On the predictive ability of mechanistic models for the Haitian cholera epidemic

Lorenzo Mari[1,2,*], Enrico Bertuzzo[2], Flavio Finger[2],
Renato Casagrandi[1], Marino Gatto[1], Andrea Rinaldo[2,3]

[1] Dipartimento di Elettronica, Informazione e Bioingegneria,
Politecnico di Milano, 20133 Milano, Italy
[2] Laboratory of Ecohydrology ECHO/IIE/ENAC,
École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland
[3] Dipartimento di Ingegneria Civile, Edile ed Ambientale,
Università di Padova, 35131 Padova, Italy
[*] Corresponding author: lorenzo.mari@polimi.it

Manuscript type Research Article

E-mail addresses lorenzo.mari@polimi.it (LM), enrico.bertuzzo@epfl.ch (EB),
flavio.finger@epfl.ch (FF), renato.casagrandi@polimi.it (RC),
marino.gatto@polimi.it (MG), andrea.rinaldo@epfl.ch (AR)

Keywords multilayer network model, spatially explicit model, ecohydrology,
epidemic forecast, model calibration, model validation

Running head Predictive models for cholera epidemics

Printed in colour Figures 1–5

ESM Figures S1–S9

Abstract

Predictive models of epidemic cholera need to resolve at suitable aggregation levels spatial data pertaining to local communities, epidemiological records, hydrologic drivers, waterways, patterns of human mobility and proxies of exposure rates. We address the above issue in a formal model comparison framework and provide a quantitative assessment of the explanatory and predictive abilities of various model settings with different spatial aggregation levels and coupling mechanisms. Reference is made to records of the recent Haiti cholera epidemics. Our intensive computations and objective model comparisons show that spatially explicit models accounting for spatial connections have better explanatory power than spatially disconnected ones for short-to-intermediate calibration windows, while parsimonious, spatially disconnected models perform better with long training sets. On average, spatially connected models show better predictive ability than disconnected ones. We suggest limits and validity of the various approaches and discuss the pathway towards the development of case-specific predictive tools in the context of emergency management.

1 Introduction

Cholera was reported in Haiti for the first time in recent history in October 2010, about nine months after the catastrophic earthquake that stroke the country and damaged its poor infrastructures for health care, water, and sanitation [1]. The source of the infection has been tracked back to the abrupt contamination of the Artibonite River from an external source, as unambiguously documented on both epidemiological and genetic grounds [2, 3, 4, 5, 6] (see also [7]). A first epidemic peak originated as cholera spread out from the Artibonite Valley. The disease was reported within weeks in all Haitian provinces, including the capital city Port-au-Prince. More than 150,000 cases and 3,500 casualties were reported by the end of 2010. A second peak (Spring 2011) was related to the revamping of disease transmission boosted by the rainy season [8]. One year after the beginning of the epidemic, the total toll of cholera in Haiti amounted to about 490,000 cases, with more than 6,200 deaths. At that time, those figures already qualified the Haitian epidemic as the largest cholera outbreak in recent history. Cholera has not disappeared yet from Haiti about four years after its appearance in the country, as shown by the increasing counts of cases and casualties. As of December 2014, more than 720,000 cases and the death of 8,700 people have been reported.

Several mathematical models of cholera transmission have been developed to describe the course of the Haitian epidemic [9, 10, 11, 12, 13, 14, 15, 16, 17, 18]. Their application to the ongoing epidemic was made possible by the immediate release of epidemiological data, initially recorded by the disease-surveillance systems set up by the Haitian government in the aftermath of the earthquake and later by the National Cholera Surveillance System [1], as well as by the widespread availability of georeferenced environmental datasets. Those models were different in assumptions, spatial resolution and degrees of

spatial coupling, but they all addressed the dynamics of susceptibles, infected individuals and bacterial concentrations in a discrete and geographically referenced set of local human communities. Mathematical modelling of the ongoing Haitian cholera epidemic has certainly provided important insights, especially concerning spatial transmission mechanisms [10, 11, 12, 14], rainfall patterns [14, 15, 17], intervention strategies [9, 10, 11, 12, 16, 17], local basic reproduction numbers [16], conditions for large-scale pathogen invasion [13], and the probability of epidemic extinction [18]. Most models provided predictions about the unfolding of the epidemic [9, 10, 14, 17, 18]. However, while predictive models have been successfully used in endemic settings [19, 20, 21, 22], the predictive power of mathematical models of cholera transmission has never been evaluated formally and systematically in an epidemic setting (but see [14] for a reassessment of early predictions of the Haitian epidemic). Bridging this gap is obviously of paramount importance to understand whether (and, in case, to what extent) predictions drawn in the very course of an outbreak (i.e. in conditions of severe data limitation and uncertainty about the relevant epidemiological processes) can be trusted, and potentially used to aid real-time emergency management, allocate health care resources and evaluate the effects of alternative intervention strategies.

In this work we aim at evaluating the predictive ability of mechanistic modelling for the Haitian cholera epidemic. To this end, we adopt the model proposed in [14] and later refined by [18]. This model has been shown to reproduce cholera transmission dynamics in Haiti in a fairly accurate and robust way. The model builds on a spatially explicit epidemiological framework [23, 24] that has already been applied to both past [25, 26] and ongoing cholera epidemics [10, 17]. To study the role of spatial settings on the predictive potential of cholera modelling, we consider two different spatial scales, namely a fine-grained subdivision of the Haitian territory into hydrological units (as in [14, 18]), and a coarse subdivision into administrative departments (e.g. as in [9, 12, 16]). The same set of epidemiological assumptions (detailed below) is retained for the two spatial scales, while different mechanisms of spatial propagation of the disease are considered. Specifically, a fine spatial resolution based on hydrological divides allows the model to account for both hydrological transport of pathogens and human mobility. Conversely, a coarser resolution based on administrative units cannot accommodate a proper description of hydrological connectivity. To evaluate the effects of the different assumptions on spatial processes that have been made in the literature at the latter spatial scale, we test a set-up of the model in which human mobility is the only driver of spatial disease spread between districts (as in [12]) and four set-ups in which no spatial transmission mechanisms are considered (i.e. each department's epidemiological dynamics is assumed to be independent of the other departments', e.g. as in [9]). Of these four set-ups, two (which differ from one another for initial conditions only) assume spatially heterogeneous epidemiological dynamics (different parameters across districts), while the other two (differing, again, for initial conditions only) assume homogeneous epidemiology (same parameters across districts).

The six model set-ups are calibrated against available epidemiological records by using progressively longer training sets. Validation (over different temporal windows) is performed to assess quantitatively the predictive ability of the calibrated models. Calibration and validation results are then analysed in a formal comparative framework to discuss strengths and limitations of the different approaches, as well as to identify the missing steps towards real-time use of the model. The intense computational effort made here is in fact meant to give practical (and possibly general) directions for management-prone modelling strategies in an actual epidemic context.

2 Methods

2.1 The model

The Haitian population is subdivided into n human communities spatially distributed within a domain that embeds both human mobility and hydrological networks (if applicable). Let $S_i(t)$, $I_i(t)$ and $R_i(t)$ be the local abundances of susceptible, symptomatic infected and recovered individuals in each node i of the network at time t, and let $B_i(t)$ be the environmental concentration of *Vibrio cholerae*. Cholera transmission dynamics can be described by the following set of coupled differential equations [14, 18]:

$$\begin{aligned}
\frac{dS_i}{dt} &= \mu(H_i - S_i) - F_i S_i + \rho R_i \\
\frac{dI_i}{dt} &= \sigma F_i S_i - (\gamma + \mu + \alpha) I_i \\
\frac{dR_i}{dt} &= (1-\sigma) F_i S_i + \gamma I_i - (\rho + \mu) R_i \\
\frac{dB_i}{dt} &= -\mu_B B_i + \frac{p}{W_i}\left[1 + \phi J_i(t)\right] I_i - l\left(B_i - \sum_{j=1}^{n} P_{ji} \frac{W_j}{W_i} B_j\right) .
\end{aligned}$$

The population of each node is assumed to be at demographic equilibrium, with μ and H_i being the human mortality rate and the population size of the local community. The force of infection F_i, which represents the rate at which susceptible individuals become infected because of contact with contaminated water, is expressed as

$$F_i = \beta \left[(1-m) \frac{B_i}{K + B_i} + m \sum_{j=1}^{n} Q_{ij} \frac{B_j}{K + B_j} \right] ,$$

where the parameter β represents the exposure rate (assumed to be constant over time; for a different hypothesis see again [18]) and the fraction $B_i/(K + B_i)$ is the probability of becoming infected because of exposure to a concentration B_i of *V. cholerae* (K being the half-saturation constant, [27]). Susceptible individuals can be exposed to the pathogen while travelling outside their home community. Parameter m

represents the community-level fraction of individuals that travel outside their home site, while Q_{ij} represents the fraction of people travelling from their home community i to destination j. The force of infection in a given node thus depends not only on the local concentration of pathogens (for a fraction $1-m$ of the susceptible hosts), but also on pathogen concentration in the connected communities (for the remaining fraction m, properly weighted by mobility patterns). Mobility fluxes are described through a gravity model [28, 29] in which the attractiveness of node j for node i is directly proportional to the population size of j and inversely proportional to the distance between the two nodes (through an exponential kernel with scale factor D), i.e.

$$Q_{ij} = \begin{cases} \frac{H_j e^{-d_{ij}/D}}{\sum_{k\neq i}^{n} H_k e^{-d_{ik}/D}} & \text{if} \quad i \neq j \\ 0 & \text{otherwise}\,. \end{cases}$$

Upon exposure to contaminated water, a fraction σ of infected individuals develops symptoms. Symptomatic infected individuals recover at rate γ, or die because of cholera or other causes at rates α or μ, respectively. Asymptomatic infected individuals shed *V. cholerae* bacteria at a much lower rate (about 1,000 times lower, see e.g. [30]) than symptomatic ones and recover more rapidly (in about one day instead of five, according to [31]). Therefore, it is reasonable to assume that their contribution to environmental contamination is negligible with respect to that of symptomatic individuals [18]. Asymptomatic infections can still result in temporary immunity [20], thus contributing to the depletion of the pool of susceptibles and affecting the rate of occurrence of symptomatic infections. This assumption translates into a flux of asymptomatic infected individuals $(1-\sigma)F_iS_i$ entering directly the recovered compartment.

