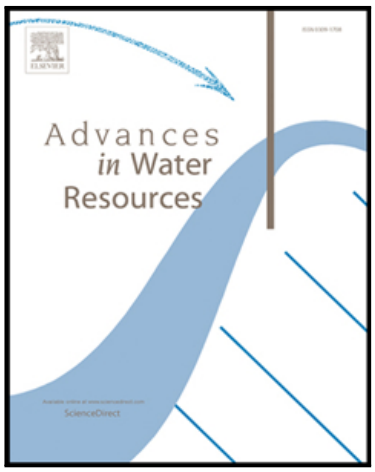

Seasonality in cholera dynamics: A rainfall-driven model explains the wide range of patterns in endemic areas

Theo Baracchini, Aaron A. King, Menno J. Bouma, Xavier Rodó, Enrico Bertuzzo, Mercedes Pascual

PII: S0309-1708(16)30705-9
DOI: 10.1016/j.advwatres.2016.11.012
Reference: ADWR 2738

To appear in: *Advances in Water Resources*

Received date: 10 May 2016
Revised date: 23 November 2016
Accepted date: 25 November 2016

Please cite this article as: Theo Baracchini, Aaron A. King, Menno J. Bouma, Xavier Rodó, Enrico Bertuzzo, Mercedes Pascual, Seasonality in cholera dynamics: A rainfall-driven model explains the wide range of patterns in endemic areas, *Advances in Water Resources* (2016), doi: 10.1016/j.advwatres.2016.11.012

Highlights

- An SIR-like model driven only by rainfall and temperature is proposed.
- We consider additional state variables for the pathogens and the volume of water.
- Rainfall and temperature explain the seasonal an interannual dynamics of cholera.
- The hydrological regime and the water reservoir influence the disease dynamics.

Seasonality in cholera dynamics: A rainfall-driven model explains the wide range of patterns in endemic areas

Theo Baracchini[a], Aaron A. King[b], Menno J. Bouma[c,d], Xavier Rodó[d], Enrico Bertuzzo[e], Mercedes Pascual[f,*]

[a]*School of Architecture, Civil and Environmental Engineering, Ecole Polytechinque Fédérale de Lausanne, Switzerland*
[b]*Department of Ecology and Evolutionary Biology, University of Michigan, MI, United States of America*
[c]*Faculty of Public Health and Policy, London School of Hygiene and Tropical Medicine, United Kingdom*
[d]*Catalan Institute of Climate Sciences, Spain*
[e]*Department of Environmental Sciences, Informatics and Statistics, University Ca' Foscari, Venice, Italy*
[f]*Department of Ecology and Evolution, University of Chicago, IL, United States of America*

Abstract

Seasonal patterns in cholera dynamics exhibit pronounced variability across geographical regions, showing single or multiple peaks at different times of the year. Although multiple hypotheses related to local climate variables have been proposed, an understanding of this seasonal variation remains incomplete. The historical Bengal region, which encompasses the full range of cholera's seasonality observed worldwide, provides a unique opportunity to gain insights on underlying environmental drivers. Here, we propose a mechanistic, rainfall-temperature driven, stochastic epidemiological model which explicitly accounts for the fluctuations of the aquatic reservoir, and analyze

*Corresponding author
Email address: pascualmm@uchicago.edu (Mercedes Pascual)

with this model the historical dataset of cholera mortality in the Bengal region. Parameters are inferred with a recently developed sequential Monte Carlo method for likelihood maximization in partially observed Markov processes. Results indicate that the hydrological regime is a major driver of the seasonal dynamics of cholera. Rainfall tends to buffer the propagation of the disease in wet regions due to the longer residence times of water in the environment and an associated dilution effect, whereas it enhances cholera resurgence in dry regions. Moreover, the dynamics of the environmental water reservoir determine whether the seasonality is unimodal or bimodal, as well as its phase relative to the monsoon. Thus, the full range of seasonal patterns can be explained based solely on the local variation of rainfall and temperature. Given the close connection between cholera seasonality and environmental conditions, a deeper understanding of the underlying mechanisms would allow the better management and planning of public health policies with respect to climate variability and climate change.

Keywords: Infectious disease, modelling, cholera, seasonality, endemic, historical dataset, Bengal

1. Introduction

Although diarrheal diseases are preventable through suitable sanitary conditions, education and hygiene [46], they remain the second leading cause of mortality and are responsible for 20% of the deaths among children under 5 years of age [10]. In particular, although the treatment of cholera today is relatively easy and affordable, the disease remains a public health threat

across the globe, and an endemic problem in the estuary of the Ganges, its native habitat.

A clear explanation for the diverse seasonal patterns of cholera outbreaks in endemic areas has remained elusive [37]. Previous studies addressing the role of climate drivers in disease dynamics have focused on interannual variability while prescribing the intra-annual seasonality [30]. The few proposed explanations for seasonality have relied on complex environmental interactions that vary with spatial location, involving regional hydrological models [6], river discharge [28, 3], sea surface temperature [17, 7], and plankton blooms [2, 15, 28]. No simple and unified mechanism based on *local* climate variables has been considered that can account for different seasonalities within a region and across different regions of the world [20]. A better understanding of seasonality in relation to climate variables would provide a basis to better understand the effects of climate variability and climate change in general.

Bangladesh and North-East India are endemic regions for cholera that harbor the causing pathogen in the environment, the bacterium *V. cholerae*. The interplay of high population density, seasonal hydroclimatology, floodplain geography and coastal ecology makes this region particularly vulnerable to periodic outbreaks [3]. This region encompasses the most heterogeneous temporal patterns of endemic cholera dynamics worldwide, ranging from a single annual peak during or preceding the rainy season to a double peak in the pre-and post-monsoon periods [34, 43] (Fig. 1). These patterns of recurrent outbreaks are still prevalent today in North-East India and Bangladesh, as well as in various other regions of the world [23, 16, 20, 22]. If they have a

common explanation based on hydrology and local climate conditions, their understanding can provide insights into the causes of endemicity and suggest modifications of the environment to mitigate or eliminate disease burden locally.

Through the analysis of a unique historical dataset containing 40 years of monthly meteorological, demographic and epidemiological records, we propose a process-based model for the population dynamics of cholera driven by local rainfall and temperature, and show that this model is able to capture the full range of seasonal patterns of this large estuarine region. The transmission model explicitly accounts for volume fluctuations of the aquatic reservoir and for the environmental bacterial concentration.

Although the crucial role played by the aquatic reservoir in the population dynamics of the disease has already been widely assessed (e.g. [15, 37, 3, 34, 15, 38, 36, 7, 1, 27, 22, 4]), this work provides the first investigation of the full range of seasonal incidence patterns with a simple process-based approach. The model combines the two opposing views on the dominant drivers behind cholera epidemiological patterns [37]: that of the "localists" supporting a dominant role of the environment and of an environmental reservoir in transmission, and that of the "contagionists", emphasizing human-to-human transmission and sanitary conditions. Mathematical models confronted to time series data provide a useful tool to examine different hypotheses concerning the climatic influences on disease dynamics, including the timing and causes of seasonal patterns [5, 15]. They further provide a basis for climate-based early warning systems, and for evaluating mitigation strategies for environmentally driven infectious diseases [38].

