Informing policy via dynamic models: Cholera in Haiti

Jesse Wheeler[1*], AnnaElaine L. Rosengart[2], Zhuoxun Jiang[1], Kevin Tan[3], Noah Treutle[1], Edward L. Ionides[1]

1 Statistics Department, University of Michigan, Ann Arbor, Michigan, USA
2 Statistics and Data Science, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA
3 Wharton Statistics and Data Science, University of Pennsylvania, Philadelphia, Pennsylvania, USA

* jeswheel@umich.edu

Abstract

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 decisions can be posed as statistical questions about scientifically motivated dynamic models. Thus, we encounter the methodological task of building credible, data-informed decisions based on stochastic, partially observed, nonlinear dynamic models. This necessitates addressing the tradeoff between biological fidelity and model simplicity, and the reality of misspecification for models at all levels of complexity. [I'VE RE-WRITTEN HERE, LOOKING TO BE MORE SPECIFIC ABOUT THE GOALS AND CONTRIBUTION OF THE PAPER.] ~~As a case study, we consider the 2010-2019 cholera epidemic in Haiti. We study three dynamic models developed by expert teams to advise on vaccination policies. We assess previous methods used for fitting and evaluating these models, and we develop data analysis strategies leading to improved statistical fit.~~ We assess current methodological approaches to these issues via a case study of the 2010-2019 cholera epidemic in Haiti. We consider three dynamic models developed by expert teams to advise on vaccination policies. We evaluate previous methods used for fitting these models, and we demonstrate modified data analysis strategies leading to improved statistical fit. Specifically, we present approaches to diagnosis of model misspecification, development of alternative models, and computational improvements in optimization, in the context of likelihood-based inference on nonlinear dynamic systems. Our workflow is reproducible and extendable, facilitating future investigations of this disease system.

Author summary

Quantitative understanding of infectious disease transmission dynamics relies upon mathematical models informed by scientific knowledge and relevant data. The models aim to provide a statistical description of the trajectory of an epidemic and its uncertainty, together with a representation of the underlying biological mechanisms. Evaluation of success at these goals is necessary in order for a model to provide a reliable tool for guiding evidence-based public policy interventions. In this article, we discuss common modeling decisions that may not provide an adequate statistical explanation of the data, and we show that attention to statistical fit can lead to more

accurate policy evaluations. We do this by conducting a re-analysis of the 2010 cholera outbreak in Haiti. Our analysis presents methodology for diagnosing how well a model describes observed data, and provides careful justification of valid conclusions from the fitted model. Objective measures are used to benchmark model fit; when these are combined with reproducibility, a framework emerges for continual improvement when revisiting the data and models. Our data analysis workflow is supported by recent advances in algorithms, software and hardware, which facilitate statistical fitting of nonlinear stochastic dynamic models to observed incidence data. However, inference for high-dimensional systems remains a methodological challenge. One of the models under consideration involves spatially coupled stochastic meta-populations, and we demonstrate how a recently developed algorithm permits likelihood-based inference and model diagnostics in this setting. We contend that raising the currently accepted standards of infectious disease modeling will result in a greater ability of scientists and policy makers to understand and respond to future infectious disease outbreaks.

Introduction

Regulation of biological populations is a fundamental topic in epidemiology, ecology, fisheries and agriculture. 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. Quantitative models for these dynamic systems offer potential for designing effective control measures. Developing and testing models for dynamic systems, 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 variations 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~~ public health policy 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. [1] 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. The data that were used for this study, and that is the subject of our reanalysis, is displayed in Fig. [1].~~ Various dynamic models were developed to study this outbreak: searching PubMed with keywords "Haiti, cholera, model" we obtained 20 deterministic models [1–20] and 8 stochastic models [21–28]. Here we focus on a multi-group modeling exercise by Lee et al. [29] 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. These modeling strategies were selected by [30] to represent the range of approaches used in the research community. We focus on Models 1–3, as the strengths and weaknesses of the agent-based modeling approach [31] are outside the scope of this article. [In addition, agent-based models were less widely used [NO OTHERS SHOW UP IN OUR LITERATURE REVIEW???].] The data that were used for this study, and that is the subject of our reanalysis, are displayed in Fig. 1.

The four independent teams were given the task of estimating the potential effect of prospective oral cholera vaccine (OCV) programs. While OCV is accepted as a safe and

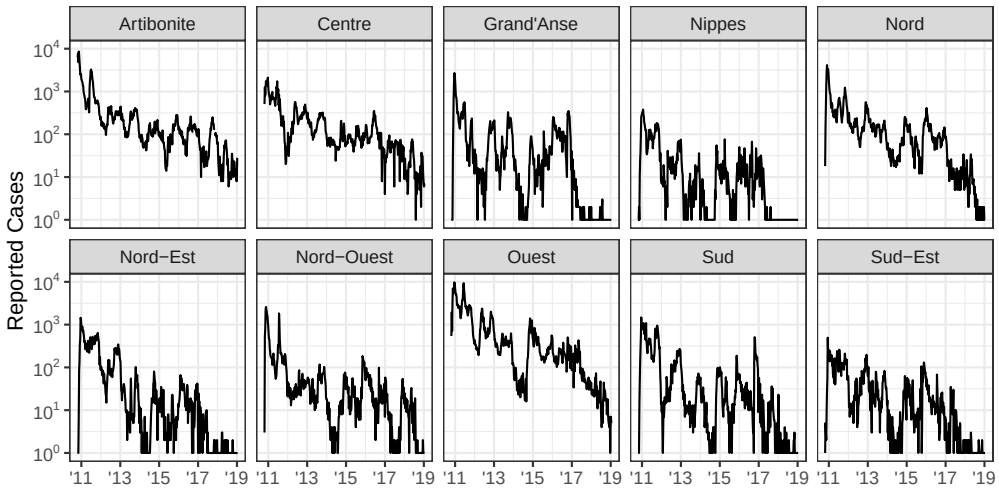

Fig 1. Weekly cholera cases. Weekly reported cholera cases in Haiti from October 2010 to January 2019 for each of the 10 administrative departments.

effective tool for controlling the spread of cholera, the global stockpile of OCV doses remains limited [32]. Advances in OCV technology and vaccine availability, however, raised the possibility of planning a national vaccination program [29]. In the study, 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. The groups largely adhered to existing guidelines for creating models to inform policy [33, 34] and, despite their autonomy, obtained a consensus that an extensive nationwide vaccination campaign would be necessary to eliminate cholera from Haiti. Their forecasts, however, were inconsistent with the prolonged period with no confirmed cholera cases between February, 2019 and September, 2022 [13]. Though cholera has recently reemerged in Haiti [35, 36], the inability to accurately forecast cholera incidence from 2019-2022—while diligently adhering to existing scientific standards for infectious disease modeling—demonstrates that existing standards of disease modeling and forecasting may be insufficient.

The discrepancy between the model-based conclusions of Lee et al. [29] and the prolonged absence of cholera in Haiti has been debated [37–40]. Suggested origins of this discrepancy include the use of unrealistic models [38] and unrealistic criteria for cholera elimination [39]. We find a more nuanced conclusion: attention to methodological details in model fitting, diagnosis and forecasting can improve each of the proposed model's ability to quantitatively describe observed data. This improved ability may result in more accurate forecasts and facilitates the exploration of model assumptions. Based on this retrospective analysis, we offer suggestions on fitting mechanistic models to dynamic systems for future studies.

~~While this work is focused on the models proposed by Lee et al. [1], our suggestions have broader relevance. To investigate the extent to which [1] is typical of the substantial body of work on Haiti cholera dynamics, we performed a literature review by searching PubMed with keywords: Haiti, cholera, model. The search resulted in 66 papers, of which 32 used dynamic models to describe the cholera epidemic in Haiti. The models make various choices on the dichotomies considered by Lee et al. [1]; deterministic versus stochastic; compartment model versus agent-based; aggregated versus spatially explicit.~~

We proceed by introducing the general modeling scheme employed by Models 1–3 and provide details of each individual model; we then describe how each model is calibrated to data, and present a systematic approach to examining and refining these models. We then use improved model fits to project cholera incidence in Haiti under various vaccination scenarios considered by Lee et al. [29]. Finally, we conclude with a discussion of the results.

Materials and methods

Mechanistic models for disease modeling

Models that focus on learning relationships between variables in a dataset are called *associative*, whereas models that incorporate a known scientific property of the system are called *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. [41] pointed out that it is naive to predict the effects of an intervention on a given system based entirely on historical associations. To successfully predict the effect of an intervention, a model should therefore both provide a quantitative explanation of existing data and should have a causal interpretation: a manipulation of the system should correspond quantitatively with the corresponding change to the model. This motivates the development of mechanistic models, which provides a statistical fit to the available data while also supporting a causal interpretation. [This paragraph and the one below seem to go well together, so perhaps this is a natural place to put this new paragraph.]

Mechanistic models representing biological phenomena are valuable for epidemiology and consequently for public health policy [42, 43]. More broadly, they have useful roles throughout biology, expecially when combined with statistical methods that properly account for stochasticity and nonlinearity [44]. In some situations, modern machine learning methods can outperform mechanistic models on epidemiological forecasting tasks [45]. The predictive skill of non-mechansitic models can reveal limitations in mechanistic models, but cannot readily replace the scientific understanding obtained by describing the biological dynamics of the system in a mathematical model.

The four mechanistic models of Lee et al. [29] were deliberately developed with limited coordination. This allows us to treat the models as fairly independently developed expert approaches to understanding cholera transmission. However, it led to differences in notation, and in subsets of the data chosen for analysis, that hinder direct comparison. Here, we have created a common notational framework that facilitates model comparison, and put all comparable model parameters—including parameters that were estimated or held constant—into Table 1. Translations back to the original notation of Lee et al. [29] are given in the supplement (S1 Table).

Each model describes the cholera dynamics as a partially observed Markov process (POMP) with a latent state vector $\boldsymbol{X}(t)$ for each continuous time point t. N observations on the system are collected at time points $t_1, \ldots, t_N$, written as $t_{1:N}$. The observation at time t_n is modeled by the random vector $\boldsymbol{Y}_n$. While the latent process exists between observation times, the value of the latent state at observations times is of particular interest. We therefore write $\boldsymbol{X}_n = \boldsymbol{X}(t_n)$ to denote the value of the latent process at the nth observation time, and $\boldsymbol{X}_{1:N}$ is the collection of latent state values for all observed time points. The observable random variables $\boldsymbol{Y}_{1:N}$ are assumed to be conditionally independent given $\boldsymbol{X}_{0:N}$. Together, with the density for the initial value of the latent state $\boldsymbol{X}_0 = \boldsymbol{X}(t_0)$, each model assumes a joint density $f_{\boldsymbol{X}_{0:N},\boldsymbol{Y}_{1:N}}\big(\boldsymbol{x}_{0:N}, \boldsymbol{y}_{1:N}; \theta\big)$, where θ is a parameter vector that indexes the model. The observed data $\boldsymbol{y}^*_{1:N}$, along with the unobserved true value of the latent state, are

Table 1. Model parameters.