Recovered individuals lose their immunity and return to the susceptible compartment at rate ρ or die at rate μ. Infected individuals showing clinical symptoms are assumed to be non-mobile. Therefore, they contribute only to the local environmental concentration of *V. cholerae* at rate p/W_i, with p being the rate at which bacteria excreted by one infected individual reach and contaminate the local water reservoir of volume W_i (assumed to be proportional to population size, i.e. $W_i = cH_i$ as in [14]). Note that for the sake of parsimony a dimensionless bacterial concentration $B_i^* = B_i/K$ can be introduced, along with a synthetic contamination rate $\theta = p/(cK)$. Bacteria die at a constant rate μ_B and undergo hydrologic dispersal at rate l. Cholera pathogens can move between nodes i and j with probability P_{ij}. We assume $P_{ij} = 1$ if j is the downstream nearest neighbour of node i and zero otherwise. To include the worsening of sanitation conditions due to rainfall-induced run-off, which causes additional loads of pathogens to enter the water reservoir because of the overflow of latrines and the washout of open-air defecation sites [8], the contamination rate is increased by local rainfall intensity $J_i(t)$ via a proportionality coefficient ϕ [14].

2.2 Model set-ups

We consider six versions of this model: fine-grained spatially connected (M1), coarse-grained spatially connected (M2), coarse-grained spatially disconnected with heterogeneous parameters (M3 and M4), and coarse-grained spatially disconnected with homogeneous parameters (M5 and M6). All of them share the same basic epidemiological assumptions outlined above. Other possible mechanisms, such as bacterial hyperinfectivity [32] or latent infected stages [11], could obviously be accounted for, yet they have not been included in the present work to avoid potential confounding effects and limit the number of candidate models. This choice also follows from a previous comparative analysis of modelling assumptions in the context of the Haitian cholera epidemic [14]. The six model set-ups considered here differ from each other as regards the spatial scale of analysis, i.e. fine-grained (M1) vs. coarse-grained (M2–M6), the inclusion of spatial coupling mechanisms, i.e. human mobility and hydrological transport (M1) vs. human mobility only (M2, in which $l = 0$) vs. no spatial coupling at all (M3–M6, in which $l = 0$ and $m = 0$), and the choice of initial conditions (spatially disconnected models M3–M6 only, see below).

The six set-ups of the model are also different in terms of parameter parsimony. As a matter of fact, spatially connected network models account for the spatial heterogeneity induced by spatial interactions, while spatially disconnected models do not. To circumvent this limitation, epidemiological parameters in the latter models are usually allowed to be different for different geographical units (see e.g. [9, 15, 16] for applications to the Haitian cholera epidemic). As a result, spatially disconnected models may be considerably less parsimonious than spatially connected ones, in which the epidemiological parameters are usually assumed to be spatially homogeneous (essentially to facilitate model calibration; see [14, 18] and previous related applications) or linked to remotely acquired proxies [26]. Therefore, georeferenced spatially disconnected models could better account for local heterogeneities in disease transmission than parsimonious spatially connected ones. To investigate whether prodigality of structural parameters is actually compensated for by improved explanatory/predictive power we thus consider two baseline disconnected models with spatially homogeneous epidemiological parameters (and, again, different assumptions regarding initial conditions). The main features of each model are summarized in Table 1.

Table 1 around here

2.3 Application to the 2010– Haitian epidemic

As stated above, the model is run at two different spatial scales. At the finest resolution (model M1), the Haitian territory is subdivided into watersheds on the basis of hydrologic divides, inferred from drainage directions extracted from a digital terrain model [DTM, see e.g. 33]. We use the DTM provided by the US Geological Survey (USGS, available online at http://nationalmap.gov/viewer.html), with a grid

resolution of 100 m and a precision of ±0.5 m in the elevation field (Fig. 1*a*). Following the procedure detailed in [14], the Haitian territory is subdivided into 365 hydrological units with an average extent of 76 km^2 (Fig. 1*b*). The use of hydrologically-defined units allows a straightforward identification of the hydrological connection from each watershed to its unique downstream neighbour (or to the ocean, for coastal watersheds). The hydrologic connectivity matrix $[P_{ij}]$ thus follows directly through nowadays standard extraction techniques [34]. Note that absorbing boundary conditions are assumed at the ocean outlets, i.e. pathogens that leave the river network from a coastal outlet cannot re-enter the system.

Figure 1 around here

At the coarsest resolution, the official Haitian subdivision in ten administrative departments is retained. Note that this scale corresponds to that at which epidemiological records, consisting of daily counts of cholera cases reported in each of the ten administrative departments, are currently made available by the Haitian Ministry of Health (available online at `http://mspp.gouv.ht/newsite/`; Fig. 1*c*). At such a coarse resolution, however, the information on hydrological connectivity is lost (models `M2`–`M6`).

Human communities are defined as the population hosted within each computational unit (watershed or department, depending on the spatial scale considered), estimated by using a remotely sensed map of population distribution produced by the Oak Ridge National Laboratory (data available online at `http://www.ornl.gov/sci/landscan/index.shtml`; Fig. 1*d*). The spatial resolution is 30×30 arc-seconds, resulting in cells of about 1 km^2. Distances d_{ij} among communities in the fine-scale model are computed using the road network provided by the OpenStreetMap project (available online at `www.openstreetmap.org`; Fig. 1*e*). Specifically, pairwise shortest distances along the road network are computed between the centroids of population distribution in each community. Great-circle distances between population centroids are used in the department-based connected model.

Systematic collection of rainfall through rain gauges has been relatively rare in post-earthquake Haiti, with on-the-ground rainfall measurements available only for Ouest (by USGS) and Sud (by Haiti Regeneration Initiative) departments (see Figure 1 in [15] for a map of the existing rain gauges). Therefore, daily rainfall $J_i(t)$ for each community i is computed from satellite data collected by the NASA-JAXA's Tropical Rainfall Measuring Mission (TRMM_3B42 precipitation estimates, see `http://trmm.gsfc.nasa.gov/` for details). Rainfall data have a spatial resolution of 0.25 degrees of latitude and longitude (Fig. 1*f*). Precipitation fields are first downscaled to the resolution of the DTM with nearest neighbour interpolation and then averaged over the watershed/department area to obtain a representative value for the whole community. Because of the lack of surface measurements, it is not possible to perform a thorough comparison between remotely sensed estimates and on-the-ground data. Comparison of the surface data from Port-à-Piment (Sud, 2010–2012 data available online at `http://blogs.cuit.columbia.edu/haitienvironment/`

environmental-monitoring/precipitation/monthly-rainfall/) to satellite estimates (at the scale of either the watershed or the administrative department that include the gauge station) suggests that remote sensing may consistently underestimate local rainfall measurements (Fig. S1 online). This bias can be partly due to the complexity of Haitian orography, which can induce significant variations in local precipitation intensities within the spatial resolution of TRMM_3B42 estimates [15]. However, satellite-based rainfall estimates can qualitatively reproduce the rainfall patterns recorded on the ground (Pearsons's $r = 0.63$ [0.59] for satellite estimates at the watershed [department] scale).

As initial conditions for model simulations we assume that, as of the date of the reported beginning of the outbreak (around October 20, 2010; $t = 0$), the local numbers of infected people $I_i(0)$ match the reported cases detailed in [3]. The locations of the first cases are tracked accurately at the fine spatial resolution (M1), while at the coarser scale (M2–M6) the reported cases can only be attributed to administrative departments. We note that just a few watersheds/departments were actually home to cholera cases as of October 20, 2010. This obviously represents an issue for the initialization of spatially disconnected models M3–M6. The lack of spatial coupling mechanisms, in fact, would keep the departments that were initially left untouched by the epidemic indefinitely in a cholera-free state. Suitable initial conditions are thus to be set also in these departments. A possible alternative to initialize the model would be to let t_0 vary across departments following the observation of the first cases. However, this would require to change calibration/validation windows across sites (see below). As such, we have decided not to follow this route. We test instead two different settings (for both spatially heterogeneous and homogeneous models), one in which one symptomatic carrier is placed in each of the cholera-free departments at $t = 0$ (M3 and M5), and one in which the initial number of infected individuals is calibrated in each department (including, for consistency, also those with cholera cases reported at the beginning of the epidemic, M4 and M6). As for the other state variables, we assume that the whole population is susceptible at the beginning of the epidemic, i.e. $S_i(0) = H_i - I_i(0)$ (and $R_i(0) = 0$), because of the lack of any pre-existing immunity, consistently with assessment of prior conditions in Haiti [3]. Local pathogen concentrations are assumed to be initially at equilibrium with the local number of infected cases, i.e. $B_i^*(0) = \theta I_i(0)/(H_i \mu_B)$ [14].

2.4 Model calibration

Some of the model parameters (namely μ, α and γ) can be reliably estimated from the literature or from epidemiological/demographic records, while numerical fitting is necessary to calibrate the remaining ones. To mimic quasi-real-time use of the model presented above in an ongoing epidemic context we use twelve different calibration windows with increasing durations – ranging from one month (November 2010, when the first data became publicly available) to one year (from November 2010 to October 2011) with a one-month step.

Model fitting is performed via Markov Chain Monte Carlo (MCMC) sampling [35, 36]. Specifically, we use the algorithm DREAM (Differential Evolution Adaptive Metropolis [37], available online at `http://jasper.eng.uci.edu/software.html`), an efficient implementation of MCMC that runs multiple different chains simultaneously to ensure global exploration of the parameter space, and adaptively tunes the scale and orientation of the jumping distribution using Differential Evolution [38] in addition to a Metropolis-Hastings update step [39, 40]. We adopt the DREAM$_{ZS}$ variant of the DREAM algorithm [41], which allows for an effective exploration of the target posterior distribution with just a few parallel chains. The algorithm is initialized with broad flat prior distributions for parameter values and is allowed to run up to convergence ($\mathcal{O}(10^5)$ iterations).

A Gaussian model is assumed for the likelihood of the epidemiological observations in MCMC sampling. Goodness of fit is measured as the residual sum of squares (RSS) between weekly reported cholera cases in each of the $n_d = 10$ Haitian departments as recorded in the epidemiological bulletins and simulated by the model over the n_w^c weeks of the calibration window, i.e.

$$RSS = \sum_{i=1}^{n_d} \sum_{j=1}^{n_w^c} \left[C_r(i,j) - C_s(i,j) \right]^2 ,$$

where $C_r(i,j)$ and $C_s(i,j)$ are the weekly reported cases in department i during week j according to records and model simulations, respectively. The estimation of weekly cases from the model output requires to compute

$$C_s(i,j) = \sigma \int_{t_j}^{t_j+\Delta t} F_i S_i dt \,,$$

where t_j marks the beginning of the j-th week and $\Delta t = 1$ week. Note that the simulation results for the fine-grained version of the model (`M1`) are computed at the watershed level and need to be upscaled to the departmental level. The upscaling procedure is performed by accounting for the fraction of population of each watershed that belongs to a given administrative department.