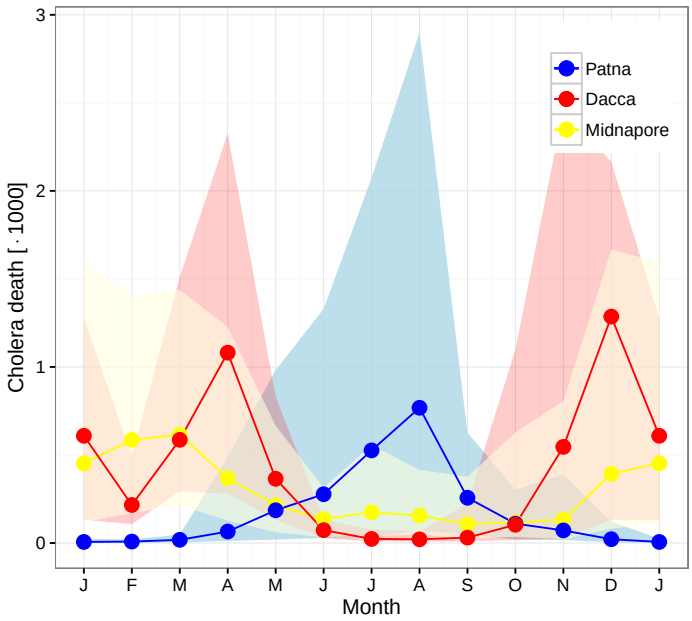

Figure 1: Three distinct cholera patterns representative of those found in the historical dataset as a whole: (1) a double peak, respectively pre- and post-monsoon, in the district of Dhaka; (2) a single wide peak from post- to pre-monsoon in the coastal district of Midnapore; and (3) a single annual peak during the monsoon in the drier north-western district of Patna. The lines correspond to the median of the monthly values, and the shaded areas, to the envelope of the data (bounded by the 10% and 90% quantiles) of the yearly values.

2. Material and Methods

2.1. The historical dataset

The former Bengal region constitutes the eastern part of the Indian sub-continent, and corresponds to the Indian state of West Bengal and the nation of Bangladesh today. It comprises the world's largest delta and is the second most densely populated region around the globe. Besides Bengal, the study

area also includes the Indian states of Assam (north-east), Bihar (north-west), Meghalaya and Tripura (east). Except for the dryer and mountainous north-western parts, this tropical and humid region is a fertile alluvial plain. The low elevation of the delta (Fig. 2) allows inland intrusions of salt water during low river discharges.

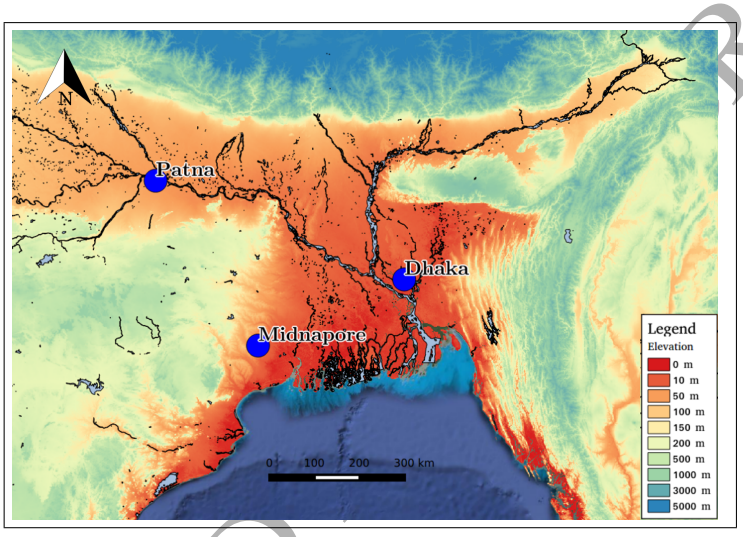

Figure 2: Elevation map of Bengal and its water bodies. Circles indicate the three representative districts presented in this paper.

An extensive data set on cholera deaths for 155 districts in 7 provinces from 1891 through 1941 was collected from the records of the sanitary commissioners of the former British East Indian province of Bengal. A decadal population census is also available for the same period. The published results for 1891, 1901, 1911, 1921, 1931 and 1941 were linearly interpolated after corrections for administrative changes. Monthly temperature and rainfall data were also obtained from the India Weather Review, Annual Reports of the Meteorological Department, Government of India, at the level of districts. A

monthly average was used for each location. Figure 3 illustrates the seasonality of these environmental variables for the three representative districts that are the focus of this study: Dhaka, Midnapore and Patna. (Although the methodology has been successfully tested for other districts, this work focuses on the three aforementioned ones whose representative seasonalities correspond to those of their respective regions, and cover the full range of observed seasonal patterns within the data set as a whole).

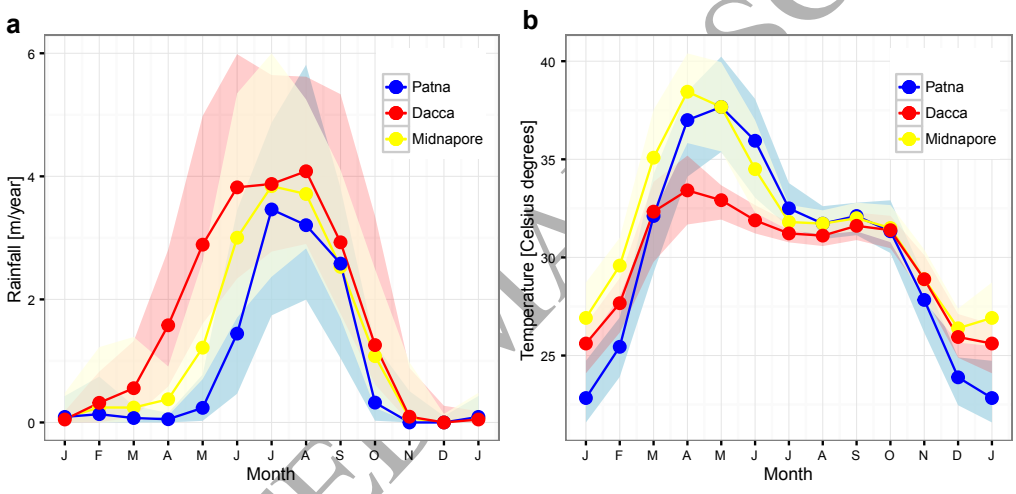

Figure 3: Seasonality of rainfall (a) and temperature (b) for three representative districts. The lines represent the median of the monthly variables across different years, whereas the shaded areas represent the envelope of the data (bounded by the 10% and 90% quantiles) of the yearly values.

2.2. Models

In this study we develop a non-linear, stochastic epidemiological model for cholera dynamics that builds upon previous efforts [30, 40, 6, 41]. The system is an expanded SIR-like model (for Susceptible-Infected-Recovered

classes of individuals) with 7 compartments (Fig. 4). The population of recovered individuals is split into 3 compartments, to change the distribution of the duration of immunity from the typical exponential of models with a single recovery compartment, to a more realistic gamma distribution (with a characteristic duration or mode). This formulation provides a more flexible and realistic biological assumption, since an exponential distribution considers an immune duration independent from the time since an individual has recovered [35]. Three compartments provide a proper trade-off between allowing the implementation of a gamma distribution while incurring a reasonable computational cost. The model has two additional state variables, not present in standard SIR formulations, for the population of pathogens in the aquatic environment and for the volume of the aquatic reservoir, respectively.

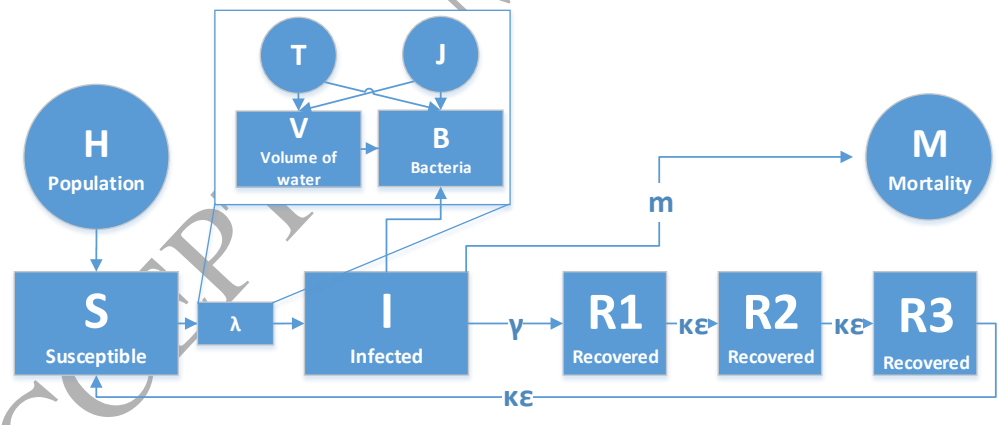

Figure 4: Diagram of the compartmental model. The rectangles correspond to the states variables and the circles, to observations that enter as environmental covariates (temperature and rainfall) or to the measurement variable (here, deaths). For simplicity, natural deaths are not included here in the diagram but are taken into account in the model.

Figure 4 depicts a diagram of this compartmental model. In this diagram, $S(t)$ denotes the number of susceptible individuals at time t, $I(t)$, the number of infections, and $R_1(t)$, $R_2(t)$, $R_3(t)$ correspond to the multiple stages of recovery. $B(t)$ gives the bacterial abundance in the aquatic reservoir, and $V(t)$, the volume of this reservoir per unit area. In addition, $H(t)$ stands for the human population entering the system (births), $M(t)$, for the individuals dying from cholera, $\lambda(t)$, for the force of infection, $T(t)$, for the local temperature, and $J(t)$, for the local rainfall. The diagram of Figure 4 can be written as the following set of coupled stochastic differential equations:

$$\frac{dS}{dt} = \kappa\epsilon R_3 + \mu H(t) + \frac{dH}{dt}(t) - (\lambda(t) + \mu)S \quad (1)$$

$$\frac{dI}{dt} = \lambda(t)S - (\gamma + m + \mu)I \quad (2)$$

$$\frac{dR_1}{dt} = \gamma I - (\kappa\epsilon + \mu)R_1 \quad (3)$$

$$\frac{dR_2}{dt} = \kappa\epsilon R_1 - (\kappa\epsilon + \mu)R_2 \quad (4)$$

$$\frac{dR_3}{dt} = \kappa\epsilon R_2 - (\kappa\epsilon + \mu)R_3 \quad (5)$$

$$\frac{dV}{dt} = J(t) - ET(T, V) - f(V) \cdot V \quad (6)$$

$$\frac{dB}{dt} = -\mu_B(T)B + p(t)[1 + \phi \cdot J(t)]I \cdot \xi(t) - f(V)B \quad (7)$$

where μ [s^{-1}] denotes the birth and mortality rate ($1/\mu$ is the life expectancy, fixed to 50 years), κ [-], the number of recovered compartments (here equal to 3), ϵ [s^{-1}], the rate of immunity loss, $dH/dt(t)$, the observed changes in population size, m [s^{-1}], the mortality rate due to disease, and γ [s^{-1}], the rate

of recovery from infection. The force of infection $\lambda(t)$ [s^{-1}] depends on the exposure rate β [s^{-1}] and on the environmental concentration of pathogens through a saturating function [15]:

$$\lambda(t) = \beta \frac{\frac{B(t)}{V(t)A}}{\frac{B(t)}{V(t)A} + K}, \tag{8}$$

where A [m^2] is the geographical area in contact with the human population, and K is the half saturation concentration [#bacteria m^{-3}].

The evolution of the volume of water per unit area V [m] is driven by the hydrological cycle (Eq. 6), namely by rainfall J [ms^{-1}], evapotranspiration ET [ms^{-1}] and drainage. Raw monthly rainfall data has been employed and interpolated to satisfy the daily time step of the model. The potential evapotranspiration (ET_p) is computed according to a re-calibrated Blaney-Criddle formula [9, 45, 44] based on historical temperature records. This modified form incorporates the new multiplicative and additive coefficients (0.35 and 2.5 respectively, in place of 0.46 and 8) re-calibrated to region-specific values by Sperna Weiland *et al.* [44]. This formula corresponds to the potential quantity of water that can be evapotranspired assuming that plants are in optimal conditions. When water availability is a limiting factor, the actual evapotranspiration ET decreases according to the following equation:

$$ET(T, V) = \begin{cases} ET_p(T) \cdot \frac{V(t)}{V_t} & if \quad V(t) < V_t \\ ET_p(T) & else, \end{cases} \tag{9}$$

with V_t a calibrated parameter acting as a threshold for potential evapotranspiration. This formulation used here allows regions with different environmental conditions to exhibit different evapotranspiration behaviors. Drainage corresponds to the flux of water leaving the area and it depends

on many factors, including soil type, topography, and the structure of the river network. Here it is modelled as a function of the volume V through a calibrated 3-parameter function describing the drainage rate $f(V)$:

$$f(V) = \delta \frac{V(t)^{\alpha}}{V(t)^{\alpha} + \tilde{V}^{\alpha}}. \tag{10}$$

The three parameters δ [s^{-1}], α [-] and $\tilde{V}$ [m] flexibly change the behavior of the drainage function and allow different temporal scales of the responses to an increasing water volume (e.g. delayed or immediate drainage). These different responses allow the representation of the hydrological characteristics of different areas (e.g. mountainous versus estuarine).

The evolution of the environmental pool of bacteria (Eq. 7) results from a balance between contamination from infected individuals, pathogen death and drainage. The net death rate μ_B [s^{-1}] is assumed to be linearly dependent on temperature:

$$\mu_B(T) = \bar{\mu_B}(1 - \varepsilon \frac{T - \bar{T}}{T_{max} - \bar{T}}), \tag{11}$$

where the temperature is in degree Celsius. The parameter $\bar{\mu_B}$ [s^{-1}] denotes the average death rate of the bacterium, ε [-], the dependency on temperature, and $\bar{T}$ [°C] and T_{max} [°C] correspond respectively to the mean and maximum temperature of the studied area over the 40 years. When ε is larger than one, the death rate can become negative which describes the possible reproduction of bacteria in the environment at high temperature. The input from infected individuals is modeled through the term $p(t)[1 + \phi \cdot J(t)]I \cdot \xi(t)$ where $p(t)$ is the *per capita* rate at which infected individuals shed bacteria

that contaminate the environmental reservoir. As the cholera time series indicate a long-term decrease in the number of deaths in some districts, we assume that sanitary conditions, represented by the parameter p [-], can potentially change and model this process through an exponential function $p(t) = p_0 e^{-d(t-\bar{t})}$, where $\bar{t}$ [s] corresponds to the middle of the simulation period and p_0 [-] and d [s^{-1}] are two calibration parameters. The contamination process is assumed to be enhanced by rainfall which can wash out contaminated sites and deliver bacteria to the water reservoir [41, 18]. This input is accounted for by the parameter ϕ [sm^{-1}] [41]. Finally, $\xi(t)$ is the process noise of the model, with $\xi(t) = \frac{dW}{dt}$ and $dW \sim \Gamma_{WhiteNoise}(\mu_W, dt)$ (μ_W equals the non-zero expected value, fixed here to 0.015 after an initial calibration of the model to the different districts). The last term, $f(V) \cdot B$, accounts for the bacteria within the water reservoir leaving the area through drainage.

By normalizing bacterial counts as $B^* = B/(KA)$, three parameters (p_0, K and A) are grouped into a single one, namely the ratio $\theta_0 = p_0/(KA)$, which reduces the number of parameters to be estimated. It follows that Eq. 7 becomes $\frac{dB^*}{dt} = -\mu_B(T)B^* + \theta[1 + \phi \cdot J(t)]I \cdot \xi(t) - f(V)B^*$, and that the force of infection is given by $\lambda(t) = \beta \frac{B^*(t)/V(t)}{B^*(t)/V(t)+1}$.

The measurement model relates the deaths generated by the process model (the above-described differential equations for the 7 compartments SIR-like model), to those observed in the data, y_n, and allows one to compute a likelihood for the model given the observations. In a monthly time step, the number of new cholera deaths in the n^{th} interval is $M_n = m \int_{(n-1)/12}^{n/12} I(t)dt$ (with t in years). The log-likelihood of each data point y_n is obtained through

a negative binomial distribution as:

$$\log(\mathcal{L}) = log(NegBinom(y_n; \rho M_n, \frac{1}{overdisp^2})), \quad (12)$$

with mean ρM_n (where ρ [-] is the reporting rate) and variance $(\rho M_n/overdisp)^2$, with *overdisp* [-] a dispersion parameter. The negative binomial distribution allows more overdispersion than that of the Poisson distribution.

2.3. Parameter inference

Parameter inference for nonlinear systems of stochastic differential equations has recently been facilitated by the development of methods for maximizing the likelihood via Iterated Filtering (MIF) [26, 25]. This frequentist method is based on a particle filter approach developed by Ionides *et al.* [25], which allows the estimation of parameters via simulation of the model (via sequential Monte Carlo). Iterating filtering allows for models with measurement error, non-stationarity, irregular sampling intervals, and the inclusion of covariates. It also allows for hidden variables, that is variables for which observations are unavailable, such as the number of susceptible individuals. Moreover, the method has the advantage of focusing adaptively on favorable regions of the state-space, and can cope with a broad range of state and noise distributions. Iterated filtering is implemented in the R statistical open-source computing environment within the package *POMP* [29]. The stochastic equations were integrated using the Euler-Maruyama algorithm. For detailed description of the fitting algorithm see [25], and for a previous application and explanation of the algorithm in the context of a climate-driven model see [33]. In this study, 15 unknown parameters are estimated using the 40 years time series of reported cholera deaths.

3. Results

3.1. Parameters estimation

Table 1 provides a summary of the fitted parameters for each district obtained after an initial broad search and an additional local refining of this search.

	Dhaka	Patna	Midnapore
V_t [m]	0.49	5.66	4.59
α [-]	19.85	4.91	14.96
$\tilde{V}$ [m]	1.87	1.27	3.42
δ [y^{-1}]	5.04	474.35	553.43
$\bar{\mu_B}$ [y^{-1}]	317.49	107.59	368.20
ε [-]	0.16	0.02	0.02
θ_0 [y^{-1}]	0.0036	0.0036	0.0877
ϕ [y/m]	0.0148	0.1646	0.0298
$1/\epsilon$ [y]	5.43	6.73	4.88
$1/\gamma$ [d]	3.01	1.57	1.20
d [y^{-1}]	0.0000	0.0109	0.0072
$overdisp$ [-]	0.82	0.79	0.52
ρ [-]	0.10	0.40	0.71
β [y^{-1}]	87.29	2.60	2.15
m [y^{-1}]	23.70	43.43	42.12

Table 1: Maximum likelihood parameters for each district.

3.2. Seasonality

Figure 5a shows the median monthly cholera deaths for 40 years of data and the corresponding simulation from 1900 to 1940 (1890 to 1930 for Midnapore). The seasonality exhibits the typical bi-modal pattern of cholera observed in the district of Dhaka. The two pre- and post-monsoon peaks fall respectively in spring and autumn, with corresponding maxima in April and December, as expected for the Classical biotype of the pathogen (the current El Tor biotype emerged later in the region). The seasonality is well captured by the model, with the peaks in phase with the data. The median of the simulations also compares well with that of the data, except for a slight underestimation of the fall peak. Although the envelope of the model does include this variability, it overestimates the winter and spring infections. The absence of reported deaths during summer is well captured by the model.

The single annual peak pattern observed during the monsoon in the north-western and drier region of Patna, is also captured by the model (Fig. 5b). The medians overlap well with the data, except for the month of August, when the simulation underestimates the observed deaths. The envelope of the data has a more negatively skewed distribution, with a sharp decrease after August, whereas the one of the model is more symmetrical. Finally, when no cases are observed between January and March, some sporadic deaths are found in the simulations based on its envelope.

The coastal area of Midnapore shows another interesting pattern, a single wider peak in the late winter-early spring. Once again, the dynamics are captured by the model. Generally, a slight underestimation is observed in the median and in the envelope in late autumn and early winter.

