Reassessment of the 2010–2011 Haiti cholera outbreak and rainfall-driven multiseason projections

Andrea Rinaldo[a,b], Enrico Bertuzzo[a], Lorenzo Mari[a], Lorenzo Righetto[a], Melanie Blokesch[c], Marino Gatto[a,d], Renato Casagrandi[d], Megan Murray[e], Silvan M. Vesenbeckh[e], and Ignacio Rodriguez-Iturbe[f,1]

[a]Laboratory of Ecohydrology, School of Architecture, Civil and Environmental Engineering, and [c]Laboratory of Molecular Microbiology, Global Health Institute, École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland; [b]Dipartimento di Ingegneria Civile, Edile ed Ambientale, Università di Padova, 35131 Padua, Italy; [d]Dipartimento di Elettronica e Informazione, Politecnico di Milano, 20133 Milan, Italy; [e]Harvard School of Public Health, Harvard University, Cambridge, MA 02138; and [f]Department of Civil and Environmental Engineering, Princeton University, Princeton, NJ 08544

Mathematical models can provide key insights into the course of an ongoing epidemic, potentially aiding real-time emergency management in allocating health care resources and by anticipating the impact of alternative interventions. We study the ex post reliability of predictions of the 2010–2011 Haiti cholera outbreak from four independent modeling studies that appeared almost simultaneously during the unfolding epidemic. We consider the impact of different approaches to the modeling of spatial spread of *Vibrio cholerae* and mechanisms of cholera transmission, accounting for the dynamics of susceptible and infected individuals within different local human communities. To explain resurgences of the epidemic, we go on to include waning immunity and a mechanism explicitly accounting for rainfall as a driver of enhanced disease transmission. The formal comparative analysis is carried out via the Akaike information criterion (AIC) to measure the added information provided by each process modeled, discounting for the added parameters. A generalized model for Haitian epidemic cholera and the related uncertainty is thus proposed and applied to the year-long dataset of reported cases now available. The model allows us to draw predictions on longer-term epidemic cholera in Haiti from multiseason Monte Carlo runs, carried out up to January 2014 by using suitable rainfall fields forecasts. Lessons learned and open issues are discussed and placed in perspective. We conclude that, despite differences in methods that can be tested through model-guided field validation, mathematical modeling of large-scale outbreaks emerges as an essential component of future cholera epidemic control.

waterborne diseases | epidemiology | ecohydrology | human mobility | Markov chain Monte Carlo

As a major cholera epidemic spread through Haiti (1–6), leading to 170,000 reported cases and 3,600 deaths at the end of 2010 (1), four independent modeling studies (7–10) appeared almost simultaneously, each predicting the subsequent course of the epidemic and/or the impact of potential management strategies.

The spate of mathematical models confirmed earlier suggestions that their impact on public health practice is gaining momentum (11). Indeed mathematical models of infectious diseases, once properly tested for reliability, can provide key insight into the course of an epidemic in time for action, thus averting deaths and reducing the number of infected patients through a sensible allocation of resources, possibly including vaccines (12, 13). Because the Haiti cholera models were published early in the course of the epidemic, which as of Oct 26, 2011 had gone on to produce an estimated 485,092 cases, 259,549 hospitalizations, and 6,712 deaths (1), the subsequent course of the epidemic allows an assessment of the reliability of the early predictions and a related discussion on the lessons learned.

Before 2010, cholera had never before been reported in Haiti (14, 15) and it was thus likely that the population had no significant prior exposure or acquired immunity to the disease. Haiti also lacked preparedness for this epidemic and suitable health infrastructure through which to combat it (2, 3). Although there is some debate as to the source, most experts agree that the first cases were autochthonous, brought into Haiti from a distant geographic source (6, 16, 17), and that these cases seeded the subsequent epidemic, which originated within the Centre department and then spread to all of the Haitian departments, exhibiting complex spatial and temporal patterns (1, 6) (Fig. S1).

Once a cholera epidemic starts, infected patients excrete huge numbers of *Vibrio cholerae* bacteria that spread either through water pathways (via active and passive dispersal) (10, 18–20) or through human mobility networks involving both susceptibles and infected individuals (8, 10, 20). Poor sanitation, which characterized Haiti after the disastrous 2010 earthquake, facilitates both types of spread and fosters the abundance of microorganisms in the water system. Some, such as *V. cholerae*, are extremely versatile and can quickly adapt to new environments. Being primarily an aquatic bacterium, *V. cholerae* can persist indefinitely in rivers, estuaries, and coastal regions without any need for human passage. The incidence of cholera in such ecosystems fluctuates as a function of climatic forces (in particular El Niño Southern Oscillation, ENSO) and changes as extensively described for the region around the Bay of Bengal (21–27) (e.g., ENSO is likely to operate in part through interannual changes in rainfall and act as a remote driver of the disease (26, 27)). Because of weak sanitary infrastructures and favorable environmental conditions, it seems therefore likely that cholera will continue to be a threat in Haiti, as well as in many developing countries (3–5).

In this paper, we examine premises, methods, and results of models of the 2010–2011 Haiti epidemic cholera, to derive lessons that directly affect the predictive value of model outputs on pathogen dispersal mechanisms, model-guided field validations, data requirements, and model identification. On the basis of this analysis, we highlight shortcomings of past approaches and discuss mechanisms of disease transmission driven by rainfall. Best-performing models are identified via rigorous criteria from available epidemiological and hydrological time series and, through them, a multiseason projection is proposed and discussed.

A First Assessment

All four models of the 2010–2011 Haiti cholera epidemic (7–10) address the coupled dynamics of susceptibles, infected individuals, and bacterial concentrations in the water reservoir in a spatially explicit setting of local human communities. The entire Haitian population was assumed to be susceptible at the outset of the epidemic. Each of the models assumed that the rate at which susceptibles become infected is dependent on the *V. cholerae* concentration in available water and, in turn, that new free-living bacteria are produced by infected individuals through fecal

contamination of water. The main differences among the models stemmed from assumptions about pathogen redistribution mechanisms among the different human communities.

We first examine the reliability of our own original scheme (7), later discussing the main differences from the other approaches (8–10) and the impact of these differences on predictions. The reference model of epidemiological dynamics and pathogen redistribution is described in (*SI Materials and Methods*) Fig. 1 compares the projected course of the epidemic as published in January 2011 with the actual reported case counts reported at the end of September 2011 (1). Highlighted (dark gray) in Fig. 1 is the dataset used for calibration, which is limited to the end of December 2010 before the first decline of the incidence of the disease. The original prediction (7) ran up to the end of May 2011 (solid line in Fig. 1). The 5-mo forecast, judged in retrospect, was quite robust and could have been used to make practical decisions and act in time. To facilitate a further assessment of our model reliability, we have now extended the original prediction to the end of September 2011 (Fig. 1, dashed line). Whereas the order of magnitude of total cumulated infections is captured up to September 2011, important features are clearly missed, such as the June–July revamping of weekly incidence (which is likely correlated to seasonal rainfall, Fig. 1, *Top*).

The various approaches dealing with the Haiti cholera outbreak (7–10) may be differentiated on the basis of their treatment of spatial transmission mechanisms. Andrews and Basu (9)

treated each Haiti administrative department independently without explicitly considering the spread of cholera among them whereas the other models (7, 8, 10) explicitly modeled inter- and intradepartmental pathogen redistribution. Different spatial resolutions also characterize the metacommunity models (10 local communities in ref. 8, 11 in ref. 9, 560 in ref. 7, and on the order of 20,000 in ref. 10).

For those studies that provided predictions of the subsequent course of the outbreak, the projections can be tested against current data, yielding a first assessment of validity and the limitations of different modeling assumptions. For example, Andrews and Basu (9) forecast a toll of 779,000 cases and 11,100 deaths from March 1 to November 30, 2011 that significantly overestimated the course of the epidemic; 324,405 new reported cases (accounted for as proposed in ref. 9) (*SI Materials and Methods*) and 2,040 deaths were actually reported between March 1 and October 26 (1). We suggest that the differences with observational data stem from the nature of the modeling assumptions that undermine the predictability of the approach. In fact, disregarding interdepartmental pathogen dispersal mechanisms implies the independent fitting of the model parameters to 11 separate departments. Therefore, model parameters are also charged with the effects of long-range transmission mechanisms, including human mobility that is empirically known to be reaching well beyond departmental domains (28). Regardless of calibration techniques, the large number of parameters (33 vs. 5 in ref. 7) likely increases the uncertainty of early projections.

The role of asymptomatic carriers may be clarified by the detailed examination of one scheme (8) that employs transmission mechanisms similar to those in ref. 7 (save for the inclusion of a human-to-human contagion) but neglects the role of inapparent infections. Asymptomatic carriers are thought to be a critical factor in cholera epidemics (29), particularly in Haiti (7, 9, 10), because of their number, unimpaired mobility, and thus major role in long- and short-range disease transmission. Inapparent infections are estimated at 20% of the total (9, 10, 29–31,). Moreover, they lead to some acquired immunity, thus temporarily reducing the number of persons in a region who are actually susceptible to the disease. The model by Tuite et al. (8) with realistic values of the basic reproduction number R_0 (*SI Materials and Methods*) fits the initial phases of the epidemic but would predict an excessive number of reported cases at later stages. To overcome this limitation, the authors propose an ad hoc sixfold reduction of the effective reproduction number in the first 3 mo of the epidemic, owing to disease-control interventions. If the compartment of susceptibles is not depleted otherwise, an equal decrease in transmission rates is implied, which seems unrealistic compared with the sanitation interventions analyzed. Adopting a model in which inapparent infections are accounted for avoids the need to force effective reproduction numbers to decrease in time because of unspecified disease-control measures (32).

An interesting mathematical cholera transmission model, individual based and of stochastic nature (10), focuses on the effects of vaccination strategies for epidemic cholera in Haiti. It addresses the same basic transmission processes as those in refs. 7 and 8, but also includes a 1- to 5-d latent period, a hyperinfective state of freshly shed bacilli, and a model of human mobility that incorporates remotely sensed population density data at 1-km^2 resolution and the localization of major rivers and highways. The study does not attempt to tune processes by matching observed space–time distributions of reported cholera cases (10) but is rather based on model parameter values and ranges from the literature. The main shortcomings of this approach include a limited capability of reproducing past observed infections and thus of reliably predicting future epidemic evolutions.

A Second Assessment

The newly available information on the Haitian outbreak allows a thorough reanalysis of primary and ancillary transmission mechanisms. The result of our reanalysis is a host of models of different complexity that are described in *SI Materials and Methods*.
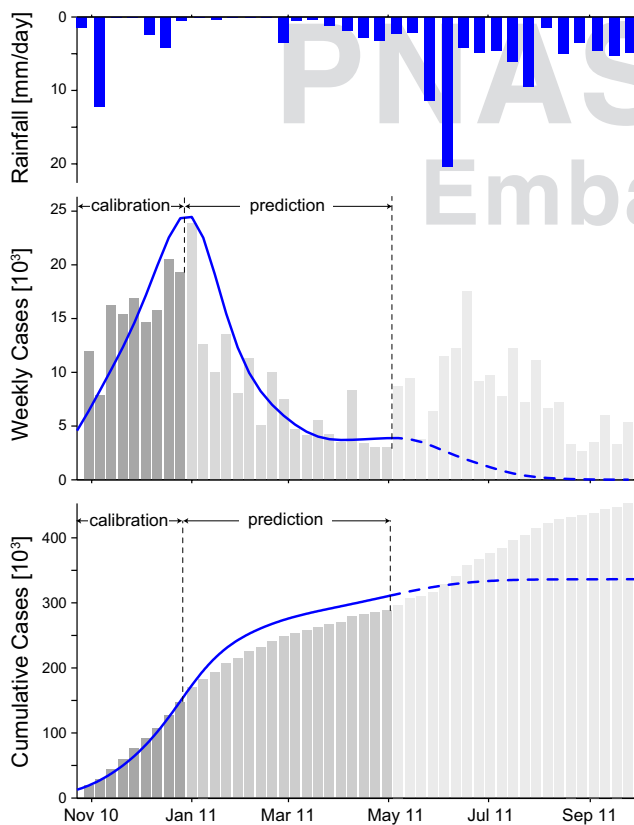

Fig. 1. (*Top*) Daily decadal rainfall intensity, averaged over the entire Haiti region (*SI Materials and Methods*). (*Middle*) Weekly reported cases (1) (gray bars) compared with the simulated incidence pattern (solid line) computed by the model in ref. 7. Data from each department were collected until September 30, 2011. The calibration dataset (dark gray) was limited to the total reported cases available until December 2010. The solid line shows the published early prediction (7) that was run until the end of May 2011. To facilitate the assessment, we have now extended the original prediction to the end of September 2011 (dashed line). (*Bottom*) Simulated and reported weekly cumulated cases.

Rinaldo et al.