We repeat the calibration procedure twelve times for each of the six model set-ups introduced above, each of which is characterized by a different number of tuning parameters. Specifically, `M1` has 9 calibration parameters (β, θ, σ, ρ, μ_B, ϕ, l, m, D), chosen as in [18] to ease across-study comparison. `M2` has 8 calibration parameters, the same as `M1` except for $l = 0$. We have decided to fit the same set of epidemiological parameters for both spatially connected (`M1` and `M2`) and spatially disconnected models (`M3`–`M6`). For disconnected models with spatially heterogeneous parameters (`M3` and `M4`) calibration is performed department-by-department. Independent calibration runs (e.g. as done in [9]) in fact represent the simplest possibility when longitudinal data at different locations are available, but results in a large number of fitting parameters, namely 60 for `M3` (β, θ, σ, ρ, μ_B, ϕ, calibrated independently for each of

the ten administrative departments) and 70 for `M4` (same as `M3`, plus the initial number of infectives $I_i(0)$ in each department). More parsimonious spatially disconnected models with heterogeneous parameters could actually be devised (see again [9]), also by constraining some epidemiological parameters across sites. Thus, the most parsimonious models are those with spatially homogeneous parameters (`M5` and `M6`): the fitting parameters of `M5` are 6 (the same as `M1` with $l = 0$ and $m = 0$), while those of `M6` are 16 (the same as `M5`, plus the initial number of infectives in each of the ten departments).

To assess the explanatory power of the different versions of the model, for each calibration window we rank their performances according to Akaike's Information Criterion (AIC) [42] suitably corrected for small sample sizes [43]. AIC is a model-selection procedure that explicitly takes into account the trade-off between model accuracy and complexity, measured as the number of free parameters Θ (i.e. the structural parameters of the model, inclusive of initial conditions for `M4` and `M6`, plus residual variance for each independent calibration run; see [43]). After calibrating `M1`–`M6` against the epidemiological data, for each best-fit model simulation we compute

$$AIC = 2\Theta + \eta \ln\left(\frac{RSS}{\eta}\right) + 2\Theta\frac{\Theta+1}{\eta-\Theta-1}\,,$$

where $\eta = n_d n_w^c$ is the number of data points in the calibration window. The model with lowest AIC score is retained as the best candidate to explain the observed epidemic patterns in the relevant calibration window, with Akaike differences $\Delta AIC > 4$ being required for a significant ranking (see again [43]). Fitting is performed also for the shortest window (one month), but `M3` and `M4` are not included in AIC scoring because in that case the number of calibration parameters exceeds the available data points.

2.5 Model validation

Model validation is performed by extending model simulations outside the calibration windows. For each calibration window we use twelve validation intervals with increasing durations (ranging from one month to one year with monthly steps), each of which starts right after the end of the relevant calibration window. Actual precipitation fields are fed to the model also during the validation periods. We simulate each model for the best-fit parameter combination found during calibration and for an ensemble of $N = 100,000$ sets from the posterior distribution of the model parameters obtained through MCMC sampling, including variance residual to account for total predictive uncertainty (observational errors are assumed to be additive and normally distributed as in [44]; see Fig. S2 for two examples of the evaluation of total predictive uncertainty). In this way, model predictions explicitly include parameter uncertainty and observational errors. Conversely, other possible sources of uncertainty, such as process noise [45], biases in the input rainfall patterns [46] and structural modelling errors [47] are not explicitly accounted for, and may thus be

reflected in increased parameter/observational uncertainty.

Predictive power is assessed by using three different indicators, namely

$$
\begin{aligned}
V_1 &= \sqrt{\frac{1}{n_d n_w^v} \sum_{i=1}^{n_d} \sum_{j=1}^{n_w^v} [C_r(i,j) - C_s(i,j)]^2}\,, \\
V_2 &= \frac{1}{n_d n_w^v} \sum_{i=1}^{n_d} \left| \sum_{j=1}^{n_w^v} C_r(i,j) - \sum_{j=1}^{n_w^v} C_s(i,j) \right|\,, \\
V_3 &= \frac{100}{n_d n_w^v} \sum_{i=1}^{n_d} \frac{\left| \sum_{j=1}^{n_w^v} C_r(i,j) - \sum_{j=1}^{n_w^v} C_s(i,j) \right|}{\sum_{j=1}^{n_w^v} C_r(i,j)}\,,
\end{aligned}
$$

where n_w^v is the number of weeks in the validation window. The first indicator (V_1) is the standard deviation of the prediction residuals and represents a natural modification of the goodness-of-fit score used in model calibration. The second and the third quantities are based on the cumulative number of cases predicted by the model in each department over the whole validation window, suitably averaged over the number of data points in the validation period to make it possible to compare the predictive power of the model over different windows. Although they may downplay temporal errors, these two quantities provide an easily readable measure of absolute (V_2) or relative (V_3) prediction errors, respectively. They are evaluated for the best-fit model simulation (V_x^{Bf}, $x = 1, 2, 3$), the ensemble of N simulations accounting for total predictive uncertainty (V_x^{Ns}; in this case the three indicators are evaluated for each simulation, then average values are computed), and the median of the model predictions, that is the median of the weekly local cases predicted in the N model runs (V_x^{Me}). This procedure is repeated for each of the six model set-ups, 12 calibration intervals and 12 validation windows.

3 Results

All the six versions of the model are able to grasp qualitatively the evolution of the epidemic during the calibration phase (Figs. 2 and S3), except for model `M5` that consistently underestimates reported cholera cases. In terms of explanatory power, models `M1` and `M2` perform similarly in the shortest calibration window, but `M1` significantly outperforms all the others for six calibration windows out of twelve (Table 2), ranging from two to seven months. Model `M6` is selected for the five longest calibration windows (8–12 months). Model `M2` is the second-best for the four shortest calibration runs, after which it is outperformed by `M6`. Model `M1` is also second-best for intermediate calibration windows (8–10 months), while `M4` is second-best for the two longest calibration windows (11–12 months). The least supported models are `M6` for the shortest calibration timespan, `M3` and `M4` for windows of 2–6 months, `M5` for windows longer than six months.

Figure 2 and Table 2 around here

As an example of the outcomes of the calibration procedure, Fig. 3 reports the best-fit parameter values of model M1 and the related uncertainties as obtained for different calibration windows. Local disease transmission, represented by the basic reproduction number ($R_0 = \sigma\beta\theta/\mu_B/(\mu + \alpha + \gamma)$; see [27]), is found to be higher for long calibration windows (panel *a*). Conversely, higher fractions of moving people, shorter movement distances and higher hydrological dispersal rates are selected for short calibration intervals (panels *b*–*c*). Also, a longer immunity is selected for short calibration intervals, while higher values of the rainfall coefficient are selected for long calibration runs. As expected, the interquantile ranges of parameter uncertainty are quite large for short calibration windows, but progressively narrow for longer intervals, thus making it difficult to get reliable estimates of some of the model parameters (e.g. R_0 and ρ) over relatively short timespans (see again panel *a*). Estimated initial conditions (M4 and M6) averaged over the different calibration runs are found to be well correlated with reported cases (Pearson's $r = 0.50$ [0.84] for M4 [M6]), although with large absolute differences (on average, $\approx 5,400$ [$\approx 4,500$] cases per department for M4 [M6]).

Figure 3 around here

Sharp differences among the six versions of the model emerge in the validation trials (Fig. 4; see also Fig. S4 for the whole sequence of validation experiments). Spatially connected models (M1 and M2), even when tuned over very short calibration windows (say 1–2 months, see Fig. 4*a*), are able to forecast at least the order of magnitude of the observed cholera cases for several weeks after the end of the calibration window. To a lesser extent, the same is true for parsimonious, spatially disconnected models M5 (one-month calibration window) and M6 (two-month calibration window). Conversely, spatially disconnected models with heterogeneous parameters (M3 and M4) largely overestimate cholera incidence. No model is actually able to predict consistently the peak of cholera cases recorded during Spring 2011.Similar patterns can be found for relatively longer calibration windows (3–4 months, Fig. 4*b*), although with better predictions of the 2011 Spring peak (especially by M6), which however are robustly achieved (most notably by spatially connected models M1 and M2) only with calibration windows as long as 6 months (Fig. 4*c*), i.e. with a lead time of about one month. Validation performances greatly improve for longer calibration windows (7–10 months, Fig. S4), especially for spatially connected models M1–M2 and parsimonious disconnected models M5–M6, which become able to predict actually observed peak and lull phases with a prediction lead time as long as one year. Note that model M1 tends to remarkably overestimate cholera incidence towards the end of the longest validation runs. Once tuned over even longer calibration windows (11–12 months, Fig. 4*d*) almost all models display good predictive ability at least for a few months after the end of the calibration

window. The Spring peak of 2012 is poorly captured by all models. The subsequent lull phase is better predicted by M2 among all.

Figure 4 around here

An exhaustive investigation of the possible combinations of calibration/validation windows for performance indicator V_1 (Fig. 5*a*–*c*) shows that the predictive ability of spatially connected models (M1 and M2) usually outperforms that of the spatially disconnected ones for short calibration windows or long validation intervals. Disconnected models with heterogeneous parameters (M3 and M4) are rarely found to be best, and only for long calibration and short validation periods. Disconnected models with homogeneous parameters (M5 and M6) may represent better predictive tools for intermediate calibration periods, or for long calibration/intermediate validation runs (best-fit parameters only). Similar patterns are found when considering V_2, although with a larger share of validation trials in which model M6 performs best (Fig. 5*d*–*f*). A clear pattern emerges for V_3: model M1 usually performs best for short calibration windows (with some exceptions), while model M6 represents the best predictive tool for long calibration windows (Fig. 5*g*–*i*). The best predictive performances are attained by either best-fit runs or median model projections for all the considered validation indicators (Figs. S5–S7).

