# Rainfall mediations in the spreading of epidemic cholera

L. Righetto[1], E. Bertuzzo[1], L. Mari[1,2], E. Schild[1], R. Casagrandi[2], M. Gatto[2], I. Rodriguez-Iturbe[3] and A. Rinaldo [1,4]

1 Laboratory of Ecohydrology, School of Architecture, Civil and Environmental Engineering, Ecole Polytechinque Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland
2 Dipartimento di Elettronica, Informazione e Bioingegneria Politecnico di Milano, 20133 Milano, Italy
3 Department of Civil and Environmental Engineering, Princeton University, Princeton, 08544 NJ, USA.
4 Dipartimento di Ingegneria Civile, Edile ed Ambientale, Universit di Padova, 35131 Padua, Italy.

**waterborne disease, hydrologic drivers epidemiology, ecohydrology,**

*Preprint submitted to Advances in Water Resources* *June 4, 2013*

## Abstract

Following the empirical evidence of a clear correlation between rainfall events and cholera resurgence that was observed in particular during the recent outbreak in Haiti, a spatially explicit model of epidemic cholera is re-examined. Specifically, we test a multivariate Poisson rainfall generator, with parameters varying in space and time, as a driver of enhanced disease transmission. The relevance of the issue relates to the key insight that predictive mathematical models may provide into the course of an ongoing cholera epidemic aiding emergency management (say, in allocating life-saving supplies or health care staff) or in evaluating alternative management strategies. Our model consists of a set of dynamical equations (SIRB-like i.e. subdivided into the compartments of Susceptible, Infected and Recovered individuals, and including a balance of Bacterial concentrations in the water reservoir) describing a connected network of human communities where the infection results from the exposure to excess concentrations of pathogens in the water. These, in turn, are driven by rainfall washout of open-air defecation sites or cesspool overflows, hydrologic transport through waterways and by mobility of susceptible and infected individuals. We perform an *a posteriori* analysis (from the beginning of the epidemic in October 2010 until December 2011) to test the model reliability in predicting cholera cases and in testing control measures, involving vaccination and sanitation campaigns, for the ongoing epidemic. Even though predicting reliably the timing of the epidemic resurgence proves difficult due to rainfall inter-annual variability, we find that the model can reasonably quantify the total number of reported infection cases in the selected time-span. We then run a multi-seasonal prediction of the course of the epidemic until December 2015, to investigate conditions for further resurgences and endemicity of cholera in the region with a view to policies which may bring to the eradication of the disease in Haiti. The projections show an endemic, seasonal pattern establishing in the region, which can be better forestalled by an improvement of the sanitation system only rather than by vaccination alone. We thus conclude that hydrologic drivers and water resources management prove central to prediction, emergency management and long-term control of epidemic cholera.

## 1. Introduction

The recent, still ongoing cholera outbreak that has struck Haiti has brought to broad public attention the magnitude of the loss of human lives and of the social and economic disruption caused, even to date, by epidemics of the disease. The global relevance of the problem and the need for a preventive assessment and control of cholera spreading is manifest also in view of other recent or ongoing outbreaks in the Congo river basin, Cuba, Sierra Leon and the Sahel region (Luque Fernandez et al., 2009; Kelvin, 2011; Nkoko et al., 2011; Al-Tawfiq and Memish, 2012).

While the role of climatic conditions, and rainfall in particular, on patterns of waterborne infections have long been studied especially in empirical frameworks (Pascual et al., 2000; Lipp et al., 2002; Altizer et al., 2006; de Magny et al., 2008; Emch et al., 2008; Koelle, 2009), hydrologically-driven, spatially explicit mathematical models of cholera epidemics have only recently been developed (Bertuzzo et al., 2008, 2010). They have been applied to study the course of the Haitian epidemic, starting from the very first months after its insurgence in late 2010 (Bertuzzo et al., 2011; Tuite et al., 2011; Chao et al., 2011), and following disease resurgence occurred in May 2011 in connection with unusually intense tropical rains (Rinaldo et al., 2012). Even though concerns for correct surveillance, monitoring and intervention planning have been on the rise in international institutions debate, regarding cholera in particular (e.g. WHO, 2011), none of these models have been utilized to date to test their effectiveness as predictive and control tools. Such models could be in principle applied, for instance, to deploy medical staff and life-saving supplies through projections of the patterns of cholera infections, and to implement pro-active rather than reactive policies as commonplace in epidemiological control strategies.

The Haitian epidemic represents more than just another test case. In fact, cholera had never been reported in Haiti before 2010 and therefore it is likely that the population had no significant prior exposure or acquired immunity to the disease, suggesting that the entire population was initially susceptible to infection. Moreover, once a cholera epidemic starts, infected patients excrete huge numbers of *Vibrio cholerae* bacteria which spread either through water pathways (via active and passive dispersal; Bertuzzo et al., 2008, 2010; Chao et al., 2011; Righetto et al., 2011; Mari et al., 2012a) or through human mobility networks involving both susceptibles and infected individuals (Tuite et al., 2011; Chao et al., 2011; Mari et al., 2012b). Thus the poor sanitation

conditions experienced especially after the disastrous 2010 earthquake that struck the island, facilitated both types of spread and fostered the abundance of microorganisms in the water system, thus rendering the Haitian outbreak exemplary (Rinaldo et al., 2012).

The Haiti epidemic also provided direct and compelling evidence relating cholera resurgence to environmental drivers, specifically to rainfall patterns. Little insight could be gained, in fact, from past empirical studies correlating rainfall to cholera cases because most, if not all, previous studies were carried out in contexts where cholera is endemic (see e.g. Lipp et al., 2002; Emch et al., 2008). In fact, reported correlations between rainfall events and resurgences – both in their sign and time lag – have been rather disparate (Ruiz-Moreno et al., 2007; Akanda et al., 2009; Luque Fernandez et al., 2009; Hashizume et al., 2010). This reflects the range of potential mechanisms through which rainfall may affect increased exposure to risk of infections (e.g. crowding effects due to flooding; raw sewage contamination of water sources; increased availability of compounds boosting *V. cholerae* survival or toxins diminishing it; increased contamination due to over-exploitation of the water reservoirs, to name a few). Rinaldo et al. (2012) have shown how such correlation could be implemented in epidemiological models by forcing the contamination of the local water reservoir through rainfall-runoff transfer of *V. cholerae* from waste- to drinking-water. In the spatially explicit framework presented in Rinaldo et al. (2012) – which includes a family of models encompassing different epidemiological and hydrological assumptions – Haiti is depicted as a network of human communities (the nodes) connected by both hydrology and human mobility (the edges). Each community is represented by a system of Ordinary Differential Equations (ODE), in which the population is divided into Susceptible ($S$), Infected ($I$) and Recovered ($R$) individuals. The evolution of the concentration of *V. cholerae* in the environmental water reservoir is also considered. Here we further extend that approach, generating scenarios of precipitation to perform epidemiological predictions and to evaluate *a priori* the impact of intervention policies.

Unlike cholera, rainfall predictions are an established endeavour (Rodriguez-Iturbe et al., 1986; Cho et al., 1987) and rainfall stochastic generators have recently been widely considered for studying precipitation patterns (Laio et al., 2001), also in the light of the inclusion of a description of superstatistics of interannual variability (Porporato et al., 2006). Here we use a Poisson generator that takes into account both the inter-annual and the spatial variabilities of rainfall intensity in order to preserve space/time correlations while

generating rainfall at local scale. The identification of statistically equivalent spatio-temporal aggregates is carried out using suitable clustering techniques. This approach allows to generate a large number of precipitation scenarios, naturally preserving the statistical properties of the rainfall dataset.

We make use of these synthetic rainfall fields to force our epidemiological model and to obtain, as a result, estimates of the strength of the disease resurgence. It should be noted that our attempt differs substantially from, say, classical hydrological predictions, as several epidemiological and social processes are acting simultaneously on top of the rainfall dynamics we try to reproduce. As the magnitude of many of these processes is often uncertain (sometimes being even difficult to identify correctly the whole set of intervening processes), epidemiological predictions are particularly challenging. Here, we perform two types of analysis: i) an *ex post* evaluation, in which calibration, validation and prediction all belong to the past course of the outbreak ; ii) multi-seasonal projections, from the current state of the epidemic to the next few years in which cholera is speculated to become endemic in the region (Mukandavire et al., 2013). The first analysis simulates real-time conditions in which short-term (a few month) scenarios of cholera resurgence are used to evaluate the performance of the model as a predictive and control tool during the very course of an epidemic. We then analyze the effect of different, alternative scenarios of intervention (sanitation and vaccination, possibly differing in timing and in spatial distribution) on the evolution of the outbreak to mimic model-guided intervention policies. We study whether the inference of the most effective policy – say, that aiming at the maximum reduction of the total number of reported cases in a given time frame – may still hold in the face of the actual development of disease resurgence. In the long-term case, the study of correlations of cyclic resurgence of the disease with the seasonal rainfall cycle matters, as the particular initial conditions that have favored the appearance of cholera in Haiti – i.e. a high number of susceptibles – will no longer apply in the future. The epidemic, in fact, can be expected to revamp in particular conditions of stress (e.g. extreme rainfall events) with an intensity that depends on the rate at which recovered individuals lose their temporary acquired immunity to the disease. This kind of analysis allows also to estimate the amount of sanitation or the extent of a vaccination campaign aimed at eradicating the disease from the region, and is deeply rooted in hydrologic sciences.

The paper is organized as follows. In Section 2 we detail the epidemiological, spatially explicit models of cholera spreading used in this work and in

section 3 we define the procedure to obtain the rainfall generation model. In Section 4 we present and discuss the results relative to both predictions and intervention efficacy evaluation, for both short-term and long-term scenarios. A set of conclusions closes then the paper.

## 2. Spatially explicit epidemiological models for the Haitian epidemic

We make use here of some of the models presented in Rinaldo et al. (2012), who have constructed a spatially explicit framework for the description of the Haitian epidemic and whose approach evolved from the first Haiti application by Bertuzzo et al. (2011). In particular, we restrict here our analysis to the two models which emerged as best performing under absorbing or diffusive boundary conditions (Rinaldo et al., 2012, see Fig. *S*8 therein). They consider $n$ communities ($i = 1, n$) spatially distributed within a given domain that embeds the hydrologic and the human mobility networks (Fig. 1). Let $S_i(t)$, $I_i(t)$ and $R_i(t)$ be the local abundances of susceptible, infected and recovered individuals in each node $i$ of the river network at time $t$, and let $B_i(t)$ be the concentration of *V. cholerae* in the water reservoir at site $i$.