Because the analysis is now extended to a 1-y time span, the loss of acquired immunity (i.e., a flux from the pool of recovered cases back to the pool of susceptible individuals) cannot be ignored. It is thus accounted for by all models compared here. We have also revised our metacommunity model of transmission among human settlements exposed to the infection by including specific mechanisms of hydrological transport (10, 18–20), as suggested by empirical observation of the downstream spreading of early infections along the Artibonite River (6). Pathogen dispersal along waterways is described (*i*) by a careful extraction of the river networks (Fig. 2) from digital terrain maps (DTMs), through suitable geomorphologic criteria (*SI Materials and Methods*), and (*ii*) as a biased random walk process on an oriented graph (18) (*Materials and Methods* and *SI Materials and Methods*). Pathogen redistribution is also enhanced by contamination of the water reservoir driven by heavy seasonal rains (*Materials and Methods*). Inclusion of such overloads was prompted by the clear empirical correlation, observed in Haiti in June and July, between weekly rainfall and enhanced infections (Fig. 1). We considered two options: an increase of contamination rates depending on rainfall intensity and a mechanistic account of the washout of open-air defecation sites by surface runoff (*SI Materials and Methods*). Human mobility patterns are explicitly modeled (8, 10, 20). Observations of fast intercatchment transmission of the infection that would not be explained by water pathway pathogen dispersal and of actual individual displacements in times of cholera support this assumption. Mobility patterns are described by a layer of nodal connectivity (Fig. 2D and *Materials and Methods*). With suitable spatial resolution, human settlements may be placed at nodes of the hydrological network (Fig. S2), and edges are measured by the distances connecting them (*SI Materials and Methods*). We assume that susceptible and infected individuals engage in short-term trips from the communities where they live toward other settlements. While traveling or commuting, susceptible individuals can be exposed to pathogens and return as infected carriers to their settlement (10, 20). Similarly, infected hosts can disseminate the disease away from their home community—in many cases infected individuals are asymptomatic and thus are not barred from their usual activities. Connectivity structures and fluxes of human mobility have long been studied in epidemiology (33, 34), often on the basis of gravity-like models where the flux between two communities owing to human mobility is proportional to the product of the respective populations and decays with the distance separating them (10, 20, 35). Our choice of a model of this kind (*SI Materials and Methods*) is indirectly supported by a recent empirical study (28) that tracked daily Haitian average movements through mobile phones to determine likely new areas for cholera outbreaks far from the site where the disease was first detected. The study proves that outbound travels from the source area are frequent and most of the country received persons from the affected area whereas the vast majority of individuals leaving the source area traveled to just a few large recipients that include surrounding communal sections.

We have also comparatively tested other disease transmission mechanisms. The revised models considered in particular enhanced community-wide transmission due to a hyperinfectious *V. cholerae* state (freshly shed cholera pathogens requiring much lower concentrations to cause infection) (9, 10, 29, 36, 37). Latent stages and human-to-human transmission have been ruled out as discussed in *SI Materials and Methods*. Other cofactors of disease transmission, judged of lesser importance, are also discussed in *SI Materials and Methods*.

Parameter calibration is performed via Markov chain Monte Carlo techniques (38, 39) (*SI Materials and Methods*). To compare the ability of different models (with different added complexity and parameters) to reproduce the spatiotemporal epidemic patterns observed in Haiti, we have ranked the performances of different candidate models according to Akaike's information criterion (AIC) (40). AIC is a model-selection procedure that explicitly takes into account the trade-off between model accuracy and complexity. The ranking is based on AIC scores that measure the goodness of fit, discounting for the different number of calibration parameters (*SI Materials and Methods*). We specifically tested four candidate models: (*i*) a simple model with no pathogen hyperinfectivity and a single water compartment (*SI Materials and Methods*), (*ii*) the same model with pathogen hyperinfectivity, (*iii*) a model with two water compartments (water reservoir and sewage system) but no hyperinfectivity, and (*iv*) the same as in *iii* but including vibrio hyperinfectivity. AIC scores are in Table S1. The optimal model is fully described in *Materials and Methods*. According to the results of model identification, bursts of infections can be best explained by accounting for larger concentrations in the water compartment due to massive pathogen loads brought by hydrologic washout. AIC not retaining the modeling of hyperinfective stages of the *V. cholerae* bacterium may be surprising given its importance in other approaches. Our result would confirm earlier remarks (41) suggesting that the timescale of hyperinfectivity is so short that all that matters for modeling purposes would be the overall rate of transmission resulting from the many mechanisms that underlie it, especially given the complexity of the spatial linkages. This result cannot be generalized, however, as discussed in *SI Materials and Methods*.
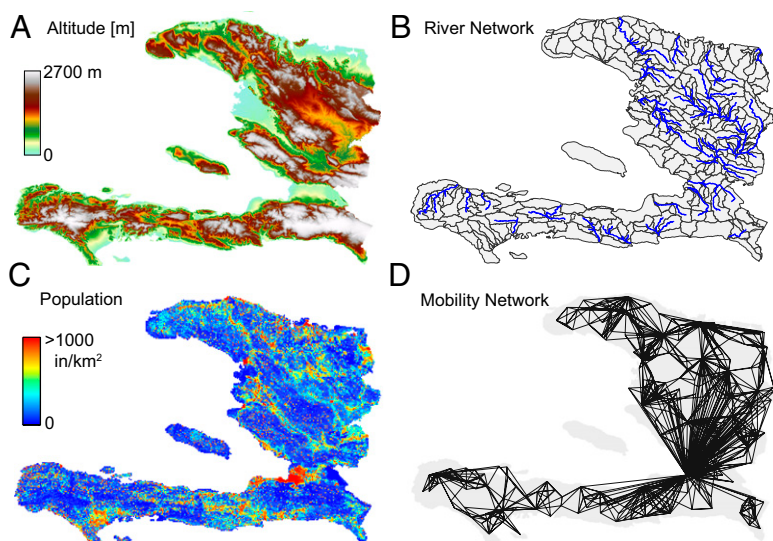
Fig. 2. (*A*) Color-coded digital terrain elevation map (DTM) of Haiti; (*B*) the subdivision of Haitian territory in hydrological units (subbasins) extracted from the DTM, as a result of the convergence of several geomorphological criteria (*SI Materials and Methods*); (*C*) spatial distribution of population density obtained by LandScan remote sensing, which is translated into a geo-referenced spatial distribution of nodes *i* endowed with population H_i (*SI Materials and Methods*); (*D*) A relevant subset of the network of human mobility, here portrayed synthetically by the four largest outbound connections for each node.

Fig. 4. Multiseasonal evolution of Haiti epidemic cholera (from October 2010 to January 2014) simulated by the best-performing model. Reported new weekly cases (gray bars) are shown along with the simulated incidence pattern (solid line). (*Upper*) Rainfall is predicted starting from October 1, 2011. The range of uncertainty due to the uncertainty in rainfall forecast (*SI Materials and Methods*) is highlighted by the shading. Red dots highlight cases reported after September and not used for calibration. Note the agreement between the epidemic fading in the data and the model projection, with the exception of an unpredicted infections peak in the late Fall of 2011. This exception is likely explained by the extreme rainfall events that occurred in the first decade of October that were missed by the rainfall patterns projected from the end of September.
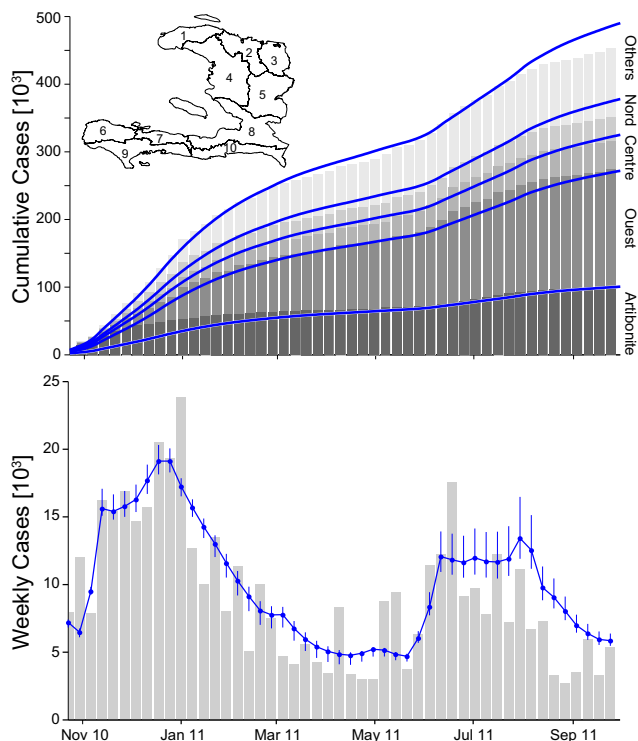
Fig. 3. Simulated evolution of the Haiti epidemic cholera from October 2010 to September 2011 by the revised complete model that includes bursts of infections caused by pathogen loads brought into the water reservoir by hydrologic washout. The final choice follows from the ranking of the performances of different candidate models according to Akaike's information criterion (40). (*Upper*) Weekly cumulated reported cases are visualized as the sum of the reported cases in each department (gray bars), fitted by the simulation of the revised model at the department level (blue solid line). The performance of the model at the department level is also shown (blue solid lines). (*Inset*) Haitian departments are listed as follows: 1, Nord-Ouest; 2, Nord; 3, Nord-Est; 4, Artibonite; 5, Centre; 6, Grand-Anse; 7, Nippes; 8, Ouest; 9, Sud; and 10, Sud-Est. (*Lower*) Evolution of reported new weekly cases (gray bars) along with the simulated incidence pattern of the revised model (solid line). Error bars highlight the range of uncertainty due to parameter estimation (described in detail in *SI Materials and Methods*). See also animation Movie S1.

Fig. 3 shows the results of the best-ranking model and its estimation uncertainty. Optimal parameters and their credible intervals are reported in Table S2. The model can better reproduce the timing and the magnitude of the epidemic in the 10 Haitian departments (Fig. S4), including the seasonal June–July resurgence, in particular in the most populated and affected regions (Artibonite and Ouest). The capabilities of describing the spatial and temporal patterns of the reported infections grew considerably with the information gained after the early predictions, although the short-term prognostic value of the early model (7) remains noteworthy. Thus, different levels of model sophistication might serve well for evolving insight into the course of an ongoing epidemic. In particular, the model including rainfall drivers and waning immunity allows us to draw predictions for long-term cholera dynamics in Haiti: Fig. 4 shows a multiseason projection up to January 2014, obtained by using suitable rainfall field predictions (*SI Materials and Methods*). We have chosen to run the example with an average duration of the acquired immunity of 3 y (it must be noted, however, that in such cases model-guided field studies on the rate of loss of acquired immunity become crucial). The related predictions are fairly consistent in suggesting significant Fall bursts of infections, stemming from seasonal rainfall and from the timing of the replenishment of the pool of susceptibles from previous infections due to immunity waning. The underestimation of the predicted
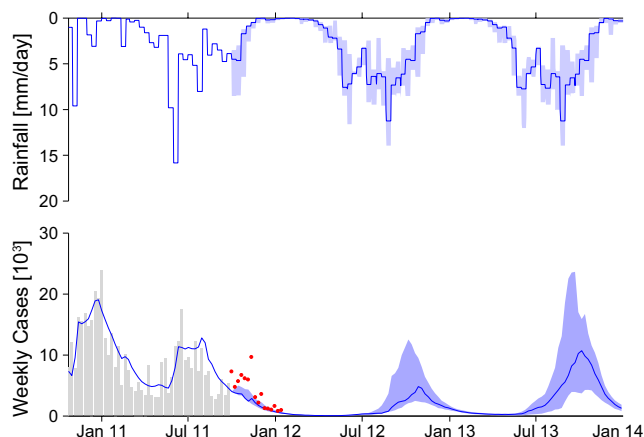
infections in the late Fall of 2011 (for which data are now available) is likely explained by the fact that rainfall patterns projected from the end of September missed the extreme rainfall events that actually occurred in the first decade of October. Note finally that this is a worst-case scenario that assumes no improvement in sanitation and ignores any decrease in exposure upon learning from past experience.

Discussion

The encouraging outcomes of early predictions of the 2010–2011 Haiti cholera outbreak and the broadened capabilities of a generalized approach based also on later observations have been highlighted. Despite their capabilities, several limitations and open issues remain toward a general predictive model of epidemic cholera.

One important limitation for long-term predictions is our relatively poor knowledge of community-wide loss rates of acquired immunity (here characterized by a deterministic rate parameter ρ) (*Materials and Methods*). Susceptibles decrease because of infections and mortality and increase through the loss of immunity of recovered cases. The dynamics of recovered individuals play a crucial role in the long run. Although at short timescales it is reasonable to assume complete immunity of recovered patients, waning or boosting of acquired immunities at longer times may have a substantial impact on epidemic dynamics. Acquired immune responses may vary in relation to age group, *V. cholerae* strain (e.g., serogroup, biotype, and serotype) (42), the severity of the contracted infection (43), and endemic vs. epidemic cholera (44). On a population level, the buildup of durable immunity usually results in a lengthening of the interepidemic intervals. However, complex cyclic climatic forcings may control interarrivals of cholera outbreaks. Unfortunately, cholera elicits only a temporary immunity (44) whose uncertainty affects the long-term replenishment of the pool of susceptibles. Thus, the fraction of the population susceptible to the disease at the onset of subsequent outbreaks will likely be predictable only with great uncertainty. Worst-case scenarios (i.e., assuming again the entire Haitian population being susceptible) seem too crude an approximation for predicting a meaningful deployment of suitable medical supplies and staff.