The values of the validation performance indicators averaged over all the combinations of calibration and validation intervals (Table 3) show that spatially connected models M1 and M2 have better overall predictive abilities than spatially disconnected ones. Averaging the validation performance indicators over the different validation windows (Fig. S8) shows that spatially connected models actually are better predictive tools for either short or long calibration intervals, but not always for calibration windows of intermediate length. By averaging the validation performance indicators over the different calibration windows (Fig. S9) it becomes evident that the predictive ability of the spatially connected models (M1 in particular) is higher than that of the spatially disconnected ones for almost all validation intervals.

Figure 5 and Table 3 around here

4 Discussion

Almost all models are able to reproduce complex spatio-temporal epidemic patterns during calibration, provided that the algorithm used for parameter fitting is fed with a sufficient amount of epidemiological data (Fig. 2). Model M5 represents an exception in this respect, most likely because of its relatively poor structural properties (lack of spatial coupling mechanisms, homogeneous parameters, fixed initial conditions). Differences in the explanatory power of the different model set-ups do emerge, though. Interestingly, in fact, spatially connected models are consistently selected for short to intermediate calibration windows,

while parsimonious, spatially disconnected models (with calibrated initial conditions, which could implicitly account for the disregarded spatial component of epidemic spread) perform better over long calibration intervals (Table 2). We argue that spatially disconnected models cannot directly account for the spatial mechanisms of pathogen dissemination (i.e. hydrological transport and human mobility), which play a crucial role in the initial phase of the outbreak, when the infection starts propagating into disease-free regions [13, 48]. Conversely, spatial coupling may become less important when looking at longer time horizons, once the epidemic has already spread all across the country. In this case, in fact, epidemiological dynamics is expected to be mostly controlled by local factors, which could indeed determine some spatial heterogeneity in the transmission processes [49]. The relatively good AIC scores of `M4` for the longest calibration runs (11–12 months) seem to support the idea that spatially disconnected models with site-specific parameters could indeed account for this spatial heterogeneity, although the calibration windows used in this work may be too short to let this pattern emerge more clearly. The outcomes of model fitting in the fine-scale spatially connected model `M1` (which provides the most detailed account of the spatial dissemination of the pathogen) further supports this explanation, with higher values of the dispersal rates/fractions selected for short calibration runs, and higher values of the basic reproduction number selected for long calibration intervals (Fig. 3). The present study thus helps clarify which modelling tool might be best suited to describe outbreak unfolding depending on the stage of development of the epidemic.

As far as accurate spatiotemporal projections are concerned (e.g. as evaluated by validation indicator V_1), spatially connected models seem to have greater predictive ability than spatially disconnected ones, especially for short calibration intervals (or long validation windows; Fig. 5). This finding may be particularly important in the early phase of an outbreak, when data are limited and the epidemic is undergoing rapid growth. We argue that the intrinsic lack of spatial coupling mechanisms and/or the prodigality of structural parameters determine the inability of spatially disconnected models to accurately project cholera dynamics in the first phase of the epidemic. On the one hand, in fact, the lack of spatial coupling mechanisms leads to biased estimations of initial conditions or local infection processes. On the other hand, the prodigality of structural parameters (which results in high uncertainty in parameter estimation) may act on top of the first shortcoming, thus leading to unsatisfying epidemiological projections even in the short run. Taken together (as in models `M3` and `M4`), these two features may determine a sizable excess of predicted cholera cases in the validation periods. Therefore, the use of spatially disconnected models with heterogeneous parameters for prediction purposes in the initial stage of an epidemic (i.e. when spatial coupling is of paramount importance and data availability is forcedly limited) seems hazardous at best. This remark is empirically confirmed by some early projections of cholera transmission dynamics in Haiti drawn from a spatially disconnected model with heterogeneous parameters published a few months after the beginning of the epidemic [9]. In that study, the authors predicted more than one million cases in the

first year of the epidemic – a figure that fortunately has not been reached even four years after the beginning of the epidemic. By contrast, the first predictions drawn from a parsimonious, spatially connected model of cholera dynamics in Haiti [10] correctly predicted the unfolding of the epidemic with a prediction lead time of five months, and were retrospectively judged quite robust in a later comparative reassessment [14]. Should coarser projections be of interest (e.g. as evaluated by validation indicator V_2, which quantifies absolute errors in terms of cumulative cases; or V_3, which measures relative errors and is less sensitive to deviations during high-prevalence periods), spatially disconnected models with homogeneous parameters would represent the tool of choice for intermediate-to-long calibration windows, while spatially connected models would still be best for short calibration runs (see again Fig. 5). This distinction could be relevant, for instance, to decision makers interested in robust projections of total epidemic evolution over decisional time scales spanning from a few weeks to a few months.

As for the prediction of epidemic peaks, model validation outcomes are far from perfect also for the best-performing models. For instance, for calibration windows shorter than four months, none of the considered models is able to robustly predict the cholera peak observed during Spring 2011, which has been linked to heavy seasonal precipitations [14]. As a matter of fact, these calibration windows correspond to the first, explosive phase of the epidemic, in which rapid transmission dynamics most likely beclouded the effect of rainfall (fairly scarce during this period, except for the passage of Hurricane Thomas at the beginning of November 2010) and possibly of other environmental drivers as well [8]. The outcomes of parameter calibration support this interpretation, with higher values of the rainfall coefficient being progressively selected for increasingly long calibration windows. The six model set-ups are not able to forecast the peak of Spring 2012 either. However, in this case, other mechanisms could have played a role. A relatively large share of cases was in fact localized in the capital city Port-au-Prince during the peak. The mismatch could have thus been caused by a poor estimation of local rainfall intensity. Satellite-based precipitation estimates may be of limited utility when looking at local features [50, 15], but represent a precious alternative to traditional ground measurements wherever the latter are rare, provided that they at least correlate with actual rainfall patterns (as our preliminary analyses seem indeed to confirm). This is true, in particular, for our country-scale model of cholera transmission, in which rainfall is considered as a forcing term for local contamination rates acting through a coefficient that has to be numerically calibrated. On the other hand, the errors possibly introduced in the model by inaccurate rainfall estimates are most likely translated into higher uncertainty in parameter values, and would perhaps deserve an explicit treatment (e.g. as in [46]; see also [14, 17, 18] on the use of stochastic rainfall generators for projections of the Haitian epidemic). All these considerations point to the importance of ground-truthing of remotely sensed rainfall patterns, and calls for a better integration of satellite estimates and co-located surface measurements. Further progress in this area will come through improved rainfall forecasting, which will also allow for real-time

testing of our models. Another possible source of error could be represented by environmental and/or social factors not accounted for in the model. The analysis of epidemiological reports at the communal level (which are recorded but not yet available to the scientific community, see e.g. [1, 8]) represents a possible way to reduce structural modelling errors. Such data could in fact shed light on highly localized transmission processes that may be difficult to describe when working at broad spatial scales [18]. However, increasing the spatial resolution of the model could make the application of a fully deterministic framework inappropriate (especially during lull phases in which disease transmission could be remarkably affected by demographic stochasticity [18]) and call for the inclusion of process noise in the modelling framework – which in turn would make numerical procedures much more involved (see e.g. [45]).

As far as long-term predictions are concerned, a possible limitation of our approach is the underlying assumption that epidemiological processes do not change over time. Rainfall is in fact the only term in the model that depends explicitly upon time. However, local exposure/contamination rates could also have changed over the timescales considered in this study (as proposed by [8], who suggested that transmission patterns may have changed significantly in time and space). A first plausible mechanism to explain this change is the increasing awareness of the population about the mechanisms of cholera transmission, as a result of the campaigns for hygiene promotion set up by Haitian authorities and non-governmental organizations [51, 49]. Increased knowledge about cholera and its transmission dynamics may have prompted behavioural changes in the population at risk of infection, thus significantly influencing cholera dynamics [52]. As an example, [18] estimated that such increased awareness may have led to a 20% reduction of exposure risk just in a few months after the beginning of the epidemic. Another possible mechanism of change is the progressive improvement of living conditions, potentially linked to the return of internally displaced people to their original households after the earthquake that stroke the country in January 2010 [53]. Structural interventions, aimed e.g. at improving access to safe water and basic sanitation, could also have helped reduce cholera transmission [54]. Because of the isolated and sporadic nature of such interventions, however, it is difficult to include their effects in a reasonably parsimonious mechanistic framework. Changes in human mobility patterns between the initial, explosive phase of the epidemic and the following phases, possibly triggered by the unfolding of the outbreak, could also have had nontrivial effects on disease dynamics [24]. All these missing features can be compensated for by biased parameter estimates, which in turn can contribute to the remarkable variations of parameters calibrated with different fitting windows.

The shortcomings of our approach do not preclude a quantitative assessment of the overall predictive ability of the different models. On average, in fact, the spatially connected models performs better than the others, although quantitative differences are relatively small for some indicators (Table 3). Some practical observations are thus in order. First, because of their added information value, spatially connected models should represent the tool of choice for accurate predictions of the spatio-temporal patterns of cholera

epidemics in their initial phases, when data are scant and the spatial spread of the pathogen is of paramount importance. Predictions drawn from current mathematical models of diseases like cholera must always be taken with care – but extra-care should be applied, in particular, whenever these predictions come from spatially disconnected models endowed with many structural parameters and calibrated over relatively short intervals. Second, although temporal, spatially aggregated models are usually much easier to fit to longitudinal data than spatiotemporal ones, spatially connected models for cholera epidemics are not necessarily doomed to be much more complex than spatially disconnected ones. This must be made clear not only to model makers, but also to consumers of models with diverse (and not always quantitative) backgrounds (e.g. general medical and public health readers as discussed in [55]). For modellers, some additional complexity clearly lies in the derivation of the connectivity structures for hydrological transport and human mobility; however, the increasing availability of georeferenced information on hydrological and transportation networks, human demography and mobility, sanitation infrastructure and treatment center distribution, coupled with objective manipulation techniques, has allowed rapid progresses in the field of spatially explicit modelling of eco-epidemiological dynamics over the past few years. For model users, spatially connected models are often only seemingly more complex than spatially disconnected ones; while the mathematical notation might indeed look complicated at first sight, they offer a natural language to describe eco-epidemiological processes that are intrinsically rooted in spatial dynamics.