### *2.1. Basic model and dynamics*

Epidemiological dynamics, pathogen transport and human mobility can be described by the following set of coupled differential equations, which includes most of the mechanisms common to the models and represents the simplest of the two models here considered:

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

The host population is assumed to be at a demographic equilibrium, where $\mu$ is the human mortality rate, $H_i$ is the population size of the local community and $\mu H_i$ a constant recruitment. The force of infection $\mathcal{F}_i(t)$, which represents the rate at which susceptible individuals become infected via contact with contaminated water, is expressed as:

$$\mathcal{F}_i(t) = \beta \left[ (1-m)\frac{B_i}{K+B_i} + m\sum_{j=1}^{n} Q_{ij}\frac{B_j}{K+B_j} \right].$$

The parameter $\beta$ represents the rate of exposure to contaminated water, and $B_i/(K+B_i)$ is the probability of becoming infected due to the exposure to a concentration $B_i$ of *V. cholerae*, $K$ being the half-saturation constant (Capasso and Paveri-Fontana, 1979; Codeço, 2001). The parameter $m$ represents the fraction of individuals that travel outside their node. Because of human mobility, infection in a given node depends on the local concentration $B_i$ for a fraction $1-m$ of the susceptible hosts and on the concentration of the destination community $B_j$ for the remaining fraction $m$. The concentrations $B_j$ are weighted according to the probability $Q_{ij}$ that an individual (a susceptible one, in this case) living in node $i$ would reach $j$ as a destination. Choosing a gravity-like pattern to describe human mobility (Erlander and Stewart, 1990), one can define connection probability as:

$$Q_{ij} = \frac{H_j e^{-d_{ij}/D}}{\sum_{k\neq i}^{n} H_k e^{-d_{ik}/D}},$$

where the attractiveness factor of node $j$ depends on its abundance, while the deterrence factor is assumed to be dependent on distance $d_{ij}$ and represented by an exponential kernel (with shape factor $D$). Infected individuals recover at a rate $\gamma$, or die for natural or cholera-induced mortality at a rate $\mu$ or $\delta$, respectively. Recovered individuals $R_i(t)$ lose their immunity and return to the compartment of susceptibles at a rate $\rho$ or die for natural mortality at a rate $\mu$. Bacterial shedding in node $i$ – quantified by the rate $p/W_i$, where $p$ is the *per-capita* contamination rate of infectives and $W_i$ is the volume of the local water reservoir (here assumed to be proportional to the population size, i.e. $W_i = cH_i$ as in Rinaldo et al., 2012) – is regulated by the total infective pool $\mathcal{G}_i(t)$ which is defined as:

$$\mathcal{G}_i(t) = (1-m)I_i + m\sum_{j=1}^{n} Q_{ji} I_j .$$

Similarly to the the force of infection, the total infective pool accounts for human mobility and in particular for infected people who travel from community $j$ to the focal community $i$.

In order to express the worsening of sanitation conditions caused by rainfall-induced runoff, which causes additional loads of pathogens to be cast into the water reservoir, the contamination rate $p$ is increased by rainfall intensity $J_i(t)$ via a coefficient $\phi$. Bacteria are also supposed to be endowed with an environmental mortality $\mu_B$ and to be transported along the hydrologic network at rate $l$. We assume that pathogens can travel from node $i$ to $j$ with probability $P_{ij}$. In this case we assume $P_{ij} = 1$ if $j$ is a downstream nearest neighbor of node $i$ and zero otherwise. We consider partially reflecting boundaries at the outlets of all the river systems to simulate specific coastal hydrologic settings, such as upstream transport (i.e. diffusion prevailing on advection in low flow conditions) or more favorable local environmental conditions (i.e. a higher salinity) causing higher viability of bacterial populations close to the coast. We fix the fraction of reflected particles at the outlets at 0.5, which corresponds to diffusive conditions.

Epidemiological records usually refer to reported disease incidence (i.e. number of reported cases per unit time) not to disease prevalence (current number of infected individuals). In order to derive disease incidence from the model one simply needs to i) compute the cumulative reported cases $C_i(t)$ by solving

$$\frac{dC_i}{dt} = \sigma \mathcal{F}_i(t) S_i \ ,$$

where $\sigma$ represents the fraction of total infected people that show severe symptoms and are thus likely reported (here $\sigma = 0.2$, PAHO, 2010, see the Appendix for details) and ii) differentiate $C_i(t)$ in time.

As in previous applications (Bertuzzo et al., 2011; Rinaldo et al., 2012), we assume that before the epidemic the whole population is susceptible, i.e. $S_i(0) = H_i$ because of the lack of any pre-existing immunity (Enserink, 2010; Walton and Ivers, 2011; Sack, 2011; Piarroux et al., 2011). The model is also initialized by some infected individuals being placed in the locations of the first reported cases (see again Piarroux et al., 2011).

### *2.2. Hyperinfectivity*

The second model we test here accounts for a hyperinfective state of *V. cholerae*, caused by passage through human intestine (Merrell and al., 2002; Alam and al., 2005), which has already been used in modeling exercises

(Hartley et al., 2006), also in the context of the Haiti epidemic (Andrews and Basu, 2011; Chao et al., 2011). Thus, an equation describing the dynamics of the hyperinfective stage of *V. cholerae* has to be added to model 1, so that such model can be modified as follows:

$$\begin{aligned}
\frac{dS_i}{dt} &= \mu(H_i - S_i) - \mathcal{F}_i^{HI}(t)S_i + \rho R_i \\
\frac{dI_i}{dt} &= \mathcal{F}_i^{HI}(t)S_i - (\gamma + \mu + \delta)I_i \\
\frac{dR_i}{dt} &= \gamma I_i - (\rho + \mu)R_i \\
\frac{d\mathcal{B}_i}{dt} &= -\xi\mathcal{B}_i - l\left(\mathcal{B}_i - \sum_{j=1}^{n} P_{ji}\frac{W_j}{W_i}\mathcal{B}_j\right) + \frac{p}{W_i}\left[1 + \phi J_i(t)\right]\mathcal{G}_i(t) \\
\frac{dB_i}{dt} &= \xi\mathcal{B}_i - \mu_B B_i - l\left(B_i - \sum_{j=1}^{n} P_{ji}\frac{W_j}{W_i}B_j\right), \qquad (2)
\end{aligned}$$

where $\mathcal{B}_i$ is the concentration of hyperinfective pathogens in the water reservoir and $\xi$ is the rate at which *V. cholerae* lose hyperinfectivity and convert to normal state (here we impose $1/\xi = 1$ day; Hartley et al., 2006). The total contact rate $\mathcal{F}_i^{HI}(t)$ has to take into account the joint effect of hyperinfective and regular *V. cholerae*, i.e.:

$$\begin{aligned}
\mathcal{F}_i^{HI}(t) &= \beta\left[(1-m)\left(\frac{B_i}{K + B_i} + \frac{\mathcal{B}_i}{K_{HI} + \mathcal{B}_i}\right) + \right. \\
&+ \left. m\sum_{j=1}^{n} Q_{ij}\left(\frac{B_j}{K + B_j} + \frac{\mathcal{B}_j}{K_{HI} + \mathcal{B}_j}\right)\right],
\end{aligned}$$

where $K_{HI}$ is the half-saturation constant for hyperinfective bacteria ($K/K_{HI} \approx 50$; Hartley et al., 2006). Analogously to model 1, reported cases for model 2 can be computed by solving $dC_i/dt = \sigma\mathcal{F}_i^{HI}(t)S_i$ and differentiating the cumulative reported cases with respect to time.

### *2.3. Modeling of interventions*

We detail here how actions of public health management may be evaluated by (and integrated in) our model. One of the possible key uses of

epidemiological, large-scale predictions, in fact, is the evaluation of the effectiveness of different intervention strategies. In the case of a waterborne disease such as cholera, two major categories of intervention can be taken into account: sanitation of the water supply system and vaccination. In the former case, several active (e.g. chlorination tablets, direct clean water supply) or pro-active (education campaigns) actions can be taken to increase the sanitation level in a country. The application of these measures decreases the probability of ingesting contaminated water and/or the probability that critical concentrations of pathogens may reach the water reservoir. They can be expressed as a decrease, respectively, of the contact rate $\beta(t)$ and/or of the contamination rate $p(t)$, which can also be spatially-distributed in order to represent localized policies. We model, instead, vaccinations as an outgoing flux of individuals being immunized from the disease from the susceptible compartment. Two doses, administered separately with a 10-14 days delay between the two, are required for current standard cholera vaccination, in which immunization builds up 7 days after the second dose (Jertborn et al., 1993). Standard cholera vaccines only grant temporary immunization, so that immunity loss has to be taken into account in the evaluation of long-term intervention scenarios. Field trials (Clemens, 1990; Girard et al., 2006) suggest parameter values of immunity loss for the only vaccine acknowledged by WHO. In particular, we model here the effects of vaccines as observed by Clemens (1990) in Bangladesh. The vaccine grants 85% immunization in the first 6 months, 60% in the following 18 months and 20% in the third and last year of (partial) immunization . This means that vaccinated people still have a finite probability of contracting the disease, and they may become fully susceptible again once their immunity is lost, after three years. We thus introduce four new state variables in our model: $S_i^v$ which quantifies the abundance of individual vaccinated whose immunity has not built up yet, and $V_{i,I/II/III}(t)$ which represent the abundance of individuals who have been immunized less than 6, between 6 and 24, and more than 24 months before time $t$, respectively. The fluxes among these compartments are illustrated in Fig. 1, panel A. We detail here the implementation of vaccinations in model 2:

$$
\begin{aligned}
\frac{dS_i}{dt} &= \mu(H_i - S_i) - \mathcal{F}_i^{HI}(t)S_i + \rho R_i - \nu_i(t)\frac{S_i}{H_i - C_i - \int_t^{t_v}\nu_i(t)dt} + \rho_{V,III}V_{i,III} \\
\frac{dS_i^v}{dt} &= -\mu S_i^v - \mathcal{F}_i^{HI}(t)S_i^v + \nu_i(t)\frac{S_i}{H_i - C_i - \int_t^{t_v}\nu_i(t)dt} - \omega S_i^v \\
\frac{dV_{i,I}}{dt} &= \omega S_i^v - (1-\psi_{V,I})\mathcal{F}_i^{HI}(t)V_{i,I} - (\rho_{V,I}+\mu)V_{i,I} \\
\frac{dV_{i,II}}{dt} &= \rho_{V,I}V_{i,I} - (1-\psi_{V,II})\mathcal{F}_i^{HI}(t)V_{i,II} - (\rho_{V,II}+\mu)V_{i,II} \\
\frac{dV_{i,III}}{dt} &= \rho_{V,II}V_{i,II} - (1-\psi_{V,III})\mathcal{F}_i^{HI}(t)V_{i,III} - (\rho_{V,III}+\mu)V_{i,III} \\
\frac{dI_i}{dt} &= \mathcal{F}_i^{HI}(t)\left[S_i + S_i^v + (1-\psi_{V,I})V_{i,I} + (1-\psi_{V,II})V_{i,II} + (1-\psi_{V,III})V_{i,III}\right] - (\gamma+\mu+\delta)I_i \\
\frac{dR_i}{dt} &= \gamma I_i - (\rho+\mu)R_i \\
\frac{d\mathcal{B}_i}{dt} &= -\xi\mathcal{B}_i - l\left(\mathcal{B}_i - \sum_{j=1}^{n} P_{ji}\frac{W_j}{W_i}\mathcal{B}_j\right) + \frac{p}{W_i}\left[1+\phi J_i(t)\right]\mathcal{G}_i(t) \\
\frac{dB_i}{dt} &= \xi\mathcal{B}_i - \mu_B B_i - l\left(B_i - \sum_{j=1}^{n} P_{ji}\frac{W_j}{W_i}B_j\right)
\end{aligned}
\tag{3}
$$

Susceptibles are recruited according to the vaccination rate $\nu_i(t)$, which represents the number of doses administered per day, possibly varying in time/space. It should be noted that the denominator $H_i - C_i - \int_t^{t_v}\nu_i(t)dt$ is the pool of potential candidates for vaccination. Because most infected individuals do not develop acute symptoms, only people whose infection has been reported ($C_i$), or to whom the vaccine has already been administered ($\int_t^{t_v}\nu_i(t)dt$), can be safely excluded from the campaign. Therefore the probability of giving vaccine to people who are actually susceptible to cholera is represented by the fraction $S_i/\left[H_i - C_i - \int_{t_v}^{t}\nu_i(t)dt\right]$. Vaccinated individuals $S_i^v$ gain immunity to the disease at rate $\omega$ (here $\omega = 0.5$ days$^{-1}$). According to the observations made by Clemens (1990), loss of immunity of vaccinated individuals progresses at rates $1/\rho_{V,I} = 1/2$ year, $1/\rho_{V,II} = 1.5$ years and $1/\rho_{V,III} = 1$ year. During these periods, these individuals develop a partial immunity, which means that they can contract the disease with probability (1-$\psi_{V,i}$), with $\psi_{V,I} = 0.85$, $\psi_{V,II} = 0.6$ and $\psi_{V,III} = 0.2$. Notice that

for this model the cumulative reported cases can be computed integrating $dC_i/dt = \sigma \mathcal{F}_i^{HI}(t)(S_i + S_i^v + (1-\psi_{V,I})V_{i,I} + (1-\psi_{V,II})V_{i,II} + (1-\psi_{V,III})V_{i,III})$.

*2.4. Parameter calibration and model selection*

While several parameters are estimated from the literature (see Table 1 for the numerical values and for the relevant references),the remaining five are obtained through calibration. Introducing the dimensionless bacterial concentrations $B^* = B/K$ and $\mathcal{B}^*/K_{HI}$ it is possible to group three parameters of model 1 and 2 in a single ratio $\theta = p/(cK)$, whose value is determined through calibration. The other four are the hydrologic transport rate $l$, the fraction of moving people $m$, the deterrence distance $D$ and the coefficient $\phi$. We calibrate the models using a Markov Chain Monte Carlo sampling algorithm (ter Braak and Vrugt, 2008, see the Appendix for details). The goodness of each single simulation is computed as the residual sum of squares (RSS) between weekly reported cholera cases in each of the ten Haitian departments as recorded in the epidemiological dataset and simulated by the model being tested. The numerical values of the best-fit parameters of both models are reported in Table 2. Model selection is then carried out using the Akaike Information Criterion (AIC, Akaike, 1974, see the Appendix).

We perform here a first test calibration run, from the start of the epidemic (23/10/2010) until 28/05/2011, for both the presented models. The results of this run, expressed with the RSS and AIC values, are shown in Table 3. We use this fitting to perform an *ex-post* evaluation of the performance of model predictions, in particular regarding the resurgence of the disease of the summer/autumn of 2011 (so until 31/12/2011).

In this first run, none of the models emerges as performing significantly better (as the Akaike difference must be $> 4$ for significance, Akaike, 1974; Burnham and Anderson, 2002; Corani and Gatto, 2007). We choose model 2 as best ranked in the first run and as the model including the higher level of detail and realism and we perform a second, long-term calibration run only for this model until 14/01/2012 and project cholera patterns up to December 2015. In this case, we also tune the parameters which play a major role in the long-term dynamics of endemic periodic resurgence: the loss of immunity rate, $\rho$, and the fraction of symptomatic infected individuals, $\sigma$. These two parameters control the rate at which the susceptible pool is emptied ($\sigma$) and replenished ($\rho$). At the beginning of an epidemic the dynamics is not constrained by the pool of susceptible which is, particularly in the case of Haiti, very large. Therefore these two parameters have been safely

assumed from literature values in the short-term calibration run. In Fig. 2 we depict the temporal sequence of calibration runs and the subsequent validation/prediction windows for each of them.

## 3. Rainfall generation patterns

In the hydrological literature, stochastic rainfall generation is often modeled as a marked Poisson process, where rainfall events are treated as a series of point events in continuous time where the associated mark represents the rainfall depth of the event (see e.g. Rodriguez-Iturbe and Porporato, 2004; Laio et al., 2001). This implies that no temporal evolution of a single event is taken into account, such that the amount of rainfall falling at a given time scale – which is usually assumed as daily, as in this paper – is modeled by a point process. The arrival of rainfall events is modeled as a Poisson process with rate $\lambda$. Therefore the inter-arrival time $\tau$ between rainfall events is exponentially distributed with mean $\sim 1/\lambda$, i.e.:

$$p(\tau) = \lambda e^{-\lambda\tau} \text{ for } \tau \geq 0. \tag{4}$$

The depth of each rainfall event is then sampled, again, from an exponential distribution, described by the following probability density function, in which $\alpha$ is the mean rainfall depth of all events:

$$p(h) = \frac{1}{\alpha} e^{-h/\alpha} \text{ for } h \geq 0. \tag{5}$$

This model has been shown to perform well in describing daily rainfall statistics (see e.g. Benjamin and Cornell, 1970) but it is better suited to applications to confined climatic regions, showing similar precipitation regimes. In Fig. 3 one can appreciate the peculiarity of the Haitian rainfall patterns in both space (panel A; mean daily rainfall for the period 1998-2012) and time (panel B: average yearly pattern for whole Haiti). Daily satellite rainfall estimates have been obtained from data collected by the NASA-JAXA's Tropical Rainfall Measuring Mission (TRMM_3B42 precipitation estimates, see `http://trmm.gsfc.nasa.gov/` for details) through the IRI/LDEO data portal set up by Columbia University (`http://iridl.ldeo.columbia.edu/`). Rainfall data are spatially distributed with the resolution of 0.25 degrees of latitude and longitude and are then downscaled at the node level with nearest neighbor interpolation.

In order to derive climatically homogeneous regions (and seasons) from data, we calculate the mean depth of observed rainfall occurrences (which expresses $\alpha$) and their mean interarrival time ($1/\lambda$, in days) at the node scale, for each month. We then perform a cluster analysis on those parameters, taking also into account the coordinates of each node to preserve spatial continuity of each cluster. Using the $k$-means clustering technique (Xu and Wunsch, 2005), we identify 20 clusters – over 12 months times 301 spatial nodes – of statistically similar monthly rainfall regimes that are contiguous in space. $k$-means clustering uses an iterative procedure which is initialized by fixing $k$ means in the space of data points. The algorithm then progressively changes their position until the distance among such centroids and each point of the cluster is minimized for all clusters (so that the contiguity of the points of each cluster is highest). As the procedure depends on the initial position of the centroids, it has been repeated $10,000$ times choosing the partition which gives the minimum distance among points belonging to the same cluster. In the case at hand, each data point of the 4-dimensional clustering space is characterized by the normalized values of: a) the monthly depth and interarrival time of rainfall events, averaged over the whole period of observation (1998-2012), in each node; b) the spatial coordinates of the node. In order to limit the number of clusters in the spatial subset of the data space, we weigh coordinate values less than rainfall statistics (in this case, the weight is equal to 0.5). Fig. 4 shows the performance of the clusterization with respect to rainfall statistics (panel A) and to the spatial distribution of clusters (panel B).

To generate the widest range of plausible scenarios of rainfall events in the Haiti territory, we also include inter-annual variability as described in Porporato et al. (2006), who observed that different yearly patterns may not be merely described by different realizations of the same stochastic process, but by explicit changes in the statistical properties of the process – the parameters $\alpha$ and $\lambda$. This procedure is carried out by assuming that the parameters of the exponential distributions of rainfall depths and inter-arrival times change from year to year and are gamma-distributed random variables (Porporato et al., 2006):

$$g_x(x) = \frac{(b_x)^{a_x}}{\Gamma(a_x)} x^{(a_x-1)} e^{-b_x x} \tag{6}$$

where $x$ is alternatively $\alpha$ or $\lambda$. The parameters $a_x$ and $b_x$ of the distributions of each cluster have been evaluated by fitting the empirical distribution

extracted from the rainfall dataset using the moments method. Finally, to generate daily rainfall events we use the following procedure:

- for each year of generation and each cluster we extract $\alpha$'s and $\lambda$'s from their respective Gamma distribution;
- for each cluster, we generate rainfall events using the corresponding $\alpha$ and $\lambda$, assigning the generated daily depth to each node belonging to the respective cluster.

The rainfall depth $h$ is then used to derive the rainfall intensity time series $J_i(t)$ which forces the epidemiological model.

## 4. Results and Discussion

### *4.1. Rainfall scenarios*

We first analyse the performance of the stochastic model for rainfall generation. Fig. 5 shows the the cumulative probability distributions and the probability density functions of the whole ensemble of inter-arrival times and of rainfall depths for the observed rainfall and for the multivariate Poisson generator, for both the global and the local scale. As in Porporato et al. (2006), the so-called "super-statistics" generator shows a good agreement with data. Moreover, we run an exercise to illustrate the improvements this model produces with respect to the 0-degree model used in Rinaldo et al. (2012), where a simple reshuffling of the 13 observed years of rainfall events was used to perform model projections. Such method is indeed practical and simple, but it falls short of generating events of intensity outside the observed realizations. Here we simulate a large number of yearly rainfall scenarios (10,000) with both models, taking as reference data-set the period 1998-2011, and we compare the results to the observed rainfall events of 2012. In Fig. 6 we show that, in many cases, observed rainfall events fall outside the $5-95^{th}$ percentile boundaries of the 0-degree model, differently from our generator. It should be noted that this is true for both models for a few major events, probably corresponding to tropical storms and hurricanes (e.g. the event at day$\simeq$ 300 is hurricane Sandy).

### *4.2. Short-term projections*

From an epidemiological perspective, we intend to test whether our models can be used as an effective tool for the prediction and the control of the

course of an ongoing epidemic – in this case, the Haitian one. To this aim, we focus on the short-term evolution of the disease, in order to simulate conditions of epidemiological emergency. For this analysis, we use as calibration time horizon the interval between the beginning of the epidemic (October 2010) and the end of May 2011, just before the June 2011 resurgence driven by the Haitian rainy season. We emphasize the fact that a limited dataset is specifically used to calibrate our models to represent a worst-case scenario, i.e. a situation in which decision makers would face limited information and likely miss key drivers. Notice in particular that the calibration window ends right before the spring rainy season and therefore it contains few indications on the effect of rainfall on the epidemic dynamics (i.e. on the magnitude of parameter $\phi$). This limited information refers to the autumn rainfalls of 2010 and is possibly clouded by the initial boost of the epidemic. All these elements make this prediction exercise particularly challenging.

We use rainfall scenarios generated by the multi-variate Poisson model presented in the previous section to force model 2 – which shows a slightly better AIC value – and simulate the course of the epidemic in the following months, until the end of 2011. In all panels of Fig. 7 we show the trajectory of the models when forced by the rainfall pattern that was actually observed in Haiti and the range of possible epidemiological scenarios emerging from the generated rainfall patterns.

The different timing of the epidemic peak – observed in June 2011 from the data and at the end of October in the model simulations – can be explained by the peculiar rainfall pattern that was observed in Haiti in 2011, with a highly concentrated event of very rare intensity at the beginning of June followed by an abnormally wet summer. We deem that the difficulty in generating such particular rainfall pattern lies in the limitedness of the dataset (15 years) available to fit the rainfall stochastic model. Simulations performed with the actually observed rainfall pattern display, in fact, a far better synchronization with reported cases (Fig.7).

An important result of our prediction effort concerns the ability of the model to grasp reasonably well the order of magnitude of the new outbreak. The cumulated reported cases between May-December 2011 amount to approximately $188,000$, while the model forced with generated rainfall events predicts in the median $230,000$ cases. One can appreciate this more clearly when looking at the course of cumulated cases, which we show in Fig. 8. This result is reassuring if one aims at using mathematical models as prediction and control tools. Knowing the order of magnitude of the upcoming

spreading of the disease in space and time, public health organization can set the scale of the reaction needed to act in time and possibly mitigate the resurgence of the epidemic.

### *4.3. Evaluation of intervention strategies*

Mathematical models like the ones presented here offer a invaluable tool to evaluate *a priori* the effectiveness of policy scenarios on the course of the epidemic. We present (Fig. 9) an assessment of different intervention scenarios in the same time span (01/06/2011-31/12/2011), using the best fit parameters of the best performing model (model 2, Table 3). This time interval, for which we know that the model is able to reproduce the expected number of cases, allows us to make quantitative considerations on the effects of possible control actions. Panels A and C of Fig. 9 show the effect of a reduction in the contact rate $\beta$, which is supposed to decrease linearly, and uniformly in the territory of Haiti, in the span of one month since the start of the campaign. Given our prediction of resurgence peaking in late October and starting in September, we set the beginning of the campaign either at 01/06/2011 or a month later, at 01/07/2011, depending on the rapidity of decision and intervention deployment. In panels B and D the effect of a spatially uniform vaccination of a fraction of the Haitian population is displayed, under the same assumptions at the start of the campaign (i.e. June or July 2011) and in the same time span of intervention (1 month). A constant vaccination effort is assumed. As the model is not able to predict accurately the temporal evolution of the cases but rather the total cumulative cases at the end of the year, one may wonder how reliable the predicted effects of the interventions are. To assess their accuracy, in Fig. 9 we compare the predicted effects of the interventions (red lines) with those obtained applying the same interventions to the model when calibrated using all the available epidemiological and rainfall data until the end of 2011 (blue lines). This benchmark calibration run provides the most reliable estimates of the effect of such interventions. Results show that the difference between prediction and benchmark in the estimates of reduction of cases remains below 10% for any type and magnitude of the interventions, provided that intervention campaigns start promptly (01/06/2011). It should be noted that, if these are delayed to one month later, the uncertainty increases, so that such estimates differ of an average value of around 15%.

The application of the two types of interventions (which could also be applied simultaneously, of course) in the field presents indeed different kinds

of difficulties. Therefore, the effort needed to implement a given policy can be different and not easily estimated. Under such premises, we can still compare the effectiveness of both sanitation and vaccination at reducing the impact of cholera resurgence on the Haiti population. In panels C and D of Fig. 9 we show the number of new cholera cases between 28/05 and 31/12/2011 normalized with respect to the "no intervention" scenario, as a function of the reduction of the contact rate (C) and of the number of vaccines deployed (D). We choose here a maximum effort of 8 million vaccinations in the latter case (corresponding to the entire Haitian population), and of 50% sanitation rate in the former. Note that such a rate implies that the probability of coming into contact with contaminated water is reduced by a half.

Timely intervention represents an essential feature of any public health policy, be it focused on sanitation or on vaccination. The average number of "avoided infections" (the difference between the sum of cases simulated from 28/05/2011 to 31/12/2011 in absence of interventions and the average number of cases when measures are taken) in fact amounts to around $230,000$ (vaccination) and $185,000$ (sanitation) when policies start being applied in June 2011, and to $187,000$ (vaccination) and $155,000$ (sanitation) for actions starting in July 2011. This means a rough average of 85% of total cases avoided should interventions have started in June, while figures drop to 70% had campaigns begun in July.

### *4.4. Multi-season projections*

Another use of the tools implemented deals with multi-seasonal projections of epidemiological predictions, which allows to estimate possible future resurgences of the disease should it become endemic (Tappero and Tauxe, 2011). Fig. 10 shows a projection of future outbreaks spanning until 31/12/2015. We perform several $(1,000)$ simulation runs, generating rainfall events and sampling from the posterior distribution of our fitting parameter set (i.e. from the last $1,000$ parameter combinations explored by the fitting algorithm), which now includes also the rate of immunity loss $\rho$ and the asymptomatic ratio $\sigma$. Our predictions show a first resurgence of the disease in 2012 and then a settling on an endemic, seasonal pattern characterized by roughly constant annual attack rate (Fig. 10). The comparison with the actual observed cases shows that, although within the 5-95th percentile range, the model generally overestimates the intensity of the new prevalence peaks. In this respect, an increased awareness regarding common sanitizing practices and the risk connected to the ingestion of contaminated water

could have led to lower contact/contamination rates and, thus, to a lower intensity of cholera resurgence. Another factor that can concur to this slight overestimation is the uncertainty related to the calibration of the duration of the immunity of Recovered individuals $1/\rho$ and the ratio of asymptomatic infections $\sigma$. In fact, the long-term calibration run converges to a duration of the immunity of $1-2$ years ($5-95th$ percentile range of the parameter posterior distribution), which causes a fast replenishment of the susceptible pool. The asymptomatic ratio $\sigma$ ranges between 0.2 and 0.25 in the $5-95th$ percentile range of the parameter posterior distribution. One can observe – see panel B – how the intensity of new outbreaks is set by the initial size of the susceptible pool and that a threshold size is needed for the outbreak to start (which reminds the concept of generalized reproduction numbers for spatially explicit models; see Gatto et al., 2012, 2013). Overall, in an epidemic management context, the results of this long-term prediction are deemed particularly valuable.

The possibility of cholera becoming endemic in Haiti and the possible strategies to eradicate it have been the subject of an intense debate (see e.g. Mukandavire et al., 2013). If one considers vaccination more easily deployed in the field in times of emergency especially when living conditions are precarious and the sanitation system is nearly absent, effects of different intervention policies on the long-term evolution of the disease must be accounted for. Fig. 11 illustrates the effects of different intervention scenarios on the predicted course of the epidemic in the period 14/01/2012-31/12/2015 (policy implementation starts on 01/08/2012 and lasts 1 year in the case of sanitation and 90 days in the case of vaccination to reflect the different effort required to implement such measures). In panel A the effect of vaccinating 1 to 4 million susceptibles is displayed. One can already appreciate the effect of a mass vaccinations campaign when carried out only once: increasing the number of vaccinations has the sole effect of delaying – and possibly exacerbating – the resurgence of the disease, even in the extreme event of vaccinating the entire population of Haiti (see inset of panel A, portraying the predicted evolution of new weekly cases until 31/12/2017). Note, however, that ours is a worst case scenario as no concurrent improvement of sanitation conditions is accounted for. In the other case (panel B), we show that a sanitation campaign, besides other positive effects brought by an increased drinking water quality, can effectively eradicate the disease from the country, provided that the final effort exceeds a certain threshold. In our analysis, the threshold lies somewhere between a $20-30\%$ reduction of the contact rate in

all nodes of the Haitian network (see the inset of panel B, showing the effect of 30% sanitation on the course of the epidemic until 31/12/2017). Focusing only on vaccination as intervention policy, the eradication of the disease could have been obtained only if periodic vaccination campaigns were set up. These periodical campaigns, however, may well imply an effort comparable to that needed to minimize exposure probabilities, say, by sanitizing the water supply system on a permanent basis. No complete eradication of the disease would be reached, however, without improving concurrently the sanitation conditions of water supply in the country. We also point to the limited stock of vaccines that is currently available (which counts less than $400,000$ doses currently; Waldor et al., 2010), which would not allow, of course, vast vaccinations campaigns such as the ones we have simulated. We remind that mass vaccinations, however, remain one of the potentially fastest interventions to be deployed in an emergency.

## 5. Conclusions

The following conclusions are worth mentioning:

- spatially explicit mathematical models provide a tool to predict and control the course of ongoing cholera epidemics. The relevance of this new class of models relates to the fact that inappropriate responses can be avoided by providing adequate and timely information to policy-makers, decision-makers, the media and the public. While several issues remain open, like the field validation of parameters defined node-by-node, major public health policy challenges like those involving limitations of human mobility, structural measures (construction of hospitals, placement of field hospitals, construction of water sanitation infrastructures) and other interventions (vaccination and/or sanitation campaigns, antibiotics administration) can be thoroughly addressed by the proposed class of models;

- rainfall patterns can be used to drive epidemiological models with realistic rainfall scenarios. We have used a Poisson generator integrated with space-time interannual variability. Poisson generators constitute a simple, synthetic way to generate rainfall patterns preserving the basic statistics of observed precipitation events (depth and interarrival times). However, they perform poorly when it comes to reproducing

widespread events, in both space and time, such as monsoons and tropical storms in general. Also, our results point out how the evaluation of epidemiological predictions must rely on more comprehensive indicators than the simple expected value, due to the unpredictability of rainfall events, especially the extreme ones. An altered intensity and frequency of these particular rainfall events may influence dramatically the conditions of resurgence of the disease;

- multi-season projection of the disease patterns can be used for the assessment of the effectiveness of control strategies. We suggest that vaccination alone, still considered in many studies as the key form of outbreak control, may be effective in the short term but would avoid resurgence of the disease only if sanitation conditions were to improve concurrently. Relative merits of the various interventions can therefore be weighed on a quantitative basis. In this respect, the increasing number of available spatially explicit mathematical models, after the first proposed by Bertuzzo et al. (2008) and particularly for the case of Haiti (Tuite et al., 2011; Chao et al., 2011), suggests that their impact on public health practice is gaining momentum;

- the application of spatially explicit models proves a powerful monitoring tool. The epidemiological framework can in fact be coupled with projections of rainfall scenarios. Short- and long-term assessments of the possible evolution of the epidemic can thus be produced and discussed. In particular, this exercise has proven useful to show the shortcomings and the existing pitfalls of our approach, but also the importance of being able to draw in advance possible scenarios of disease resurgence with reasonable accuracy, especially in capturing the total number of cases in a limited timespan.

## 6. Acknowledgements

This article is dedicated to the caring memory of Erika Schild, formerly MS student at EPFL, who passed away during a stage in Haiti dedicated to her thesis on modelling cholera epidemics. The authors wish to express their deepest regret. LR, EB, LM and AR acknowledge the support provided by ERC Advanced Grant program through the project RINEC-227612 and by the SFN/FNS projects 200021 124930/1 and CR2312 138104/1. MG and AR acknowledge the support from the SFN/FNS project IZK0Z2 139537/1

for international cooperation. IRI acknowledges the support of the James S. McDonnell Foundation through a grant for Studying Complex Systems (220020138).

## 7. Appendix

The optimization approach for parameter estimation is based on Markov Chain Monte Carlo (MCMC) sampling, that is a family of methods allowing for the exploration of the posterior probability density function of a desired probability distribution (in our case, the joint probability distribution of the set of tuning parameters (Gilks et al., 1995). Specifically, we use the $\text{DREAM}_{\text{ZS}}$ variant of the DREAM (Differential Evolution Adaptive Metropolis) algorithm (ter Braak and Vrugt, 2008), which makes use of sampling from past states visited by the Markov chains and of a snooker update step (in addition to parallel update steps) to generate candidate points in each individual chain – thus reducing the number of parallel chains needed for an effective exploration of the posterior distribution while at the same time increasing the diversity of candidate points (Vrugt et al., 2009). 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). From an operational perspective, we first calibrate independently each candidate model against the epidemiological data available for the Haiti cholera epidemic.

To assess whether and how more complex models are better suited to describe the evolution of the Haiti cholera outbreak we rank the performances of the two candidate models according to Akaike's Information Criterion (AIC Akaike, 1974). AIC is a model-selection procedure that explicitly takes into account the trade-off between model accuracy and complexity, measured as the number $\Theta$ of free parameters (i.e. the structural parameters for each candidate model, plus one residual variance parameter; see Burnham and Anderson, 2002; Corani and Gatto, 2007). For each best-fit model we then compute

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

where $\eta$ is the number of data points ($\eta = n_d n_w$, $n_d = 10$ and $n_w = 31$ – in the short term calibration run – and $n_w = 65$ – in the long term run – being the numbers of administrative departments and weeks from the onset of the epidemic, respectively), and $RSS$ is the value of the residual sum of squares

computed on the basis of model results and the epidemiological record, i.e.

$$RSS = \sum_{i=1}^{n_d} \sum_{j=1}^{n_w} \left[ \mathcal{N}(i,j) - \hat{\mathcal{N}}(i,j) \right]^2 ,$$

where $\mathcal{N}(i,j)$ and $\hat{\mathcal{N}}(i,j)$ are the new reported cases at the department scale evaluated from the weekly epidemiological bulletins and best-fit model simulations, respectively. Note that model results are given at the watershed level. Therefore, they have to be up-scaled to the Department level for comparison with the available epidemic data. Should of course less aggregated data be available, the current procedure would require no coarse graining of computed results. The up-scaling procedure is performed here by accounting for the fraction of population of each watershed that belongs to a given Department. RSS values and AIC scores for both models are reported in Table 3 for both calibration periods. Results show that the performance of models 1 and 2 cannot be ranked with high significance in the first case (Akaike, 1974; Burnham and Anderson, 2002; Corani and Gatto, 2007).

## 8. References


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

Akanda, A. S., Jutla, S., Islam, S., 2009. Dual peak cholera transmission in Bengal Delta: a hydroclimatological explanation. Geophysical Research Letters 36, L19401.

Al-Tawfiq, J. A., Memish, Z. A., 2012. The Hajj: updated health hazards and current recommendations for 2012. Eurosurveillance 17 (41), 6–10.

Alam, A., al., 2005. Hyperinfectivity of human-passaged V. *cholerae* can be modeled by growth in the infant mouse. Infection and Immunity 73, 6674–6679.

Altizer, S., Dobson, A., Hosseini, P., Hudson, P., Pascual, M., Rohani, P., 2006. Seasonality and the dynamics of infectious diseases. Ecology Letters 9 (4), 467–484.

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

Benjamin, J., Cornell, C., 1970. Probability, Statistics and Decision for Civil Engineers. New York, McGraw-Hill.

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

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

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

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

Capasso, V., Paveri-Fontana, S., 1979. A mathematical model for the 1973 cholera epidemic in the European Mediterranean region. Révue Epidemiologique et Santé Publique 27, 121–132.

Chao, D. L., Halloran, M. E., Longini, Jr., I. M., 2011. Vaccination strategies for epidemic cholera in Haiti with implications for the developing world. Proceedings of the National Academy of Sciences of the United States of America 108 (17), 7081–7085.

Cho, H., Fritsch, M., Gupta, V., Rodrguez-Iturbe, I., Taqqu, M., 1987. Rainfall Fields: Estimation Analysis and Prediction. Special Issue of the Journal of Geophysical Research.

CIA, 2009. CIA's World Factbook, available online at https://www.cia.gov/library/publications/the-world-factbook/index.html.

Clemens, J. E., 1990. Field trial of oral cholera vaccines in Bangladesh: results from three-year follow-up. Lancet 335, 270–273.

740 Codeço, C., 2001. Endemic and epidemic dynamics of cholera: the role of the
741 aquatic reservoir. BMC Infectious Diseases 1 (1).

742 Corani, G., Gatto, M., 2007. Structural risk minimization: a robust method
743 for density-dependence detection and model selection. Ecography 30, 400–
744 416.

745 de Magny, G. C., Murtugudde, R., Sapiano, M. R. P., Nizam, A., Brown,
746 C. W., Busalacchi, A. J., Yunus, M., Nair, G. B., Gil, A. I., Lanata, C. F.,
747 Calkins, J., Manna, B., Rajendran, K., Bhattacharya, M. K., Huq, A.,
748 Sack, R. B., Colwell, R. R., 2008. Environmental signatures associated
749 with cholera epidemics. Proceedings of the National Academy of Science
750 of the United States of America 105 (46), 17676–17681.

751 Emch, M., Feldacker, C., Islam, M. S., Ali, M., 2008. Seasonality of cholera
752 from 1974 to 2005: A review of global patterns. International Journal of
753 Health Geographics 7, 31.

754 Enserink, M., 2010. Infectious diseases: Haiti's outbreak is latest in cholera's
755 new global assault. Science 330, 738–739.

756 Erlander, S., Stewart, N. F., 1990. The Gravity Model in Transportation
757 Analysis – Theory and Extensions. VSP Books, Zeist, The Netherlands.

758 Gatto, M., Mari, L., Casagrandi, R., Bertuzzo, E., Righetto, L., Rodriguez-
759 Iturbe, I., Rinaldo, A., 2012. Generalized reproduction numbers and the
760 prediction of patterns in waterborne disease. Proceedings of the National
761 Academy of Sciences of The United States of America 109 (48), 19703–
762 19708.

763 Gatto, M., Mari, L., Casagrandi, R., Bertuzzo, E., Righetto, L., Rodriguez-
764 Iturbe, I., Rinaldo, A., 2013. Spatially explicit conditions for waterborne
765 pathogen invasion. The American Naturalist, in press.

766 Gilks, W. R., Richardson, S., Spiegelharter, D. J., 1995. Markov Chain Monte
767 Carlo in Practice. Chapman and Hall, New York, USA.

768 Girard, M., Steele, D., Chaignat, M.L. Kieny, M., 2006. A review of vaccine
769 research and development: human enteric infections. Vaccine 24, 2732–
770 2750.

Hartley, D., Morris, J., Smith, D. L., 2006. Hyperinfectivity: A critical element in the ability of *V. cholerae* to cause epidemics? PLoS Medicine 3, 63–69.

Hashizume, M., Faruque, A. S. G., Wagatsuma, Y., Hayashi, T., Armstrong, B., 2010. Cholera in Bangladesh: climatic components of seasonal variation. Epidemiology 21 (5), 706–710.

Jertborn, M., Svennerborn, A., Holmgren, J., 1993. Evaluation of different immunization schedules for oral cholera B subunit-whole cell vaccine in Swedish volunteers. Vaccine 11, 1007–1012.

Kelvin, A. A., 2011. Cholera outbreak in the Republic of Congo, the Democratic Republic of Congo, and cholera worldwide. Journal of Infection in Developing Countries 5 (10), 688–691.

Koelle, K., 2009. The impact of climate on the disease dynamics of cholera. Clinical Microbiology And Infection 15, 29–31.

Koelle, K., Rodo, X., Pascual, M., Yunus, M., Mustafa, G., 2005. Refractory periods and climate forcing in cholera dynamics. Nature 436, 696–700.

Laio, F., Porporato, A., Ridolfi, R., Rodriguez-Iturbe, I., 2001. Plants in water-controlled ecosystems: Active role in hydrologic processes and response to water stress. Probabilistic soil moisture dynamics. Advances in Water Resources 24, 707–723.

Lipp, E., Huq, A., Colwell, R., 2002. Effects of global climate on infectious disease: The cholera model. Clinical Microbiology Reviews 15 (4), 757-770.

Luque Fernandez, M. A., Bauernfeind, A., Diaz Jimenez, J., Linares Gil, C., El Omeiri, N., Herrera Guibert, D., 2009. Influence of temperature and rainfall on the evolution of cholera epidemics in Lusaka, Zambia, 2003-2006: analysis of a time series. Transactions of the Royal Society of Tropical Medicine and Hygiene 103 (2), 137–143.

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

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

Merrell, D.S., Butler, S.M., Qadri, F., Dolganov, N.A., Alam, A., Cohen, M.B., Calderwood, S.B., Schoolnik, G.K., Camilli, A. 2002. Host-induced epidemic spread of the cholera bacterium. Nature 417, 642–645.

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

Nkoko, D. B., Giraudoux, P., Plisnier, P.-D., Tinda, A. M., Piarroux, M., Sudre, B., Horion, S., Tamfum, J.-J. M., Ilunga, B. K., Piarroux, R., 2011. Dynamics of cholera outbreaks in Great Lakes Region of Africa, 1978-2008. Emerging Infectious Diseases 17 (11), 2026–2034.

PAHO, 2010. Eoc situation report #17. Tech. rep., Pan-American Health Organization, Regional Office of the World Health Organization, online at http://new.paho.org/hq/images/Atlas_IHR/CholeraHispaniola/atlas.html.

Pascual, M., Rodo, X., Ellner, S., Colwell, R., Bouma, M., 2000. Cholera dynamics and El Niño-Southern oscillation. Science 289, 1766–1769.

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

Porporato, A., Vico, G., Fay, P., 2006. Superstatistics of hydro-climatic fluctuations and interannual ecosystem productivity. Geophysical Research Letters 33, L15402.

Righetto, L., Bertuzzo, E., Casagrandi, R., Gatto, M., Rodriguez-Iturbe, I., Rinaldo, A., 2011. Modelling human movement in cholera spreading along fluvial systems. Ecohydrology 4 (1), 49–55.

Rinaldo, A., Bertuzzo, E., Mari, L., Righetto, L., Blokesch, M., Gatto, M., Casagrandi, R., Murray, M., Vesenbeckh, S., Rodriguez-Iturbe, I., 2012. Early predictions of the 2010-2011 Haiti cholera outbreak: Assessments

and perspectives. Proceedings of the National Academy of Science of the United States of America 109 (17), 6602–6607.

Rodriguez-Iturbe, I., Cox, D., Eagleson, P., 1986. Spatial modeling of total storm rainfall. Proceedings of the Royal Society of London Series A-Mathematical and Physical Sciences 403 (1824), 27–50.

Rodriguez-Iturbe, I., Porporato, A., 2004. Ecohydrology of Water-Controlled Ecosystems: Soil Moisture and Plant Dynamics. Cambridge University Press, Cambridge, UK.

Ruiz-Moreno, D., Pascual, M., Bouma, M., Dobson, A., Cash, B., 2007. Cholera seasonality in Madras (1901-1940): dual role for rainfall in endemic and epidemic regions. Ecohealth 4 (1), 52–62.

Sack, D., 2011. How many cholera deaths can be averted in Haiti? Lancet 377, 1214–1216.

Tappero, Tauxe, 2011. Lessons learned during public health response to cholera epidemic in haiti and the dominican republic. Tech. rep., United States Center for Disease Control.

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

Tuite, A., Tien, J., Eisenberg, M., Earn, D., Ma, J., Fisman, D., 2011. 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.

Vrugt, J. A., ter Braak, C. J. F., Diks, C. G. H., Robinson, B. A., Hyman, J. M., Higdon, D., 2009. 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.

Waldor, M. K., Hotez, P. J., Clemens, J. D., 2010. A national cholera vaccine stockpile - A new humanitarian and diplomatic resource. New England Journal of Medicine 363 (24), 2279–2282.

Table 1: Estimated parameter values, relevant units, and cited literature for both calibration runs (short-term and long-term) and both models (including hyper-infectivity or not). Note that $\rho$ and $\sigma$ are fitted in the longer calibration run.

| **Parameter** | **Units** | **Value** | **References** |
|---|---|---|---|
| $\mu$ | $\text{day}^{-1}$ | $1/(61 \cdot 365)$ | CIA (2009) |
| $\beta$ | $\text{day}^{-1}$ | 1 | Codeço (2001) |
| $\rho$ | $\text{day}^{-1}$ | $1/(3 \cdot 365)$ | Koelle et al. (2005) |
| $\gamma$ | $\text{day}^{-1}$ | 0.2 | Codeço (2001) |
| $\delta$ | $\text{day}^{-1}$ | $4 \cdot 10^{-3}$ | PAHO (2010) |
| $\mu_B$ | $\text{day}^{-1}$ | 0.2 | Codeço (2001) |
| $\sigma$ | - | 0.2 | WHO (2011) |


Walton, D., Ivers, L., 2011. Responding to cholera in post-earthquake Haiti. The New England Journal of Medicine 364, 3–5.

WHO, 2011. Annual report on cholera, available online at http://www.who.int/cholera/statistics/en/. Tech. rep., World Health Organization.

Xu, R., Wunsch, D., 2005. Survey of clustering algorithms. IEEE Transactions on Neural Networks 16 (3), 645–678.

Table 2: Fitted parameter values and relevant units for both calibration runs and both models. Note that $\rho$ and $\sigma$ are not fitted but taken from literature in the shorter calibration run.

| | | Short-term | | Long-term |
|---|---|---|---|---|
| **Parameter** | **Units** | *Model 1* | *Model 2* | *Model 2* |
| $\theta$ | $\text{day}^{-1}$ | 0.11 | 0.37 | 0.51 |
| $l$ | $\text{day}^{-1}$ | 0.70 | 1.37 | 2.54 |
| $m$ | - | 0.05 | 0.06 | 0.05 |
| $D$ | km | 95.47 | 50.82 | 150.33 |
| $\phi$ | $\text{day mm}^{-1}$ | $6.23 \cdot 10^{-2}$ | $4.77 \cdot 10^{-2}$ | $2.43 \cdot 10^{-2}$ |
| $\rho$ | $\text{day}^{-1}$ | - | - | $2.6 \cdot 10^{-3}$(1.04 years) |
| $\sigma$ | - | - | - | 0.21 |

Table 3: Number of calibrated parameters and AIC scores for the two tested models (see text for technical details on the candidate models) in the calibration runs 23/10/2010-28/05/2011 and 23/10/2010-14/01/2012. Columns indexed with $\Delta_{AIC}$ show Akaike differences (with respect to the best-ranked model), which must be larger than 4 for significance (Akaike, 1974; Burnham and Anderson, 2002; Corani and Gatto, 2007).

| | | Short-term | | | Long-term |
|---|---|---|---|---|---|
| **Model** | $\Theta$ | $RSS$ | $AIC$ | $\Delta_{AIC}$ | $RSS$ |
| model 1 | 5 | $1.905 \cdot 10^8$ | 4267 | 1 | – |
| model 2 | 5 | $1.902 \cdot 10^8$ | 4266 | – | $3.34 \cdot 10^8$ |

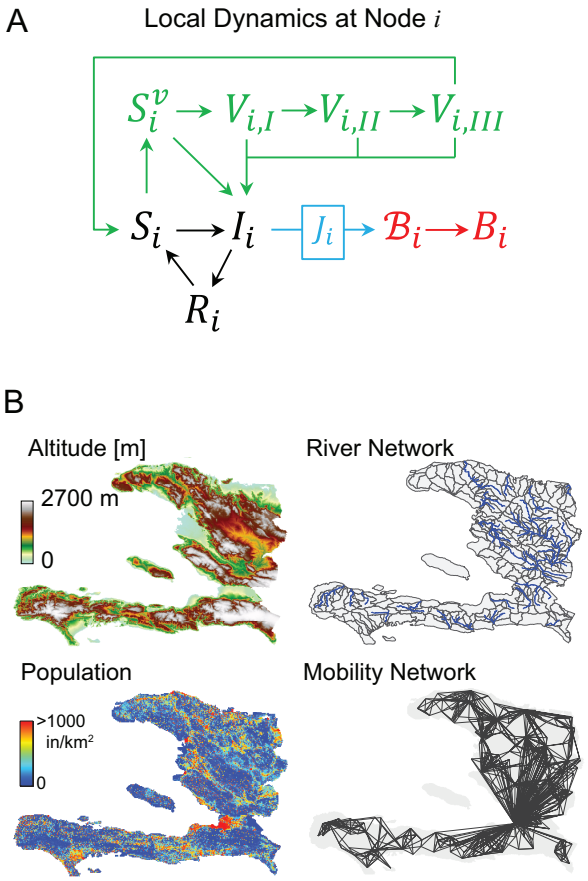


Figure 1: Panel A. Block diagram of the $i$-th site epidemiological model, including the hyperinfective state $\mathcal{B}_i$ and the implementation of vaccination (see subsection 2.3). Panel B: Spatial databases. In clock-wise order: color-coded Digital Terrain elevation Map (DTM) of Haiti; the subdivision of Haitian territory in hydrological units (sub-basins) extracted from the DTM; a relevant subset of the network of human mobility, here portrayed synthetically by the four largest outbound connections for each node; spatial distribution of population density obtained by LandScan project, which is translated into a georeferenced spatial distribution of nodes $i$ endowed with population $H_i$. Partially redrawn from Rinaldo et al. (2012).