In a similar vein, the minimalist assumption of a constant ratio of asymptomatic to symptomatic infections during the course of

the epidemic, to which modeling results prove quite sensitive (32), is clearly an approximation. Extended epidemiological evidence and relatedly improved modeling are needed. Also, individuals with blood group O are more prone to severe cholera symptoms than other blood group individuals, although the mechanism underlying this association is still under debate (45, 46). Hence, the distribution of blood groups among the population might need to be considered where significant differences emerge. Besides blood group, moreover, the susceptibility to cholera depends on local intestinal immunity (from previous exposure or vaccination), bacterial load and intrinsic host factors such as stomach pH (gastric acid provides a barrier) (47) whose community-wide evaluations are difficult. One further issue concerns the modeling of the rapid patient discharge from treatment centers. Infected patients are released when fewer than three watery stools over the past 6 h are observed. This action is reasonable, given the need for space in emergency hospitals, but poses the problem that released patients are treated as recovered. However, they still excrete significant quantities of vibrios and thus contribute to enhanced spread of the disease, a process that is not accounted for in any of the proposed models. Although backtracking the mobility of early discharges seems possible on the basis of treatment center acceptance data, a large-scale application seems impractical. A related issue concerns the actual sanitation within treatment centers, e.g., the isolation of their latrines from the local water cycle especially during acute phases of the epidemic. From a modeling viewpoint, all this information is currently combined into calibration parameters at the loss of predictive power.

Another issue concerns the proper depletion of the pool of susceptibles due to intervention strategies. In this context, two main interventions have been discussed: vaccination and the extended use of antibiotics (48, 49). Whereas the idea of mass vaccination in Haiti has split experts (13), the impact of parameter uncertainty on the effects of vaccinations is certainly in need of an assessment. A more uniform opinion concerns the use of antibiotics. With respect to the latter, the World Health Organization generally recommends the administration of antibiotics only to severe cases of cholera (50). However, soon after the onset of the cholera epidemic in Haiti, many researchers demanded a widening of these regulations toward an extended use of antibiotics (49, 51). In favor of a broader use of antibiotics to all hospitalized cases, regardless of the severity of the symptoms, is the shorter duration of the acute phase and the reduced shedding of *V. cholerae* bacteria into the environment (52). The new policies therefore recommend also treating moderately ill patients with antibiotics (53). Significant differences in the use of antibiotics existed among treatment centers. The most extreme examples were health partners who used antibiotics in a prophylactic manner to protect family/community members affiliated with cholera patients (54). The most common antibiotic used in Haiti is a single dose of doxycycline, which proved very effective for cholera treatment (55). However, drastic increases of doxycycline/tetracycline-resistant isolates were recently reported, referring to a cholera outbreak in Zambia (56) where a clear correlation was observed between the development of resistant strains and large-scale use of antibiotics (either for treatment or for prophylaxis) (56). In fact, many drug-resistant strains of *V. cholerae* have been described within recent years due to the spread of antibiotic resistance genes by horizontal gene transfer (HGT) (for recent review see ref. 57). That antibiotics themselves, as well as signals within *V. cholerae*'s environmental niche, induce diverse mechanisms of HGT (58–61) strengthens the fear of the development of antibiotic resistance in Haiti. It is therefore crucial to monitor the resistance pattern of the circulating *V. cholerae* strains. Regardless of deeper questions about possible development of specific resistance by the bacterial strains, we note that only very detailed information—possibly beyond reach—about clinical practice in space and time would allow us to better frame the related parameters of the transmission model (*Materials and Methods*).

In conclusion, weak sanitary infrastructures and favorable environmental conditions will likely lead to indefinite persistence of the pathogen within Haiti, suggesting that cholera will continue to be a threat for many years. Thus, we deem it significant that reliable predictions of epidemic cholera within Haiti could be offered by spatially explicit quantitative disease transmission models. Such models could thus be effectively used in the very course of an outbreak to allocate healthcare resources. To some extent, and with greater uncertainty, they could also possibly evaluate alternative strategies of emergency management. Although deeper integration of epidemiological, socioeconomical, and environmental data is needed to clarify issues still open, and despite differences in methods that can be tested through model-guided field validation, mathematical modeling of large-scale outbreaks emerges as an essential component of the control of future cholera epidemics and of modern epidemiology.

Materials and Methods

We describe in what follows only the best-ranked model according to the AIC (see *SI Materials and Methods* for a detailed description of all candidate models). 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 concentrations of *V. cholerae* in the water reservoir. We consider n metacommunities ($i = 1, n$) spatially distributed within a given domain that embeds the hydrologic and the human mobility networks. Epidemiological dynamics and pathogen transport can be described by the following set of coupled differential equations:

$$\frac{dS_i}{dt} = \mu(H_i - S_i) - \mathcal{F}_i(t)S_i + \rho R_i \qquad [1]$$

$$\frac{dI_i}{dt} = \mathcal{F}_i(t)S_i - (\gamma + \mu + \alpha)I_i \qquad [2]$$

$$\frac{dR_i}{dt} = \gamma I_i - (\rho + \mu)R_i \qquad [3]$$

$$\frac{dB_i}{dt} = -\mu_B B_i - l\left(B_i - \sum_{j=1}^{n} P_{ji}\frac{W_j}{W_i}B_j\right) + \frac{p}{W_i}[1 + \phi J_i(t)]\mathcal{G}_i(t). \qquad [4]$$

The evolution of the susceptible compartment (Eq. 1) is a balance between population demography and infections due to contact with pathogens infesting the water reservoir. The host population is assumed to be at a demographic equilibrium, where μ is the human mortality rate and H_i is the size of the local community. The total contact rate $\mathcal{F}_i(t) = \beta((1-m)B_i/(K+B_i) + m\sum_{j=1}^{n} Q_{ij}B_j/(K+B_j))$ accounts for both local (first-term) and mobility-related (second-term) disease transmission. The parameter β represents the rate of exposure to contaminated water (possibly varying in time and/or space to account for heterogeneous sanitation conditions and related control strategies) (7), whereas $B/(K+B)$ is the probability of becoming infected due to the exposure to a concentration B of vibrios, with K being the half-saturation constant (62). Human mobility patterns are defined according to a connection matrix in which individuals leave their original node (say i) with a probability m, reach their target location (say j) with a probability Q_{ij}, and then come back to node i. Here, we have opted for a gravity-like model of connections Q_{ij} that decays exponentially with nodal distance; i.e., $Q_{ij} = H_j e^{-d_{ij}/D} / (\sum_{k\neq i}^{N} H_k e^{-d_{ik}/D})$, where d_{ij} is the (shortest-path) distance between nodes i and j, and D is the deterrence cutoff distance. The dynamics of the infected compartment (Eq. 2) are a balance between newly infected individuals and losses due to recovery and natural/pathogen-induced mortality. Infected individuals recover at a rate γ or die of natural or cholera-induced mortality at a rate μ or α, respectively. The dynamics of recovered individuals are considered here (Eq. 3), because usually waterborne diseases confer just temporary immunity (25). The parameter ρ quantifies the loss of acquired immunity and the related replenishment of the susceptible compartment. Infected individuals contribute to the concentration of free-living vibrios in the water reservoir at a rate p/W_i, where p is the contamination rate and W_i (with $W_i = cH_i$) (7, 19) is the volume of water in the reservoir. To account for both local and mobility-related pathogen dispersion, here we actually consider the total infective pool $\mathcal{G}_i(t)$; i.e., $\mathcal{G}_i(t) = (1-m)I_i + m\sum_{j=1}^{n} Q_{ji}I_j$. To address the seasonal revamping of infections, we assume that the baseline contamination rate can be increased by local rainfall intensities $J_i(t)$ via a runoff coefficient ϕ

(*SI Materials and Methods*). We further assume that the vibrios diluted in the water reservoir undergo hydrological dispersal at rate l. The spread of pathogens over the river network is described as a biased random walk process on an oriented graph in which pathogens move from node i to node j of the hydrological network with probability P_{ij} (18, 63). The transport process is assumed to be conservative, i.e., $\sum_{j=1}^{n} P_{ij} = 1$, except for the network outlets where absorbing boundary conditions are imposed. Finally, the bacteria in the water reservoir are assumed to die at a constant rate μ_B. As initial conditions, we impose that, as of October 18, 2010 ($t = 0$), we have $S_i(0) = H_i$ except for the nodes i where $I_i(0)$ is given, matching the reported cases as detailed in ref. 6. Also, $R_i(0) = 0$ for all nodes i, and $B_i(0)$ is in equilibrium with the local number of infected cases; i.e., $B_i(0) = pI_i(0)/(W_i\mu_B)$. Several model parameters can be estimated from the literature (Table S2). Five parameters are obtained through calibration of contrasting model simulations with the reported cumulative cases in each Haitian department as recorded in the epidemiological dataset (*SI Materials and Methods*). These parameters are the ratio $\theta = p/(Kc)$, the hydrological dispersal rate l, the human mobility rate m, the average deterrence distance D, and the contamination parameter ϕ (*SI Materials and Methods*).