All these observations point consistently to the importance of being spatially explicit (and connected) when dealing with the early phases of a cholera epidemic. As a matter of fact, when only limited data are available, information about spatial coupling mechanisms proves essential for making reasonable predictions. One question could arise related to what spatial scale of analysis should be used. Our analysis has showed, in fact, that a clear distinction emerges between spatially disconnected vs. connected models, while the predictive performances of the latter appear to be quite independent of the spatial scale of analysis. A possible explanation of this result is that river transport has played a limited role in disease propagation at the country scale after the initial contamination of the Artibonite River [8], and that neglecting hydrological connectivity can be compensated for by increased human mobility in model `M2` (estimates of m are indeed systematically larger in `M2` than in `M1`; see also [10], in which a fine-grained, spatially connected model without hydrological transport was successfully applied to the Haitian epidemic). An alternative explanation is the spatial resolution of the available epidemiological data, which are currently aggregated for administrative departments (i.e. the scale used in the coarse spatially connected model and in the spatially disconnected ones). As such, the current lack of more detailed epidemiological records may preclude a fair assessment of the actual explanatory/predictive power of the fine-scale spatially connected model. We thus suggest that the spatial scale of analysis should be suitably chosen so as to i) match (at least) the resolution of the available epidemiological records; ii) allow for the description of processes that

are deemed important for the epidemiological dynamics (such as hydrological transport, whose inclusion usually requires a finer scale of analysis than that dictated by administrative boundaries); iii) match the expectations of decision makers with respect to the level of spatial detail required in the epidemiological projections (i.e. prediction of large-scale patterns vs. local features) and in the assessment of intervention strategies. In this respect, coarse-grained spatially connected models (see e.g. [12]) could represent a good trade-off between spatial accuracy, implementation effort and computational requirements, and could be used effectively to deploy robust epidemic predictions relatively soon after the beginning of an outbreak. These findings may have implications that extend beyond cholera dynamics, possibly being relevant to other diseases, not necessarily waterborne. As an example, multi-layer network models like the one presented in this work could be used, with the necessary modifications, to study vector-borne epidemics, although in this case a proper characterization of the spatiotemporal patterns of vector movement may not be straightforward (see e.g. [56]). We also note that in endemic settings, provided that space-time data are available, other mathematical tools could be more appropriate than those used in this work, including time-series analysis [19, 57, 58, 21], spatially implicit [20] and Markov chain models [22, 59].

In conclusion, the wealth of data gathered during the Haiti cholera epidemic [3, 1, 8] paves the way for a new generation of epidemiological models, which will be required to accommodate real-time assimilation of epidemiological, hydrological and ecological information, as well as reliable projections of rainfall patterns, so as to improve epidemic forecasts and the evaluation of alternative interventions strategies (possibly within an adaptive management scheme; see e.g. [60, 61]). The inclusion of these features will turn tools that at their current stage are mostly descriptive into full-fledged decision-support systems for the prediction of the residual evolution of the epidemic and the design of optimal (ideally, in a multicriterial sense; e.g. [62]) intervention strategies. Such quantitative decision support tools will inform decision makers towards sustainable choices, i.e. towards the design of public health policies and sanitary interventions linked to the territory where the measures are to be implemented. Real-time support to epidemic management will also allow timely decisions and a quantitative assessment of alternative intervention strategies, thus possibly contributing to the optimization of sanitary and humanitarian efforts.

Acknowledgements

The Authors wish to thank Professor Mercedes Pascual (University of Michigan) and two anonymous reviewers for their insightful comments, and Dr. Jasper A. Vrugt (University of California, Irvine) for making available the source code of the DREAM$_{ZS}$ software. LM, EB and AR acknowledge the support provided by the European Research Council (ERC) advanced grant program through the project "River networks as ecological corridors for species, populations and waterborne disease" (RINEC 227612). LM,

EB, FF and AR acknowledge the support from the Swiss National Science Foundation (SNF/FNS) project "Dynamics and controls of large-scale cholera outbreaks" (DYCHO CR23I2_138104). LM and RC are also grateful to the students of Alta Scuola Politecnica for promoting stimulating discussions during the annual courses on infectious disease dynamics.

References

[1] E.J. Barzilay, N. Schaad, R. Magloire, K.S. Mung, J. Boncy, G.A. Dahourou, E.D. Mintz, M.W. Steenland, J.F. Vertefeuille, and J.W. Tappero. Cholera surveillance during the Haiti epidemic – The first 2 years. *New England Journal of Medicine*, 368:599–609, 2013.

[2] C.S. Chin, J. Sorenson, J.B. Harris, W.P. Robins, R.C. Charles, R.R. Jean-Charles, J. Bullard, D.R. Webster, A. Kasarskis, P. Peluso, E.E. Paxinos, Y. Yamaichi, S.B. Calderwood, J.J. Mekalanos, E.E. Schadt, and M.K. Waldor. The origin of the Haitian cholera outbreak strain. *The New England Journal of Medicine*, 364:33–42, 2011.

[3] R. Piarroux, R. Barrais, B. Faucher, R. Haus, M. Piarroux, J., Gaudart, R. Magloire, and D. Raoult. Understanding the cholera epidemic, Haiti. *Emerging Infectious Diseases*, 17:1161–1168, 2011.

[4] R.S. Hendriksen, L.B. Price, J.M. Schupp, J.D. Gillece, R.S. Kaas, D.M. Engelthaler, V. Bortolaia, T. Pearson, A.E. Waters, B.P. Upadhyay, S.D. Shrestha, S. Adhikari, G. Shakya, P.S. Keim, and F.M. Aarestrup. Population genetics of *Vibrio cholerae* from Nepal in 2010: evidence on the origin of the Haitian outbreak. *mBio*, 2:e00157, 2011.

[5] R.R. Frerichs, P.S. Keim, R. Barrais, and R. Piarroux. Nepalese origin of cholera epidemic in Haiti. *Clinical Microbiology and Infection*, 18:E158–E163, 2012.

[6] M. Eppinger, T. Pearson, S.S.K. Koenig, O. Pearson, N. Hicks, S. Agrawal, F. Sanjar, K. Galens, S. Daugherty, J. Crabtree, R.S. Hendriksen, L.B. Price, B.P. Upadhyay, G. Shakya, C.M. Fraser, J. Ravel, and P.S. Keim. Genomic epidemiology of the Haitian cholera outbreak: a single introduction followed by rapid, extensive, and continued spread characterized the onset of the epidemic. *mBio*, 5:e01721–14, 2014.

[7] F.D. Orata, P.S. Keim, and Y. Boucher. The 2010 cholera outbreak in Haiti: How science solved a controversy. *PLoS Pathogens*, 10:e1003967, 2014.

[8] J. Gaudart, S. Rebaudet, R. Barrais, J. Boncy, B. Faucher, M. Piarroux, R. Magloire, G. Thimothe, and R. Piarroux. Spatio-temporal dynamics of cholera during the first year of the epidemic in Haiti. *PLoS Neglected Tropical Diseases*, 7:e2145, 2013.

[9] J.R. Andrews and S. Basu. Transmission dynamics and control of cholera in Haiti: an epidemic model. *Lancet*, 377:1248–1252, 2011.

[10] E. Bertuzzo, L. Mari, L. Righetto, M. Gatto, R. Casagrandi, I. Rodriguez-Iturbe, and A. Rinaldo. Prediction of the spatial evolution and effects of control measures for the unfolding Haiti cholera outbreak. *Geophysical Research Letters*, 38:L06403, 2011.

[11] D.L. Chao, M.E. Halloran, and I.M. Longini. Vaccination strategies for epidemic cholera in Haiti with implications for the developing world. *Proceedings of the National Academy of Sciences USA*, 108:7081–7085, 2011.

[12] A.L. Tuite, J. Tien, M. Eisenberg, D.J.D. Earns, J. Ma, and D.N. Fisman. Cholera epidemic in Haiti, 2010: using a transmission model to explain spatial spread of disease and identify optimal control interventions. *Annals of Internal Medicine*, 154:593–601, 2011.

[13] M. Gatto, L. Mari, E. Bertuzzo, R. Casagrandi, L. Righetto, I. Rodriguez-Iturbe, and A. Rinaldo. Generalized reproduction numbers and the prediction of patterns in waterborne disease. *Proceedings of the National Academy of Sciences USA*, 48:19703–19708, 2012.

[14] A. Rinaldo, E. Bertuzzo, L. Mari, L. Righetto, M. Blokesch, M. Gatto, R. Casagrandi, M. Murray, S. Vesenbeckh, and I. Rodriguez-Iturbe. Reassessment of the 2010–2011 Haiti cholera outbreak and rainfall-driven multiseason projections. *Proceedings of the National Academy of Sciences USA*, 109:6602–6607, 2012.

[15] M.C. Eisenberg, G. Kujbida, A.R. Tuite, D.N. Fisman, and J.H. Tien. Examining rainfall and cholera dynamics in Haiti using statistical and dynamic modeling approaches. *Epidemics*, 5:197–207, 2013.

[16] Z. Mukandavire, D.L. Smith, and J.G. Morris. Cholera in Haiti: Reproductive numbers and vaccination coverage estimates. *Nature Scientific Reports*, 3:997, 2013.

[17] L. Righetto, E. Bertuzzo, L. Mari, E. Schild, R. Casagrandi, M. Gatto, I. Rodriguez-Iturbe, and A. Rinaldo. Rainfall mediations in the spreading of epidemic cholera. *Advances in Water Resources*, 60:34–46, 2013.

[18] E. Bertuzzo, F. Finger, L. Mari, M. Gatto, and A. Rinaldo. On the probability of extinction of the Haiti cholera epidemic. *Stochastic Environmental Research and Risk Assessment*, 28:in press, 2014.

[19] M. Pascual, X. Rodó, S.P. Ellner, R.R. Colwell, and M.J. Bouma. Cholera dynamics and El Niño Southern Oscillation. *Science*, 289:1766–1769, 2000.

[20] A.A. King, E.L. Ionides, M. Pascual, and M.J. Bouma. Inapparent infections and cholera dynamics. *Nature*, 454:877–880, 2008.

[21] M. Pascual, L.F. Chaves, B. Cash, X. Rodó, and M. Yunus. Predicting endemic cholera: the role of climate variability and disease dynamics. *Climate Research*, 36:131–140, 2008.

[22] R.C. Reiner, A.A. King, M. Emch, M. Yunus, A.S.G. Faruque, and M. Pascual. Highly localized sensitivity to climate forcing drives endemic cholera in a megacity. *Proceedings of the National Academy of Sciences USA*, 109:2033–2036, 2012.

[23] E. Bertuzzo, R. Casagrandi, M. Gatto, I. Rodriguez-Iturbe, and A. Rinaldo. On spatially explicit models of cholera epidemics. *Journal of the Royal Society Interface*, 7:321–333, 2010.