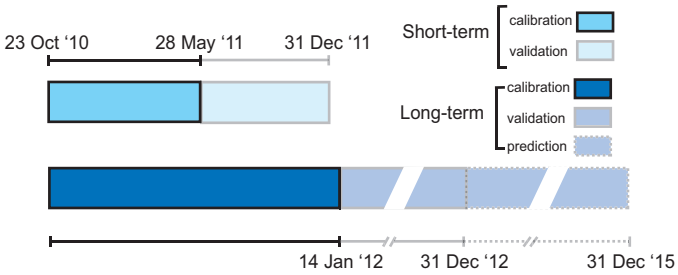


Figure 2: GANTT diagram of the calibration runs and validation/prediction windows.

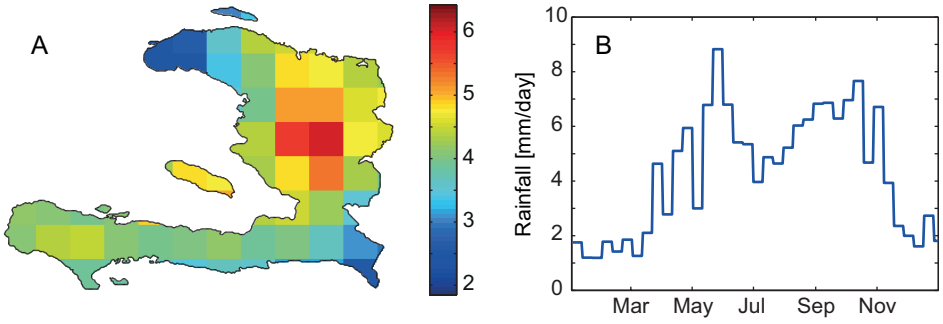


Figure 3: Panel A: Raster map of Haiti in which each pixel shows the average depth of rainfall events in the period 1998-2012 (source: `http://iridl.ldeo.columbia.edu/`; resolution: 0.25 degrees); Panel B: yearly time series of rainfall intensity averaged over the whole territory of Haiti and over the whole period of observation; resolution: 10 days.

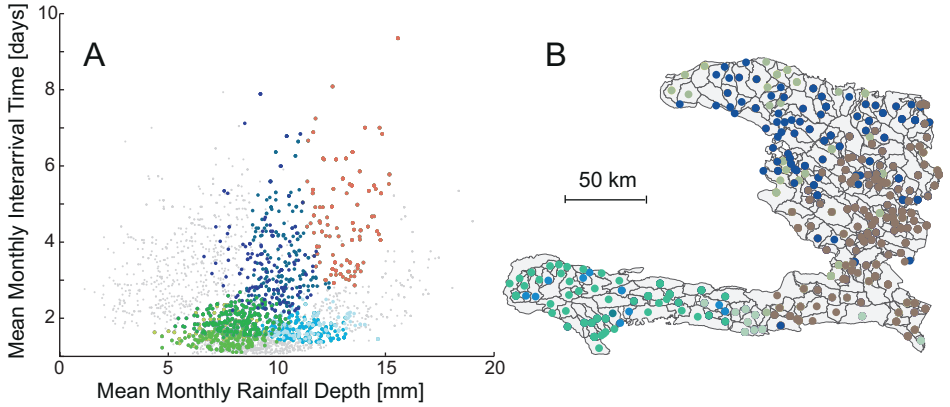


Figure 4: Panel A: Scatter plot in which each point represents the average, in the period 1998-2012, of mean monthly depth/interarrival time of rainfall events at the node level; colored dots belong to the same cluster (we show here the aggregates that emerge for the month of June). Panel B: Distribution of clusters in space, in the month of June.

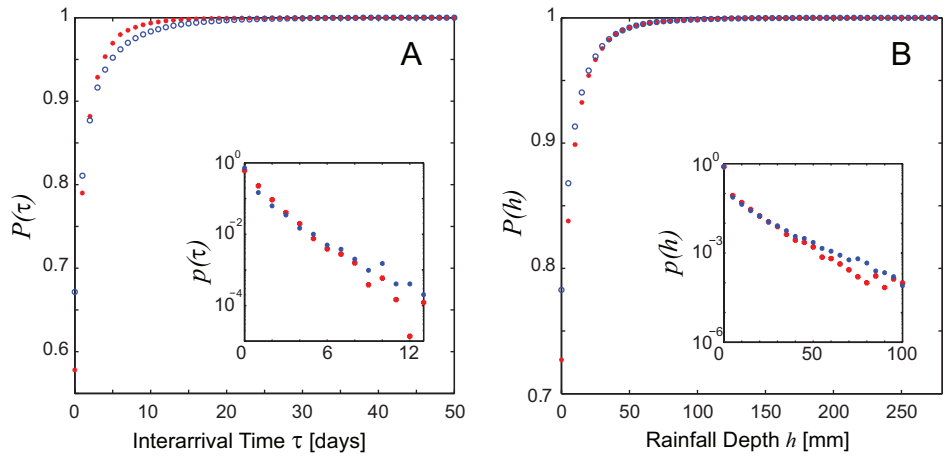


Figure 5: Distribution of the inter-arrival times (A) and precipitation depths (B) of rainfall events at the Haiti scale, derived from the multi-variate (red stars) Poisson process or from the data (blue circles). In the inset we show the distributions at the cluster scale (one the clusters containing the node representing Port-au-Prince is taken as example), as outcomes of generated rainfall (red dots) or data values (blue dots). While the cumulated probability distribution function is shown at the global scale, we choose to display the probability distribution function at the cluster level, for graphical clarity.

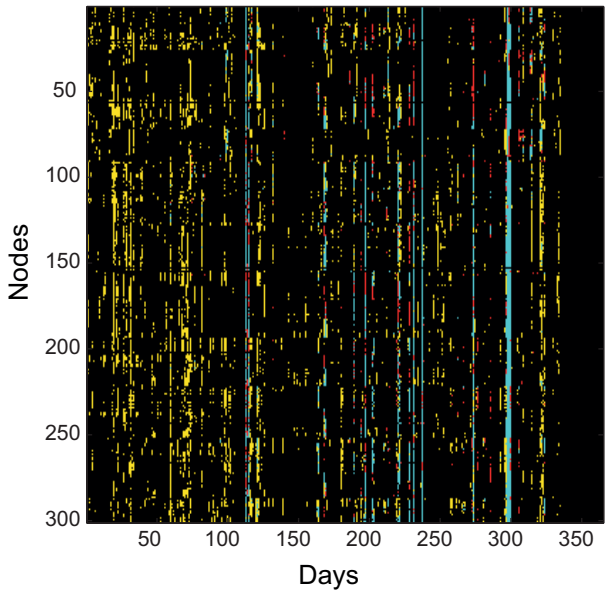


Figure 6: Grid showing a comparison between the rainfall events observed in 2012 and the rainfall events generated with the 0-degree model of Rinaldo et al. (2012) and the Poisson generator presented here, in the space of nodes and Julian days. Black pixels represent observed events that are contained in the $5 - 95^{th}$ percentile range of both distributions of generated rainfall events. Blue pixels show events that do not belong to such range in either distributions. Red/yellow pixels show events that fall outside the boundaries of this range of the distribution obtained with the Poisson generator/the 0-degree model.

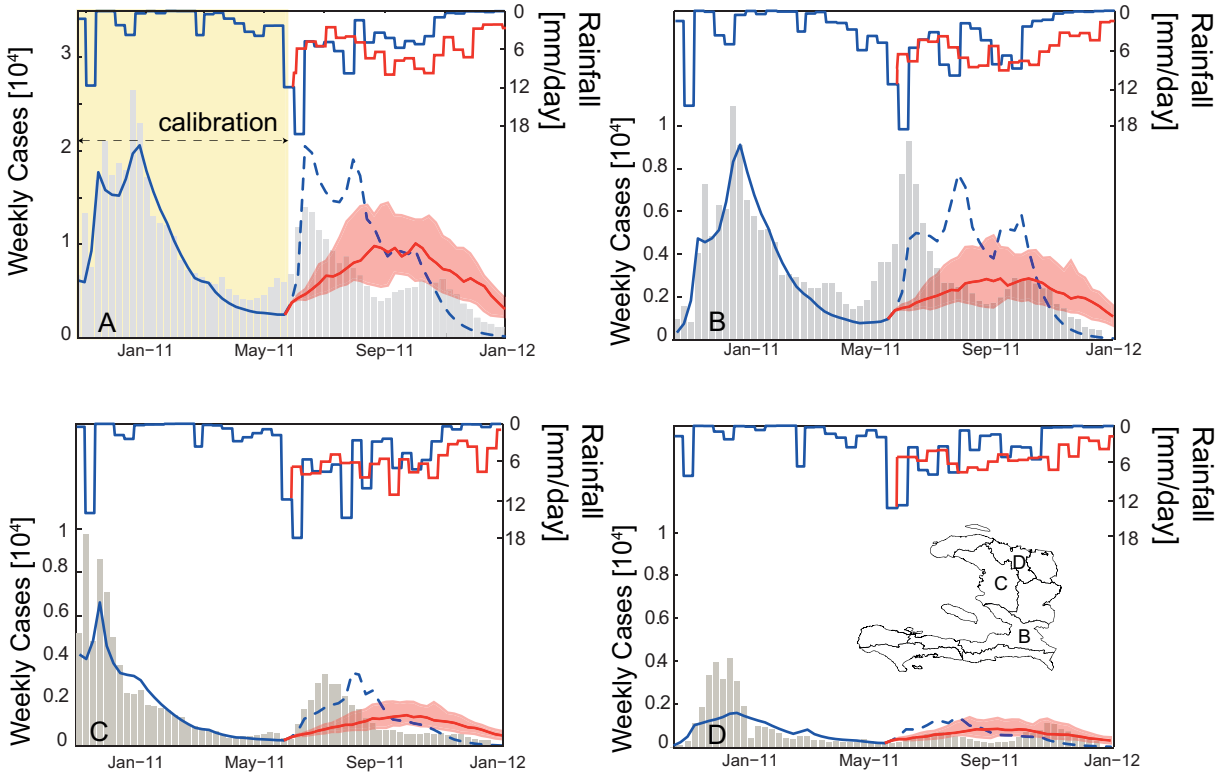


Figure 7: Rainfall and epidemiological patterns for the period 23/10/2010-31/12/2011; blue solid lines refer to observed rainfall patterns (decadal and averaged over the whole Haiti, upper part) and to the corresponding model outputs as new weekly cholera cases (blue dashed lines in the validation period); red solid lines show instead one realization in 1000 generated rainfall patterns between 28/05/2011 and 31/12/2011 and the median of the corresponding model outputs (the shaded range also shows the 25th-75th percentile span); grey bars depict the weekly reported cases over the simulation horizon. Panel A shows results at the country scale, while patterns for the most populated departments are shown in panels B-D (Ouest, Artibonite, Nord respectively; see inset in panel D). Parameters are taken from the best fit in the short term calibration run.

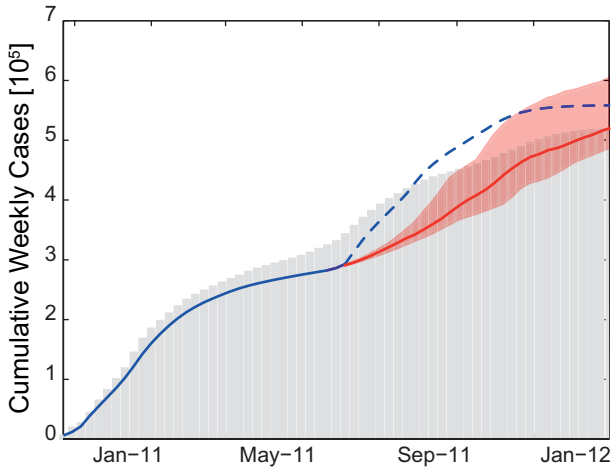


Figure 8: Cumulative weekly cases, reported (gray bars) and simulated by model 2. Blue solid/dashed line: simulation with the observed rainfall pattern; red solid line (median)/shaded range (25th-75th percentile): simulation with generated rainfall.

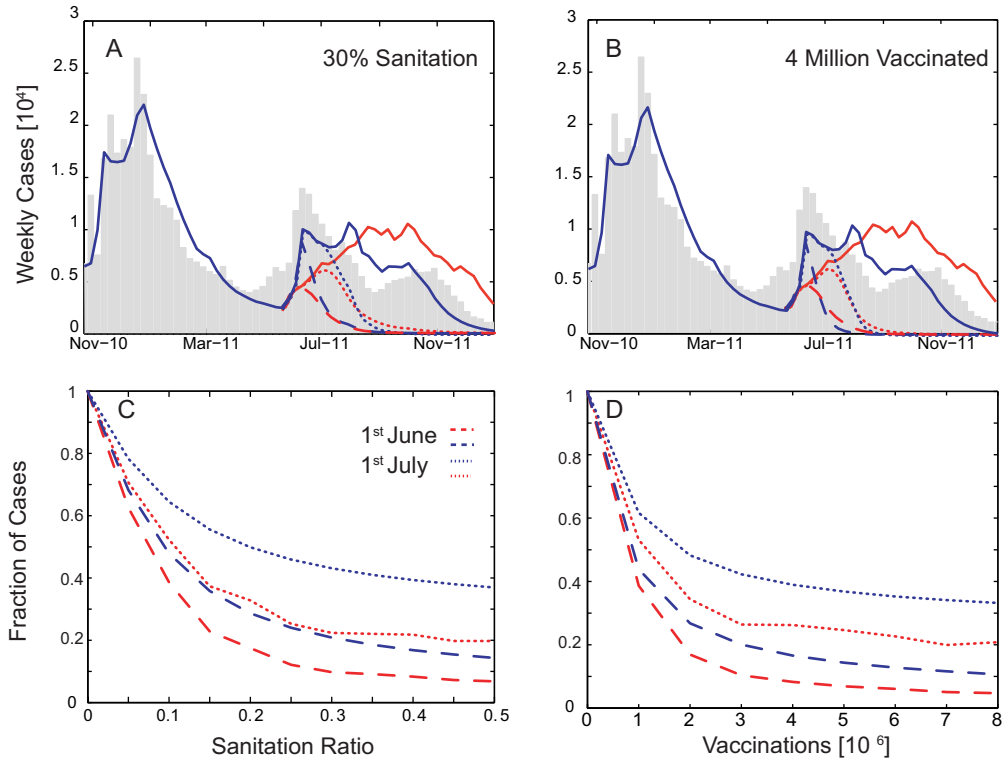


Figure 9: Effect of intervention policies on the predicted course of the epidemic between 28/05 and 31/12/2011 (blue lines: simulations with the model calibrated until 31/12/2011 and with observed rainfall as input; red lines: median of simulations with generated rainfall). A) Effect of a reduction of the 30% of the contact rate $\beta$, implemented in one month starting from the 1st of June (dashed blue/red lines) or the 1st of July (dotted blue/red lines) 2011. B) Effect of a vaccination of 4 million individuals, implemented in one month starting from the 1st of June (dashed blue/red lines) or the 1st of July (dotted blue/red lines) 2011. C-D) Number of new cases normalized with respect to the cases simulated in absence of interventions in the period 28/05-31/12/2011 as a function of the reduction of the contact rate $\beta$ (panel C) and as a function of the number of vaccinated individuals (panel D).

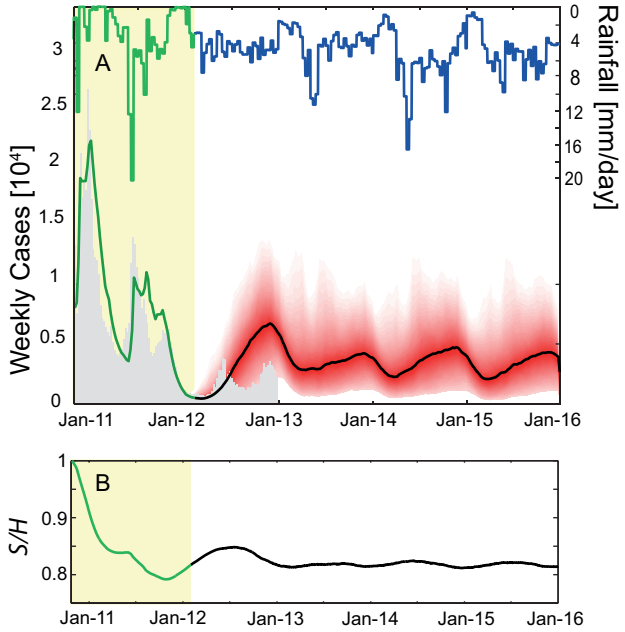


Figure 10: Panel A: multi-seasonal projection of the course of the epidemic in the period 15/01/2012-31/12/2015. Calibration is performed between 23/10/2010 and 14/01/2012 (shaded yellow). New weekly cases in Haiti in the calibration (green solid line) and in the prediction (black solid line/shaded range for the median/5th-95th percentile range) periods. Gray bars show the reported cases used for calibration and the cases registered until 31/12/2012. One rainfall realization is also shown in the upper part (green: observed; blue: generated pattern. Panel B: temporal evolution of the median ratio between susceptibles and total population.

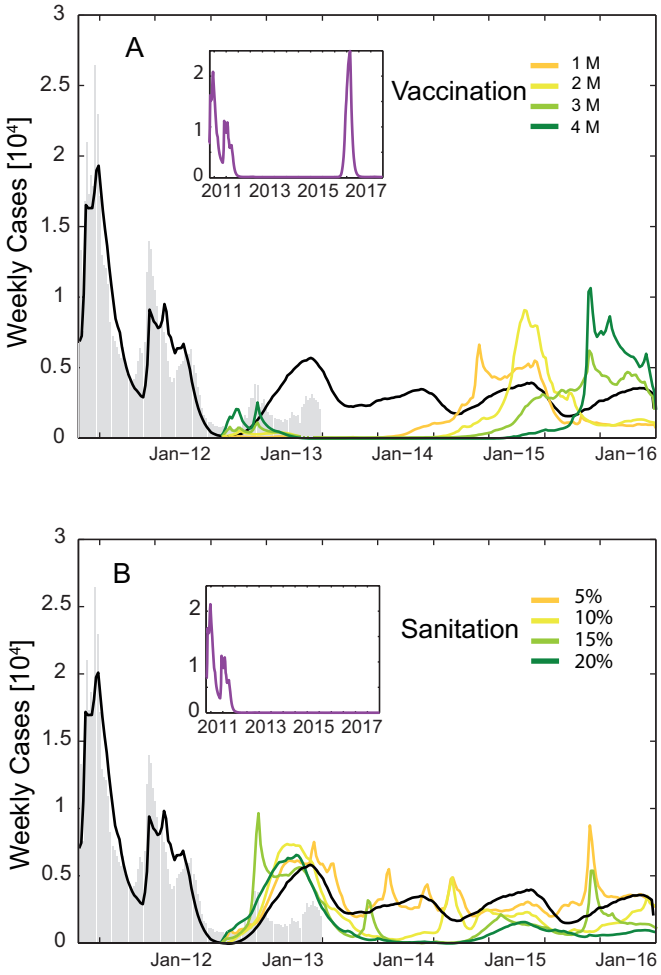


Figure 11: Effect of intervention policies on the predicted course of the epidemic between 08/2012 and 12/2015. The black solid line shows the median value of the prediction with no planned intervention. Other solid lines (from yellow to green) show the average predicted pattern of new weekly cases with increasing effort of vaccination (A) or sanitation (B). In the inset of both panels the lilac solid line shows the same pattern until 12/2017, for 8 million vaccinations (A) or 30% sanitation (B). Notice that small differences before the implementation of interventions may arise from different realizations of the stochastic rainfall scenarios.