1. Pan American Health Organization (2011) Haiti cholera outbreak data. Available at http://new.paho.org/hq/images/Atlas_IHR/CholeraHispaniola/atlas.html. Accessed December 10, 2011.
2. Butler D (2010) Cholera tightens grip on Haiti. *Nature* 468:483–484.
3. Walton DA, Ivers LC (2011) Responding to cholera in post-earthquake Haiti. *N Engl J Med* 364:3–5.
4. Sack DA (2011) How many cholera deaths can be averted in Haiti? *Lancet* 377:1214–1216.
5. Friedrich MJ (2011) Haiti cholera outbreak. *JAMA* 305:2402.
6. Piarroux R, et al. (2011) Understanding the cholera epidemic, Haiti. *Emerg Infect Dis* 17:1161–1168.
7. Bertuzzo E, et al. (2011) Prediction of the spatial evolution and effects of control measures for the unfolding Haiti cholera outbreak. *Geophys Res Lett* 38:L06403.
8. Tuite A, et al. (2011) Cholera epidemic in Haiti, 2010: Using a transmission model to explain spatial spread of disease and identify optimal control interventions. *Ann Intern Med* 154:593–601.
9. Andrews JR, Basu S (2011) Transmission dynamics and control of cholera in Haiti: An epidemic model. *Lancet* 377:1248–1255.
10. Chao DL, Halloran ME, Longini IM, Jr. (2011) Vaccination strategies for epidemic cholera in Haiti with implications for the developing world. *Proc Natl Acad Sci USA* 108:7081–7085.
11. Clark J, et al. (2001) Ecological forecasts: An emerging imperative. *Science* 293:657–660.
12. Chaignat C, et al. (2008) Cholera in disasters: Do vaccines prompt new hopes? *Expert Rev Vaccines* 7:431–435.
13. Cyranoski D (2011) Cholera vaccine plan splits experts. *Nature* 469:273–274.
14. Ackers M, et al. (1998) Are there national risk factors for epidemic cholera? The correlation between socioeconomic and demographic indices and cholera incidence in Latin America. *Int J Epidemiol* 27:330–334.
15. Anonymous ; The Lancet Infectious Diseases (2010) As cholera returns to Haiti, blame is unhelpful. *Lancet Infect Dis* 10:813.
16. Chin C, et al. (2011) The origin of the Haitian cholera outbreak strain. *N Engl J Med* 364:33–42.
17. Hendriksen R, et al. (2011) Population genetics of Vibrio cholerae from Nepal in 2010: Evidence on the origin of the Haitian outbreak. *MBio* 2:e00157-11.
18. Bertuzzo E, et al. (2008) On the space-time evolution of a cholera epidemic. *Water Resour Res* 44:W01424.
19. Bertuzzo E, et al. (2010) On spatially explicit models of cholera epidemics. *J R Soc Interface* 7:321–333.
20. Mari L, et al. (2012) Modelling cholera epidemics: The role of waterways, human mobility and sanitation. *J. R. Soc. Interface* 9:376–388.
21. Huq A, et al. (1995) Coexistence of V. cholerae O1 and O139 Bengal in plankton in Bangladesh. *Lancet* 345:1249.
22. Colwell RR (1996) Global climate and infectious disease: The cholera paradigm. *Science* 274:2025–2031.
23. Pascual M, Rodó X, Ellner SP, Colwell R, Bouma MJ (2000) Cholera dynamics and El Niño southern oscillation. *Science* 289:1766–1769.
24. Pascual M, Bouma MJ, Dobson AP (2002) Cholera and climate: Revisiting the quantitative evidence. *Microbes Infect* 4:237–245.
25. Koelle K, Rodó X, Pascual M, Yunus M, Mostafa G (2005) Refractory periods and climate forcing in cholera dynamics. *Nature* 436:696–700.
26. Pascual M, Chaves LF, Cash B, Rodó X, Yunus M (2008) Predicting endemic cholera: The role of climate variability and disease dynamics. *Research* 36:131–140.
27. Cash D, et al. (2010) Disentangling the impact of ENSO and Indian ocean variability on the regional climate of Bangladesh: Implications for cholera risk. *J Clim* 23:2817–2831.
28. Bengtsson L, et al. (2011) Improved response to disasters and outbreaks by tracking population movements with mobile phone network data: A post-earthquake geospatial study in Haiti. *PLoS Med* 8:e1001083.
29. King AA, Ionides EL, Pascual M, Bouma MJ (2008) Inapparent infections and cholera dynamics. *Nature* 454:877–880.
30. Longini IM, Jr., et al. (2007) Controlling endemic cholera with oral vaccines. *PLoS Med* 4:e336.
31. Weil A, et al. (2009) Clinical outcomes in household contacts of patients with cholera in Bangladesh. *Clin Infect Dis* 49:1473–1479.
32. Rinaldo A, et al. (2011) A transmission model of the 2010 cholera epidemic in Haiti. *Ann Intern Med* 155:403–404.
33. Eubank S, et al. (2004) Modelling disease outbreaks in realistic urban social networks. *Nature* 429:180–184.
34. Riley S (2007) Large-scale spatial-transmission models of infectious disease. *Science* 316:1298–1301.
35. Erlander S, Stewart NF (1990) *The Gravity Model in Transportation Analysis – Theory and Extensions* (VSP Books, Zeist, The Netherlands).
36. Merrell DS, et al. (2002) Host-induced epidemic spread of the cholera bacterium. *Nature* 417:642–645.
37. Hartley DM, Morris JG, Smith DL (2006) Hyperinfectivity: A critical element in the ability of Vibrio cholerae to cause epidemics? *PLoS Med* 3:63–69.
38. Vrugt JA, Robinson BA (2007) Improved evolutionary optimization from genetically adaptive multimethod search. *Proc Natl Acad Sci USA* 104:708–711.
39. ter Braak CJF, Vrugt JA (2008) Differential evolution Markov chain with snooker updater and fewer chains. *Stat Comput* 18:435–446.
40. Akaike H (1974) A new look at the statistical model identification. *IEEE Trans Automat Contr* 19:716–723.
41. Pascual M, Koelle K, Dobson AP (2006) Hyperinfectivity in cholera: A new mechanism for an old epidemiological model? *PLoS Med* 3:e280.
42. Ali M, et al. (2011) Natural cholera infection-derived immunity in an endemic setting. *J Infect Dis* 204:912–918.
43. Cash R, et al. (1974) Response of man to infection with Vibrio cholerae. 1. Clinical, serologic and bacteriologic responses to a known inoculum. *J Infect Dis* 129:45–52.
44. Longini I, et al. (2002) Epidemic and endemic cholera trends over a 33-year period in Bangladesh. *J Infect Dis* 186:246–251.
45. Chaudhuri A, De S (1977) Cholera and blood-groups. *Lancet* 2:404.
46. Harris J, et al. (2005) Blood group, immunity, and risk of infection with Vibrio cholerae in an area of endemicity. *Infect Immun* 73:7422–7427.
47. Sack DA, Sack RB, Nair GB, Siddique AK (2004) Cholera. *Lancet* 363:223–233.
48. Ivers LC, Farmer P, Almazor CP, Léandre F (2010) Five complementary interventions to slow cholera: Haiti. *Lancet* 376:2048–2051.
49. Farmer P, al. (2011) Meeting cholera's challenge to Haiti and the world: A joint statement on cholera prevention and care. *PLoS Negl Trop Dis* 5:e1145.
50. WHO (2011) Fact sheet N.107 – August 2011. Technical report (World Health Organization, Regional Office for the Eastern Mediterranean). Available at http://www.who.int/mediacentre/factsheets/fs107/en/index.html. Accessed November 1, 2011.
51. Nelson EJ, Nelson DS, Salam MA, Sack DA (2011) Antibiotics for both moderate and severe cholera. *N Engl J Med* 364:5–7.
52. Harris J, et al. (2010) Cholera's western front. *Lancet* 376:1961–1965.
53. Centers for Disease Control and Prevention (2010) Defeating cholera: Clinical presentation and management for Haiti cholera outbreak. Technical report. Available at http://www.cdc.gov/haiticholera/clinicalmanagement/pdf/clinicalmanagement.pdf. Accessed November 1, 2011.
54. Ministère de la Santé Publique et de la Population (MSPP) and the Regional Office of the World Health Organization for the Americas (PAHO) (2011) Cholera and post-earthquake response in Haiti. Health Cluster Bulletin 17, Friday, February 4, Technical report. Available at http://new.paho.org/blogs/haiti/?p=1770. Accessed November 1, 2011.
55. Sack DA, Islam S, Rabbani H, Islam A (1978) Single-dose doxycycline for cholera. *Antimicrob Agents Chemother* 14:462–464.
56. Mwansa J, et al. (2007) Multiply antibiotic-resistant Vibrio cholerae O1 biotype El Tor strains emerge during cholera outbreaks in Zambia. *Epidemiol Infect* 135:847–853.
57. Kitaoka M, Miyata ST, Unterweger D, Pukatzki S (2011) Antibiotic resistance mechanisms of Vibrio cholerae. *J Med Microbiol* 60:397–407.
58. Beaber JW, Hochhut B, Waldor MK (2004) SOS response promotes horizontal dissemination of antibiotic resistance genes. *Nature* 427:72–74.
59. Guerin E, et al. (2009) The SOS response controls integron recombination. *Science* 324:1034.
60. Meibom KL, Blokesch M, Dolganov NA, Wu CY, Schoolnik GK (2005) Chitin induces natural competence in Vibrio cholerae. *Science* 310:1824–1827.
61. Suckow G, Seitz P, Blokesch M (2011) Quorum sensing contributes to natural transformation of Vibrio cholerae in a species-specific manner. *J Bacteriol* 193:4914–4924.
62. Codeço CT (2001) Endemic and epidemic dynamics of cholera: The role of the aquatic reservoir. *BMC Infect Dis* 1:1.
63. Bertuzzo E, Maritan A, Gatto M, Rodriguez-Iturbe I, Rinaldo A (2007) River networks and ecological corridors: Reactive transport on fractals, migration fronts, hydrochory. *Water Resour Res* 43:W04419.

Supporting Information

Rinaldo et al.Rinaldo et al. 10.1073/pnas.1203333109

SI Materials and Methods

In what follows we provide a complete derivation of the mathematical treatment summarized in *Materials and Methods* in the main text. We first describe the basic model used for the early prediction (1). Then we show how it has been modified to include information gained during the course of the Haitian epidemic. We also discuss possible refinements of both epidemiological processes at the local scale and cholera spreading mechanisms.

The Original Model. The first modeling aspect that had to be faced concerned the choice of spatially explicit or implicit approaches. In the geographic scales incurred in the Haitian case, pathogen transport timescales (and thus epidemic timescales; Fig. S1) are shown (2) to be larger than timescales characteristic of the local infection dynamics (Fig. 1 of the main text), thus theoretically requiring spatially explicit schemes (3).

We consider n communities ($i = 1, n$) spatially distributed within a given domain that embeds the hydrologic and the human mobility networks. Let $S_i(t)$ and $I_i(t)$ be the local abundances of susceptible and infected individuals in each node i of the river network at time t, and let $B_i(t)$ be the concentration of *Vibrio cholerae* in the water reservoir at site i. Epidemiological dynamics and pathogen transport were described in the earliest prediction (1) by the following set of coupled differential equations:

$$\frac{dS_i}{dt} = \mu(H_i - S_i) - \beta \frac{B_i}{K + B_i} S_i$$

$$\frac{dI_i}{dt} = \beta \frac{B_i}{K + B_i} S_i - (\gamma + \mu + \alpha) I_i \qquad \textbf{[S1]}$$

$$\frac{dB_i}{dt} = \frac{p}{W_i} I_i - \mu_B B_i - l\left(B_i - \sum_{j=1}^{n} Z_{ji} \frac{W_j}{W_i} B_j\right).$$

The host population is assumed to be at a demographic equilibrium, where μ is the human mortality rate and H_i is the population size of the local community. In the early application (1) it was assumed that $S_i(0) = H_i$ because of the lack of any preexisting immunity (2, 4–6). The parameter β 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 vibrios, K being the half-saturation constant (7, 8). Infected individuals recover at a rate γ or die of natural or cholera-induced mortality at a rate μ or α, respectively. Infected individuals contribute to the concentration of free-living vibrios at a rate p/W_i, where p is the local rate at which bacteria excreted by one infected individual reach and contaminate the local water resource reservoir W_i (with $W_i = cH_i$). Bacteria die at a constant rate μ_B and undergo dispersal at rate l. Given the nearly complete lack of detailed information, the probability Z_{ij} of pathogen movement from node i to node j of the network was defined as $Z_{ij} = H_j e^{-d_{ij}/D} / \sum_{k \neq i}^{n} H_k e^{-d_{ik}/D}$, where d_{ij} is the distance between nodes i and j, and D is the distance parameter of the exponential kernel.

Lessons Learned from the Course of the Haitian Epidemic. The newly available information on the Haitian outbreak helped to deepen our understanding of primary and ancillary transmission mechanisms. Therefore, we pursue here modifications of model S1 as presented in ref. 1 to include mechanisms that could not be properly parameterized at the time of the early predictions. In particular:

i) Empirical evidence highlights the role of river networks in the early transmission of the disease (2). We therefore choose to explicitly address the hydrologic transport of cholera pathogens (3, 9–12).

ii) Proxies provided by mobile phone calls document widespread host movements in Haiti, suggesting their role in the sustained spreading of the disease (13). We therefore choose to include a framework for the description of human mobility (11, 12).

iii) The analysis is now extended to a 1-y timespan. We therefore choose to include the loss of acquired immunity, i.e., a flux from the pool of recovered back to the compartment of susceptible individuals whose basic timescales (14) are comparable to the simulated timespan.

Significant correlations are observed between seasonal rainfall patterns and the resurgence of the epidemic (15) (Fig. 1 of main text). We therefore include a description of increased water contamination due to rainfall-mediated pathogen overloads.

From a technical perspective, to account for all of the above mechanisms (Fig. S2*A*) model S1 can be modified as follows:

$$\frac{dS_i}{dt} = \mu(H_i - S_i) - \mathcal{F}_i(t) S_i + \rho R_i$$

$$\frac{dI_i}{dt} = \mathcal{F}_i(t) S_i - (\gamma + \mu + \alpha) I_i$$

$$\frac{dR_i}{dt} = \gamma I_i - (\rho + \mu) R_i \qquad \textbf{[S2]}$$

$$\frac{dB_i}{dt} = -\mu_B B_i - l\left(B_i - \sum_{j=1}^{n} P_{ji} \frac{W_j}{W_i} B_j\right) + \frac{p}{W_i}[1 + \phi J_i(t)] \mathcal{G}_i(t).$$

A detailed account of the newly introduced mechanisms follows.
Hydrological transport of vibrios. As in model S1 the evolution of bacterial concentration in the water reservoir is a balance between vibrio excretion/dilution, natural mortality, and spatial dissemination. However, the fourth equation of model S2 explicitly accounts for two different pathogen dispersal pathways, respectively through either hydrological transport over river networks or human mobility (Fig. S2). Specifically, hydrological transport can be described, in the general case, by a biased random walk process on an oriented graph (9, 10). Pathogens have a mobility rate l and move from node i to node j of the hydrological network with probability P_{ij}. For the present scale of analysis, we assume $P_{ij} = 1$ if j is a downstream nearest neighbor of node i and 0 otherwise, thus reducing the transport scheme to a deterministic advection-dominated process.

From an operational point of view, the computational domain is subdivided into watersheds on the basis of hydrologic divides. Human communities are defined by the population hosted within each watershed. These communities can thus be hierarchically organized within a river network, with the hydrologic connectivity matrix P among them following directly. The extraction of channeled and unchanneled fluvial basin features and catchment divide delineation from remotely sensed and objectively manipulated geomorphic information is a well-established procedure in hydrology (16–20). Hydrologic divides can be inferred from drainage directions extracted from digital terrain models (DTM). Raw DTM data consist of discretized elevation fields on a lattice whose size is typically $\mathcal{O}(10 - 100)$ m depending on the detail

required. In particular, the Haitian stream networks were derived from US Geological Survey (USGS) digital elevation maps (available online at http://seamless.usgs.gov), which have a grid resolution of 100 m in the planform and a precision of ±0.5 m in the elevation field. The first step consists of the determination of the unique steepest-descent flow path from each pixel to the sea. It is then possible to delineate the river basins, defined as the set of pixels that drain into the same point of the coastline. With this procedure 9,790 different basins are obtained (Fig. S3A), whose areas range from 1 km^2 to >5,000 km^2 (Artibonite). To subdivide the system into coherent hydrologic units endowed with approximately the same drainage area, two further manipulations are necessary, namely (i) to subdivide (according to catchment divides) the largest watersheds into several subwatersheds and (ii) to aggregate basins with a small catchment area. We choose the watershed area such that the whole system is represented by ∼300–400 units. This number is deemed reasonable to achieve a suitably detailed spatial resolution while limiting the computational time of each model run to the order of seconds. Fig. S3B shows an intermediate step and the final representation of the Haiti territory subdivided into 366 hydrologic entities.

To determine the number of people hosted by each hydrologic unit we use a remotely sensed map of population distribution provided by the Oak Ridge National Laboratory (21). The data represent the ambient population distribution (average over 24 h). The spatial resolution is 30 × 30 arc-seconds (or 0.083 × 0.083 decimal degrees), which represents, at that latitude, cells of ∼1 km^2. The nodes of our models are identified as the centroids of the population distribution of each watershed.

Human mobility patterns. Human mobility is described by a network in which the nodes correspond to those of the hydrological layer (Fig. S2C) and edges are defined by connections among communities. We assume that all individuals (susceptibles, infectives, and recovered) can undertake short-term trips from the communities where they live toward other settlements. Note, however, that only the movement of susceptibles and infectives is important for disease dynamics. In fact, while traveling or commuting, susceptible individuals can be exposed to pathogens and return as infected carriers to the settlement where they usually live. Similarly, infected hosts can disseminate the disease away from their home community. It should be remembered that in many cases infected individuals are asymptomatic and thus are not barred from their usual activities by the presence of the pathogen in their intestine. Human mobility patterns are defined according to a connection matrix in which individuals leave their original node (say i) with a probability m, reach their target location (say j) with a probability Q_{ij}, and then come back to node i.

Human mobility is accounted for in the first, second, and fourth equations of model S2. In the revised model, in fact, a contact rate

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

is defined to account for both local (first-term) and mobility-related (second-term) infections. $\mathcal{F}_i(t)$ is in fact the total contact rate of the disease, potentially depending on all bacteria present in the system, i.e., bacteria that might be taken up at node i by susceptibles who do not move from their home site or at each of the other nodes j by those susceptibles who do move away from it. In the same way, to account for both local and mobility-related pathogen shedding by infected individuals, a total infective pool

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

which describes infected individuals who are actually active at node i at time t, is actually considered.

Topological and transition probability structures Q_{ij} for human mobility networks used in epidemiology are quite varied (22, 23). They can be based on "attractivity" like in gravity models (24), on the actual transportation network (25), or on theoretical network models (26, 27). Most reported mobility networks, however, refer to specific settings, possibly empirically derived for developed countries, and should be thoughtfully adapted to different socioeconomic and environmental contexts. Gravity models, in particular, have long been applied in the epidemiological literature to describe the impact of human mobility on the emergence of a suite of diseases (24, 28), including influenza (29), HIV (30), and measles (31).

From a mathematical perspective, a gravity-like model for human mobility can be defined by specifying 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 attractivity factor of node j is subsumed by its size, whereas the deterrence factor is assumed to be dependent on distance and represented by an exponential kernel (with shape factor D). Note that the same formalism was applied in ref. 1. However, there the gravity-like model represented a synthetic description of all vibrio transport mechanisms, whereas here, in the present approach, it is used to characterize the effects of human mobility alone (hydrological transport being modeled separately, see section above).

To compute mobility fluxes, distances among nodes have also to be determined. As clear from Haitian topology, in this context it is not realistic to consider the Euclidean distance between any two generic points (e.g., nodes separated by the Caribbean Sea). For this reason, we first created a graph where only four nearest neighbors of each node are connected. Distances on this graph are computed as Euclidean distances between nodes (note that for the Haitian context we intentionally did not use other metrics, say via roads or railways). Then, distances between two generic nodes i and j are computed as the length of the shortest path, allowed by the graph just described, that connects i and j. Shortest paths are computed through the standard Dijkstra algorithm.

Significantly, in a recent empirical study (13), daily average movements of mobile phones have been tracked in Haiti to determine likely new areas for cholera outbreaks out from the site where the disease was first detected. The results of the analysis show that travels out of the source area are frequent and most of the country has received persons from the affected area. Interestingly, the vast majority of persons leaving the source area travel to just a few large recipients that include surrounding communal sections. Therefore, we maintain that a gravity-like network model would do well in reproducing such a mobility pattern.

Loss of acquired immunity. The loss of acquired immunity is explicitly accounted for in the first (susceptible pool) and third (recovered individuals) equations of model S2. We assume that immunity from cholera is lost at a rate ρ, thus leading to a replenishment of the susceptible compartment. The parameter ρ is initially set (*Discussion* in the main text) at $1/\rho \approx 3$ (years) (14). We investigated the sensitivity of model results with respect to variations of this parameter through a dedicated sensitivity analysis (see below).

Rainfall as a driver of increased water contamination. To address the seasonal revamping of infections, we assume that the baseline contamination rate p can be increased by rainfall [whose intensity is denoted by $J_i(t)$ in model S2] through a synthetic runoff coefficient φ. This increase is in turn due to extra loads of pathogens brought into the water reservoir by seasonal rainfall, which can wash out large loads of pathogens from open-air defecation sites into waterways through surface runoff. This infection mechanism is speculated to be properly described neither by the mass balance equation of the water reservoir (in times of heavy rainfall one would reasonably expect the water reservoir to increase in volume, thus decreasing its concentration at constant pathogen loads) nor by empirically augmented exposure probabilities, but rather from added pathogen loads to the current reservoir volume. Within the Haitian context, the clear empirical correlation between weekly rainfall and new reported infection cases—started at the end of May, 2011 (15)—leaves little doubt about the origins of the transmission (Fig. 1 in the main text).

Decadal rainfall estimates for the 2010–2011 period have been obtained from data collected by the National Aeronautics and Space Administration and Japan Aerospace Exploration Agency Tropical Rainfall Measuring Mission (TRMM_3B42 precipitation estimates, see http://trmm.gsfc.nasa.gov/ for details) through the Famine Early Warning Systems Network data portal set up by USGS (http://earlywarning.usgs.gov/fews/). Rainfall data are spatially distributed with the resolution of 0.25° of latitude and longitude. Data from 2002 to 2011 were available. Rainfall fields for the specific epidemic period simulated have been interpolated to obtain the time series of the average rainfall intensity for each hydrologic unit, $J_i(t)$.

As reported in the main text, we projected the epidemic forecast to multiple seasons. To do that, we needed to generate a suitable projection of rainfall fields, which has been generated as follows. We used the 10-y record of decadal precipitations (with the available resolution). To run the model up to January 2014 we need at least 3 y of generated rainfall starting from October 2011. Each of these 3 y of fluctuating, spatially distributed rainfall fields is sampled randomly (with replacement) from the 10-y dataset available. As a result, each sequence of generated rainfall is a standard bootstrapping at a yearly timescale of the observed data. We carried out 1,000 Monte Carlo-like realizations of rainfall generation to assess the uncertainty in the forecasts of cholera incidence due to the uncertainty in rainfall projections, estimated from the bootstrapping procedure. Although hydrologically a gross predictive tool, the rainfall model devised produces realistically correlated (in time and space) and obviously credible rainfall fields (because they actually occurred).

Testing Alternative Mechanisms of Cholera Transmission.

Although model S2 is considerably more complex than model S1, it still provides a basic description of the processes leading to new infections and water contamination. One could thus wonder whether a refined formalization of such processes could possibly lead to a better characterization of the observed spatiotemporal patterns of cholera spread—obviously at the expense of dealing with more complicated mathematical models.

Hyperinfectivity. A somewhat debated topic in the literature concerns the description of hyperinfective bacterial stages (32). In fact, experimental evidence exists, suggesting that the passage of *V. cholerae* through the human intestine increases pathogen infectivity (32, 33). Increased infectivity reportedly lasts from a few hours up to 1 d. At later times, bacteria enter a stage characterized by normal infectivity. A mathematical model accounting for the hyperinfective bacterial state was proposed a few years ago (34) and, since then and despite being subject to some criticisms (35), hyperinfectivity has often been accounted for in cholera modeling exercises, including some recent attempts to describe the Haiti epidemic (36, 37). Our modeling framework

can be effortlessly extended to include a hyperinfective bacterial stage into the general architecture of cholera transmission (Fig. S2A). To that end, the mathematical description of pathogen dynamics in the water reservoir has to be broken into two separate differential equations, respectively for hyperinfective and regular bacteria; i.e.,

$$\frac{d\mathcal{B}_i}{dt} = -\xi\mathcal{B}_i - l\left(\mathcal{B}_i - \sum_{j=1}^{n} P_{ji}\frac{W_j}{W_i}\mathcal{B}_j\right) + \frac{p}{W_i}\mathcal{G}_i(t)$$

$$\frac{dB_i}{dt} = \xi\mathcal{B}_i - \mu_B B_i - l\left(B_i - \sum_{j=1}^{n} P_{ji}\frac{W_j}{W_i}B_j\right),$$

where $\mathcal{B}_i$ is the concentration of hyperinfective pathogens in the water reservoir and ξ is the rate at which vibrios lose hyperinfectivity [here we assume $1/\xi = 1$ (day)]. Obviously, the total contact rate $\mathcal{F}_i(t)$ has to be modified accordingly as

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

where K_{HI} is the half-saturation constant for hyperinfective bacteria ($K/K_{HI} \approx 50$) (34).

A mechanistic description of water contamination. A driver of water contamination that certainly plays a key role in the spreading of the disease is rainfall. By extending the basic model (Fig. S2A), we can specifically focus on the assumption that overall water contamination increases during rainfall events owing to extra loads of pathogens brought into the water reservoir by seasonal rainfalls. In the fourth equation of model S2 it is in fact assumed that a local rainfall event of intensity $J_i(t)$ directly enhances water contamination through a synthetic runoff coefficient φ. It is interesting to see whether other formalizations, possibly aimed at a mechanistic description of the relevant contamination processes, could be engineered. In particular, here we describe how increasing water contamination during rainfall events can be derived formally from a detailed description of realistic contamination pathways.

Let $B_i(t)$ and $B_i^s(t)$ be bacterial concentrations respectively in the water reservoir and in the sewage system of node i at time t. The sewage system is intended in its broadest possible definition, including the lack of it thereof, i.e., open-air defecation sites. We assume that (i) excreted vibrios actually reach the sewage system and (ii) the pathogens contained in the sewage system can be released to the water reservoir because of leakage/washing out (at a rate that can depend upon rainfall intensity). With these hypotheses in mind we can thus write the following submodel for the balance of *V. cholerae* concentrations,

$$\frac{dB_i^s}{dt} = -\mu_B^s B_i^s + \frac{\epsilon}{W_i^s}\mathcal{G}_i(t) - f(J_i(t))B_i^s$$

$$\frac{dB_i}{dt} = -\mu_B B_i + f(J_i(t))B_i^s - l\left(B_i - \sum_{j=1}^{n} P_{ji}\frac{W_j}{W_i}B_j\right),$$

where ϵ represents the vibrio excretion rate of infected individuals, μ_B^s is the bacterial mortality within the sewage system, W_i^s is the volume of water stored in the sewage system ($W_i^s = c^s H_i$), and $f(J_i(t))$ is a function describing the flux of pathogens from the sewage system to the water reservoir owing to a rainfall event at the ith node characterized by intensity $J_i(t)$. Note that the vibrios contained in the sewage system are assumed not to

undergo hydrological dispersal. For the sake of simplicity we further assume a linear relationship between rainfall intensity and pathogen flux from the sewage system to the water reservoir; i.e., $f(J_i(t)) = f_0 + f_1 J_i(t)$, where f_0 and f_1 are positive constants.

Local Basic Reproductive Numbers R_0. A linear stability analysis of the complete model carried out for $l = 0$, $m = 0$, and $J = 0$ shows that, given an initial condition of the type $I_i(0) = \delta$, $S_i(0) = H_i - \delta$, $R_i(0) = 0$, and $B_i(0) = 0$ (where δ is a small positive number), a local epidemic outbreak can occur in node i only if

$$R_0 = \frac{\beta p H_i}{K W_i \mu_B (\gamma + \mu + \alpha)} > 1,$$

where R_0 is the value of the local basic reproductive number of the disease. Technically, $R_0 > 1$ is the local condition for the disease-free equilibrium to be unstable (8).

Model Selection. To assess whether and how more complex models are better suited to describe the evolution of the Haiti cholera outbreak, we can rank the performances of different candidate models according to Akaike's information criterion (AIC) (38). AIC is a model-selection procedure that explicitly takes into account the trade-off between model accuracy and complexity, measured as the number Θ of free parameters (i.e., the structural parameters for each candidate model, plus one residual variance parameter) (39, 40).

We specifically test four candidate models:

i) Model S2 as presented above.
ii) Same as in *i* but accounting for pathogen hyperinfectivity, i.e.,

$$\frac{dS_i}{dt} = \mu(H_i - S_i) - \mathcal{F}_i^{HI}(t)S_i + \rho R_i$$

$$\frac{dI_i}{dt} = \mathcal{F}_i^{HI}(t)S_i - (\gamma + \mu + \alpha)I_i$$

$$\frac{dR_i}{dt} = \gamma I_i - (\rho + \mu)R_i$$

$$\frac{d\mathcal{B}_i}{dt} = -\xi \mathcal{B}_i - l\left(\mathcal{B}_i - \sum_{j=1}^{n} P_{ji}\frac{W_j}{W_i}\mathcal{B}_j\right) + \frac{p}{W_i}[1 + \phi J_i(t)]\mathcal{G}_i(t)$$

$$\frac{dB_i}{dt} = \xi \mathcal{B}_i - \mu_B B_i - l\left(B_i - \sum_{j=1}^{n} P_{ji}\frac{W_j}{W_i}B_j\right). \qquad [S3]$$

iii) Same as in *i* but accounting for two water compartments (water reservoir and sewage system), i.e.,

$$\frac{dS_i}{dt} = \mu(H_i - S_i) - \mathcal{F}_i(t)S_i + \rho R_i$$

$$\frac{dI_i}{dt} = \mathcal{F}_i(t)S_i - (\gamma + \mu + \alpha)I_i$$

$$\frac{dR_i}{dt} = \gamma I_i - (\rho + \mu)R_i$$

$$\frac{d\mathcal{B}_i^s}{dt} = -\mu_B^s B_i^s + \frac{\epsilon}{W_i^s}\mathcal{G}_i(t) - f(J_i(t))B_i^s$$

$$\frac{dB_i}{dt} = f(J_i(t))B_i^s - \mu_B B_i - l\left(B_i - \sum_{j=1}^{n} P_{ji}\frac{W_j}{W_i}B_j\right). \qquad [S4]$$

iv) Same as in *iii* but accounting also for hyperinfectivity, i.e.,

$$\frac{dS_i}{dt} = \mu(H_i - S_i) - \mathcal{F}_i^{HI}(t)S_i + \rho R_i$$

$$\frac{dI_i}{dt} = \mathcal{F}_i^{HI}(t)S_i - (\gamma + \mu + \alpha)I_i$$

$$\frac{dR_i}{dt} = \gamma I_i - (\rho + \mu)R_i$$

$$\frac{d\mathcal{B}_i^s}{dt} = -\xi \mathcal{B}_i^s + \frac{\epsilon}{W_i^s}\mathcal{G}_i(t) - f(J_i(t))\mathcal{B}_i^s$$

$$\frac{dB_i^s}{dt} = \xi \mathcal{B}_i^s - \mu_B^s B_i^s - f(J_i(t))B_i^s \qquad [S5]$$

$$\frac{d\mathcal{B}_i}{dt} = f(J_i(t))\mathcal{B}_i^s - \xi \mathcal{B}_i - l\left(\mathcal{B}_i - \sum_{j=1}^{n} P_{ji}\frac{W_j}{W_i}\mathcal{B}_j\right)$$

$$\frac{dB_i}{dt} = \xi \mathcal{B}_i + f(J_i(t))B_i^s - \mu_B B_i + -l\left(B_i - \sum_{j=1}^{n} P_{ji}\frac{W_j}{W_i}B_j\right).$$

From an operational perspective, we first calibrate independently each candidate model against the epidemiological data available for the Haiti cholera epidemic (next section). For each best-fit model we then compute

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

where η is the number of data points ($\eta = n_d n_w$, where $n_d = 10$ and $n_w = 49$ are 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.,

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

where $\mathcal{C}(i,j)$ and $\hat{\mathcal{C}}(i,j)$ are the cumulative 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 the population of each watershed that belongs to a given department. RSS values and AIC scores for all models are reported in Table S1. Model S2 is ranked first in the model selection procedure and is thus chosen as reference model.

Parameter Calibration. Whereas several parameters are estimated from the literature (1) (see Table S2 for the numerical values and relevant references), others are obtained through calibration. The set of tuning parameters is peculiar to each candidate model, specifically $\theta = p/(cK)$, l, m, D, and ϕ for models S2 and S3 and θ, l, m, D, f_0, and f_1 for models S4 and S5.

The optimization approach is based on Markov chain Monte Carlo (MCMC) sampling, which 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) (41). Specifically, we use the differential evolution adaptive Metropolis (DREAM) algorithm (42), an efficient implementation of MCMC that runs multiple different chains simultaneously to ensure global explo-

ration of the parameter space and adaptively tunes the scale and orientation of the jumping distribution using differential evolution (43) and a Metropolis–Hastings update step (44, 45). More specifically, we adopt the DREAM$_{ZS}$ variant of the DREAM algorithm, which also 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 (46). 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].

The goodness of each single simulation is computed as the residual sum of squares between weekly reported cholera cases in each Haitian department as recorded in the epidemiological dataset and simulated by the model being tested. Specifically, the cumulative number of reported cases $C_i(t)$ in each network node i at time t can be directly computed by solving either

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

for models S2 and S4 or

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

for models S3 and S5. The numerical values of the best-fit parameters of the best-ranked model and their 5–95% confidence intervals are reported in Table S2. Note that with optimal parameters, the inferred values of the basic local reproductive number R_0 would have been uniformly set at $R_0 = 3.47$.

In addition to the best-fit results reported in Fig. 3 of the main text, Fig. S4 shows the time series of the new weekly cholera cases for the 10 Haitian departments.

Sensitivity Analysis. To disentangle the impact of each parameter of model S2 on the intensity of the epidemic outbreak we performed a sensitivity analysis of the model outcomes with respect to variations of the parameter values. In particular, we allowed the parameters to vary (±20% variation with respect to the best-fit parameter set) one by one through repeated model runs. We then computed the variations of simulated total cholera incidence with respect to the best-fit simulation (Fig. S5) (11, 36). Note that this sensitivity analysis is different from the uncertainty of model simulations reported in Fig. 3 of the main text, which is obtained by sampling 1,000 times the posterior distribution obtained via MCMC and recomputing reported cases.

We also performed a sensitivity analysis of the results obtained with the best ranked model (model S2) for a broader range of variation of the rate of loss of acquired immunity (ρ), a key parameter controlling the number of reinfections. The range investigated is $1/\rho = 1 - 5$ (years) [our reference value is $1/\rho = 3$ (years), Table S2]. Results are reported in Fig. S6. The impact on the simulation proves indeed small for the highest loss rates ρ. The smallest loss rate shows instead a significant effect on the simulations. However, such a low rate would have implied a large number of reinfections, unreported to date. From the results of 2010–2011, one is therefore allowed to infer that the 1-y immunity loss scenario is unrealistic. Therefore, the analysis is deemed supportive of our estimate of $1/\rho = 3$ (years) (Table S2).

SI Discussion

Assessment of the Predictive Power of Model S2.
This section provides an example of the predictive power of the reference model S2. The calibration procedure described before has been performed using all of the epidemic data available up to the end of September 2011. To assess the capability of the model to project the future evolution of an epidemic, we recalibrate the model using a subset of the epidemic data and we evaluate the agreement between the actual evolution of the epidemic after the calibration period and the one hindcast by the model (model validation). In particular, Fig. S7 shows the weekly cases of cholera predicted for the period June–September 2011, using epidemiological information up to the end of May 2011. The uncertainty related to parameter estimation in the validation period is higher—as expected—than that for the one obtained using the whole set of epidemic data (Fig. 3 of the main text). However, the model is able to capture the resurgence of the infection in correspondence to the heavy rainfalls that occurred in June and July 2011. The latter result is deemed relevant, in particular because of the absence of significant periods of heavy rainfall during the calibration period. However, note that the hindcast just presented cannot be considered as a real prediction because model S2 has been previously selected with the full knowledge of the epidemiological data. Moreover, during the hindcasting period the model is forced by the observed rainfall. Future applications of the model as a predictive tool will have to deal also with the forecasting of precipitation, possibly within a stochastic framework that might allow an ensemble of predictions.

Transmission Processes Judged of Lesser Importance.
Some cofactors of disease transmission have been judged of lesser importance without formal AIC screening. They are as follows:

i) Latency: Some models of cholera epidemic (e.g., ref. 37) include another compartment where susceptibles experience a 1- to 5-d latent period of disease incubation before their transferral to the compartment of infectives. The weekly timescales of reported cases preclude a significant calibration of the rate of release from the latent stage, and thus the related effect is included in the other transmission parameters.

ii) Human-to-human transmission: This effect has been modeled only in a few studies (e.g., ref. 47). Direct person-to-person transmission has been documented in particular cases, typically for household members of patients with cholera who are naturally at increased risk of infection, e.g., from food preparation by infected individuals (48) or contamination of water storage containers (49). This effect is accounted for in our model(s) by the direct injection of freshly shed bacteria into the water reservoir that conceptualizes all of the above.

iii) The implications of the stigma of cholera among certain population groups: In fact, reported cases often occur far from the closest treatment center (as archived, georeferenced disinfection sites of households of deceased patients suggest), implying less prompt treatment and possibly higher mortality.

The role of internally displaced people (IDP): The concentration of people in provisional tents that have by necessity replaced previous households after the disastrous January 2010 earthquake may result in higher than usual contact rates (so-called "packing" effects reported in endemic regions, say during monsoons) yet facing better sanitation and water supply conditions than their former neighborhoods where fewer nongovernmental organizations (NGOs) have been operating.

Higher than expected mortality of reported cholera cases, especially in the early phases of the epidemic, possibly due to social unrest (e.g., road blocks preventing timely treatment).

The link between school terms and the resurgence of infections, of possible relevance because children are argued to be less exposed to infection especially from food while in school.

The removal of services (e.g., in IDP camps) in relation to exit strategies of NGOs.

The impact of decaying attention toward hygiene along with the fading of the acute phase of the epidemic.

A further possible refinement could be concerned with a detailed account of the dynamics of the local water reservoir W_i. Whether it is wise to allow W_i to fluctuate in time depends on the setting. Coupled with model S2 (or more complex accounts), however, a rich range of highly nontrivial epidemic behaviors is shown to depend on the dynamics of the water reservoir (50). In the Haitian case, however, it is argued that the time variation of the water reservoir is not a major factor to determine increased infections. We therefore simplify hydrological dynamics by assuming that $dW_i/dt \approx 0$; i.e., W_i is assumed constant and functionally dependent on the population H_i; i.e., $W_i = cH_i$, where c is a suitable constant (3, 10, 11).

Assessment of the Estimates by Andrews and Basu (36). Andrews and Basu (36) fitted their model against a base-case estimate for the total number of cases for each department. These base-case estimates were obtained by dividing the reported hospitalized cases by 0.38, the highest ratio of hospitalized cases to total cases reported from any departments. They forecast a toll of 779,000 cases and 11,100 deaths from March 1 to November 30, 2011. The related observational numbers from March 1 to October 26 yield 232,452 equivalent reported cases (i.e., the reported hospitalized cases divided by 0.38) and 2,040 deaths (15).

Boundary Conditions. The above models (Eqs. **S2–S5**) were solved with absorbing boundary conditions at sea for bacterial transport. These conditions imply that the bacterial contents are discharged at sea and do not reenter the system at any time, thus implying that coastal waters act as a sink of pathogens. One

could argue, however, that coastal waters may act differently from an absorbing boundary because *V. cholerae* can thrive in saline environments as documented in settings prone to endemic cholera like in the Bay of Bengal (51). Moreover, inbound salt wedge propagation or tidal ebb flows may affect bacterial transport and induce an upstream flux of pathogens, possibly enhanced by passive transport, resulting in higher local concentrations (from where they affect the whole domain). Technically, this process can be modeled in several manners. A simple option consists of assuming the outlet acts as a diffusive, as opposed to absorbing, barrier where one assumes that half of the outgoing bacterial flux to the sea reenters the outlet node. Alternatively, one might set the reentering flux to an outflux fraction to be calibrated. Such assumptions, however, should be the subject of specific model-guided field validation requiring field work on the hydroclimatically forced ecology of the bacterium in crucial coastal–riverine transition spots. The detailed relevant hydrodynamics should also be pointed out.

For the sake of example, we ran the exercise of model calibration and model selection under the new assumption of diffusive barrier. Results show that the best-performing model in this case is model S3, which includes hyperinfectivity. Fig. S8 shows a comparison of the best-performing models under the assumption of absorbing (blue) or diffusive (red) boundaries. Best-fit parameters obviously change and with them the estimation of the relative importance of different transmission mechanisms for the infection. It is important to note, however, that our modeling framework can consistently reproduce the main features of observed epidemic patterns under both assumptions and, most importantly, forecast similar multiseasonal epidemic trajectories.

1. Bertuzzo E, et al. (2011) Prediction of the spatial evolution and effects of control measures for the unfolding Haiti cholera outbreak. *Geophys Res Lett* 38:L06403.
2. Piarroux R, et al. (2011) Understanding the cholera epidemic, Haiti. *Emerg Infect Dis* 17:1161–1168.
3. Bertuzzo E, et al. (2010) On spatially explicit models of cholera epidemics. *J R Soc Interface* 7:321–333.
4. Enserink M (2010) Infectious diseases. Haiti's outbreak is latest in cholera's new global assault. *Science* 330:738–739.
5. Walton DA, Ivers LC (2011) Responding to cholera in post-earthquake Haiti. *N Engl J Med* 364:3–5.
6. Sack DA (2011) How many cholera deaths can be averted in Haiti? *Lancet* 377: 1214–1216.
7. Capasso V, Paveri-Fontana SL (1979) A mathematical model for the 1973 cholera epidemic in the European Mediterranean region. *Rev Epidemiol Sante Publique* 27:121–132.
8. Codeço CT (2001) Endemic and epidemic dynamics of cholera: The role of the aquatic reservoir. *BMC Infect Dis* 1:1.
9. Bertuzzo E, Maritan A, Gatto M, Rodriguez-Iturbe I, Rinaldo A (2007) River networks and ecological corridors: Reactive transport on fractals, migration fronts, hydrochory. *Water Resour Res* 43:W04419.
10. Bertuzzo E, et al. (2008) On the space-time evolution of a cholera epidemic. *Water Resour Res* 44:W01424.
11. Mari L, et al. (2012) Modelling cholera epidemics: The role of waterways, human mobility and sanitation. *J R Soc Interface* 9:376–388.
12. Mari L, et al. (2012) On the role of human mobility in the spread of cholera epidemics: Towards an epidemiological movement ecology. *Ecohydrology* 6, in press, doi: 10.1002/eco.262.
13. Bengtsson L, et al. (2011) Improved response to disasters and outbreaks by tracking population movements with mobile phone network data: A post-earthquake geospatial study in Haiti. *PLoS Med* 8:e1001083.
14. Koelle K, Rodó X, Pascual M, Yunus M, Mostafa G (2005) Refractory periods and climate forcing in cholera dynamics. *Nature* 436:696–700.
15. Pan American Health Organization (2011) Haiti cholera outbreak data. Available at http://new.paho.org/hq/images/Atlas_IHR/CholeraHispaniola/atlas.html. Accessed Dec 10, 2011.
16. Band L (1986) Topographic partition of watersheds with digital elevation models. *Water Resour Res* 22:15–24.
17. Montgomery D, Dietrich W (1988) Where do channels begin? *Nature* 336:232–234.
18. Montgomery DR, Dietrich WE (1992) Channel initiation and the problem of landscape scale. *Science* 255:826–830.
19. Tarboton D (1997) A new method for the determination of flow directions and contributing areas in grid digital elevation models. *Water Resour Res* 33:309–319.
20. Rodriguez-Iturbe I, Rinaldo A (1997) *Fractal River Basins. Chance and Self-Organization* (Cambridge Univ Press, New York).
21. Oak Ridge National Laboratory (2011) Haiti population data. Available at http://www.ornl.gov/sci/landscan/index.shtml. Accessed January 1, 2011.
22. Eubank S, et al. (2004) Modelling disease outbreaks in realistic urban social networks. *Nature* 429:180–184.
23. Riley S (2007) Large-scale spatial-transmission models of infectious disease. *Science* 316:1298–1301.
24. Erlander S, Stewart NF (1990) *The Gravity Model in Transportation Analysis – Theory and Extensions* (VSP Books, Zeist, The Netherlands).
25. Colizza V, Barrat A, Barthélemy M, Vespignani A (2006) The role of the airline transportation network in the prediction and predictability of global epidemics. *Proc Natl Acad Sci USA* 103:2015–2020.
26. Bollobas L (2000) *Random Graphs* (Cambridge Univ Press, Cambridge, UK).
27. Watts DJ, Strogatz SH (1998) Collective dynamics of 'small-world' networks. *Nature* 393:440–442.
28. Longini IM, Jr. (1988) A mathematical model for predicting the geographic spread of new infectious agents. *Math Biosci* 90:367–383.
29. Eggo RM, Cauchemez S, Ferguson NM (2011) Spatial dynamics of the 1918 influenza pandemic in England, Wales and the United States. *J R Soc Interface* 8:233–243.
30. Thomas R (1999) Reproduction rates in multiregion modeling systems for HIV / AIDS. *J Reg Sci* 39:359–385.
31. Ferrari MJ, et al. (2008) The dynamics of measles in sub-Saharan Africa. *Nature* 451: 679–684.
32. Merrell DS, et al. (2002) Host-induced epidemic spread of the cholera bacterium. *Nature* 417:642–645.
33. Alam A, et al. (2005) Hyperinfectivity of human-passaged V. cholerae can be modeled by growth in the infant mouse. *Infect Immun* 73:6674–6679.
34. Hartley DM, Morris JG, Smith DL (2006) Hyperinfectivity: A critical element in the ability of Vibrio cholerae to cause epidemics? *PLoS Med* 3:63–69.
35. Pascual M, Koelle K, Dobson AP (2006) Hyperinfectivity in cholera: A new mechanism for an old epidemiological model? *PLoS Med* 3:e280.
36. Andrews JR, Basu S (2011) Transmission dynamics and control of cholera in Haiti: An epidemic model. *Lancet* 377:1248–1255.
37. Chao DL, Halloran ME, Longini IM, Jr. (2011) Vaccination strategies for epidemic cholera in Haiti with implications for the developing world. *Proc Natl Acad Sci USA* 108:7081–7085.
38. Akaike H (1974) A new look at the statistical model identification. *IEEE Trans Automat Contr* 19:716–723.
39. Burnham K, Anderson D (1983) *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach* (Springer, New York).
40. Corani G, Gatto M (2007) Structural risk minimization: A robust method for density-dependence detection and model selection. *Ecography* 30:400–416.
41. Gilks WR, Richardson S, Spiegelharter DJ (1995) *Markov Chain Monte Carlo in Practice* (Chapman & Hall, New York).

42. ter Braak CJF, Vrugt JA (2008) Differential evolution Markov chain with snooker updater and fewer chains. *Stat Comput* 18:435–446.
43. Storn R, Price K (1997) Differential evolution – a simple and efficient heuristic for global optimization over continuous spaces. *J Glob Optim* 11:341–359.
44. Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E (1953) Equations of state calculations by fast computing machines. *J Chem Phys* 21:1087–1092.
45. Hastings WK (1970) Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* 57:97–109.
46. Vrugt JA, et al. (2009) Accelerating Markov chain Monte Carlo simulation by differential evolution with self-adaptive randomized subspace sampling. *Int J Nonlin Sci Numer Simul* 10:271–288.

47. Tuite A, et al. (2011) Cholera epidemic in Haiti, 2010: Using a transmission model to explain spatial spread of disease and identify optimal control interventions. *Ann Intern Med* 154:593–601.
48. Holmberg S, et al. (1984) Foodborne transmission of cholera in Micronesian households. *Lancet* 323:325–328.
49. Weil A, et al. (2009) Clinical outcomes in household contacts of patients with cholera in Bangladesh. *Clin Infect Dis* 49:1473–1479.
50. Righetto L, et al. (2012) The role of aquatic reservoir fluctuations in long-term cholera patterns. *Epidemics* 4:33–42.
51. Lipp EK, Huq A, Colwell RR (2002) Effects of global climate on infectious disease: The cholera model. *Clin Microbiol Rev* 15:757–770.
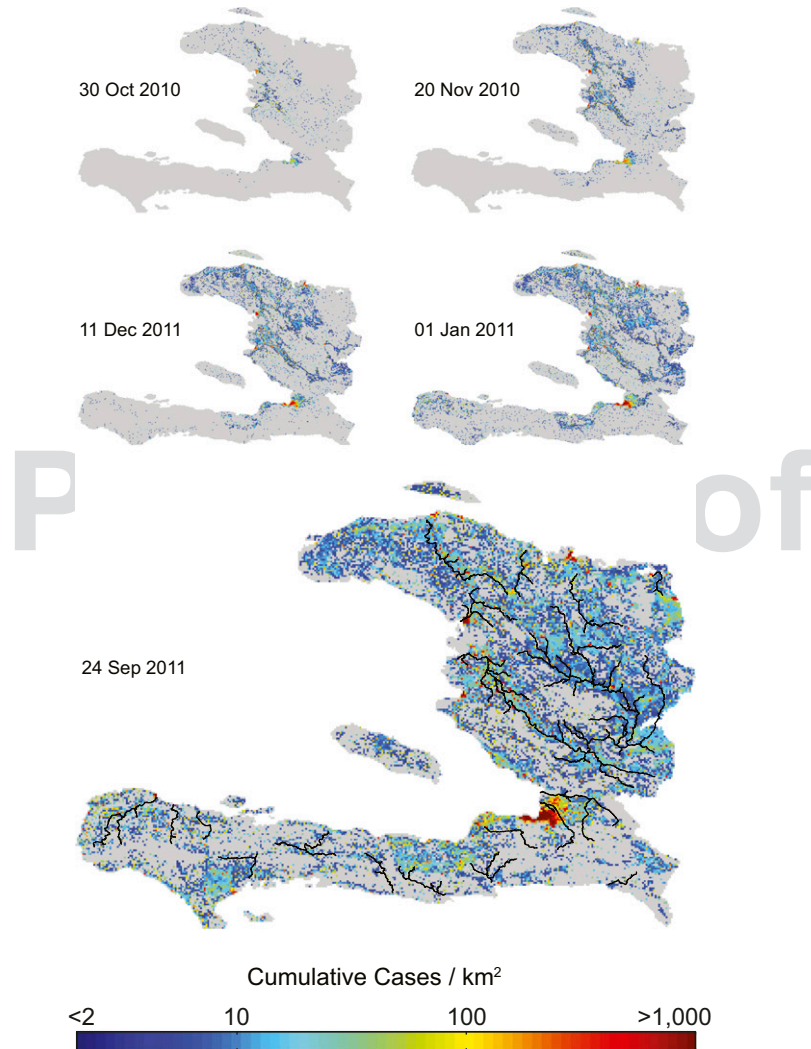
Fig. S1. Map of the spatiotemporal evolution of cumulative reported cholera cases at the country scale (15). Infections are projected at the (pixel) scale of the population database (Fig. 2*C* of the main text) with the assumption of spatially uniform incidence within departments. The early propagation of the disease from the Centre department mainly along the Artibonite River is clearly shown. The subsequent, fast outbreak in the most densely populated region (Port-au-Prince) is also evident.
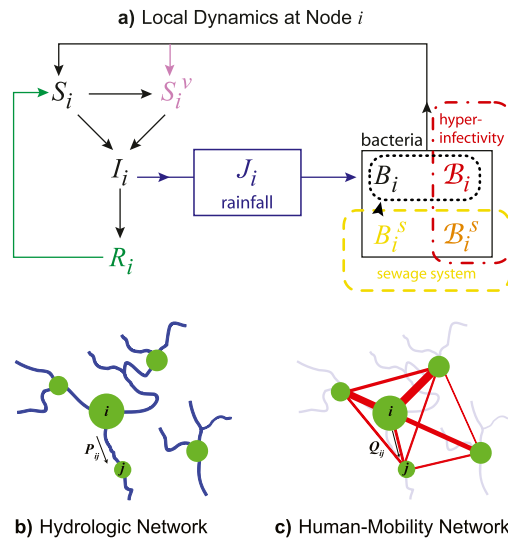
Fig. S2. Elements for a revised model. (*A*) Block diagram of the *i*th site epidemiological model (model S2 with its extensions described in text). (*B* and *C*) Schematic representations of hydrologic (*B*) and human mobility (*C*) networks. Human communities of different sizes are assumed to be concentrated in nodes (green circles). Connections between nodes can be due to either hydrological pathways (i.e., river branches, *B*), which determine active and passive fluvial transport of pathogens, or human displacements (*C*).
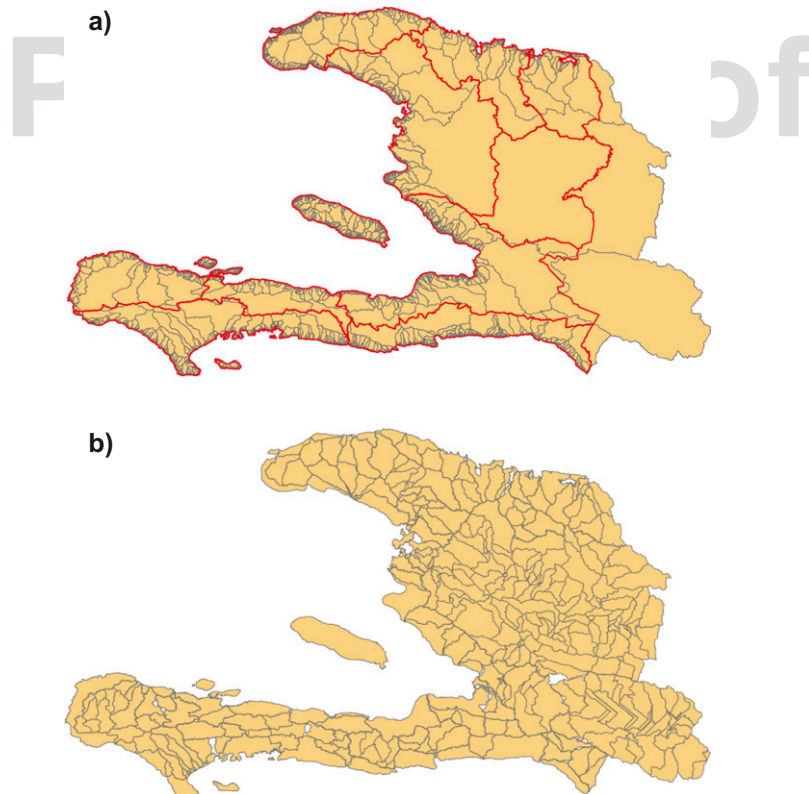
Fig. S3. Spatial distribution of computational nodes. (*A*) A comparative analysis of Haitian river catchments and administrative departments. In this map, the divides of 9,790 natural hydrologic basins from the extracted drainage directions are shown (gray), together with 10 administrative departments of Haiti (delimited in red). Catchments across the Haitian–Dominican Republic boundaries must be considered in their entirety because all parts of a watershed are connected by the river network. A suite of small to very small coastal basins is extracted by the automatic procedure. (*B*) Final map of Haiti after a suitable aggregation of the hydrologic nodes on which the transmission model is based.
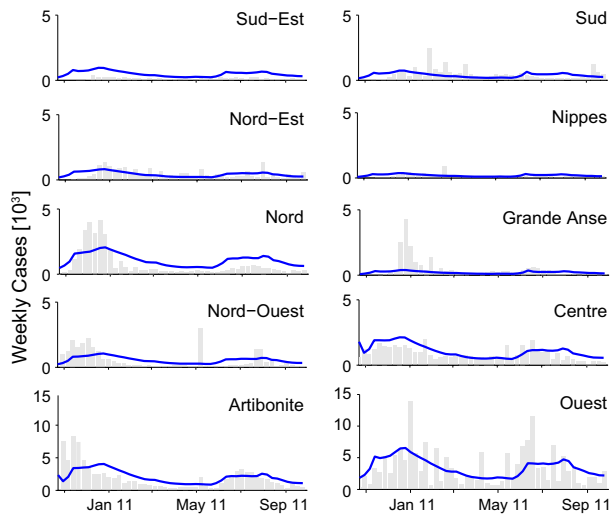
Fig. S4. Comparison of the temporal evolution of reported (gray bars) and simulated (solid blue line) new weekly cases for the Haitian departments. *Bottom, Left* and *Right* (corresponding to the most affected departments), has a different vertical scale.
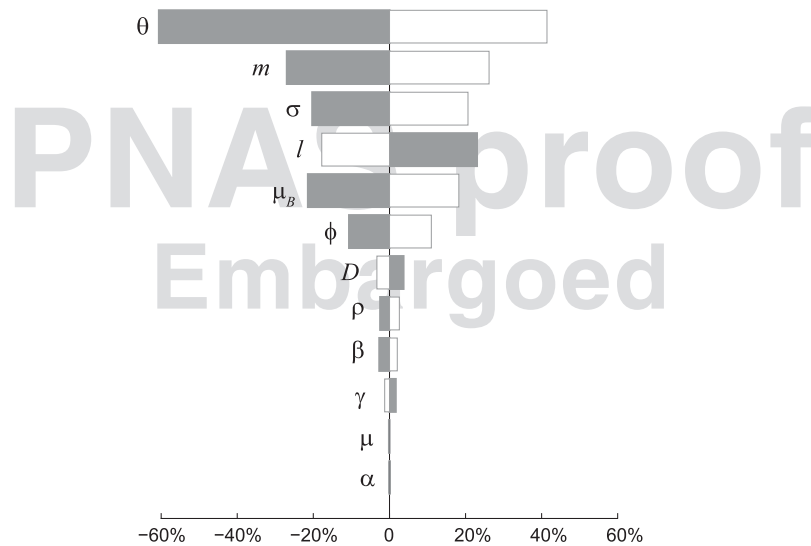

Fig. S5. Effects of parameter variations on the simulation of model S2. Shown are variations of total cholera incidence produced by ±20% variations of the parameters reported in Table S2. Shaded or open bars represent, respectively, negative or positive variations of the relevant parameters.
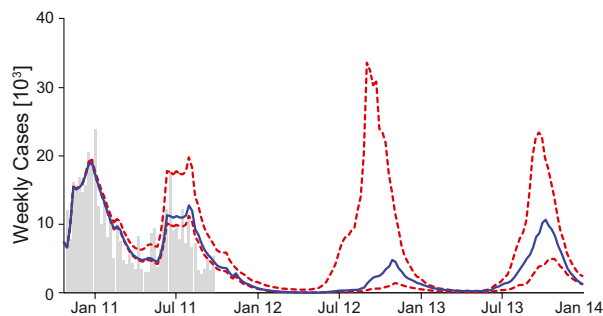

Fig. S6. Effects of the variation of the rate of loss of acquired immunity ρ on the simulation of weekly cases for model S2. The solid line is the simulation reported in Fig. 4 of the main text. The upper (red dashed) line corresponds to the fastest loss of immunity [$1/\rho = 1$ (years)], whereas the lower red dashed line corresponds to $1/\rho = 5$ (years).
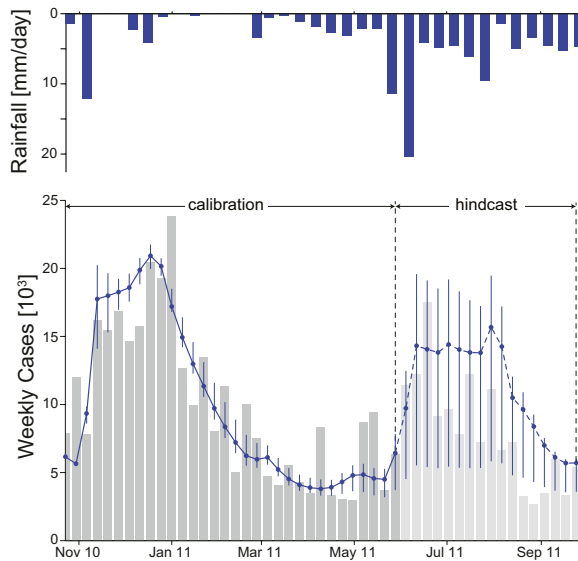
Fig. S7. Predictive ability of model S2. (*Upper*) Daily decadal rainfall intensity, averaged over the entire Haiti region. (*Lower*) Weekly reported cases (15) (gray bars) compared with the simulated incidence pattern (solid and dashed lines) computed by model S2, calibrated on a subset of the data available to date. Data from each department are collected until September 30, 2011. The calibration dataset (dark gray) was extended to the total reported cases available until the end of May, 2011. The validation dataset is in light gray. The dashed line shows the hindcasting provided by model S2, which was run until the end of October, 2011. Error bars are chosen as the 5th and 95th percentiles of the Monte Carlo distribution of simulated incidence obtained by suitably sampling the posterior parameter distribution from the MCMC procedure adopted (*Parameter Calibration* section).
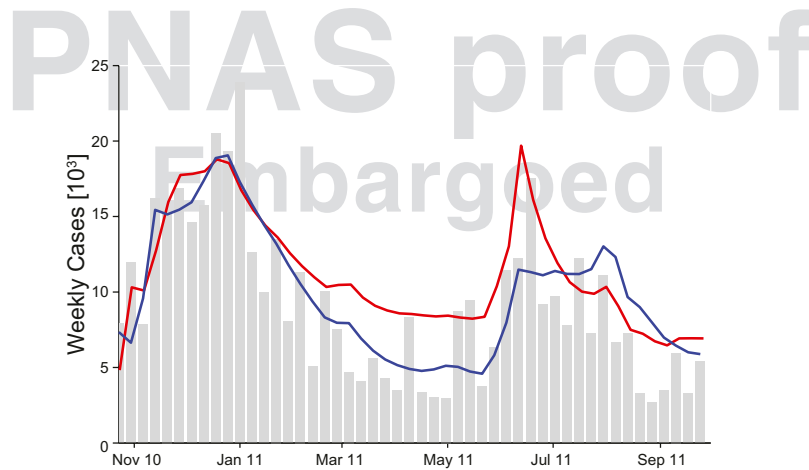

Fig. S8. Comparison of the best-performing models under the assumption of absorbing (blue) and diffusive (red) boundary conditions. The optimal model is model S2 or model S3 for adsorbing or diffusive boundaries, respectively.

Table S1. Number of calibrated parameters and AIC scores for the four tested models

| Model	No. parameters	RSS	AIC	Δ_{AIC}
Model S2	5	$2.52 \cdot 10^{10}$	8,713	—
Model S3	5	$3.06 \cdot 10^{10}$	8,806	93
Model S4	6	$2.58 \cdot 10^{10}$	8,725	12
Model S5	6	$2.78 \cdot 10^{10}$	8,762	49

See text for technical details on the candidate models. The last column shows Akaike differences (with respect to the best-ranked model), which must be >4 for significance (38–40).

Table S2. Estimated (*Upper* section) and fitted (*Lower* section) parameter values and their 5–95% credible intervals, for the best-ranked model (model S2) and relevant units

Parameter	Units	Value	References
μ	d^{-1}	$1/(61 \cdot 365)$	(1)
β	d^{-1}	1.0	(2–5)
ρ	d^{-1}	$1/(3 \cdot 365)$	(6)
γ	d^{-1}	0.20	(2–4, 7, 8)
α	d^{-1}	$4.0 \cdot 10^{-3}$	(9)
μ_B	d^{-1}	0.2	(2–4, 7)
σ	—	0.2	(8, 10–12)
θ	—	0.141 (0.136–0.171)	—
l	d^{-1}	1.83 (1.64–3.20)	—
m	—	0.69 (0.68–0.70)	—
D	km	100 (95–110)	—
φ	d/mm	4.83 (4.71–5.76)$\cdot 10^{-2}$	—

For the estimated parameters see the references in the last column.

1. Central Intelligence Agency (2009) *2009 CIA's World Factbook*. Available at https://www.cia.gov/library/publications/the-world-factbook/index.html. Accessed November 1, 2011.
2. Codeço CT (2001) Endemic and epidemic dynamics of cholera: The role of the aquatic reservoir. *BMC Infect Dis* 1:1.
3. Bertuzzo E, et al. (2008) On the space-time evolution of a cholera epidemic. *Water Resour Res* 44:W01424.
4. Mari L, et al. (2012) Modelling cholera epidemics: The role of waterways, human mobility and sanitation. *J R Soc Interface* 9:376–388.
5. Tuite A, et al. (2011) Cholera epidemic in Haiti, 2010: Using a transmission model to explain spatial spread of disease and identify optimal control interventions. *Ann Intern Med* 154: 593–601.
6. Koelle K, Rodó X, Pascual M, Yunus M, Mostafa G (2005) Refractory periods and climate forcing in cholera dynamics. *Nature* 436:696–700.
7. Bertuzzo E, et al. (2011) Prediction of the spatial evolution and effects of control measures for the unfolding Haiti cholera outbreak. *Geophys Res Lett* 38:L06403.
8. Andrews JR, Basu S (2011) Transmission dynamics and control of cholera in Haiti: An epidemic model. *Lancet* 377:1248–1255.
9. Pan American Health Organization (2011) Haiti cholera outbreak data. Available at http://new.paho.org/hq/images/Atlas_IHR/CholeraHispaniola/atlas.html. Accessed Dec 10, 2011.
10. WHO (2010) *Prevention and Control of Cholera Outbreaks: WHO Policy and Recommendations*. Available at: http://www.who.int/cholera/technical/prevention/control/en/index.html. Accessed November 1, 2011.
11. Longini IM, Jr., et al. (2007) Controlling endemic cholera with oral vaccines. *PLoS Med* 4:e336.
12. Chao DL, Halloran ME, Longini IM, Jr. (2011) Vaccination strategies for epidemic cholera in Haiti with implications for the developing world. *Proc Natl Acad Sci USA* 108:7081–7085.
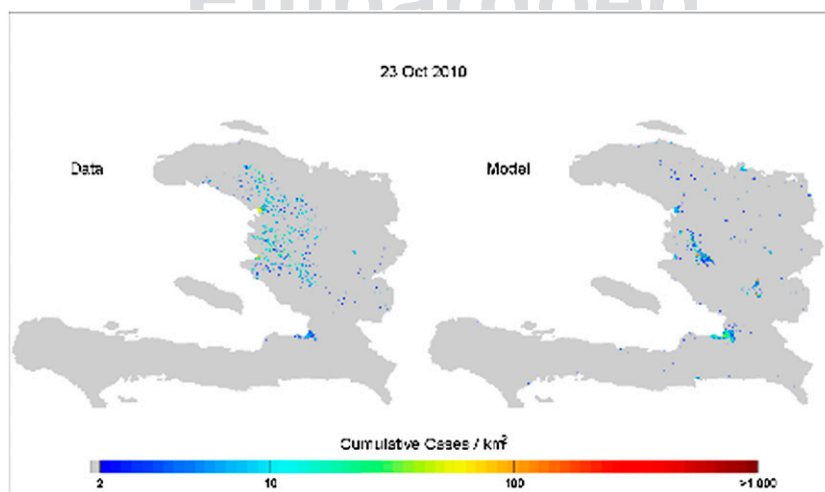
Movie S1. Comparative analysis of data and the best-performing model results. (*Left*) Sequence of maps of the spatiotemporal distribution of cumulative reported cholera cases for Haiti (October 2010, September 2011). Infections are projected at the (pixel) scale of the population database (see Fig. 2*C*) with the assumption of spatially uniform incidence within Departments. The early propagation of the disease from the Centre Department mainly along the Artibonite river is clearly shown. The subsequent fast outbreak experienced in the most densely populated region (Port-au-Prince) is also evident. (*Right*) Model results plotted in the same spatial and temporal scales of the observational data. Infections are projected at the (pixel) scale of the population database with the assumption of spatially uniform incidence within modelled local communities (watersheds, Fig. S3*B*). Besides a slight overestimation of the infections in the early phases of the outbreak in the southern Departments, the capabilities of the modeling tool are deemed remarkable.

Movie S1