[24] L. Mari, E. Bertuzzo, L. Righetto, R. Casagrandi, M. Gatto, I. Rodriguez-Iturbe, and A. Rinaldo. On the role of human mobility in the spread of cholera epidemics: towards an epidemiological movement ecology. *Ecohydrology*, 5:531–540, 2012.

[25] E. Bertuzzo, M. Gatto, A. Maritan, S. Azaele, I. Rodriguez-Iturbe, and A. Rinaldo. On the space-time evolution of a cholera epidemic. *Water Resources Research*, 44:W01424, 2008.

[26] L. Mari, E. Bertuzzo, L. Righetto, R. Casagrandi, M. Gatto, I. Rodriguez-Iturbe, and A. Rinaldo. Modelling cholera epidemics: the role of waterways, human mobility and sanitation. *Journal of the Royal Society Interface*, 9:376–388, 2012.

[27] C. T. Codeço. Endemic and epidemic dynamics of cholera: the role of the aquatic reservoir. *BMC Infectious Diseases*, 1:1, 2001.

[28] S. Erlander and N. F. Stewart. *The Gravity Model in Transportation Analysis – Theory and Extensions*. VSP Books, Zeist, The Netherlands, 1990.

[29] J. Truscott and N. M. Ferguson. Evaluating the adequacy of gravity models as a description of human mobility for epidemic modelling. *PLoS Computational Biology*, 8:e1002699, 2012.

[30] J.B. Kaper, J.G. Morris, and M.M Levine. Cholera. *Clinical Microbiology Reviews*, 8:48–86, 1995.

[31] E.J. Nelson, J.B. Harris, J.G. Morris, S.B Calderwood, and A. Camilli. Cholera transmission: the host, pathogen and bacteriophage dynamic. *Nature Reviews Microbiology*, 7:693–702, 2009.

[32] D.M. Hartley, J.G. Morris, and D.L. Smith. Hyperinfectivity: a critical element in the ability of *Vibrio cholerae* to cause epidemics? *PLoS Medicine*, 3:63–69, 2006.

[33] D.G. Tarboton. A new method for the determination of flow directions and contributing areas in grid digital elevation models. *Water Resources Research*, 33:309–319, 1997.

[34] I. Rodriguez-Iturbe and A. Rinaldo. *Fractal River Basins. Chance and Self-Organization.* Cambridge University Press, New York, US, 1997.

[35] W.R. Gilks, S. Richardson, and D.J. Spiegelharter. *Markov Chain Monte Carlo in Practice.* Chapman and Hall, New York, USA, 1995.

[36] G. Hamra, R. MacLehose, and D. Richardson. Markov chain Monte Carlo: an introduction for epidemiologists. *International Journal of Epidemiology*, 42:627–634, 2013.

[37] C.J.F. ter Braak and J.A. Vrugt. Differential Evolution Markov Chain with snooker updater and fewer chains. *Statistics and Computing*, 18:435–446, 2008.

[38] R. Storn and K. Price. Differential Evolution – a simple and efficient heuristic for global optimization over continuous spaces. *Journal of Global Optimization*, 11:341–359, 1997.

[39] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, and E. Teller. Equations of state calculations by fast computing machines. *Journal of Chemical Physics*, 21:1087–1092, 1953.

[40] W.K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57:97–109, 1970.

[41] J.A. Vrugt, C.J.F. ter Braak, H.V. Gupta, and B.A. Robinson. Accelerating Markov Chain Monte Carlo simulation by Differential Evolution with self-adaptive randomized subspace sampling. *International Journal of Nonlinear Sciences and Numerical Simulation*, 10:271–288, 2009.

[42] H. Akaike. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19:716–723, 1974.

[43] K.P. Burnham, D.R., and Anderson. *Model Selection and Multimodel Inference: a Practical Information-Theoretic Approach.* Springer-Verlag, New York, N.Y., USA, 2002.

[44] J.A. Vrugt, C.J.F. ter Braak, C.G.H. Diks, B.A. Robinson, J.M. Hyman, and D. Higdon. Equifinality of formal (DREAM) and informal (GLUE) Bayesian approaches in hydrologic modeling? *Stochastic Environmental Research and Risk Assessment*, 23:1011–1026, 2009.

[45] E.L. Ionides, C. Bret'ø, and A.A. King. Inference for nonlinear dynamical systems. *Prooceedings of the National Academy of Sciences USA*, 103:18438–18443, 2006.

[46] J.A. Vrugt, C.J.F. ter Braak, M.P. Clark, J.M. Hyman, and B.A. Robinson. Treatment of input uncertainty in hydrologic modeling: doing hydrology backwards with Markov chain Monte Carlo simulation. *Water Resources Research*, 44:W00B09, 2008.

[47] J.A. Kopec, P. Finès, D.G. Manuel, D.L. Buckeridge, W.M. Flanagan, J. Oderkirk, M. Abrahamowicz, S. Harper, B. Sharif, A. Okhmatovskaia, E.C. Sayre, M.M. Rahman, and M.C. Wolfson. Validation of population-based disease simulation models: a review of concepts and methods. *BMC Public Health*, 10:710, 2010.

[48] M. Gatto, L. Mari, E. Bertuzzo, R. Casagrandi, L. Righetto, I. Rodriguez-Iturbe, and A. Rinaldo. Spatially explicit conditions for waterborne pathogen invasion. *The American Naturalist*, 182:328–346, 2013.

[49] S. Rebaudet, P. Gazin, R. Barrais, S. Moore, E. Rossignol, N. Barthelemy, J. Gaudart, J. Boncy, R. Magloire, and R. Piarroux. The dry season in Haiti: a window of opportunity to eliminate cholera. *PLoS Currents Outbreaks*, 2013.

[50] G.J. Huffman, R.F. Adler, D.T. Bolvin, G. Gu, E.J. Nelkin, K.P. Bowman, Y. Hong, E.F. Stocker, and D.B. Wolff. The TRMM multisatellite precipitation analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. *Journal of Hydrometeorology*, 8:38–55, 2007.

[51] V.E.M.B. de Rochars, J. Tipret, M. Patrick, L. Jacobson, K.E. Barbour, D. Berendes, D. Bensyl, C. Frazier, J.W. Domercant, R. Archer, T. Roels, J.W. Tappero, and T. Handzel. Knowledge, attitudes, and practices related to treatment and prevention of cholera, Haiti, 2010. *Emerging Infectious Diseases*, 17:2158–2161, 2011.

[52] A. Carpenter. Behavior in the time of cholera: Evidence from the 2008–2009 cholera outbreak in Zimbabwe. In W.G. Kennedy, N. Agarwal, and S.J. Yang, editors, *Social Computing, Behavioral-Cultural Modeling and Prediction*, volume 8393 of *Lecture Notes in Computer Science*, pages 237–244. Springer International Publishing, 2014.

[53] L. Bengtsson, X. Lu, A. Thorson, R. Garfield, and J. von Schreeb. Improved response to disasters and outbreaks by tracking population movements with mobile phone network data: a post-earthquake geospatial study in Haiti. *PLoS Medicine*, 8:e1001083, 2011.

[54] B. Hubbard, G. Lockhart, R.J. Gelting, and F. Bertrand. Development of Haiti's rural water, sanitation and hygiene workforce. *Journal of Water, Sanitation and Hygiene for Development*, 4:159–163, 2014.

[55] S. Basu and J.R. Andrews. Complexity in mathematical models of public health policies: A guide for consumers of models. *PLoS Medicine*, 10:e1001540, 2013.

[56] A. Wesolowski, N. Eagle, A.J. Tatem, D.L. Smith, A.M. Noor, R.W. Snow, and C.O. Buckee. Quantifying the impact of human mobility on malaria. *Science*, 6104:267–270, 2012.

[57] K. Koelle and M. Pascual. Disentangling extrinsic from intrinsic factors in disease dynamics: a nonlinear time series approach with an application to cholera. *The American Naturalist*, 163:901–913, 2004.

[58] D. Ruiz-Moreno, M. Pascual, M. Bouma, A. Dobson, and B. Cash. Cholera seasonality in Madras: dual role for rainfall in endemic and epidemic regions. *EcoHealth*, pages 52–62, 2007.

[59] F. Finger, A. Knox, E. Bertuzzo, L. Mari, D. Bompangue, M. Gatto, I. Rodriguez-Iturbe, and A. Rinaldo. Cholera in the Lake Kivu region (DRC): integrating remote sensing and spatially explicit epidemiological modeling. *Water Resources Research*, 50:5624–5637, 2014.

[60] C.S. Holling. *Adaptive Environmental Assessment and Management*. John Wiley & Sons, Chichester, UK, 1978.

[61] C. Allan and G.H. Stankey. *Adaptive Environmental Management: a Practitioner's Guide*. Springer/CSIRO Publishing, Collingwood, Australia, 2009.

[62] V. Belton and T. Stewart. *Multiple Criteria Decision Analysis: an Integrated Approach*. Kluver Academic Publishers, Dordrecht, the Netherlands, 2002.

Tables

Table 1: Summary of the main characteristics of the six model set-ups.

model	spatial scale	spatial coupling	epidemiological parameters	initial conditions	calibration parameters
M1	watershed	hydrological transport and human mobility	spatially homogeneous	estimated	9
M2	department	human mobility	spatially homogeneous	estimated	7
M3	department	none	spatially heterogeneous	estimated	60
M4	department	none	spatially heterogeneous	calibrated	70
M5	department	none	spatially homogeneous	estimated	6
M6	department	none	spatially homogeneous	calibrated	16

Table 2: AIC differences for the six tested set-ups of the model for different calibration interval lengths. Asterisk (*) indicates that the candidate set-up cannot be safely discarded ($\Delta AIC < 4$). Double asterisk (**) indicates that the model has not been included in model selection because the number of structural parameters exceeds the number of available data points.

calibration length (months)	M1	M2	M3	M4	M5	M6
1	0.0	1.3*	—**	—**	29.3	55.4
2	0.0	59.7	711.1	1552.1	96.3	113.7
3	0.0	69.6	302.6	406.7	143.4	157.9
4	0.0	80.5	234.7	284.2	209.7	102.4
5	0.0	82.5	223.9	213.6	214.0	19.0
6	0.0	165.0	246.1	263.1	261.4	54.2
7	0.0	95.0	247.6	87.3	254.6	9.3
8	7.6	58.6	175.5	115.7	182.6	0.0
9	23.5	140.5	159.0	24.2	210.9	0.0
10	44.3	149.4	172.2	55.5	244.2	0.0
11	43.3	132.5	175.3	6.5	291.7	0.0
12	60.4	141.5	234.6	20.7	307.1	0.0

Table 3: Validation performance indicators averaged over all combinations of calibration and validation windows. See text for details on model validation and the computation of the different indicators.