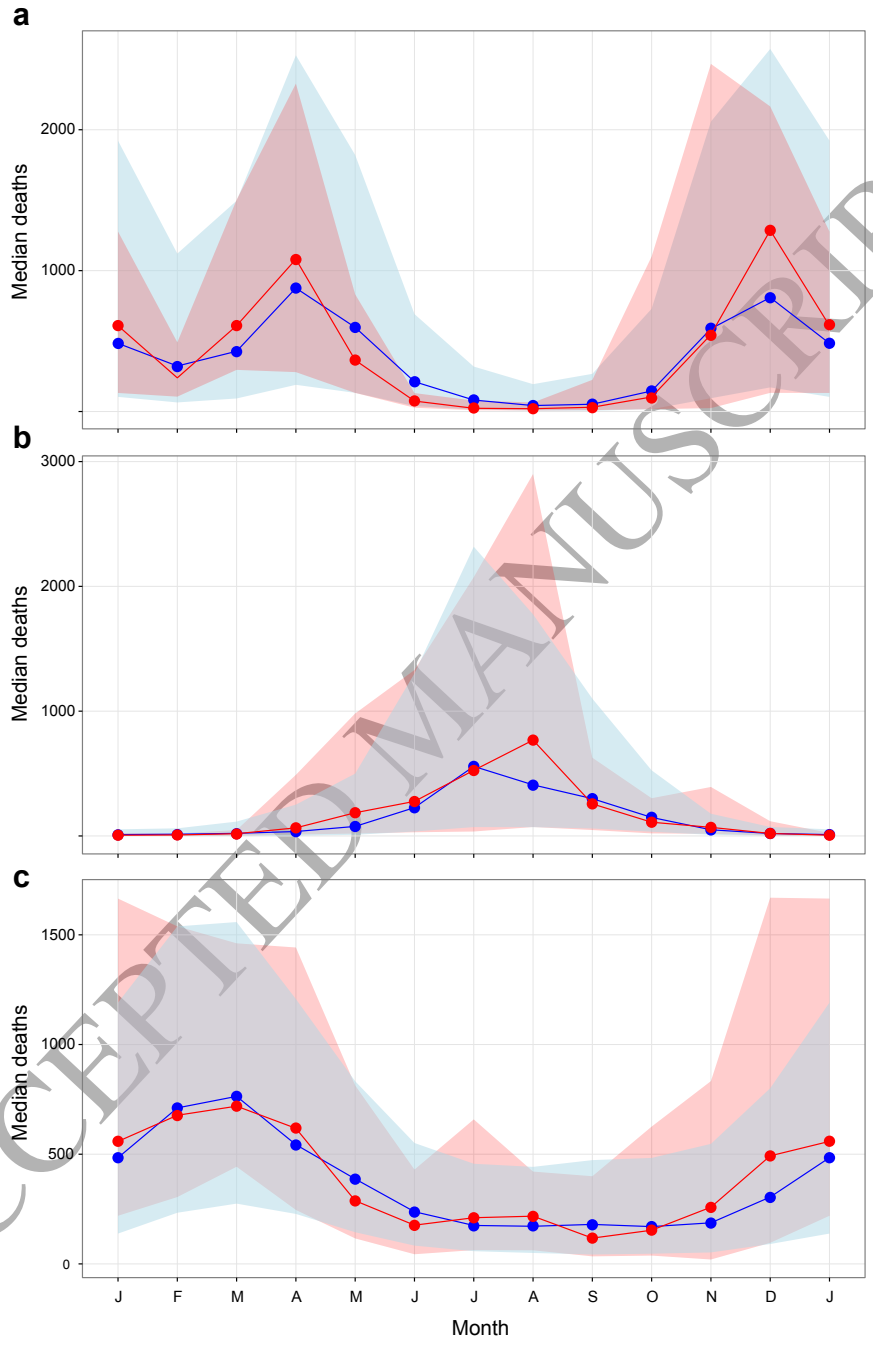

Figure 5: Cholera seasonality for the districts of Dhaka (a), Patna (b) and Midnapore (c). The median (solid lines) and envelope (shaded areas, monthly median of the 10% and 90% quantiles of the 250 simulated distributions) are shown for cholera mortality, to compare observations (in red) to model simulations (the result of 250 runs with the MLE parameter set) (blue).

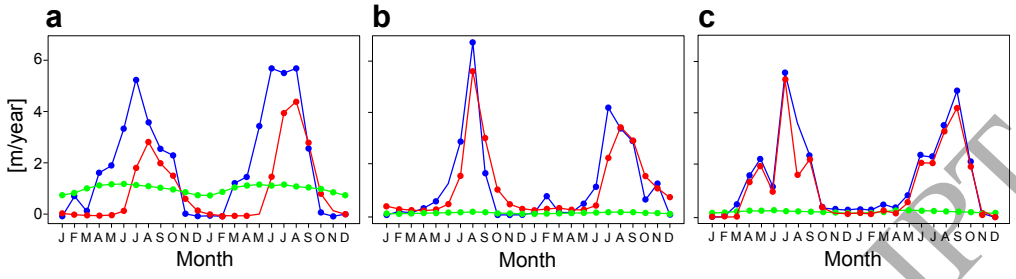

Figure 6: Hydrological fluxes for Dhaka (a), Patna (b) and Midnapore (c) for 2 representative years (respectively 1917-1918, 1914-1915 and 1902-1903). Rainfall (blue), drainage f(V) (red), and evapotranspiration (green).

Figure 6 shows the different fluxes controlling the water reservoir state (rainfall, evapotranspiration, and drainage) over a period of 2 years. The model suggests a much lower evapotranspiration in Patna than in the estuarine region of Dhaka. Interestingly, the drainage has a faster response and a behaviour that closely tracks rainfall in the dry-northern district, whereas a delay is present in the wet-southern areas together with lower values. Midnapore shows an intermediate pattern with low evapotranspiration, a fast drainage increase after a rainfall event, followed by a faster decay than that in Patna.

3.3. Interannual variability

Figure 7 compares the time series of the data to those from the simulations. (We note that these values do not represent next step prediction but the result of a set of 40-year simulations from estimated initial conditions). For the districts of Dhaka and Patna, the median of the model captures partially but not fully the interannual variation. Dhaka is more subject to frequent large outbreaks, and only a few of these are fully captured by the

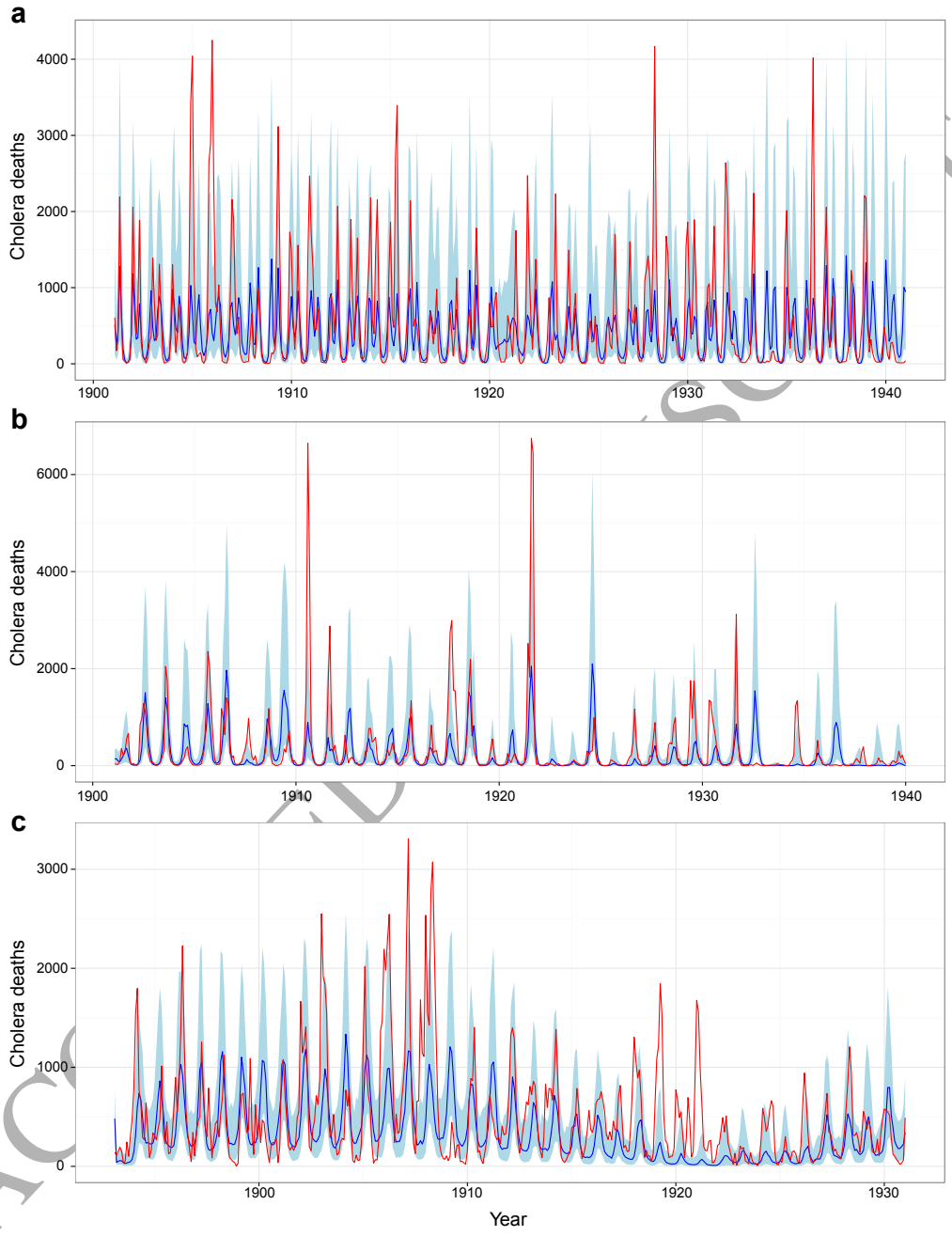

Figure 7: Time-series of the median of 250 simulations (blue) with the envelope (light blue, bounded by the 10% and 90% quantiles of the 250 simulated distributions), together with the cholera mortality data (red) for Dhaka (a), Patna (b) and Midnapore (c).

median of the simulations. Patna exhibits less frequent violent epidemics, although two of them, in 1910 and 1921, are of particular intensity (exceeding 6500 monthly cholera deaths). Almost every important outbreak is within the envelope of the model, suggesting that the model is capable of producing those behaviors, but that stochasticity determines their exact timing and the resulting variation results in the lower median. Finally, it is worth mentioning that although average mortality appears more constant over time for Dhaka, a slight downward trend is observed for Patna as reflected in parameter d of Table 1. For Midnapore, the results are less clear, as important outbreaks between 1900 and 1910 together with a temporary phase of milder infections in 1922-1925 give an impression of a downward trend in time. This trend is reflected in parameter d, which is slightly positive.

To assess quantitatively the interannual variation of the data, Singular Spectrum Analysis (SSA), a statistical method decomposing the time series into (orthogonal) principal components, was used to remove the seasonal component of the time series, to extract the interannual variation ([8]; see [39, 42] for examples of applications in epidemiology in the context of climate variability). Subsequent Fourier analysis of the interannual component identified dominant periods of the anomalies in reported deaths of 4.2 and 7.8 years for Dhaka, 4.2 and 6 years for Patna, and 7.8 years for Midnapore. No evident link with the periodicity of the anomalies in the rainfall or temperature could be detected.

4. Discussion

The proposed model explains the first seasonal outbreak in Dhaka by the increase in temperature and associated drier conditions of spring, which in turn increase pathogen concentration in the aquatic reservoir. This finding offers an alternative hypothesis to that of Akanda *et al.* [2], who propose that the first peak is mainly modulated by coastal hydroclimatic conditions (salinity, plankton abundance) and the intrusion of salt water inland, during periods of low river discharge (spring). Here, hydrological conditions alone suffice to explain this characteristic pattern in Dhaka, and the full variation of seasonalities across the extensive Bengal region.

Moreover, for Dhaka, the important summer rains would induce a dilution effect, presumably lowering incidence, as suggested by Emch *et al.* [21] for cholera in Bangladesh. The peak stream flow observed in June creates important inundations spreading the pathogen across the landscape. Given the presence of water bodies in this estuarine region and the low drainage suggested by the model, conditions of large scale contamination would be expected, with the bacterial population thriving locally without being washed out from the area. This persistence would set the stage for a new outbreak once the rainy season is over, the concentration of pathogen increases, and the susceptible pool is replenished. This explanation is in accordance with the more complex hypotheses in the literature (e.g. [28]) relying on important discharges during the monsoon, lower salinity levels and pH, and high nutrient loads of the water sources, which in turn favor plankton blooms and bacterial growth. Finally the decline in cholera infection observed in January and February is found to be temperature related, as suggested by Pascual *et*

al. [37]. However, the bacteria can survive through the winter in the aquatic reservoir and be ready to initiate a new outbreak the following year [24, 14].

Rainfall in the model is found to buffer the propagation of the disease in wet regions due to a dilution effect, while enhancing cholera resurgence in dry regions. The more important drainage rate found in the dryer district of Patna suggests higher discharges, possibly leading to the breakdown of sanitary conditions and the boosting of transmission. This completely opposite pattern to that of Dhaka is consistent with the observation that "overall water levels matter and appear to determine whether the effect of rainfall is positive or negative" [37]. It further emphasizes the importance of the hydrological regime and of the water reservoir to cholera dynamics.

The model is also able to capture some of the interannual variability of cholera based on rainfall and temperature. Although particularly explosive outbreaks are above the median of the simulations, these anomalies do fall within the envelopes of the model.

The results of Singular Spectrum Analysis suggest a role of stochasticity in explaining the timing of these abnormally large outbreaks, at the same time that they also indicate the existence of regularity in the form of some detected periodicity above one year. Indeed the periodicity found in the anomalies (of the interannual component) of reported deaths implies an interplay with other climatic or demographic events. Interestingly, the periodicity roughly corresponds to the dominant frequency of El Niño (about 1/4 years^{-1}), the most important driver of interannual climate variability on a global scale. This is in accordance with other findings [39, 42], where the authors conclude that cholera dynamics are associated with a remote forcing

by ENSO (El Niño Southern Oscillation). For example, after the warming of the Pacific, changes in cloud cover, evaporation, and increased heat flux can be observed a few months later in the Bay of Bengal, thus linking general climate to local variables impacting cholera [31]. Other studies also found a link between ENSO and the regional climate of Bangladesh by studying changes in the monsoon circulation over the area, their associated precipitations changes and the possible implications for cholera incidence [11, 12, 13]. Consistent with our findings, an influence of ENSO on cholera would have been weaker than in more recent decades, as it was previously described as non-stationary in time, and was mainly observed for the more recent decades and between 1900 and 1940 exclusively for the spring-peak (February to June) [7]. Nevertheless, ENSO would have exerted an influence on the climate of the Indian Ocean during the colonial period.

Patna shows a decline of both the reported and simulated cases over time. Although several hypotheses can be formulated to explain this long term trend, not much can be done to assess them. One explanation would be a change of the reporting rate over time (with changes in administration, demography, etc.). Cholera mortality rates in hospitals, for example, are known to have decreased over this period [32]. Also, an improvement of sanitation in Bengal, reducing cholera prevalence and deaths, is the most likely explanation. Regardless, the long-term trend is well captured by the model through the parameter d (Eq. 7).

Besides its application to the three chosen districts, the approach has general applicability to other locations within the larger region. Other districts were fitted successfully with the same model, including Chittagong and

Parganas. For the coastal district of Parganas, we obtained consistent results to those of its close neighbour Midnapore, whose dynamics are also similar (Parganas exhibits some differences, including a lower count of summer infections and a strong decreasing trend over time). For Northern districts, such as Lakhimpur, their strongly epidemic dynamics with intermittent outbreaks were only partially captured by the model. This kind of district would require an extension of the model that explicitly incorporates extinctions and re-invasions.

Importantly the seasonal patterns considered here are still observed today in the Bengal region, in Bangladesh and North-East India [23, 16, 20], and in other regions of the world, as described in the global review of seasonal cholera patterns for the period between 1974 and 2005 by Emch *et al.* [20]. For example, cholera infections peak during the rainy season in the Philippines, Costa Rica, Lesotho, and Gambia [23, 16]; they peak during the summer in South America [34], and after the rainy season in Amazonia, Brazil [15]. Furthermore the rather unique double peak of historical Dhaka for the classical biotype, has been observed also for the more recent El Tor biotype, and for the temporarily emergent strain, *Vibrio cholerae* O139, in 1993 in Bangladesh [19].

5. Conclusion

For two hundred years, an explanation for the range of seasonal patterns in cholera based on local and simple environmental drivers has remained elusive. Despite numerous studies of the association between climate variability and incidence, no unified mechanisms explaining the temporal patterns in en-

demic regions have been proposed for cholera. Because the ecology of *Vibrio cholerae* and the relative importance of its different transmission pathways (human-to-human and environmental-to-human) are not fully understood, there has been a sense that simple environmental drivers cannot explain the diverse seasonal patterns of the disease. This study shows that a mechanistic model including the explicit influence of rainfall and temperature is capable to capture the full range of cholera seasonal patterns present in the historical Bengal region.

Based on an SIR-like model with additional compartments for the water volume and the pathogen concentration, insights were gained on the conditions creating endemicity and variation in seasonal patterns. In particular, the hydrological regime proved to be a dominant driver determining the seasonal dynamics, with rainfall exerting different effects in different regions. Specifically, rainfall can enhance transmission in dry regions, while buffering the propagation of the disease in wet regions due to a dilution effect. Such opposite influences indicate that overall water levels matter and act in complex ways to determine whether the effect of rainfall is positive or negative. Persistence of the disease is enabled by the environmental reservoir, which underlies endemicity.

Although cholera today does no longer exert the global death toll it once did, it remains responsible for substantial public health burdens in Bangladesh and many developing countries. The dynamics behind its seasonality have been shown to be closely associated with climate and environmental variability. An understanding of environmental influences based on hydrology could contribute to the better management and planning of public

health policies. Informing those capabilities in this way has become today of paramount importance, given on-going changes in climate, including extremes, and their expected impact on the population dynamics of infectious diseases. The changing environment, as the result not just of climate but also urbanization and higher population densities, will lead to new societal and scientific challenges in disease prevention and mitigation strategies.

Acknowledgments

EB acknowledges the support from the Swiss National Science Foundation (SNF/FNS) project "Dynamics and controls of large-scale cholera outbreaks" (DYCHO CR23I2 138104). MP and AAK are grateful for the support of the National Oceanic and Atmospheric Administration's Program on Oceans and Health (Grant # F020704), and earlier support by the National Science Foundation (NSF) through the program on the Ecology of Infectious Diseases (Grant # 0430120). XR acknowledges the project New Indigo ("Intestinal parasites in northern India", PCIN-2013-038). The authors acknowledge Leonard Evequoz for preliminary efforts on data and modelling. This research was supported in part through computational resources and services provided by Advanced Research Computing at the University of Michigan, Ann Arbor.

References

[1] Acosta, C. J., Galindo, C. M., Kimario, J., Senkoro, K., Urassa, H., Casals, C., Corachn, M., Eseko, N., Tanner, M., Mshinda, H., 2001. Cholera outbreak in southern Tanzania: risk factors and patterns of transmission. Emerging infectious diseases 7 (3), 583.

[2] Akanda, A. S., Jutla, A. S., Alam, M., de Magny, G. C., Siddique, A. K., Sack, R. B., Huq, A., Colwell, R. R., Islam, S., 2011. Hydroclimatic influences on seasonal and spatial cholera transmission cycles: Implications for public health intervention in the Bengal Delta: Hydroclimatic Influences on Seasonal Cholera. Water Resources Research 47 (3).

[3] Akanda, A. S., Jutla, A. S., Islam, S., 2009. Dual peak cholera transmission in Bengal Delta: A hydroclimatological explanation. Geophysical Research Letters 36 (19).

[4] Ali, M., Emch, M., Donnay, J.-P., Yunus, M., Sack, R. B., 2002. Identifying environmental risk factors for endemic cholera: a raster GIS approach. Health & place 8 (3), 201–210.

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

[6] Bertuzzo, E., Mari, L., Righetto, L., Gatto, M., Casagrandi, R., Rodriguez-Iturbe, I., Rinaldo, A., 2012. Hydroclimatology of dual-peak annual cholera incidence: Insights from a spatially explicit model. Geophysical Research Letters.

[7] Bouma, M. J., Pascual, M., 2001. Seasonal and interannual cycles of endemic cholera in Bengal 1891-1940 in relation to climate and geography. In: The Ecology and Etiology of Newly Emerging Marine Diseases. Springer, pp. 147–156.

[8] Broomhead, D. S., King, G. P., Jun. 1986. Extracting qualitative dynamics from experimental data. Physica D: Nonlinear Phenomena 20 (2), 217–236.

[9] Brouwer, C., Heibloem, M., 1986. Irrigation Water Management: Irrigation Water Needs. Food and Agriculture Organization - Natural Resources Management and Environment Department.

[10] Bryce, J., Boschi-Pinto, C., Shibuya, K., Black, R. E., 2005. WHO estimates of the causes of death in children. The Lancet 365 (9465), 1147–1152.