Mechanism	Model 1	Model 2	Model 3
Infection (day)	$\mu_{IR}^{-1} = 2.0^{\dagger}$ (8)	$\mu_{IR}^{-1} = 7.0^{\dagger}$ (16)	$\mu_{IR}^{-1} = 5.0^{\dagger}$ (28)
Latency (day)	$\mu_{EI}^{-1} = 1.4^{\dagger}$ (7)	$\mu_{EI}^{-1} = 1.3^{\dagger}$ (15)	
Seasonality	$\beta_{1:6} = (1.4, 1.2, 1.1, 1.1, 1.4, 1.0)$ (4) $\zeta = -0.04^{*}$ (34)	$a = 0.4^{\dagger}$ (13) $\phi = 0.97^{*}$ (13)	$a = 0.99$ (32) $r = 0.843$ (32)
Immunity (yr)	$\mu_{RS}^{-1} = 8.0^{\dagger}$ (9)	$\mu_{RS}^{-1} = 1.4\times10^{11}$ (17) $\omega_1^{-1} = 1.0^{\dagger}$ (19) $\omega_2^{-1} = 5.0^{\dagger}$ (20)	$\mu_{RS}^{-1} = 8.0^{\dagger}$ (30)
Birth/death (yr^{-1})	$\mu_S = 10^{-2} \times 2.23^{\dagger}$ (11) $\delta = 10^{-3} \times 7.5^{\dagger}$ (11)		$\delta = 10^{-2} \times 1.59^{\dagger}$ (29) $\delta_C = 1.46^{\dagger}$
Sympt. frac.	$f_z(t) = c\vartheta^{*}(t - \tau_d)^{\dagger}$ (6-7)	$f = 0.2^{\dagger}$ (15)	$f = 0.25^{\dagger}$ (27)
Asympt. infectivity	$\epsilon = 0.05^{\dagger}$ (3)	$\epsilon = 0.001^{\dagger}$ (13) $\epsilon_W = 10^{-7\dagger}$ (21)	$\epsilon = 1^{\dagger}$ (25) $\epsilon_W = 0.030$ (32)
Human to human	$\beta_{1:6}$ as above (3)	$\beta = 5.97 \times 10^{-15}$ (13)	$\beta_{1:10} = (0.52, 0.06, 0.30, 0.20, 0.45, 0.35, 0.27, 0.11, 0.20, 0.26) \times 10^{-6}$ (25)
Water to human		$W_{\text{sat}} = 10^{5\dagger}$ (13) $\beta_W = 1.1$ (13)	$\beta_{W_{1:10}} = (2.91, 12.19, 14.93, 15.19, 3.11, 19.11, 6.33, 0.59, 7.32, 6.91)$ (25)
Human to water		$\mu_W = 9320$ (21)	$\mu_W = 6.61 \times 10^{-5}$ (32)
Water survival (wk)		$\delta_W^{-1} = 3^{\dagger}$ (22)	$\delta_W^{-1} = 0.13$ (33)
Mixing exponent	$\nu = 0.98$ (3)		
Process noise ($wk^{1/2}$)	$\sigma_{\text{proc}} = (0.30, 0.34)^{*}$ (3)		$\sigma_{\text{proc}} = 0.031$ (27)
Reporting rate	$\rho = 0.679$ (S15)	$\rho = 0.20^{\dagger}$ (S16)	$\rho = 0.99$ (S19)
Observation variance	$\psi = (279.15, 78.33)$ (S15)	$\psi = 1.319$ (S16)	$\psi = 105.43$ (S19)
Initial Values	$I_{0,0} = 7298$ $E_{0,0} = 350$		$I_{0,0}^{3,4} = (31, 28)^{*}$ (S20)
Hurricane Parameters			$\beta_{W3,9}^{hm} = (25.63, 34.28)^{*}$ (25) $h_{3,9}^{hm} = (87.56, 79.92)^{*}$ (25)

References to the relevant equation are given in parentheses. Parameters that were fixed and not calibrated using the data are indicated with †; all fixed parameters values were chosen to match the fixed parameter values of [29]. Parameters that were added during our re-analysis and were not considered by Lee et al. are indicated with *. Translations back into the notation of [29] are given in S1 Table.

modeled as a realization of this joint distribution.

Because of the probabilistic nature of both the unobserved latent state and the observable random variables, it is possible to consider various marginal and conditional densities of these two jointly random vectors. An important example is the marginal density of the observed random vector $\boldsymbol{Y}_{1:N}$, evaluated at the observed data $\boldsymbol{y}^*_{1:N}$, as shown in Eq. (1):

$$f_{\boldsymbol{Y}_{1:N}}\big(\boldsymbol{y}^*_{1:N};\theta\big) = \int f_{\boldsymbol{X}_{0:N},\boldsymbol{Y}_{1:N}}\big(\boldsymbol{x}_{0:N},\boldsymbol{y}^*_{1:N};\theta\big)d\boldsymbol{x}_{0:N}. \tag{1}$$

When treated as a function of the parameter vector θ, this marginal density is called the *likelihood function*, which is the basis of likelihood based statistical inference.

Using the conditional independence of $\boldsymbol{Y}_{1:N}$ given $\boldsymbol{X}_{0:N}$ and the Markov property of $\boldsymbol{X}_{0:N}$, the joint density can be refactored into the useful form given in Eq. (2):

$$f_{\boldsymbol{X}_{0:N},\boldsymbol{Y}_{1:N}}\big(\boldsymbol{x}_{0:N},\boldsymbol{y}_{1:N};\theta\big) = f_{\boldsymbol{X}_0}\big(\boldsymbol{x}_0;\theta\big)\prod_{n=1}^{N} f_{\boldsymbol{X}_n|\boldsymbol{X}_{n-1}}\big(\boldsymbol{x}_n|\boldsymbol{x}_{n-1};\theta\big)f_{\boldsymbol{Y}_n|\boldsymbol{X}_n}\big(\boldsymbol{y}_n|\boldsymbol{x}_n\big). \tag{2}$$

This factorization is useful because it demonstrates that POMP models may be completely described using three parts: the *initialization model* for the latent states $f_{\boldsymbol{X}_0}\big(\boldsymbol{x}_0;\theta\big)$; the *one-step transition density*, or *the process model* $f_{\boldsymbol{X}_n|\boldsymbol{X}_{n-1}}\big(\boldsymbol{x}_n|\boldsymbol{x}_{n-1};\theta\big)$; and the *measurement model* $f_{\boldsymbol{Y}_n|\boldsymbol{X}_n}\big(\boldsymbol{y}_n|\boldsymbol{x}_n\big)$. In the following subsections, we describe Models 1–3 in terms of these three components. The latent state vector $\boldsymbol{X}(t)$ for each model consists of individuals labeled as susceptible (S), infected (I), asymptomatically infected (A), vaccinated (V), and recovered (R), with various sub-divisions sometimes considered. The observable random vector $\boldsymbol{Y}_{1:N}$ represents the random vector of cholera incidence data for each model; Models 2 and 3 have metapopulation structure, meaning that each individual is a member of a spatial unit, denoted by a subscript $u \in 1{:}U$, in which case we denote the observed data for each unit using $\boldsymbol{y}^*_{1:N} = y^*_{1:N,1:U}$. Here, the spatial units are the $U = 10$ Haitian administrative départements (henceforth anglicized as departments).

While the complete model description is scientifically critical, as well as necessary for transparency and reproducibility, the model details are not essential to our methodological discussions of how to diagnose and address model misspecification with the purpose of informing policy. A first-time reader may choose to skim through the rest of this section, and return later. Additional details about the numeric implementation of these models are provided in a supplemental text (S1 Text). While each of the dynamic models considered in this manuscript can be fully described using the mathematical equations provided in the following section, diagrams of dynamic systems can be helpful to understand the equations. For this reason, we provide flow chart diagrams for Models 1–3 in supplement figures (S1 Fig, S2 Fig and S3 Fig).

Model 1

The latent state vector $\boldsymbol{X}(t) = \big(S_z(t), E_z(t), I_z(t), A_z(t), R_z(t), z \in 0:Z\big)$ describes susceptible, latent (exposed), infected (and symptomatic), asymptomatic, and recovered individuals in vaccine cohort z at time t. Here, $z = 0$ corresponds to unvaccinated individuals, and $z \in 1{:}Z$ describes hypothetical vaccination programs. The force of infection is

$$\lambda(t) = \Big(\sum_{z=0}^{Z} I_z(t) + \epsilon\sum_{z=0}^{Z} A_z(t)\Big)^{\nu}\frac{d\Gamma(t)}{dt}\beta(t)/N, \tag{3}$$

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

$$\log \beta(t) = \sum_{j=1}^{6} \beta_j s_j(t). \tag{4}$$

The process noise $d\Gamma(t)/dt$ is multiplicative Gamma-distributed white noise, with infinitesimal variance parameter σ^2_{proc}. Lee et al. [29] included process noise in Model 3 but not in Model 1, i.e., they fixed $\sigma^2_{\text{proc}} = 0$. Gamma white noise in the transmission rate gives rise to an over-dispersed latent Markov process [46] which has been found to improve the statistical fit of disease transmission models [47, 48].

For any time point in $t_{1:N}$, the process model $f_{\boldsymbol{X}_n|\boldsymbol{X}_{n-1}}\big(\boldsymbol{x}_n|\boldsymbol{x}_{n-1};\theta\big)$ is defined by describing how individuals move from one latent state compartment to another. Per-capita transition rates are given in Eqs. (5)-(12):

$$\begin{aligned}
\mu_{S_zE_z} &= \lambda(t), && (5)\\
\mu_{E_zI_z} &= \mu_{EI}\big(1-\vartheta_z(t)\big), && (6)\\
\mu_{E_zA_z} &= \mu_{EI}\,\vartheta_z(t), && (7)\\
\mu_{I_zR_z} &= \mu_{A_zR_z} = \mu_{IR}, && (8)\\
\mu_{R_zS_z} &= \mu_{RS}, && (9)\\
\mu_{S_0S_z} &= \mu_{E_0E_z} = \mu_{I_0I_z} = \mu_{A_0A_z} = \mu_{R_0R_z} = \eta_z(t), && (10)\\
\mu_{S_z\bullet} &= \mu_{E_z\bullet} = \mu_{I_z\bullet} = \mu_{A_z\bullet} = \mu_{R_z\bullet} = \delta, && (11)\\
\mu_{\bullet S_0} &= \mu_S, && (12)
\end{aligned}$$

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 $\bullet$ modeling entry into the study population due to birth or immigration, and exit from the study population due to death or immigration. Thus, $\mu_{A\bullet}$ is a rate of exiting the study population from compartment A and $\mu_{\bullet B}$ is a rate of entering the study population into compartment B.

In Model 1, 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. (3). In Eqs. (7) and (6) 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 $\vartheta_z(t) = c\,\vartheta^*(t-\tau_z)$ where $\vartheta^*(t)$ is efficacy at time t since vaccination for adults, taken from [29], Table S4, and $c = \big(1-(1-0.4688)\times 0.11\big)$ 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 modeled as equal between single and double dose until 52 weeks after vaccination, at which point the single dose becomes ineffective.

The latent state vector $\boldsymbol{X}(t)$ is initialized by setting the counts for each compartment and vaccination scenario $z \neq 0$ as zero, and introducing initial-value parameters $I_{0,0}$ and $E_{0,0}$ such that $R_0(0) = 0$, $I_0(0) = \text{Pop} \times I_{0,0}$, $E_0(0) = \text{Pop} \times E_{0,0}$ and $S_0(0) = \text{Pop} \times (1 - I_{0,0} - E_{0,0})$, where Pop is the total population of Haiti. The measurement model describes reported cholera cases at time point n come from a negative binomial distribution, where only a fraction (ρ) of new weekly cases are reported. More details about the initialization model $f_{\boldsymbol{X}_0}\big(\boldsymbol{x}_0;\theta\big)$ and the *measurement model* $f_{\boldsymbol{Y}_n|\boldsymbol{X}_n}\big(\boldsymbol{y}_n|\boldsymbol{x}_n\big)$ for Models 1–3 are provided a supplement text (S2 Text and S3 Text).

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.

Like Model 1, the process model $f_{\boldsymbol{X}_n|\boldsymbol{X}_{n-1}}\big(\boldsymbol{x}_n|\boldsymbol{x}_{n-1};\theta\big)$ is primarily defined via the description of movement of individuals between compartments, however Model 2 also includes a dynamic description of a latent bacterial compartment as well. 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^A_{uz}. The force of infection depends on both direct transmission and an aquatic reservoir, $W_u(t)$, and is given by

$$\lambda_u(t) = 0.5\big(1 + a\cos(2\pi t + \phi)\big)\frac{\beta_W\, W_u(t)}{W_{\text{sat}} + W_u(t)} + \beta\left\{\sum_{z=0}^{4} I_{uz}(t) + \epsilon\sum_{z=0}^{4} A_{uz}(t)\right\}. \quad (13)$$

The latent state is therefore described by the vector $\boldsymbol{X}(t) = \big(S_{uz}(t), E_{uz}(t), I_{uz}(t), A_{uz}(t), R_{uz}(t), R^A_{uz}(t), W_u, u \in 1{:}U, z \in 0{:}4\big)$. The cosine term in Eq. (13) accounts for annual seasonality, with a phase parameter ϕ. The Lee et al. [29] implementation of Model 2 fixes $\phi = 0$.

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

$$\begin{aligned}
\mu_{S_{uz}E_{uz}} &= \vartheta_z\,\lambda, && (14)\\
\mu_{E_{uz}I_{uz}} &= f\mu_{EI}, \quad \mu_{E_{uz}A_{uz}} = (1-f)\mu_{EI}, && (15)\\
\mu_{I_{uz}R_{uz}} &= \mu_{A_{uz}R^A_{uz}} = \mu_{IR}, && (16)\\
\mu_{R_{uz}S_{uz}} &= \mu_{R^A_{uz}S_{uz}} = \mu_{RS}, && (17)\\
\mu_{S_{uz}S_{vz}} &= \mu_{E_{uz}E_{vz}} = \mu_{I_{uz}I_{vz}} = \mu_{A_{uz}A_{vz}} = \mu_{R_{uz}R_{vz}} = \mu_{R^A_{uz}R^A_{vz}} = T_{uv}, && (18)\\
\mu_{S_{u1}S_{u0}} &= \mu_{S_{u3}S_{u0}} = \omega_1, && (19)\\
\mu_{S_{u2}S_{u0}} &= \mu_{S_{u4}S_{u0}} = \omega_2, && (20)\\
\mu_{\bullet W_u} &= \mu_W\left\{\sum_{z=0}^{4} I_{uz}(t) + \epsilon_W\sum_{z=0}^{4} A_{uz}(t)\right\}, && (21)\\
\mu_{W_u\bullet} &= \delta_W, && (22)\\
\mu_{W_uW_v} &= w_r T^W_{uv}. && (23)
\end{aligned}$$

In Eq. (18) the spatial coupling is specified by a gravity model,

$$T_{uv} = v_{\text{rate}} \times \frac{\text{Pop}_u\text{Pop}_v}{D^2_{uv}}, \quad (24)$$

where Pop_u is the mean population for department u, D_{uv} is a distance measure estimating average road distance between randomly chosen members of each population, and $v_{\text{rate}} = 10^{-12}\,\text{km}^2\text{yr}^{-1}$ was treated as a fixed constant. In Eq. (23), T^W_{uv} 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 Eq. (13). In Eq. (14), ϑ_z denotes the vaccine efficacy for each vaccination campaign $z \in Z$, with $\vartheta_0 = 1$, $\vartheta_1 = 1 - 0.429q$, $\vartheta_2 = 1 - 0.519q$, $\vartheta_3 = 1 - 0.429$, and $\vartheta_4 = 1 - 0.519$ Here, $q = 0.4688$ represents the reduced efficacy of the vaccination for children under the age of five years, and the values 0.429 and 0.519 are the median effectiveness of one and two doses over their effective period respectively, according to Table S4 in the supplement material of Lee et al. [29]. Because vaccine efficacy remains constant, individuals in this model transition from a vaccinated compartment to the susceptible compartment at the end of the vaccine coverage period.

The starting value for each element of the latent state vector $\boldsymbol{X}(0)$ are set to zero except for $I_{u0}(0) = y_u^*(0)/\rho$ and $R_{u0}(0) = \text{Pop}_u - I_{u0}(0)$, where $y_u^*(0)$ is the reported number of cholera cases in department u at time $t = 0$. Reported cases are described using a log-normal distribution, with the log-scale mean equal to the reporting rate ρ times the number of newly infected individuals. See the section describing how Model 2 is fit to data and Sec. S3.2 of the supplement for more details.

Model 3

The latent state is described as $\boldsymbol{X}(t) = \big(S_{uz}(t), I_u(t), A_u(t), R_{uzk}(t), W_u(t), u \in 0{:}U, z \in 0{:}4, k \in 1{:}3\big)$. 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 processes model $f_{\boldsymbol{X}_n|\boldsymbol{X}_{n-1}}\big(\boldsymbol{x}_n|\boldsymbol{x}_{n-1};\theta\big)$ describes the movement of individuals between latent compartments, as well as the birth and death process of local, unobserved bacterial compartments $W_u(t)$. The force of infection is

$$\lambda_u(t) = \Big(\beta_{W_u} + 1_{(t\geq t_{hm})}\beta_{W_u}^{hm} e^{-h_u^{hm}(t-t_{hm})}\Big)\,\frac{W_u(t)}{1+W_u(t)} + \beta_u \sum_{v\neq u}\big(I_v(t) + \epsilon A_v(t)\big), \quad (25)$$

where t_{hm} is the time Hurricane Matthew struck Haiti [49], and $1_{(A)}$ is the indicator function for event A. In [29], $\beta_{W_u}^{hm}$ and h_u^{hm} were set to zero for all u; the need to account for the effect Hurricane Matthew had on cholera transmission for this model is explored in Sec. S5 of the supplement.

Per-capita transition rates are used for both compartments representing human counts and the aquatic reservoir of bacteria; these rates are given in Eqs. (26)–(33).

$$\begin{aligned}
\mu_{S_{uz}I_u} &= f\,\lambda_u\big(1-\vartheta_{uz}(t)\big)\,d\Gamma/dt, && (26)\\
\mu_{S_{uz}A_u} &= (1-f)\,\lambda_u\big(1-\vartheta_{uz}(t)\big)\,d\Gamma/dt, && (27)\\
\mu_{I_uR_{uz1}} &= \mu_{A_uR_{uz1}} = \mu_{IR}, && (28)\\
\mu_{I_uS_{u0}} &= \delta + \delta_C, \;\; \mu_{A_uS_{u0}} = \delta && (29)\\
\mu_{R_{uz1}R_{uz2}} &= \mu_{R_{uz2}R_{uz3}} = 3\mu_{RS}, && (30)\\
\mu_{R_{uzk}S_{u0}} &= \delta + 3\mu_{RS}\,\mathbf{1}_{\{k=3\}}, && (31)\\
\mu_{\bullet W_u} &= \big[1 + a\big(J_u(t)\big)^r\big]\text{Den}_u\,\mu_W\big[I_u(t) + \epsilon_W A_u(t)\big], && (32)\\
\mu_{W_u\bullet} &= \delta_W. && (33)
\end{aligned}$$

As with Model 1, $d\Gamma_u(t)/dt$ is multiplicative Gamma-distributed white noise in Eqs. (26) and (27). In Eq. (32), $J_u(t)$ is a dimensionless measurement of precipitation that has been standardized by dividing the observed rainfall at time t by the maximum recorded rainfall in department u during the epidemic, and Den_u is the 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 Eq. (29) to account for cholera induced deaths among infected individuals. All deaths are balanced by births into the susceptible compartment in Eqs. (29) and (31), thereby maintaining constant population in each department.

Similar to Model 1, there are no distinct compartments for individuals under five years of age, and the vaccination efficacy is taken as a age adjusted weighted average of the efficacy for individuals both over and under five years of age: $\vartheta_{uz}(t) = c\vartheta^*(t - \tau_{uz})$, where τ_{uz} is the vaccination time for unit u and vaccination campaign z. The value c and the function ϑ^* are equivalent to those described in the Model 1 description.

Latent states are initialized using an approximation of the instantaneous number of infected, asymptomatic, and recovered individuals at time t_0 by using the first week of cholera incidence data. Specifically, we set $I_{u0}(0) = \frac{y^*_{1u}}{\rho(\delta+\delta_C+\mu_{IR})}$, $A_{u0}(0) = \frac{1-f}{f} I_{u0}(0)$, $R_{u0k} = y^*_{1u} - I_{u0}(0) - A_{u0}(0)$, and we initialize $W_u(0)$ by enforcing the rainfall dynamics supposed by the one step transition model; all other compartments that represent population counts are set to zero at time t_0. For each unit u with zero case counts at time t_1, this initialization scheme results in having zero individuals in the Infected and Asymptomatic compartments, as well as having no bacteria in the aquatic reservoir. In reality, it is plausible that some bacteria or infected individuals were present in unit u but went unreported. Therefore, for departments with zero case counts in week 1, we consider estimating the number of infected individuals rather then treating this value as a constant (see supplement Sec. S5 for more details). Finally, reported cholera cases are modeled using a negative binomial measurement model with mean equal to a fraction (ρ) of individuals in each unit who develop symptoms and seek healthcare (see Sec. S3.3 of the supplement).

Model Fitting

Each of the three models considered in this study describes cholera dynamics as a partially observed Markov process (POMP), with the understanding that the deterministic Model 2 is a degenerate case of a stochastic model. Each model is indexed by a parameter vector, θ, and different values of θ can result in qualitative differences in the predicted behavior of the system. Therefore, the choice of θ used to make inference about the system can greatly affect model-based conclusions. Elements of θ can be fixed at a constant value based on scientific understanding of the system, but parameters can also be calibrated to data by maximizing a measure of congruency between the observed data and the model's mechanistic structure. Calibrating model parameters to observed data does not guarantee that the resulting model successfully approximates real-world mechanisms, since the model description of the dynamic system may be incorrect and do not change as the model is calibrated to data. However, the congruency between the model and observed data serves as a proxy for the congruency between the model and the true underlying dynamic system. As such, it is desirable to obtain the best possible fit of the proposed mechanistic structure to the observed data.

In this article we ~~calibrate~~ follow [29] by calibrating the parameters of each of our models ~~by maximizing the likelihood~~ using maximum likelihood, as described in Eq. (1). The likelihood for each of the fitted models—and the corresponding AIC values for model comparisons that include an adjustment for the number of calibrated parameters—is provided in Table 2. In the following subsections we describe in detail our approach to calibrating the three proposed mechanistic models to observed cholera incidence data. [I'VE STARTED TO PUT IN SOME DISCUSSION OF BAYESIAN METHODS HERE, THOUGH IT COULD ALSO GO IN THE DISCUSSION] The main alternative to maximum likelihood estimation is Bayesian inference via Markov

chain Monte Carlo, used to analyze the Haiti cholera epidemic by [3, 7, 13, 19, 20, 23, 25–27]. Alternatively, some investigations obtained parameters from previous sources without calibrating to data.

Table 2. AIC values for Models 1–3 and their benchmarks.

	Model 1	Model 2	Model 3
Log-likelihood	−2728.1 (−3030.9)[1]	−21957.3 (−29367.4)	−17361.9 (−33832.6)[2]
Number of Fit Parameters	15 (20)	6 (6)	34 (29)
AIC	5486.3 (6101.8)[1]	43926.5 (58746.9)	34791.8 (67723.2)[2]
Benchmark AIC	5585.3	36961.0	35945.2

Values in parentheses are corresponding values using Lee et al. [29] parameter estimates.
[1]The reported likelihood is an upper bound of the likelihood of the Lee et al. [29] model.
[2]Lee et al. [29] fit Model 3 to a subset of the data (March 2014 onward, excluding data from Ouest in 2015-2016). On this subset, their model has a likelihood of −9721.2. On this same subset, our model has a likelihood of −7213.5. Details of estimating the likelihood of the Lee et al. [29] models is provided in S4 Text See Sec. S6 of the supplement material for more details on estimating the likelihood of the Lee et al. [29] models.

Calibrating Model 1 Parameters

Model 1 was developed with the intent of providing a description of cholera dynamics that is scientifically plausible rather than one that is statistically convenient. This results in the inability to obtain a closed form expression of the non-linear joint model density—described in Eq. (2). Therefore in order to perform likelihood based inference on this model, we are restricted to use parameter estimation techniques that have the *plug-and-play* property, which is that the fitting procedure only requires the ability to simulate the latent process rather than evaluating transition densities [48, 50]; in the context of the notation and definitions employed in this article, this means that we only require the ability to simulate from $f_{\boldsymbol{X}_0}(\boldsymbol{x}_0;\theta)$ and $f_{\boldsymbol{X}_n|\boldsymbol{X}_{n-1}}(\boldsymbol{x}_n|\boldsymbol{x}_{n-1};\theta)$ rather than needing to evaluate these densities. Plug-and-play algorithms include Bayesian approaches like ABC and PMCMC [51, 52], but here we use algorithms that enable maximum likelihood estimation. To our knowledge, the only plug-and-play methods that can be used to maximize the likelihood for POMP models of this complexity are iterated filtering algorithms [53], which modify the well-known *particle filter* [54].

The particle filter, also referred to as sequential Monte Carlo, is a simulation based method that is frequently used in Bayesian inference to approximate the posterior distribution of latent states. This algorithm can also be used to accurately approximate the log-likelihood of a POMP model, defined as the integral in Eq. (1). Iterated filtering—algorithms such as IF2 [53]—extend the particle filter 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 iteration, and decreasing the random walk variance at each step. With a sufficient number of iterations, the resulting parameter values converge to a region of the parameter space that maximizes the model likelihood.

The ability to maximize the likelihood allows for likelihood-based inference, such as performing statistical tests for potential model improvements. We demonstrate this capability by proposing a log-linear trend ζ in transmission in Eq. (4):

$$\log \beta(t) = \sum_{j=1}^{6} \beta_s s_j(t) + \zeta \bar{t}, \tag{34}$$

where $\bar{t} = \frac{t-(t_N+t_0)/2}{t_N-(t_N+t_0)/2}$, so that $\bar{t} \in [-1, 1]$. 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 [55], may have substantially reduced cholera transmission over time [56].

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 95% confidence interval via a Monte Carlo Adjusted Profile (MCAP) [57]. Lee et al. [29] implemented Model 1 by fitting two distinct phases: an epidemic phase from October 2010 through March 2015, and an endemic phase from March 2015 onward. We similarly allow the re-estimation of process and measurement overdispersion parameters (σ^2_{proc} and ψ), and require that the latent Markov process $X(t)$ carry over from one phase into the next. The resulting 95% confidence interval for ζ is $(-0.098, -0.009)$, with the full results displayed in Fig. 2. These results are suggestive that the inclusion of a trend in the transmission rate improves the quantitative ability of Model 1 to describe the observed data. The maximum likelihood estimate for the trend corresponds with a 7.3% reduction to the transmission rate over the course of the outbreak, with a 95% confidence interval (1.8%, 17.9%) for the overall reduction in transmission.. The reported results for Model 1 in the remainder of this article were obtained with the inclusion of the parameter ζ.

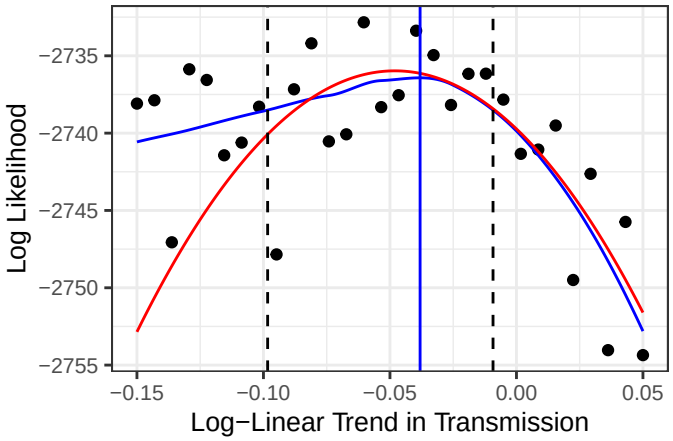

Fig 2. Confidence interval for the log-linear trend in transmission. Monte Carlo adjusted profile (MCAP) of ζ for Model 1. The blue curve is the MCAP, the vertical blue line indicates the MLE, and the vertical dashed lines indicate the 95% confidence interval.

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 competitive fit compared to associative benchmarks, 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. The statistical evidence of a trend in transmission rate in this model could be explained by any trending variable (such as hygiene improvements, or changes in population behavior), resulting in confounding from collinear 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 do not provide enough information to pin down the mechanism.

We implemented Model 1 using the `pomp` package [58], relying heavily on the `pomp` source code provided by Lee et al. [29]. Both analyses used the `mif2` implementation of the IF2 algorithm to estimate θ by maximum likelihood. One change we made in the statistical analysis that led to larger model likelihoods was increasing the computational effort in the numerical maximization. While IF2 enables parameter estimation for a large class of models, the theoretic ability to maximize the likelihood depends on asymptotics in both the number of particles and the number of filtering iterations. Many Monte Carlo replications are then required to quantify and further reduce the error. The large increase in the log-likelihood for Model 1 (see Table 2) can primarily be attributed to increasing the computational effort in fitting the model; this result highlights the importance of carefully determining the necessary computational effort needed to maximize model likelihoods and acting accordingly. Simulations from the initial conditions of our fitted model are plotted against the observed incidence data in Fig. 3.

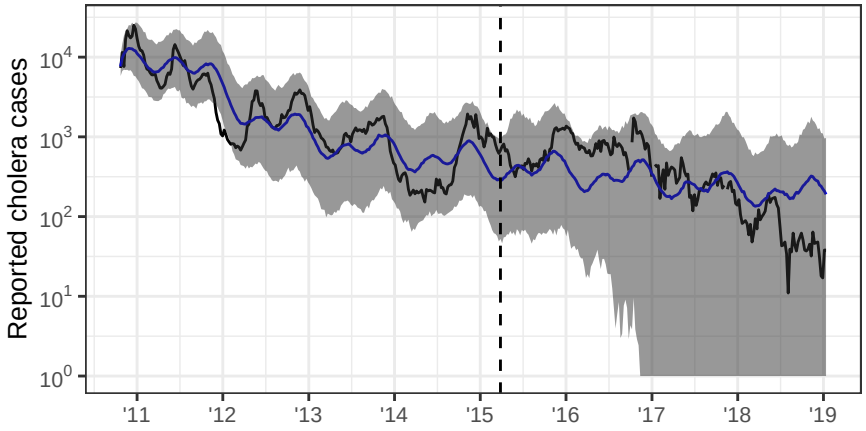

Fig 3. Simulations from Model 1 compared to reported cholera cases. The black curve is observed data, the blue curve is median of 500 simulations from the fitted model, and the vertical dashed line represents break-point when parameters are refit.

Calibrating Model 2 Parameters

Model 2 is a deterministic compartmental model defined by a set of coupled differential equations. The use of deterministic compartment models have a long history in the field of infectious disease epidemiology [59–61], and can be justified by asymptotic

considerations in a large-population limit [62, 63]. Because the process model of Model 2 is deterministic, maximum likelihood estimation reduces to a least squares calculation when combined with a Gaussian measurement model (see Sec. S3.2 of the supplement material for details). Lee et al. [29] 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. The inclusion of a change-point in model states and parameters increases the flexibility of the model and hence the ability to fit the observed data. The increase in model flexibility, however, results in hidden states that are inconsistent between model phases. The inclusion of a model break-point by Lee et al. [29] is perhaps due to a challenging feature of fitting a deterministic model via least squares: discrepancies between model trajectories and observed case counts in highly infectious periods of a disease outbreak will result in greater penalty than the discrepancies between model trajectories and observed case counts in times of relatively low infectiousness, resulting in a bias towards accurately describing periods of high infectiousness. This issue is particularly troublesome for modeling cholera dynamics in Haiti: the inability to accurately fit times of low infectiousness may result in poor model forecasts, as few cases of cholera were observed in the last few years of the epidemic.

To combat this issue, we fit the model to log-transformed case counts, since the log scale stabilizes the variation during periods of high and low incidence. An alternative solution is to change the measurement model to include overdispersion, as was done in Models 1 and 3. This permits the consideration of demographic stochasticity, which is dominant for small infected populations, together with log scale stochasticity (also called multiplicative, or environmental, or extra-demographic) which is dominant at high population counts. Here we chose to fit the model to transformed case counts rather than adding overdispersion to the measurement model with the goal of minimizing the changes to the model proposed by Lee et al. [29].

We implemented this model using the `spatPomp` `R` package [64]. The model was then fit using the subplex algorithm, implemented in the `subplex` package [65]. A comparison of the trajectory of the fitted model to the data is given in Fig. 4.

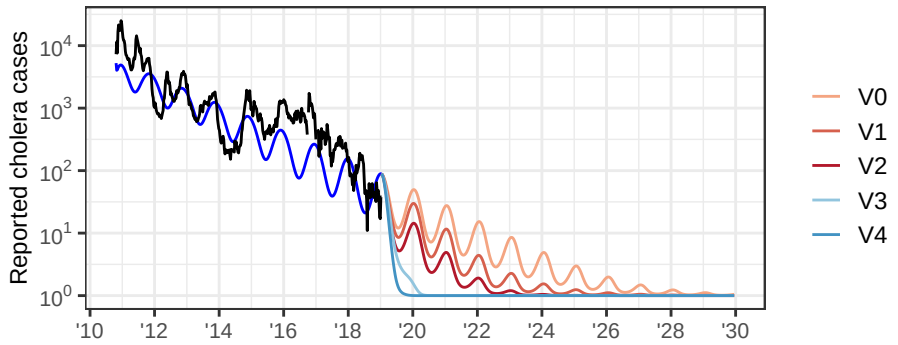

Fig 4. Simulated trajectory of Model 2. The black line shows the nationally aggregated weekly cholera incidence data. The blue curve from 2012-2019 is the trajectory of the calibrated version of Model 2. Projections under the various vaccination scenarios, which are discussed in detail in the **Forecasts** subsection are also included.

Calibrating Model 3 Parameters

Model 3 describes cholera dynamics in Haiti using a metapopulation model, where the hidden states in each administrative department has an effect on the dynamics in other

departments. The decision to address metapopulation dynamics using a spatially explicit model, rather than to aggregate over space, is double-edged. Evidence for the former approach has been provided in previous studies [66], including the specific case of heterogeneity between Haitian departments in cholera transmission [15]. However, a legitimate preference for simplicity can support a decision to consider nationally aggregated models [34, 67]. One issue that arises when fitting spatially explicit models is that parameter estimation techniques based on the particle filter become computationally intractable as the number of spatial units increases. This is a result of the approximation error of particle filters growing exponentially in the dimension of the model [68, 69].

To avoid the approximation error present in high-dimensional models, Lee et al. [29] simplified the problem of estimating the parameters of Model 3 by creating an approximate version of the model where the units are independent given the observed data. Reducing a spatially coupled model to individual units in this fashion requires special treatment of any interactive mechanisms between spatial units, such as found in Eq. (25). Because the simplified, spatially-decoupled version of Model 3 implemented in [29] relies on the observed cholera cases, the calibrated model cannot readily be used to obtain forecasts. Therefore, in order to obtain model forecasts, Lee et al. [29] used the parameters estimates from the spatially-decoupled approximation of Model 3 to obtain forecasts using the fully coupled version of the model. This approach of model calibration and forecasting avoids the issue of particle depletion, but may also be problematic. One concern is that cholera dynamics in department u are highly related to the dynamics in the remaining departments; calibrating model parameters while conditioning on the observed cases in other departments may therefore lead to an over-dependence on observed cholera cases. Another concern is that the two versions of the model are not the same, resulting in sub-optimal parameter estimates for the spatially coupled model, as parameters that maximize the likelihood of the decoupled model almost certainly do not maximize the likelihood of the fully coupled model. These two concerns may explain the unrealistic forecasts and low likelihood of Model 3 of Lee et al. [29] (see Table 2).

At the time Lee et al. [29] conducted their study, there was no known algorithm that could readily be used to maximize the likelihood of an arbitrary meta-population POMP model with coupled spatial dynamics, which justifies the spatial decoupling approximation that was used to calibrate model parameters. For our analysis, we calibrate the parameters of the spatially coupled version of Model 3 using the iterated block particle filter (IBPF) algorithm [70]. This algorithm extends the work of Ning and Ionides [71], who provided theoretic justification for the version of the algorithm that only estimates unit-specific parameters. The IBPF algorithm enables us to directly estimate the parameters of models describing high-dimensional partially-observed nonlinear dynamic systems via likelihood maximization. The ability to directly estimate parameters of Model 3 is responsible for the large increase in model likelihoods reported in Table 2. Simulations from the fitted model are displayed in Fig. 5.

Model Diagnostics

The goal of parameter calibration—whether done using Bayesian or frequentist methods—is to find the best description of the observed data under the assumptions of the model. Obtaining the best fitting set of parameters for a given model does not, however, guarantee that the model provides an accurate representation of the system under investigation. Model misspecification, which may be thought of as the omission of a mechanism in the model that is an important feature of the dynamic system, is inevitable at all levels of model complexity. To make progress, while accepting proper limitations, one must bear in mind the much-quoted observation of George Box [72] that

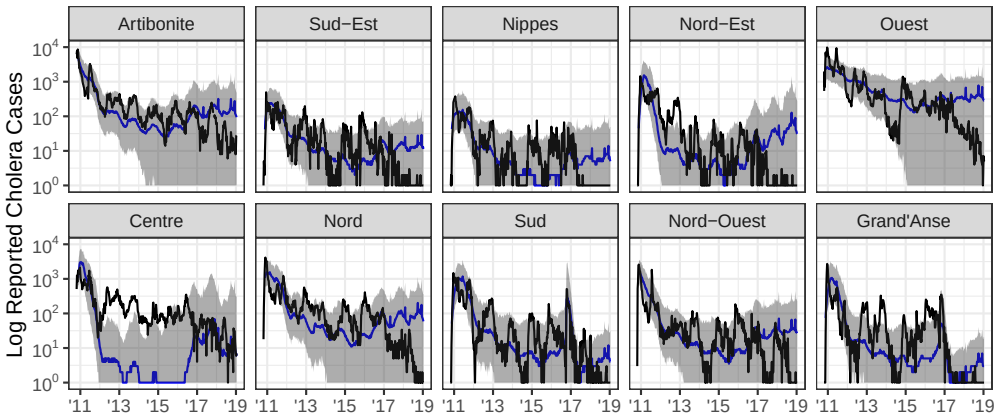

Fig 5. Simulations from Model 3 compared to reported cholera cases. Simulations from initial conditions using the spatially coupled version of Model 3. The black curve represents true case count, the blue line the median of 500 simulations from the model, and the gray ribbons representing 95% confidence interval.

"all models are wrong but some are useful." Beyond being good practical advice for applied statistics, this assertion is relevant for the philosophical justification of statistical inference as severe testing [73]. In this section, we discuss some tools for diagnosing mechanistic models with the goal of making the subjective assessment of model "usefulness" more objective. To do this, we will rely on the quantitative ability of the model to match the observed data, which we call the model's *goodness-of-fit*, with the guiding principle that a model which cannot adequately describe observed data may not be reliable for useful purposes. Goodness-of-fit may provide evidence supporting the causal interpretation of one model versus another, 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 simulations from the fitted model to the observed data. Visual inspection may indicate defects in the model, or may suggest that the observed data are a plausible realization of the fitted model. While visual comparisons can be informative, they provide only a weak and informal measure of the goodness-of-fit of a model. The study by Lee et al. [29] provides an example of this: their models and parameter estimates resulted in simulations that visually resembled the observed data, yet resulted in model likelihoods that were considerably smaller than likelihoods that can be achieved via the likelihood based optimization techniques that were used (see Table 2). Alternative forms of model validation should therefore be used in conjunction with visual comparisons of simulations to observed data.

Another approach is to compare a quantitative measure of the model fit (such as MSE, predictive accuracy, or model likelihood) among all proposed models. These comparisons, which provide insight into how each model performs relative to the others, are quite common [7, 25]. To calibrate relative measures of fit, it is useful to compare against a model that has well-understood statistical ability to fit data, and we call this model a *benchmark*. Standard statistical models, interpreted as associative models without requiring any mechanistic interpretation of their parameters, provide suitable benchmarks. Examples include linear regression, auto regressive moving average time series models, or even independent and identically distributed measurements. Benchmarks enable us to evaluate the goodness of fit that can be expected of a suitable mechanistic model.

Associative models are not constrained to have a causal interpretation, and typically are designed with the sole goal of providing a statistical fit to data. Therefore, we should not require a candidate mechanistic model to beat all benchmarks. However, a mechanistic model which falls far short against benchmarks is evidently failing to explain some substantial aspect of the data. A convenient measure of fit should have interpretable differences that help to operationalize the meaning of far short. Ideally, the measure should also have favorable theoretical properties. Consequently, we focus on log-likelihood as a measure of goodness of fit, and we adjust for the degrees of freedom of the models to be compared by using the Akaike information criterion (AIC) [74].

In some cases, a possible benchmark model could be a generally accepted mechanistic model, but often no such model is available. Because of this, we use a simple negative binomial model with an auto regressive mean as our associative benchmark; this model is described in (35).

$$Y_n|Y_{n-1} \sim \mathrm{NB}\big(\alpha + \beta Y_{n-1}, \varphi\big), \tag{35}$$

where $\mathrm{E}\big(Y_n|Y_{n-1}\big) = \alpha + \beta Y_{n-1}$, and $\mathrm{Var}\big(Y_n|Y_{n-1}\big) = \mathrm{E}\big(Y_n|Y_{n-1}\big) + \mathrm{E}\big(Y_n|Y_{n-1}\big)^2/\varphi$. In the case where the case counts are large, an alternative benchmark recommended by He et al. [48] is a log-linear Gaussian ARMA model; the theory and practice of ARMA models is well developed, and these linear models are appropriate on a log scale due to the exponential growth and decay characteristic of biological dynamics. We chose to use the auto regressive negative binomial model, however, because the large number of weeks with zero recorded cholera cases in department level data makes a benchmark based on a continuous distribution problematic. Likelihoods of Models 1–3 and their respective ARMA benchmark models are provided in Table 2.

It should be universal practice to present measures of goodness of fit for published models, and mechanistic models should be compared against benchmarks. In our literature review of the Haiti cholera epidemic, no non-mechanistic benchmark models were considered in any of the 32 papers that used dynamic models to describe cholera in order to obtain scientific conclusions. Including benchmarks would help authors and readers to detect and confront any major statistical limitations of the proposed mechanistic models. In addition, the published goodness of fit provides a concrete point of comparison for subsequent scientific investigations. When combined with online availability of data and code, objective measures of fit provide a powerful tool to accelerate scientific progress, following the paradigm of the *common task framework* [75].

The use of benchmarks may also be beneficial when developing models at differing spatial scales, where a direct comparison between model likelihoods is meaningless. In such a case, a benchmark model can be fit to each spatial resolution being considered, and each model compared to their respective benchmark. Large advantages (or shortcomings) in model likelihood relative to the benchmark for a given spatial scale that are not present in other spatial scales may provide weak evidence for (or against) the statistical fit of models across a range of spatial resolutions.

Comparing model log-likelihoods to a suitable benchmark may not be sufficient to identify all the strengths and weaknesses of a given model. Additional techniques include the inspection of conditional log-likelihoods of each observation given the previous observations in order to understand how well the model describes each data point (Sec. S5). Other tools include plotting the effective sample size of each observation [76]; plotting the values of the hidden states from simulations (S5 Text); and comparing summary statistics of the observed data to simulations from the model [66, 77].

Corroborating Fitted Models with Scientific Knowledge

The resulting mechanisms in a fitted model can be compared to current scientific knowledge about a system. Agreement between model-based inference and our current understanding of a system may be taken as a confirmation of both model-based conclusions and our scientific understanding. On the other hand, comparisons may generate unexpected results that have the potential to spark new scientific knowledge [78].

In the context of our case study, we demonstrate how the fit of Model 1 corroborates other evidence concerning the role of rainfall in cholera epidemics. Specifically, we examine the results of fitting the flexible cubic spline term in Model 1 (Eqs. (3)–(4)). The cubic splines permit flexible estimation of seasonality in the force of infection, $\beta(t)$. 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 is consistent with previous studies finding that rainfall played an important role in cholera transmission in Haiti [6, 79]. The estimated seasonality also features an increased transmission rate during the fall, which was noticed at an earlier stage of the epidemic [7]; the high transmission rate in the fall may also be a result of the increase transmission that occurred in the fall of 2016, when hurricane Matthew struck Haiti [49].

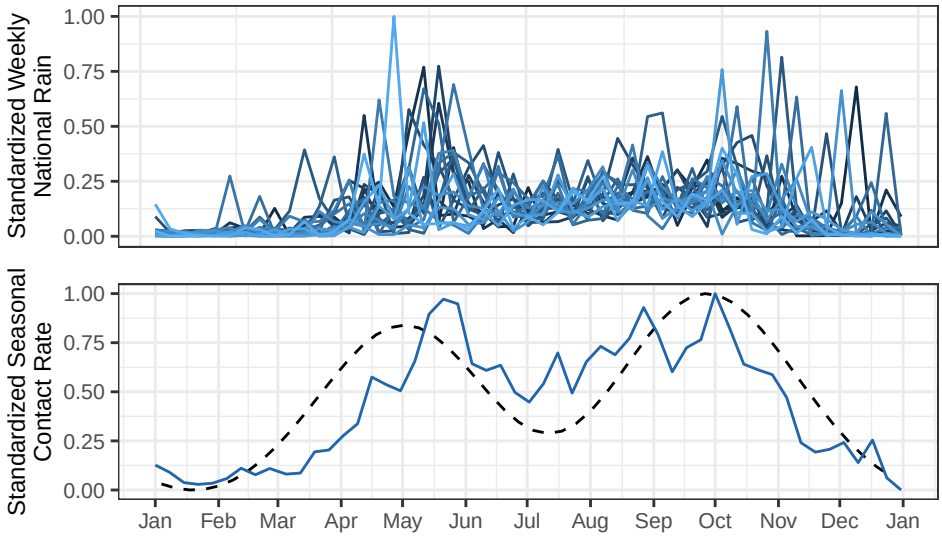

Fig 6. Seasonality of Model 1 transmission compared to rainfall data. (Top) weekly rainfall in Haiti, lighter colors representing more recent years. (Bottom) estimated seasonality in the transmission rate (dashed line) plotted alongside mean rainfall (solid line). The outsided effect of rainfall in the fall may be due to Hurricane Matthew, which struck Haiti in October of 2016 and resulted in an increase of cholera cases in the nation.

For any model-based inference, it is important to recognize and assess the modeling simplifications and assumptions that were used in order to arrive at the conclusions. In epidemiological studies, for example, quantitative understanding of individual-level processes may not perfectly match model parameters that were fit to population-level case counts, even when the model provides a strong statistical fit [48]. This makes direct interpretation of estimated parameters delicate.

Our case study provides an example of this in the parameter estimate for the duration of natural immunity due to cholera infection, μ_{RS}^{-1}. Under the framework of Model 2, the best estimate for this parameter is 1.4×10^{11} yr, suggesting that

individuals have effectively permanent immunity to cholera once infected. To interpret this result, we bear in mind that the data ranged from 2010-2019, and therefore estimates of immunity longer than 10 yr—the upper end of previous estimates of natural immunity [80]—effectively result in the same model dynamics. The depletion of susceptible individuals may also be attributed to confounding mechanisms—such as localized vaccination programs and non-pharmaceutical interventions that reduce cholera transmission [13, 56]—that were not accounted for in the model. Perhaps the best interpretation of the estimated parameter, then, is that under the modeling framework that was used, the model most adequately describes the observed data by having a steady decrease in the number of susceptible individuals. The weak statistical fit of Model 2 compared to a log-linear benchmark (see Table 2) cautions us against drawing quantitative conclusions from this model. A model that has a poor statistical fit may nevertheless provide a useful conceptual framework for thinking about the system under investigation. However, a claim that the model has been validated against data should be reserved for situations where the model provides a statistical fit that is competitive against alternative explanations.

A model which aspires to provide quantitative guidance for assessing interventions should provide a quantitative statistical fit for available 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 measurements of these variables [81]. In the context of our case study, by reconstructing the latent states of Model 3, we notice that the calibrated model favors higher levels of cholera transmission than what was typically observed in the incidence data (S5 Text). This result hints at the possibility of model mispecification, and should prompt us to be sceptical of the reliability of forecasts from this model. Similarly, parameters that have been calibrated to data should make sense in the context of alternative lines of evidence about the phenomena being modeled, while making allowance for the possibility that the interpretations of parameters may vary when modeling across differing spatial scales.

Results

Forecasts

Forecasts are an attempt to provide an accurate estimate of the future state of a system based on currently available data, together with an assessment of uncertainty. Forecasts from mechanistic models that are compatible with current scientific understanding may also provide estimates of the future effects of potential interventions. Further, they may enable real-time testing of new scientific hypotheses [82].

Recent information about a dynamic system should be more relevant for a forecast than older information. This assertion may seem self-evident, but it is not the case for deterministic models, for which the initial conditions together with the parameters are sufficient for forecasting, and so recent data do not have special importance. Epidemiological forecasts based on deterministic models are not uncommon in practice, despite their limitations [66]. That may explain why Lee et al. [29] chose to obtain forecasts by simulating the calibrated models forward from initial conditions. We obtain forecasts by simulating future values using latent states that are consistent with the most recent observations; this is done by drawing latent states at the last observation time (t_N) from the filtering distribution $f_{\boldsymbol{X}_N|\boldsymbol{Y}_{1:N}}(\boldsymbol{x}_N|\boldsymbol{y}^*_{1:N};\hat{\theta})$. The decision not to do this partially explains the unsuccessful forecasts of Lee et al. [29]: their Table S7 shows that the subset of their simulations which were consistent with observing zero cases in

2019 also accurately predicted the prolonged absence of detected cholera.

Uncertainty in just a single parameter can lead to drastically different forecasts [34]. Therefore, parameter uncertainty should also be considered when obtaining model forecasts to influence policy. If a Bayesian technique is used for parameter estimation, a natural way to account for parameter uncertainty is to obtain simulations from the model where each simulation is obtained using parameters drawn from the estimated posterior distribution. For frequentist inference, one possible approach is obtaining model forecasts from various values of θ, where the values of θ are sampled proportionally according to their corresponding likelihoods [66] (S6 Text). Both of these approaches share the similarity that parameters are chosen for the forecast approximately in proportion to their corresponding value of the likelihood function, $f_{\boldsymbol{Y}_{1:N}}(\boldsymbol{y}^*_{1:N};\theta)$. In this analysis, we do not construct forecasts accounting for parameter uncertainty as our focus is on the estimation and diagnosis of mechanistic models. Furthermore, we use the projections from a single point estimate to highlight the deficiency of deterministic models that the only variability in model projections is a result of parameter uncertainty, which can lead to over-confidence in forecasts [66].

The primary forecasting goal of Lee et al. [29] was to investigate the potential consequences of vaccination interventions on a system to inform policy. One outcome of their study include estimates for the probability of cholera elimination under several possible vaccination scenarios. Mimicking their approach, we define cholera elimination as an absence of cholera infections for at least 52 consecutive weeks, and we provide forecasts under the following vaccination scenarios:

- $V0$: No additional vaccines are administered.
- $V1$: Vaccination limited to the departments of Centre and Artibonite, deployed over a two-year period.
- $V2$: Vaccination limited to three departments: Artibonite, Centre, and Ouest deployed 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. Estimates of the probability of cholera elimination can therefore be obtained as the proportion of simulations from these models that result in cholera elimination. The results of these projections are summarized in Figs. 7.

Probability of elimination estimates of this form are not meaningful for deterministic models, as the trajectory of these models only represent the mean behavior of the system rather than individual potential outcomes. We therefore do not provide probability of elimination estimates under Model 2, but show trajectories under the various vaccination scenarios using this model (Fig. 4).

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 [34]. This article demonstrates that fitting appropriate scientific models to guide policy is a challenging statistical task. In our case study, we found that additional attention to statistical details could have resulted in improved statistical fits to the observed incidence data, leading to improvements in the accuracy of the resulting policy guidance.

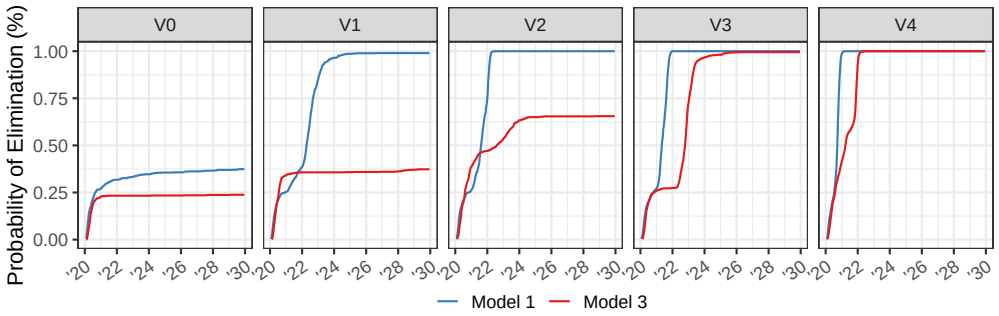

Fig 7. Simulated probability of elimination using Models 1 and 3. Probability of cholera elimination, defined as having zero cholera infectious for at least 52 consecutive weeks, based on 10 year simulations from calibrated versions of Models 1 and 3. Compare to Fig. 3A of Lee et al. [29].

We used the same data and models, and even much of the same code, as Lee et al. [29], and yet ended up with drastically different conclusions. We acknowledge the benefit of hindsight: our demonstration of a statistically principled route to obtain better-fitting models resulting in more robust insights does not rule out the possibility of discovering other models that fit well yet predict poorly.

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. Researchers should check that the computationally intensive numerical calculations are carried out adequately. Comparison against benchmarks and alternative model specifications should be considered to evaluate the statistical goodness-of-fit. Once that is accomplished, care is required to assess what causal conclusions can properly be inferred given the possibility of alternative 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 [34, 78, 83] should motivate us to follow best practices in data analysis, rather than avoiding the challenge.

Reproducibility and Extendability

Lee et al. [29] published their code and data online, and this reproducibility facilitated our work. 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 not be excessively burdensome. Scientific results are only trustworthy so far as they can be critically questioned, and an extendable analysis should facilitate such examination [84].

We provide a strong form of reproducibility, as well as extendability, by developing our analysis in the context of a software package, `haitipkg`, written in the R language [85]. Using a software package mechanism supports documentation, standardization and portability that promote extendability. In the terminology of Gentleman and Temple Lang [84], 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 [84] 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).

Supporting information

S1 Fig. Model 1. Flow chart representation of Model 1.

S2 Fig. Model 2. Flow chart representation of Model 2.

S3 Fig. Model 3. Flow chart representation of Model 3.

S1 Table Notation conversion table. Conversions between the notation used here and the notation of Lee et al. (2020) [29].

S1 Text Model details. In depth description of the Markov chain and differential equation interpretations of compartment flow rates, as well as details on the numeric implementation of these models.

S2 Text Initialization models. Additional details of the initialization model that were used for Models 1–3.

S3 Text Measurement models. Additional details of the measurement model that were used for Models 1–3.

S4 Text Likelihood of Lee et al. models. Details of how the log-likelihood of the Lee et al. models were calculated.

S5 Text Calibrating Model 3. Additional details on the procedures for fitting and diagnosing Model 3.

S6 Text Accounting for parameter uncertainty A description of a empirical-Bayes approach that can be used to account for parameter uncertainty.

Acknowledgments

This work was supported by National Science Foundation grants DMS-1761603 and DMS-1646108. The authors would like to thank Mercedes Pascual and Betz Halloran for helpful discussions. Laura Matrajt provided additional data for the Model 2 analysis.

References

1. Kirpich A, Weppelmann TA, Yang Y, Ali A, Morris JG Jr, Longini IM. Cholera Transmission in Ouest Department of Haiti: Dynamic Modeling and the Future of the Epidemic. PLOS Neglected Tropical Diseases. 2015;9(10):1–12. doi:10.1371/journal.pntd.0004153.
2. Kirpich A, Weppelmann TA, Yang Y, Morris Jr JG, Longini Jr IM. Controlling cholera in the Ouest Department of Haiti using oral vaccines. PLOS Neglected Tropical Diseases. 2017;11(4):e0005482.

3. Andrews JR, Basu S. Transmission Dynamics and Control of Cholera in Haiti: an Epidemic Model. The Lancet. 2011;377(9773):1248–1255.

4. Botelho C, Kong JD, Lucien MA, Shuai Z, Wang H. A mathematical model for Vibrio-phage interactions. Mathematical Biosciences and Engineering. 2021;18(3).

5. Fitzgibbon WE, Morgan JJ, Webb GF, Wu Y. Modelling the aqueous transport of an infectious pathogen in regional communities: application to the cholera outbreak in Haiti. Journal of the Royal Society Interface. 2020;17(169):20200429.

6. Eisenberg MC, Kujbida G, Tuite AR, Fisman DN, Tien JH. Examining Rainfall and Cholera Dynamics in Haiti using Statistical and Dynamic Modeling Approaches. Epidemics. 2013;5(4):197–207.

7. Rinaldo A, Bertuzzo E, Mari L, Righetto L, Blokesch M, Gatto M, et al. Reassessment of the 2010-2011 Haiti Cholera Outbreak and Rainfall-Driven Multiseason Projections. Proceedings of the National Academy of Sciences. 2012;109(17):6602–6607.

8. Chao DL, Halloran ME, Longini Jr IM. Vaccination strategies for epidemic cholera in Haiti with implications for the developing world. Proceedings of the National Academy of Sciences. 2011;108(17):7081–7085.

9. Date KA, Vicari A, Hyde TB, Mintz E, Danovaro-Holliday MC, Henry A, et al. Considerations for oral cholera vaccine use during outbreak after earthquake in Haiti, 2010- 2011. Emerging infectious diseases. 2011;17(11):2105.

10. Lin J, Xu R, Tian X. Transmission dynamics of cholera with hyperinfectious and hypoinfectious vibrios: mathematical modelling and control strategies. Mathematical Biosciences and Engineering. 2019;16(5):4339–4358.

11. Collins OC, Duffy KJ. Mathematical analyses on the effects of control measures for a waterborne disease model with socioeconomic conditions. Journal of Computational Biology. 2021;28(1):19–32.

12. Akman O, Schaefer E. An evolutionary computing approach for parameter estimation investigation of a model for cholera. Journal of biological dynamics. 2015;9(1):147–158.

13. Trevisin C, Lemaitre JC, Mari L, Pasetto D, Gatto M, Rinaldo A. Epidemicity of Cholera Spread and the Fate of Infection Control Measures. Journal of the Royal Society Interface. 2022;19(188):20210844.

14. Mavian C, Paisie TK, Alam MT, Browne C, Beau De Rochars VM, Nembrini S, et al. Toxigenic Vibrio cholerae evolution and establishment of reservoirs in aquatic ecosystems. Proceedings of the National Academy of Sciences. 2020;117(14):7897–7904.

15. Collins OC, Govinder KS. Incorporating Heterogeneity into the Transmission Dynamics of a Waterborne Disease Model. Journal of Theoretical Biology. 2014;356:133–143.

16. Kelly Jr MR, Tien JH, Eisenberg MC, Lenhart S. The impact of spatial arrangements on epidemic disease dynamics and intervention strategies. Journal of biological dynamics. 2016;10(1):222–249.

17. Capone F, De Cataldis V, De Luca R. Influence of diffusion on the stability of equilibria in a reaction–diffusion system modeling cholera dynamic. Journal of mathematical biology. 2015;71:1107–1131.

18. Bengtsson L, Gaudart J, Lu X, Moore S, Wetter E, Sallah K, et al. Using mobile phone data to predict the spatial spread of cholera. Scientific reports. 2015;5(1):8923.

19. Mari L, Bertuzzo E, Finger F, Casagrandi R, Gatto M, Rinaldo A. On the predictive ability of mechanistic models for the Haitian cholera epidemic. Journal of the Royal Society Interface. 2015;12(104):20140840.

20. Gatto M, Mari L, Bertuzzo E, Casagrandi R, Righetto L, Rodriguez-Iturbe I, et al. Generalized reproduction numbers and the prediction of patterns in waterborne disease. Proceedings of the National Academy of Sciences. 2012;109(48):19703–19708.

21. Tuite AR, Tien J, Eisenberg M, Earn DJ, Ma J, Fisman DN. Cholera Epidemic in Haiti, 2010: Using a Transmission Model to Explain Spatial Spread of Disease and Identify Optimal Control Interventions. Annals of internal medicine. 2011;154(9):593–601.

22. Mukandavire Z, Smith DL, Morris Jr JG. Cholera in Haiti: reproductive numbers and vaccination coverage estimates. Scientific reports. 2013;3(1):997.

23. Lewnard JA, Antillón M, Gonsalves G, Miller AM, Ko AI, Pitzer VE. Strategies to prevent cholera introduction during international personnel deployments: a computational modeling analysis based on the 2010 Haiti outbreak. PLoS medicine. 2016;13(1):e1001947.

24. Kunkel A, Lewnard JA, Pitzer VE, Cohen T. Antimicrobial resistance risks of cholera prophylaxis for United Nations peacekeepers. Antimicrobial agents and chemotherapy. 2017;61(8):10–1128.

25. Sallah K, Giorgi R, Bengtsson L, Lu X, Wetter E, Adrien P, et al. Mathematical Models for Predicting Human Mobility in the Context of Infectious Disease Spread: Introducing the Impedance Model. International Journal of Health Geographics. 2017;16(1):1–11.

26. Azman AS, Luquero FJ, Rodrigues A, Palma PP, Grais RF, Banga CN, et al. Urban cholera transmission hotspots and their implications for reactive vaccination: evidence from Bissau city, Guinea bissau. PLoS neglected tropical diseases. 2012;6(11):e1901.

27. Kühn J, Finger F, Bertuzzo E, Borgeaud S, Gatto M, Rinaldo A, et al. Glucose-but not rice-based oral rehydration therapy enhances the production of virulence determinants in the human pathogen Vibrio cholerae. PLoS neglected tropical diseases. 2014;8(12):e3347.

28. Azman AS, Luquero FJ, Ciglenecki I, Grais RF, Sack DA, Lessler J. The impact of a one-dose versus two-dose oral cholera vaccine regimen in outbreak settings: A modeling study. PLoS Medicine. 2015;12(8):e1001867.

29. Lee EC, Chao DL, Lemaitre JC, Matrajt L, Pasetto D, Perez-Saez J, et al. Achieving Coordinated National Immunity and Cholera Elimination in Haiti Through Vaccination: A Modelling Study. The Lancet Global Health. 2020;8(8):e1081–e1089.

30. Li R, Pei S, Chen B, Song Y, Zhang T, Yang W, et al. Substantial Undocumented Infection Facilitates the Rapid Dissemination of Novel Coronavirus (SARS-CoV-2). Science. 2020;368(6490):489–493.

31. Tracy M, Cerdá M, Keyes KM. Agent-Based Modeling in Public Health: Current Applications and Future Directions. Annual Review of Public Health. 2018;39(1):77–94.

32. Pezzoli L. Global Oral Cholera Vaccine Use, 2013–2018. Vaccine. 2020;38:A132–A140.

33. Behrend MR, Basáñez MG, Hamley JID, Porco TC, Stolk WA, Walker M, et al. Modelling for Policy: The Five Principles of the Neglected Tropical Diseases Modelling Consortium. PLoS Neglected Tropical Diseases. 2020;14(4):1–17.

34. Saltelli A, Bammer G, Bruno I, Charters E, Di Fiore M, Didier E, et al. Five Ways to Ensure that Models Serve Society: a Manifesto. Nature. 2020;582:428–484.

35. Rubin DHF, Zingl FG, Leitner DR, Ternier R, Compere V, Marseille S, et al. Reemergence of Cholera in Haiti. New England Journal of Medicine. 2022;doi:10.1056/NEJMc2213908.

36. Pan American Health Organization. Cholera Epidemic in Hispaniola 2023 - Situation Report 19; 2023. Available from: `https://www.paho.org/en/documents/cholera-outbreak-hispaniola-2023-situation-report-19`.

37. Francois J. Cholera Remains a Public Health Threat in Haiti. The Lancet Global Health. 2020;8(8):e984.

38. Rebaudet S, Gaudart J, Piarroux R. Cholera in Haiti. The Lancet Global Health. 2020;8(12):e1468.

39. Henrys JH, Lerebours G, Achille MA, Moise K, Raccurt C. Cholera in Haiti. The Lancet Global Health. 2020;8(12):e1469.

40. Lee EC, Ternier R, Lessler J, Azman AS, Ivers LC. Cholera in Haiti–Authors' Reply. The Lancet Global Health. 2020;8(12):e1470–e1471.

41. Lucas RE, et al. Econometric Policy Evaluation: A Critique. In: Carnegie-Rochester Conference Series on Public Policy. vol. 1; 1976. p. 19–46.

42. Lofgren ET, Halloran ME, Rivers CM, Drake JM, Porco TC, Lewis B, et al. Mathematical models: A key tool for outbreak response. Proceedings of the National Academy of Sciences of the USA. 2014;111(51):18095–18096.

43. McCabe R, Donnelly CA. Disease transmission and control modelling at the science–policy interface. Interface Focus. 2021;11(6):20210013.

44. May RM. Uses and abuses of mathematics in biology. science. 2004;303(5659):790–793.

45. Lau MS, Becker A, Madden W, Waller LA, Metcalf CJE, Grenfell BT. Comparing and linking machine learning and semi-mechanistic models for the predictability of endemic measles dynamics. PLoS computational biology. 2022;18(9):e1010251.

46. Bretó C, Ionides EL. Compound Markov Counting Processes and their Applications to Modeling Infinitesimally Over-Dispersed Systems. Stochastic Processes and their Applications. 2011;121:2571–2591.

47. Stocks T, Britton T, Höhle M. Model Selection and Parameter Estimation for Dynamic Epidemic Models via Iterated Filtering: Application to Rotavirus in Germany. Biostatistics. 2020;21(3):400–416.

48. He D, Ionides EL, King AA. Plug-and-Play Inference for Disease Dynamics: Measles in Large and Small Towns as a Case Study. Journal of the Royal Society Interface. 2010;7:271–283.

49. Ferreira S. Cholera threatens Haiti after Hurricane Matthew. BMJ. 2016;355:i5516.

50. Bretó C, He D, Ionides EL, King AA. Time Series Analysis via Mechanistic Models. Annals of Applied Statistics. 2009;3:319–348.

51. Toni T, Welch D, Strelkowa N, Ipsen A, Stumpf MPH. Approximate Bayesian Computation Scheme for Parameter Inference and Model Selection in Dynamical Systems. Journal of the Royal Society Interface. 2009;6:187–202.

52. Andrieu C, Doucet A, Holenstein R. Particle Markov chain Monte Carlo methods. Journal of the Royal Statistical Society, Series B (Statistical Methodology). 2010;72:269–342.

53. Ionides EL, Nguyen D, Atchadé Y, Stoev S, King AA. Inference for Dynamic and Latent Variable Models via Iterated, Perturbed Bayes Maps. Proceedings of the National Academy of Sciences of the USA. 2015;112(3):719—-724.

54. Arulampalam MS, Maskell S, Gordon N, Clapp T. A Tutorial on Particle Filters for Online Nonlinear, Non-Gaussian Bayesian Tracking. IEEE Transactions on Signal Processing. 2002;50:174 – 188.

55. Rebaudet S, Bulit G, Gaudart J, Michel E, Gazin P, Evers C, et al. The Case-Area Targeted Rapid Response Strategy to Control Cholera in Haiti: A Four-Year Implementation Study. PLoS Neglected Tropical Diseases. 2019;13(4):e0007263.

56. Rebaudet S, Dély P, Boncy J, Henrys JH, Piarroux R. Toward Cholera Elimination, Haiti. Emerging Infectious Diseases. 2021;27(11):2932.

57. Ionides EL, Breto C, Park J, Smith RA, King AA. Monte Carlo Profile Confidence Intervals for Dynamic Systems. Journal of the Royal Society Interface. 2017;14:1–10.

58. King AA, Nguyen D, Ionides EL. Statistical Inference for Partially Observed Markov Processes via the R Package pomp. Journal of Statistical Software. 2016;69:1–43.

59. Kermack WO, McKendrick AG. A Contribution to the Mathematical Theory of Epidemics. Proceedings of the Royal Society of London, Series A. 1927;115(772):700–721.

60. Brauer F. Mathematical Epidemiology: Past, Present, and Future. Infectious Disease Modelling. 2017;2(2):113–127.

61. Giordano G, Blanchini F, Bruno R, Colaneri P, Di Filippo A, Di Matteo A, et al. Modelling the COVID-19 Epidemic and Implementation of Population-Wide Interventions in Italy. Nature Medicine. 2020;26(6):855–860.

62. Dadlani A, Afolabi RO, Jung H, Sohraby K, Kim K. Deterministic Models in Epidemiology: From Modeling to Implementation. arXiv:200404675. 2020;.

63. Ndii MZ, Supriatna AK. Stochastic Mathematical Models in Epidemiology. Information. 2017;20:6185–6196.

64. Asfaw K, Park J, King AA, Ionides EL. Partially Observed Markov Processes with Spatial Structure via the R Package spatPomp. arXiv:210101157v3. 2023;.

65. King AA, Rowan T. `subplex`: Unconstrained Optimization using the Subplex Algorithm. R package, available at `https://cranr-projectorg/web/packages/subplex`. 2020;.

66. King AA, Domenech de Cellès M, Magpantay FM, Rohani P. Avoidable Errors in the Modelling of Outbreaks of Emerging Pathogens, with Special Reference to Ebola. Proceedings of the Royal Society B: Biological Sciences. 2015;282(1806):20150347.

67. Green KC, Armstrong JS. Simple Versus Complex Forecasting: The Evidence. Journal of Business Research. 2015;68(8):1678–1685.

68. Rebeschini P, van Handel R. Can Local Particle Filters Beat the Curse of Dimensionality? The Annals of Applied Probability. 2015;25(5):2809–2866.

69. Park J, Ionides EL. Inference on High-Dimensional Implicit Dynamic Models using a Guided Intermediate Resampling Filter. Statistics & Computing. 2020;30:1497–1522.

70. Ionides EL, Ning N, Wheeler J. An Iterated Block Particle Filter for Inference on Coupled Dynamic Systems with Shared and Unit-Specific Parameters. Statistica Sinica,. 2022; p. pre–published online.

71. Ning N, Ionides EL. Iterated Block Particle Filter for High-dimensional Parameter Learning: Beating the Curse of Dimensionality. arXiv:211010745. 2021;.

72. Box GE. Robustness in the Strategy of Scientific Model Building. In: Robustness in Statistics. Elsevier; 1979. p. 201–236.

73. Mayo DG. Statistical Inference as Severe Testing. Cambridge: Cambridge University Press; 2018.

74. Akaike H. A New Look at the Statistical Model Identification. IEEE Transactions on Automatic Control. 1974;19(6):716–723.

75. Donoho D. 50 Years of Data Science. Journal of Computational and Graphical Statistics. 2017;26(4):745–766.

76. Liu JS. Monte Carlo Strategies in Scientific Computing. New York: Springer; 2001.

77. Wood SN. Statistical Inference for Noisy Nonlinear Ecological Dynamic Systems. Nature. 2010;466:1102–1104.

78. Ganusov VV. Strong Inference in Mathematical Modeling: a Method for Robust Science in the Twenty-First Century. Frontiers in Microbiology. 2016;7:1131.

79. Lemaitre J, Pasetto D, Perez-Saez J, Sciarra C, Wamala JF, Rinaldo A. Rainfall as a Driver of Epidemic Cholera: Comparative Model Assessments of the Effect of Intra-Seasonal Precipitation Events. Acta Tropica. 2019;190:235–243.

80. King AA, Ionides EL, Pascual M, Bouma MJ. Inapparent Infections and Cholera Dynamics. Nature. 2008;454:877–880.

81. Grad YH, Miller JC, Lipsitch M. Cholera Modeling: Challenges to Quantitative Analysis and Predicting the Impact of Interventions. Epidemiology. 2012;23(4):523.

82. Lewis ASL, Rollinson CR, Allyn AJ, Ashander J, Brodie S, Brookson CB, et al. The Power of Forecasts to Advance Ecological Theory. Methods in Ecology and Evolution,. 2022; p. pre–published online.

83. Ioannidis JP, Cripps S, Tanner MA. Forecasting for COVID-19 has Failed. International Journal of Forecasting. 2020;.

84. Gentleman R, Temple Lang D. Statistical Analyses and Reproducible Research. Journal of Computational and Graphical Statistics. 2007;16(1):1–23.

85. R Core Team. R: A Language and Environment for Statistical Computing; 2022. Available from: `https://www.R-project.org/`.