	M1	M2	M3	M4	M5	M6
$\langle V_1^{Bf} \rangle$	659	674	5711	3788	780	1467
$\langle V_1^{Ns} \rangle$	963	1122	2192	1985	1195	1420
$\langle V_1^{Me} \rangle$	674	712	1428	1084	836	801
$\langle V_2^{Bf} \rangle$	317	323	2326	1484	451	700
$\langle V_2^{Ns} \rangle$	552	705	1055	1073	760	759
$\langle V_2^{Me} \rangle$	323	382	734	640.2	488	447
$\langle V_3^{Bf} \rangle$	6.7	6.3	54.2	12.8	7.3	7.2
$\langle V_3^{Ns} \rangle$	13.0	15.6	19.6	19.7	15.8	14.0
$\langle V_3^{Me} \rangle$	6.5	7.3	12.1	9.5	7.8	6.7

Figure legends

Fig. 1. Data for the Haitian cholera epidemic model. (*a*) Digital terrain model; (*b*) geomorphological subdivision in hydrological units and main river networks; (*c*) administrative departments and cumulative cholera cases reported in the first year of the epidemic; (*d*) high-resolution population distribution map; (*e*) main road network infrastructure; (*f*) remotely sensed rainfall intensity (average 1998–2012). See text for data sources and details on spatial manipulation of the georeferenced datasets.

Fig. 2. Fitting of models `M1`–`M6` for different calibration windows. Shown are results for 2- (*a*), 4- (*b*), 6- (*c*) and 12-month (*d*) calibration windows. Black dots indicate total weekly incidence in data, while coloured lines represent best-fit model simulations. Data and results are aggregated at the country level for an easier visual reference, but model fitting is performed at a higher resolution, i.e. that of the administrative departments shown in Fig. 1*c*. See Fig. S3 for the whole set of calibration intervals (1–12 months with monthly steps, starting November 1st, 2010).

Fig. 3. Parameter values of model `M1` for different calibration windows (1–12 months). (*a*) Basic reproduction number (which depends on parameters as detailed in the text) and duration of acquired immunity [years]; (*b*) community-average fraction of mobile susceptible individuals and scale factor of the exponential kernel of human mobility [km]; (*c*) pathogen's hydrological dispersal rate [day^{-1}] and rainfall coefficient [day mm^{-1}]. Lines and shaded areas represent best-fit model simulations and 5–95% credible intervals of parameter uncertainty. Non-calibrated parameters set as in [14]: $\gamma = 0.2$, $\alpha = 4.0\ 10^{-3}$, $\mu = 4.5\ 10^{-5}$ (all in [day^{-1}]).

Fig. 4. Model validation for different durations of the calibration window. Shown are results for 2- (*a*), 4- (*b*), 6- (*c*) and 12-month (*d*) calibration windows. Validation intervals span from one month to one year, and begin right after the end of the relevant calibration window. Black dots indicate total weekly incidence in data, while coloured lines represent best-fit model simulations (thick line) or the medians (thin) of model predictions. See Fig. S4 for the whole set of calibration intervals.

Fig. 5. Quantitative assessment of validation results. Panels *a* to *i* show the best performing model for each combination of calibration/validation windows for different validation performance indicators (labels). Raw values of the validation indicators are given in Figs. S5–S7.

Figures

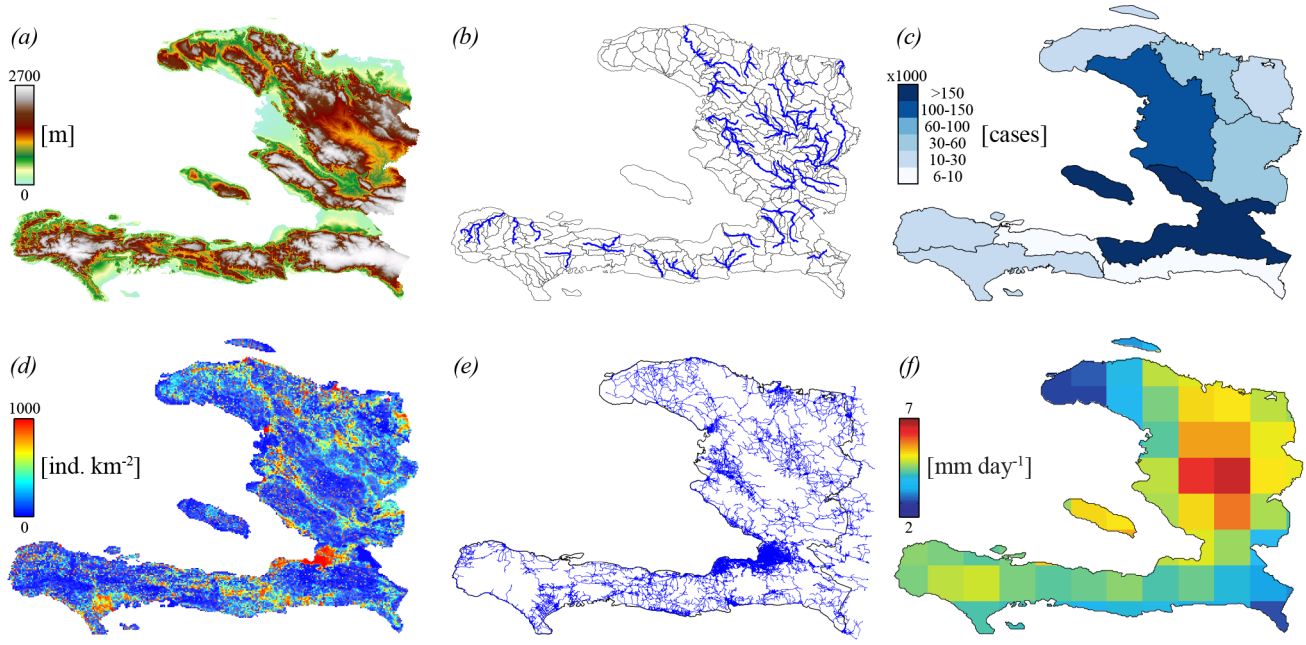

Figure 1:

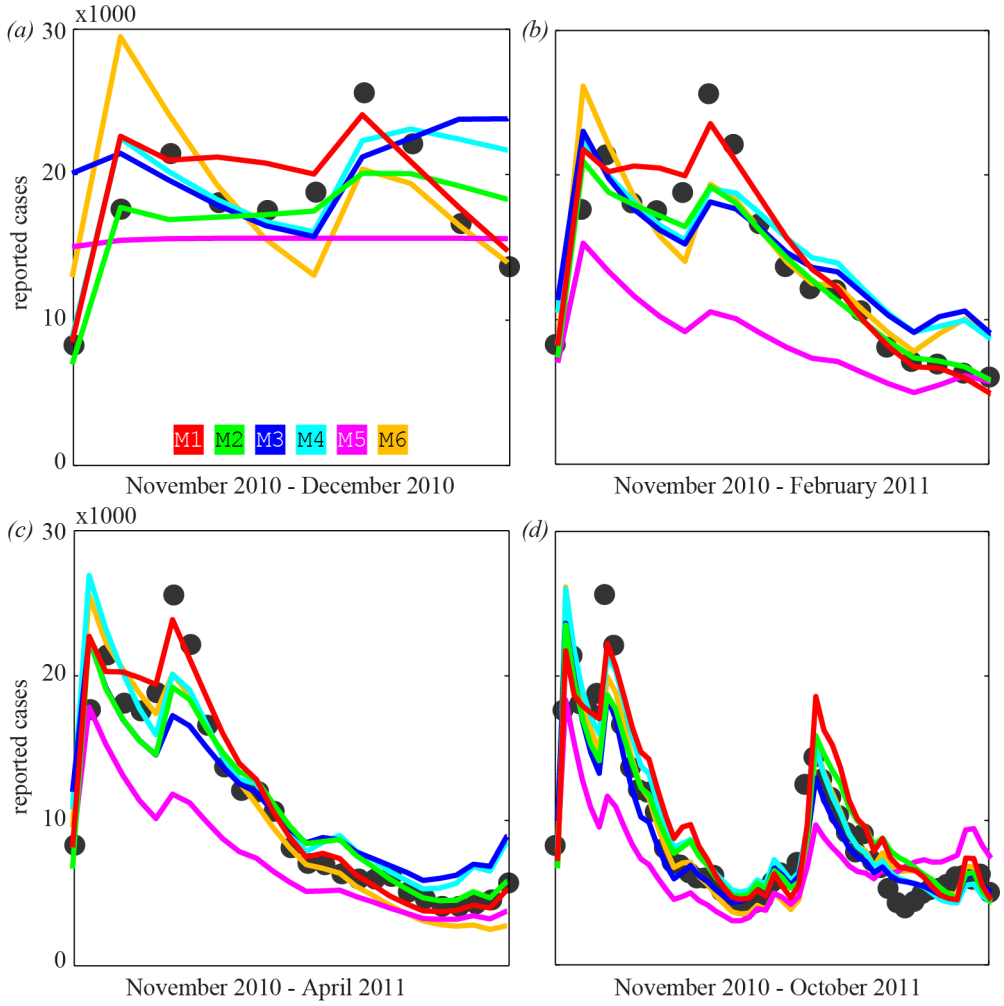
(a)
x1000
30
20
10
0
reported cases
M1
M2
M3
M4
M5
M6
November 2010 - December 2010
(b)
November 2010 - February 2011
(c)
x1000
30
20
10
0
reported cases
November 2010 - April 2011
(d)
November 2010 - October 2011

Figure 2:

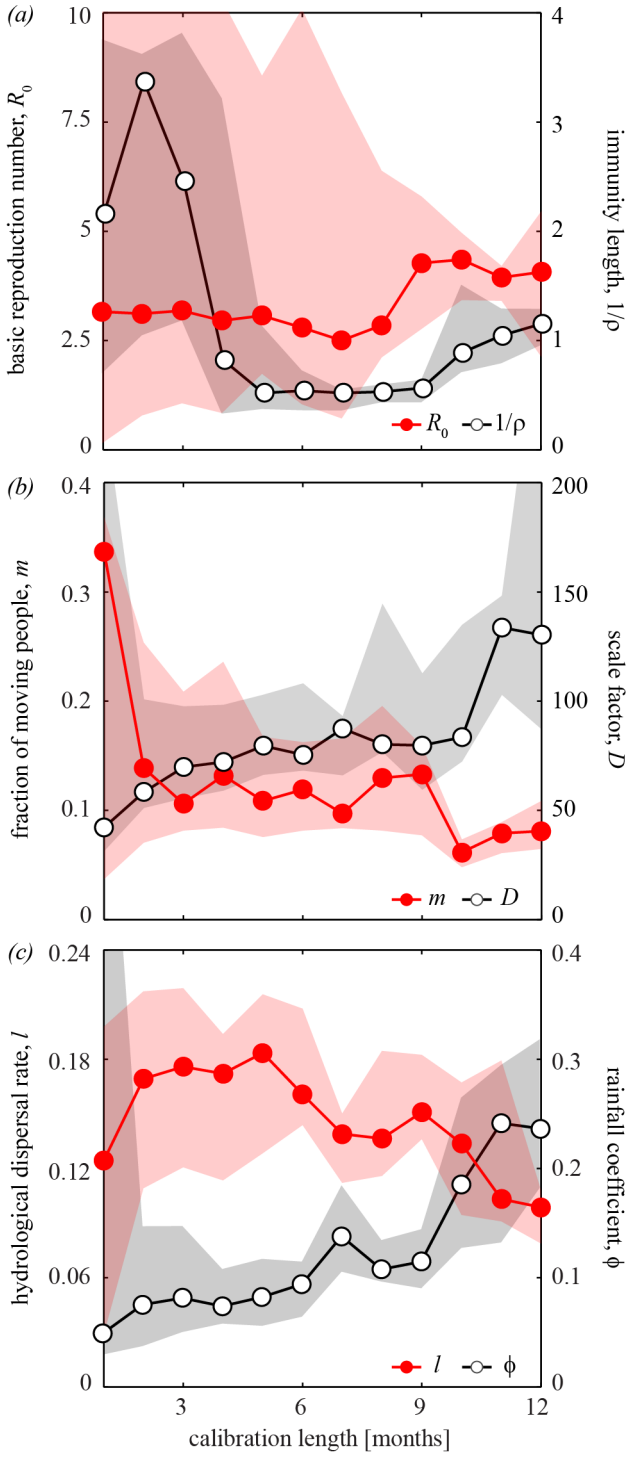
(a)
basic reproduction number, R_0
immunity length, 1/ρ
R_0
1/ρ
(b)
fraction of moving people, m
scale factor, D
m
D
(c)
hydrological dispersal rate, l
rainfall coefficient, ϕ
l
ϕ
calibration length [months]

Figure 3:

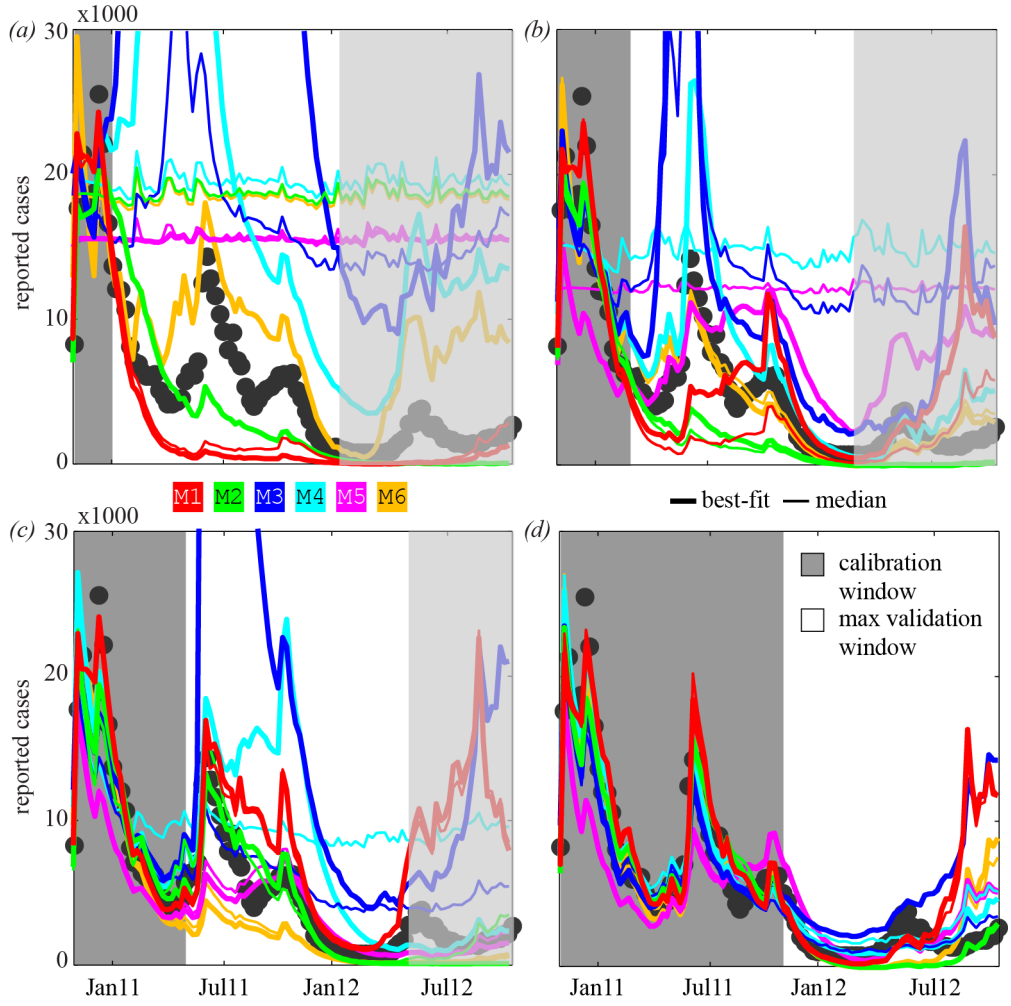
(a)
x1000
30
20
10
0
reported cases
M1
M2
M3
M4
M5
M6
(b)
best-fit
median
(c)
x1000
30
20
10
0
reported cases
Jan11
Jul11
Jan12
Jul12
(d)
calibration window
max validation window
Jan11
Jul11
Jan12
Jul12

Figure 4:

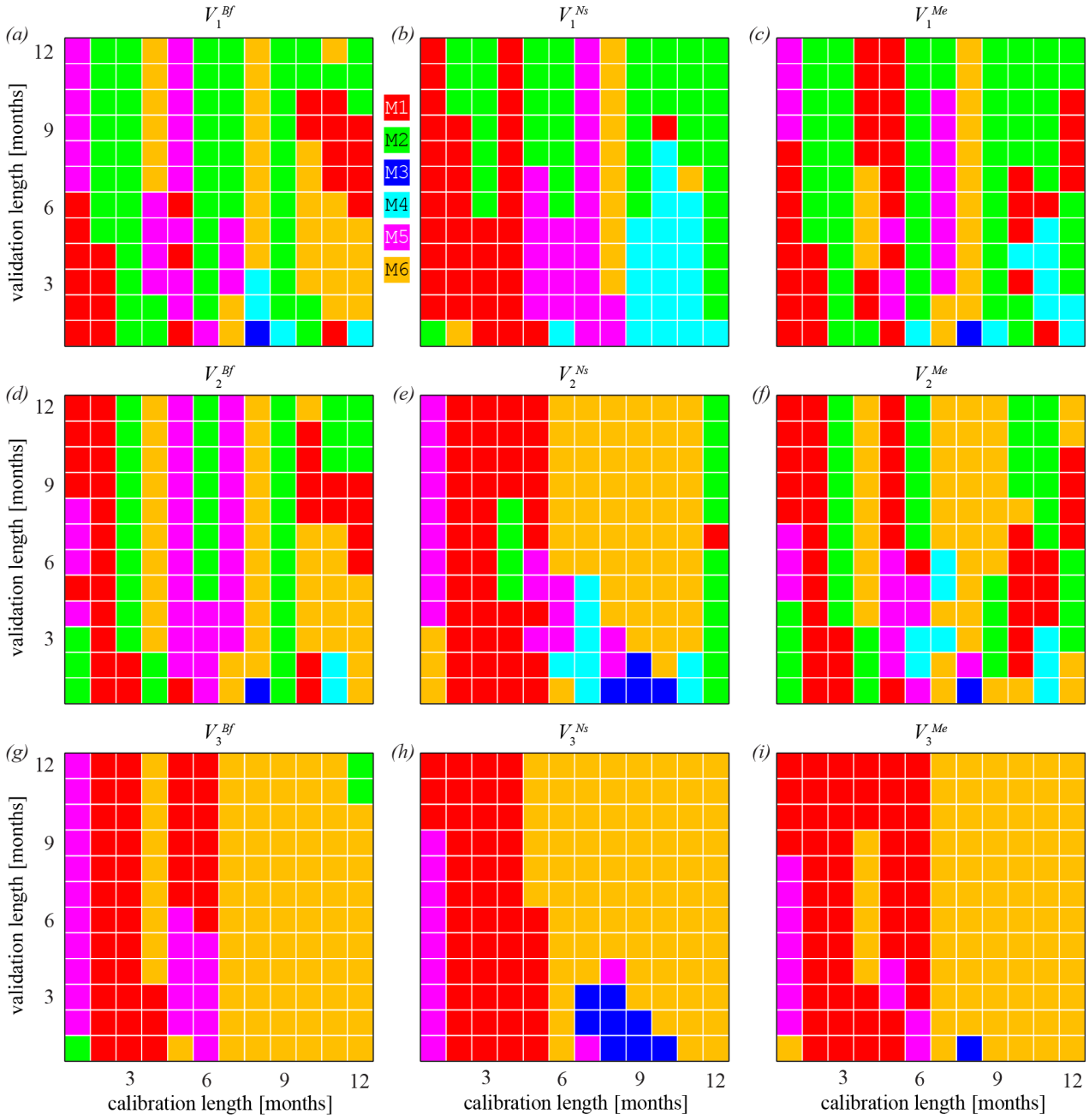
(a)
V_1^{Bf}
(b)
V_1^{Ns}
(c)
V_1^{Me}
(d)
V_2^{Bf}
(e)
V_2^{Ns}
(f)
V_2^{Me}
(g)
V_3^{Bf}
(h)
V_3^{Ns}
(i)
V_3^{Me}
M1
M2
M3
M4
M5
M6
validation length [months]
calibration length [months]
3
6
9
12

Figure 5: