATIVE MODELS, IN WHICH CASE THEY CANNOT BE EXPECTED TO PROVIDE QUANTITATIVELY ACCURATE PREDICTIONS.]

[Anna: I added some discussion on this in the supplement under Model 1 replication. Does this answer your concern? (see original comment in Rnw file)]

4. Results. [HERE'S AN INITIAL ATTEMPT TO EXPLAIN MY POINT BELOW FOR THE MS. IT IS HARD TO MAKE CONCRETE POINTS ON THIS SORT OF TOPIC. A model which aspires to provide quantitative guidance for assessing interventions should provide a quantitative statistical fit for previous data. However, strong statistical fit does not guarantee a correct causal structure: it does not even necessarily require the model to assert a causal explanation. A causal interpretation is strengthened by corroborative evidence. For example, reconstructed latent variables (such as numbers of susceptible and recovered individuals) should make sense in the context of alternative measurments of these variables; parameter values which fit the data should make sense in the context of alternative lines of evidence about the phenomena being modeled. The model is necessarily a simplified approximation of a complex process, and so we cannot expect our models (describing and fitted to population-level processes)

to perfectly match quantitative understanding of individual-level processes. Such discrepancies could lead to biases when interventions (modeled as effects on individuals) are included in the population-level, unless there are historical occurrences of the intervation which enable the consequence to be calibrated at the population level. IF WE WANT TO HEAD IN THIS DIRECTION, WE COULD EXPLAIN HOW THESE PRINCIPLES PLAY OUT IN OUR DATA ANALYSIS.]

[TO AVOID BEING TOO NEGATIVE, WE ALSO COULD EXPLAIN THE LOGIC OF HOW POSITIVE RESULTS CAN LEGITIMATELY BE CONCLUDED. If a mechanistic model including a feature (such as a representation of a mechanism, or the inclusion of a covariate) fits better than mechanistic models without that feature, and also has competititive fit compared to associative models, this may be taken as evidence supporting the scientific relevance of the feature. As for any analysis of observational data, we must be alert to the possibility of confounding. For a covariate, this shows up in a similar way to regression analysis: the covariate under investigation could be a proxy for some other unmodeled or unmeasured covariate. For a mechanism, the model feature could in principle explain the data by helping to account for some different unmodeled phenomenon. Assessing potential confounding is part of building an argument for a causal interpretation of a fitted model. In the context of our analysis, the estimated trend in transmission rate could be explained by any trending variable (such as hygiene improvements, or changes in population behavior), resulting in confounding from colinear covariates. Alternatively, the trend could be attributed to a decreasing reporting rate rather than decreasing transmission rate, resulting in confounded mechanisms. The robust statistical conclusion is that a model which allows for change fits better than one which does not—we argue that a decreasing transmission rate is a plausible way to explain this, but the incidence data themselves may not have enough information to pin down the mechanism.]

Because mechanistic models provide a scientific description of the dynamic system, a properly fit model allows a researcher to perform inference on the system in question. [FIT IS A MEASURE OF ASSOCIATION, AND ASSOCIATION DOES NOT NECESSARILY IMPLY CAUSATION. SO WE HAVE TO WORD THIS CAREFULLY. IT CAN BE A RATHER DELICATE QUESTION TO ESTABLISH THE PROPER CAUSAL INTERPRETATION OF A FITTED MODEL, AND THIS MAY NOT MATCH ONE'S INITIAL GUESS (E.G., THE MEASLES CASE STUDY IN SISMID). IT WOULD BE INTERESTING TO DISCUSS IN PERSON HOW TO THREAD THIS NEEDLE.] Examples are as diverse as estimating the effective reproductive number (R_0) of an infectious disease (He, Ionides and King, 2010) or investigating interactions between pedestrians and autonomous vehicles (Domeyer et al., 2022). Here, we demonstrate this ability by investigating the seasonal transmission trends in Model 1. Model 1 includes a flexible cubic spline term in Eq. (1)–(2) that allows for seasonality in the force of infection. This flexibility allows the model to learn any seasonal trends in cholera infections that may be present in the data. After fitting the model, we explore potential patterns in the seasonal transmission rate by plotting the average value of β in a typical year. Fig. 6 shows that the estimated seasonal transmission rate β mimics the rainfall dynamics in Haiti, despite Model 1 not having access to rainfall data. This result provides evidence that rainfall is an important driver of cholera infections in Haiti.

An additional benefit of mechanistic modelling is that they permit the simulation of potential interventions on a system, as was the primary goal of Lee et al. (2020a). Outcomes of their study include estimates for the probability of cholera elimination and cumulative number of cholera infections under several possible vaccination scenarios.
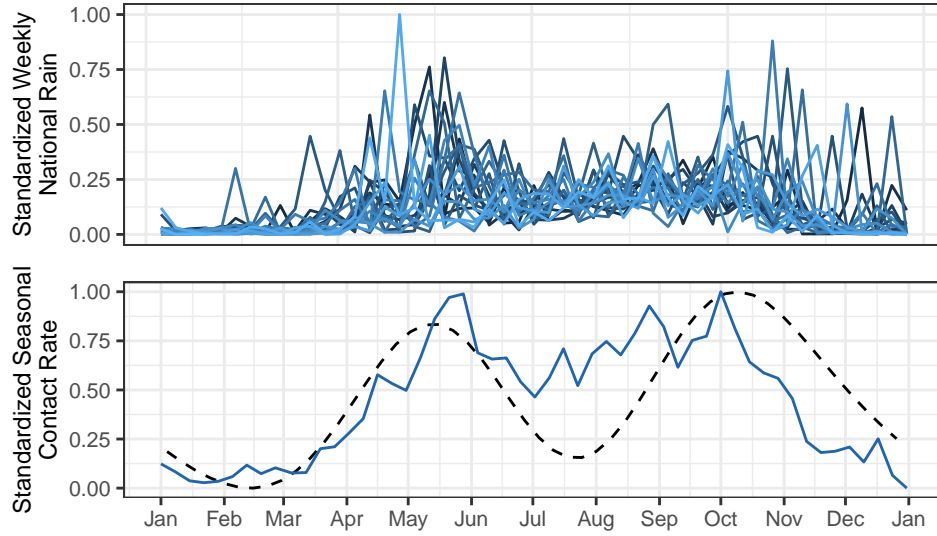
FIG 6. (Top) weekly rainfall in Haiti, lighter colors representing more recent years. (Bottom) estimated seasonality in the transmission rate (dashed line) plotted alongside mean rainfall (solid line).

Mimicking their efforts, we define cholera elimination as having less than one infection of cholera over at least 52 consecutive weeks, and provide projections under the following vaccination scenarios: Rewrite definition: cholera elimination defined as less than one cholera infection for at least fifty two consecutive weeks in the ten years after the vaccinations? [Good catch, TODO: is it necessary to say that it stayed away? Does it make a difference?]

$V0$: The default vaccination scenario, which is that no additional vaccinations are deployed.

$V1$: Vaccination limited to the departements of Centre and Artibonite, and vaccinations deployed over a two-year period.

$V2$: Vaccination limited to three departements: Artibonite, Centre, and Ouest over a two-year period.

$V3$: Countrywide vaccination implemented over a five-year period.

$V4$: Countrywide vaccination implemented over a two-year period.

Simulations from probabilistic models (Models 1 and 3) represent possible trajectories of the dynamic system under the scientific assumptions of the models. In this case study, estimates of the probability of cholera elimination can therefore be obtained as the proportion of simulations from the fitted model that result in cholera elimination. The results of these projections are summarized in Figs. 7–10. These results suggest that cholera elimination was likely, even without intense vaccination efforts. This conclusion is consistent with the fact that cholera was eradicated from Haiti without additional vaccination efforts (Ferguson, 2022).

In additional to probability of elimination estimates, we are also able to estimate the cumulative number of infections under each vaccination scenario from February 2019 – February 2024. Notably, the median number of cumulative cholera infections under the no-vaccination scenario using Models 1 and 3 were $2,058$ and $529,175$, respectively. So far, these estimates are far more consistent with the observed number of reported cholera cases than the corresponding estimates from Lee et al. (2020a), which were approximately $400,000$ and $1,000,000$.
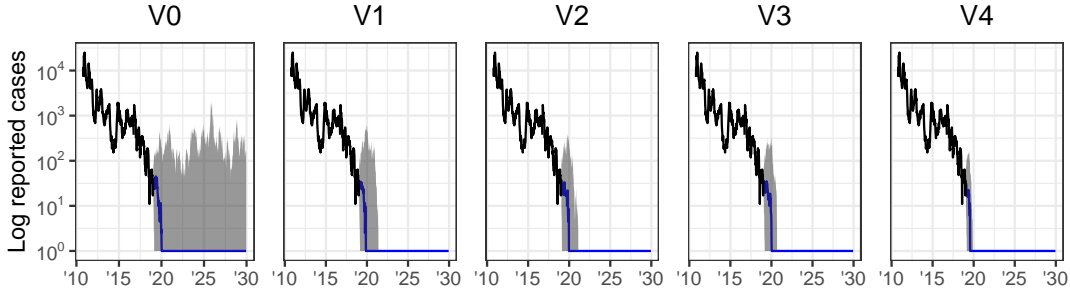
FIG 7. *Simulations of Model 1 under each vaccination scenario. Blue line indicates the median of model simulations, and ribbon represents 95% of simulations. The various vaccination campaigns made no practical difference in the median scenario, but a drastic difference in the extreme cases.*
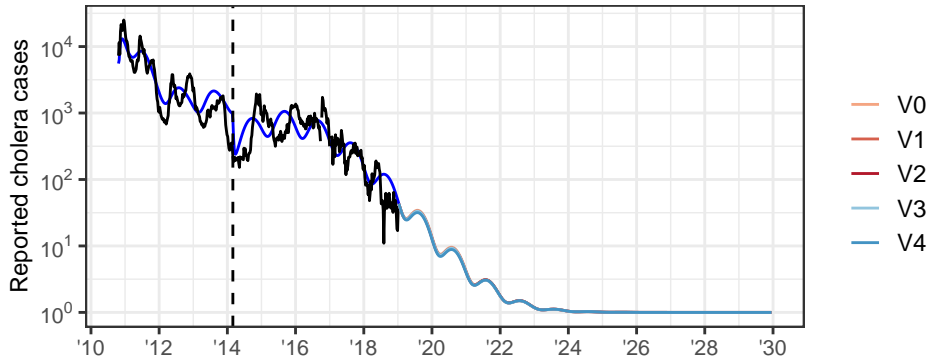

FIG 8. *Simulated trajectory of Model 2 (blue curve) and projections under the various vaccination scenarios. Reported cholera incidence is shown in black. The dashed vertical line shows where the endemic phase starts and parameter values are re-estimated. Note that the various vaccination scenarios had little effect on the projected cholera cases. This is because the primary effect of vaccines in Model 2 is to remove individuals from the susceptible pool, yet reported number of cholera infections were low and decreasing at the start of the vaccination campaigns (only 37 cases were reported in all of Haiti), so this had very little effect on cholera transmission.*

Probability of elimination estimates of this form are far less meaningful for deterministic models. This is because the data are not assumed to be draws from the model; the model is instead intended only to provide understanding about the overall trend of the epidemic. We therefore do not provide probability of elimination estimates under Model 2. Still, projected trajectories using Model 2 are consistent with those found with Models 1 and 3, and suggest that cholera was in the process of being eliminated from Haiti.

Despite these results suggesting that cholera was likely to be eliminated from Haiti without an increase in vaccination effort, we in no way suggest that vaccinations shouldn't be used to combat cholera—or other infectious diseases—in other parts of the world. In fact, our analysis shows that an increase in vaccination does drastically reduce the negative effects of a worst-case scenario.

5. Discussion. The ongoing global COVID-19 pandemic has provided a clear example on how government policy may be affected by the conclusions of scientific models.
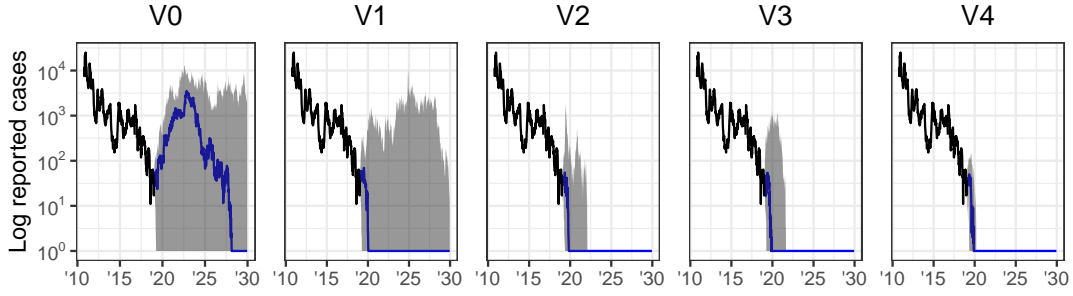
FIG 9. *Simulations of Model 3 under each vaccination scenario. Blue line indicates the median of model simulations, and ribbon represents* 95% *of simulations. The various vaccination campaigns made no practical difference in the median scenario, but a drastic difference in the extreme cases.*
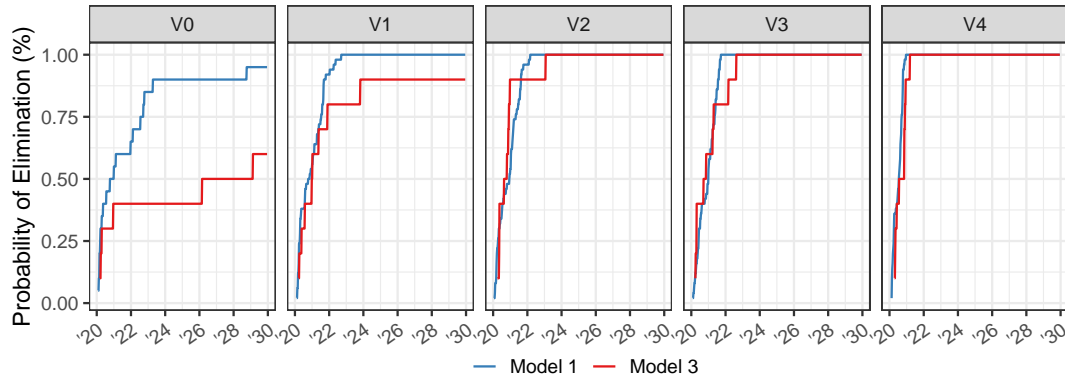

FIG 10. *Probability of elimination across simulations for a 10 year period. Compare to Figure 3A of Lee et al. (2020a).*

This article demonstrates that fitting appropriate scientific models remains a challenging statistical task, and therefore great care is needed when fitting scientific models for policy recommendations. We provided a few suggestions that may aid the fitting of mechanistic models such as comparing model likelihoods to a benchmark. Improved model fits allows for meaningful statistical inference that may provide valuable insight on a dynamic system in question and may improve the accuracy of model-based projections. Caution is nonetheless needed when making policy based on modelling conclusions, as model misspecification may invalidate conclusions.

Various have suggestions been made about why Lee et al. (2020a) failed to accurately predict the eventual eradication of cholera from Haiti. Proposals include model misspecification, overly difficult elimination criteria, and a potential conflict of interest (Rebaudet, Gaudart and Piarroux, 2020; Henrys et al., 2020). Here, we instead argue that careful attention to important statistical details could have correctly resulted in the conclusion of imminent cholera elimination. We acknowledge that we have the benefit of hindsight: our demonstration of a statistically principled route to obtain better-fitting models with more accurate predictions does not rule out the possibility of discovering other models that fit well yet predict poorly.

We used the same data and models, and even much of the same code, as Lee et al. (2020a), and yet ended up with drastically different conclusions. At a minimum, we have

Mechanism	Model 1		Model 2		Model 3	
Infection (day)	$\mu_{IR}^{-1} = 2.0$	(6)	$\mu_{IR}^{-1} = 7.0$	(14)	$\mu_{IR}^{-1} = 2.0$	(26)
Latency (day)	$\mu_{EI}^{-1} = 1.4$	(5)	$\mu_{EI}^{-1} = 1.3$	(13)	—	
Seasonality	$\beta_{1:6} = (1.4, 1.2,$ $1.2, 1.1, 1.4, 0.9)$	(2)	$a = 0.4$	(11)	$a = 19.49$ $r = 167675.97$	(31)
Immunity (year)	$\mu_{RS}^{-1} = 8.0$	(7)	$\mu_{RS}^{-1} = 5.0$ $\omega_1^{-1} = 1.0$ $\omega_2^{-1} = 0.2$	(15) (17) (17)	$\mu_{RS}^{-1} = 8.0$	(29)
Vaccine efficacy	—		$\theta_{1:4} = (0.80, 0.76$ $0.57, 0.48)$	(12)	$\eta_{ud}(t)$	
Birth/death (year)	$\mu_S^{-1} = 44.9$ $\delta^{-1} = 134.2$	(9)	—		$\delta^{-1} = 63.0$	(27)
Symptomatic frac.	$f_z(t) = c\theta^*(t - \tau_d)$	(4-5)	$f = 0.2$	(13)	$f = 1.00$	(25)
Asympt. infectivity	$\epsilon = 0.05$	(1)	$\epsilon = 0.001$ $\epsilon_W = 10^{-7}$	(11) (19)	$\epsilon = 1$ $\epsilon_W = 0.010$	(23) (31)
Human to human	$\beta_{1:6}$ as above	(1)	$\beta = (4.5 \times 10^{-15},$ $9 \times 10^{-16})$	(11)	$\beta_{1:10} = (2.16, 1.02,$ $0.35, 0.22, 2.18, 0.59,$ $1.44, 2.09, 0.44, 0.28)$ $\times 10^{-6}$	(23)
Water to human	—		$W_{\text{sat}} = 10^5$ $\beta_W = (5.95, 1.15)$	(11)	$W_{\text{sat}} = 1$ $\beta_{W_{1:10}} = (1.23, 5.21,$ $7.23, 8.31, 1.35, 8.28,$ $2.73, 0.24, 3.54, 3.91)$	(23)
Human to water	—		$\mu_W = (695,$ $4.24 \times 10^{-11})$	(19)	$\mu_W = 2.71 \times 10^{-5}$	(31)
Water survival (wk)	—		$\delta_W^{-1} = (4.14 \times 10^{11},$ $34.5)$	(20)	$\delta_W^{-1} = 0.36$	(32)
Mixing exponent	$\nu = 0.98$	(1)	—		—	
Proc. noise (wk$^{1/2}$)	$\sigma_{\text{proc}} = (0.32, 0.35)$	(1)	—		$\sigma_{\text{proc}} = 0.032$	(25)
Reporting rate	$\rho = (0.577, 0.833)$	(??)	$\rho = 0.20$	(??)	$\rho = 0.89$	(??)
Obs. overdispersion	$\psi = (398.45, 65.71)$	(??)			$\psi = 117.11$	(??)

TABLE 2

References to the relevant equation are given in parentheses. Parameters in blue were fixed based on scientific reasoning and not fitted to the data. [N] denotes parameters added during our re-analysis, not considered by Lee et al. Translations back into the notation of Lee et al. (2020a) are given in Table S1.

shown that the conclusions are sensitive to details in how the data analysis is carried out, and that attention to statistical fit (including numerical issues such as likelihood maximization) can lead to improved policy guidance.

We acknowledge there are limitations to this study; one example was the inability to fit model parameters to the fully-coupled SpatPOMP version of Model 3. Promising theoretical and methodological developments (Ning and Ionides, 2021; Ionides, Ning and Wheeler, 2022) based on the Block Particle Filter (Rebeschini and van Handel, 2015) may potentially be used to fit the SpatPOMP version of Model 3 in future work.

Inference for mechanistic time series models offers opportunities for understanding and controlling complex dynamic systems. This case study has investigated issues requiring attention when applying powerful new statistical techniques that can enable statistically efficient inference for a general class of partially observed Markov process models. Care must be taken to ensure that the computationally intensive numerical calculations are carried out adequately. Once that is accomplished, care is required to assess what causal conclusions can properly be inferred given the possibility of alter-

native explanations consistent with the data. Studies that combine model development with thoughtful data analysis, supported by a high standard of reproducibility, build knowledge about the system under investigation. Cautionary warnings about the difficulties inherent in understanding complex systems (Saltelli et al., 2020; Ioannidis, Cripps and Tanner, 2020) should motivate us to follow best practices in data analysis, rather than avoiding the challenge.

5.1. *Reproducibility and Extendability.* Lee et al. (2020a) published their code and data online, and this reproducibility facilitated our work. By design, the models were coded and analyzed independently, leading to differing implementation decisions. Robust data analysis requires not only reproducibility but also extendability: if one wishes to try new model variations, or new approaches to fitting the existing models, or plotting the results in a different way, this should be not excessively burdensome. Scientific results are only trustworthy so far as they can be critically questioned, and an extendable analysis should facilitate such examination (Gentleman and Temple Lang, 2007).

We provide a strong form of reproducibility, as well as extendability, by developing our analysis in the context of an `R` package, `haitipkg`. Using a software package mechanism supports documentation, standardization and portability that promote extendability. In the terminology of Gentleman and Temple Lang (2007), the source code for this article is a *dynamic document* combining code chunks with text. In addition to reproducing the article, the code can be extended to examine alternative analysis to that presented. The dynamic document, together with the R packages, form a *compendium*, defined by Gentleman and Temple Lang (2007) as a distributable and executable unit which combines data, text and auxiliary software (the latter meaning code written to run in a general-purpose, portable programming environment, which in this case is R).

SUPPLEMENTARY MATERIAL

Eliminating cholera in Haiti: Supplement
This document contains additional details for Models 1–3, as well as a translation table that facilitates comparisons between these models and those described in Lee et al. (2020a). The supplement also demonstrates our capability to faithfully replicate the results of Lee et al. (2020a).

haitipkg
This `R` package is contained in a public GitHub repository: zjiang2/haitipkg. The package contains all of the data and code used to create and fit the models, as well as other useful functions that were used in this article.

jesseuwheeler/haiti
This GitHub repository contains the `.Rnw` files that were used to create this document and the supplement material.

REFERENCES

ANDRIEU, C., DOUCET, A. and HOLENSTEIN, R. (2010). Particle Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)* **72** 269–342.

ARULAMPALAM, M. S., MASKELL, S., GORDON, N. and CLAPP, T. (2002). A Tutorial on Particle Filters for Online Nonlinear, Non-Gaussian Bayesian Tracking. *IEEE Transactions on Signal Processing* **50** 174 – 188.

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.

ASFAW, K., PARK, J., HO, A., KING, A. A. and IONIDES, E. L. (2021). Partially observed Markov processes with spatial structure via the R package spatPomp. *arXiv:2101.01157*.

BEHREND, M. R., BASÁÑEZ, M.-G., HAMLEY, J. I. D., PORCO, T. C., STOLK, W. A., WALKER, M., DE VLAS, S. J. and FOR THE NTD MODELLING CONSORTIUM (2020). Modelling for policy: The five principles of the Neglected Tropical Diseases Modelling Consortium. *PLOS Neglected Tropical Diseases* **14** 1-17. https://doi.org/10.1371/journal.pntd.0008033

BRAUER, F. (2017). Mathematical epidemiology: Past, present, and future. *Infectious Disease Modelling* **2** 113-127. https://doi.org/10.1016/j.idm.2017.02.001

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. https://doi.org/10.1016/j.spa.2011.07.005

BRETÓ, C., IONIDES, E. L. and KING, A. A. (2020a). Panel Data Analysis via Mechanistic Models. *Journal of the American Statistical Association* **115** 1178-1188. https://doi.org/10.1080/01621459.2019.1604367

BRETÓ, C., IONIDES, E. L. and KING, A. A. (2020b). panelPomp: Statistical Inference for PanelPOMPs (Panel Partially Observed Markov Processes) R package version 0.10.0.2.

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.

DADLANI, A., AFOLABI, R. O., JUNG, H., SOHRABY, K. and KIM, K. (2020). Deterministic Models in Epidemiology: From Modeling to Implementation. https://doi.org/10.48550/ARXIV.2004.04675

DOMEYER, J. E., LEE, J. D., TOYODA, H., MEHLER, B. and REIMER, B. (2022). Driver-Pedestrian Perceptual Models Demonstrate Coupling: Implications for Vehicle Automation. *IEEE Transactions on Human-Machine Systems* 1-10. https://doi.org/10.1109/THMS.2022.3158201

EVENSEN, G. (2009). The ensemble Kalman filter for combined state and parameter estimation. *IEEE Transactions on Control Systems* **29** 83–104.

FERGUSON, S. (2022). Haiti's winning the fight against cholera. *Forbes*.

FRANCOIS, J. (2020). Cholera remains a public health threat in Haiti. *The Lancet Global Health* **8** e984.

GENTLEMAN, R. and TEMPLE LANG, D. (2007). Statistical analyses and reproducible research. *Journal of Computational and Graphical Statistics* **16** 1–23.

GREEN, K. C. and ARMSTRONG, J. S. (2015). Simple versus complex forecasting: The evidence. *Journal of Business Research* **68** 1678-1685. Special Issue on Simple Versus Complex Forecasting. https://doi.org/10.1016/j.jbusres.2015.03.026

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.

HENRYS, J. H., LEREBOURS, G., ACHILLE, M. A., MOISE, K. and RACCURT, C. (2020). Cholera in Haiti. *The Lancet Global Health* **8** e1469.

IOANNIDIS, J. P., CRIPPS, S. and TANNER, M. A. (2020). Forecasting for COVID-19 has failed. *International Journal of Forecasting*.

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. https://doi.org/10.48550/ARXIV.2206.03837

IONIDES, E. L., NGUYEN, D., ATCHADÉ, Y., STOEV, S. and KING, A. A. (2015). Inference for dynamic and latent variable models via iterated, perturbed Bayes maps. *Proceedings of the National Academy of Sciences of the USA* **112** 719—724. https://doi.org/10.1073/pnas.1410597112

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.

KERMACK, W. O. and MCKENDRICK, A. G. (1927). A contribution to the mathematical theory of epidemics. *Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character* **115** 700–721.

KING, A. A. and ROWAN, T. (2020). subplex: Unconstrained Optimization using the Subplex Algorithm R package version 1.6.

KING, A. A., IONIDES, E. L., BRETÓ, C. M., ELLNER, S. and KENDALL, B. (2009). pomp: Statistical inference for partially observed Markov processes. R package, available at http://cran.r-project.org/web/packages/pomp.

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** 20150347.

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** e1081–e1089.

LEE, E. C., TERNIER, R., LESSLER, J., AZMAN, A. S. and IVERS, L. C. (2020b). Cholera in Haiti–Authors' reply. *The Lancet Global Health* **8** e1470–e1471.

LUCAS, R. E. et al. (1976). Econometric policy evaluation: A critique. In *Carnegie-Rochester conference series on public policy* **1** 19–46.

NDII, M. Z. and SUPRIATNA, A. K. (2017). Stochastic mathematical models in epidemiology. *Information* **20** 6185–6196.

NING, N. and IONIDES, E. L. (2021). Iterated Block Particle Filter for High-dimensional Parameter Learning: Beating the Curse of Dimensionality. https://doi.org/10.48550/ARXIV.2110.10745

PARK, J. and IONIDES, E. L. (2020). Inference on high-dimensional implicit dynamic models using a guided intermediate resampling filter. *Statistics & Computing* **30** 1497–1522.

REBAUDET, S., GAUDART, J. and PIARROUX, R. (2020). Cholera in Haiti. *The Lancet global health* **8** e1468.

REBAUDET, S., BULIT, G., GAUDART, J., MICHEL, E., GAZIN, P., EVERS, C., BEAULIEU, S., ABEDI, A. A., OSEI, L., BARRAIS, R. et al. (2019). The case-area targeted rapid response strategy to control cholera in Haiti: a four-year implementation study. *PLoS neglected tropical diseases* **13** e0007263.

REBAUDET, S., DÉLY, P., BONCY, J., HENRYS, J. H. and PIARROUX, R. (2021). Toward Cholera Elimination, Haiti. *Emerging infectious diseases* **27** 2932.

REBESCHINI, P. and VAN HANDEL, R. (2015). Can local particle filters beat the curse of dimensionality? *The Annals of Applied Probability* **25** 2809–2866.

SALTELLI, A., BAMMER, G., BRUNO, I., CHARTERS, E., DI FIORE, M., DIDIER, E., NELSON ESPELAND, W., KAY, J., LO PIANO, S., MAYO, D. et al. (2020). Five ways to ensure that models serve society: a manifesto.

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** 400–416.

TONI, T., WELCH, D., STRELKOWA, N., IPSEN, A. and STUMPF, M. P. H. (2009). Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. *Journal of the Royal Society Interface* **6** 187-202.

TRACY, M., CERDÁ, M. and KEYES, K. M. (2018). Agent-Based Modeling in Public Health: Current Applications and Future Directions. *Annual Review of Public Health* **39** 77-94. PMID: 29328870. https://doi.org/10.1146/annurev-publhealth-040617-014317

VARGHESE, A., KOLAMBAN, S., SHERIMON, V., LACAP, E. M., AHMED, S. S., SREEDHAR, J. P., AL HARTHI, H., SHUAILY, A. and SALIM, H. (2021). SEAMHCRD deterministic compartmental model based on clinical stages of infection for COVID-19 pandemic in Sultanate of Oman. *Scientific Reports* **11** 1–19.