[11] Cash, B. A., Rod, X., Kinter, J. L., Sep. 2008. Links between Tropical Pacific SST and Cholera Incidence in Bangladesh: Role of the Eastern and Central Tropical Pacific. Journal of Climate 21 (18), 4647–4663.

[12] Cash, B. A., Rod, X., Kinter, J. L., Apr. 2009. Links between Tropical Pacific SST and Cholera Incidence in Bangladesh: Role of the Western Tropical and Central Extratropical Pacific. Journal of Climate 22 (7), 1641–1660.

[13] Cash, B. A., Rod, X., Kinter, J. L., Yunus, M., May 2010. Disentangling the Impact of ENSO and Indian Ocean Variability on the Regional Climate of Bangladesh: Implications for Cholera Risk. Journal of Climate 23 (10), 2817–2831.

[14] Chun, J., Grim, C. J., Hasan, N. A., Lee, J. H., Choi, S. Y., Haley, B. J., Taviani, E., Jeon, Y.-S., Kim, D. W., Lee, J.-H., Brettin, T. S., Bruce, D. C., Challacombe, J. F., Detter, J. C., Han, C. S., Munk,

A. C., Chertkov, O., Meincke, L., Saunders, E., Walters, R. A., Huq, A., Nair, G. B., Colwell, R. R., 2009. Comparative genomics reveals mechanism for short-term and long-term clonal transitions in pandemic Vibrio cholerae. Proceedings of the National Academy of Sciences 106 (36), 15442–15447.

[15] Codeço, C. T., 2001. Endemic and epidemic dynamics of cholera: the role of the aquatic reservoir. BMC infectious diseases 1 (1), 1.

[16] Collins, A. E., 1996. The geography of cholera. In: Drasar, P. B. S., Forrest, D. B. D. (Eds.), Cholera and the Ecology of Vibrio cholerae. Springer Netherlands, pp. 255–294, dOI: 10.1007/978-94-009-1515-2_8.

[17] Colwell, R. R., 1996. Global climate and infectious disease: The cholera paradigm. Science 274 (5295), 2025–2031.

[18] Eisenberg, M., Kujbida, G., Tuite, A., Fisman, D., Tien, J., 2013. Examining rainfall and cholera dynamics in haiti using statistical and dynamic modeling approaches. Epidemics 5 (4), 197–207.

[19] Emch, M., Ali, M., Feb. 2001. Spatial and Temporal Patterns of Diarrheal Disease in Matlab, Bangladesh. Environment and Planning A 33 (2), 339–350.

[20] Emch, M., Feldacker, C., Islam, M. S., Ali, M., 2008. Seasonality of cholera from 1974 to 2005: a review of global patterns. International Journal of Health Geographics 7 (1), 31.

[21] Emch, M., Feldacker, C., Yunus, M., Streatfield, P. K., Thiem, V. D., Canh, D. G., Ali, M., 2008. Local environmental predictors of cholera in

485 Bangladesh and Vietnam. American Journal of Tropical Medicine and
486 Hygiene 78 (5), 823–832.

487 [22] Emch, M., Yunus, M., Escamilla, V., Feldacker, C., Ali, M., 2010. Local
488 population and regional environmental drivers of cholera in Bangladesh.
489 Environmental Health 9 (1), 2.

490 [23] Gangarosa, E. J., Mosley, W. H., 1974. Epidemiology and surveillance
491 of cholera. Cholera. Philadelphia: WB Saunders, 381–403.

492 [24] Huq, A., Sack, R. B., Nizam, A., Longini, I. M., Nair, G. B., Ali, A.,
493 Morris, J. G., Khan, M. N. H., Siddique, A. K., Yunus, M., Albert,
494 M. J., Sack, D. A., Colwell, R. R., 2005. Critical Factors Influencing
495 the Occurrence of Vibrio cholerae in the Environment of Bangladesh.
496 Applied and Environmental Microbiology 71 (8), 4645–4654.

497 [25] Ionides, E. L., Bhadra, A., Atchad, Y., King, A., 2011. Iterated filtering.
498 The Annals of Statistics 39 (3), 1776–1802.

499 [26] Ionides, E. L., Bret, C., King, A. A., 2006. Inference for nonlinear
500 dynamical systems. Proceedings of the National Academy of Sciences
501 103 (49), 18438–18443.

502 [27] Jutla, A., Whitcombe, E., Hasan, N., Haley, B., Akanda, A., Huq, A.,
503 Alam, M., Sack, R. B., Colwell, R., 2013. Environmental Factors Influ-
504 encing Epidemic Cholera. American Journal of Tropical Medicine and
505 Hygiene 89 (3), 597–607.

506 [28] Jutla, A. S., Akanda, A. S., Griffiths, J. K., Colwell, R., Islam, S., Jan.

2011. Warming Oceans, Phytoplankton, and River Discharge: Implications for Cholera Outbreaks. The American Journal of Tropical Medicine and Hygiene 85 (2), 303–308.

[29] King, A. A., Ionides, E. L., Breto, C., Kendall, E., Lavine, M., Reuman, D. C., Wearing, H., Wood, S. N., 2013. Package 'pomp'.

[30] King, A. A., Ionides, E. L., Pascual, M., Bouma, M. J., 2008. Inapparent infections and cholera dynamics. Nature 454 (7206), 877–880.

[31] Klein, S. A., Soden, B. J., Lau, N.-C., 1999. Remote Sea Surface Temperature Variations during ENSO: Evidence for a Tropical Atmospheric Bridge. Journal of Climate 12 (4), 917–932.

[32] Koelle, K., Pascual, M., Jun. 2004. Disentangling Extrinsic from Intrinsic Factors in Disease Dynamics: A Nonlinear Time Series Approach with an Application to Cholera. The American Naturalist 163 (6), 901–913.

[33] Laneri, K., Bhadra, A., Ionides, E. L., Bouma, M., Dhiman, R. C., Yadav, R. S., Pascual, M., Sep. 2010. Forcing Versus Feedback: Epidemic Malaria and Monsoon Rains in Northwest India. PLOS Comput Biol 6 (9), e1000898.

[34] Lipp, E. K., Huq, A., Colwell, R. R., 2002. Effects of Global Climate on Infectious Disease: the Cholera Model. Clinical Microbiology Reviews 15 (4), 757–770.

[35] Lloyd, A. L., May 2001. Destabilization of epidemic models with the

inclusion of realistic distributions of infectious periods. Proceedings of the Royal Society B: Biological Sciences 268 (1470), 985–993.

[36] Miller, C., Feachem, R., Drasar, B., 1985. Cholera epidemiology in developed and developing countries: New thoughts on transmission, seasonality, and control. The Lancet 325 (8423), 261–263.

[37] Pascual, M., Bouma, M. J., Dobson, A. P., 2002. Cholera and climate: revisiting the quantitative evidence. Microbes and infections 4 (2), 237–245.

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

[39] Pascual, M., Rodó, X., Ellner, S. P., Colwell, R., J., B. M., 2000. Cholera Dynamics and El Nino-Southern Oscillation. Science 289 (5485), 1766–1769.

[40] Righetto, L., Casagrandi, R., Bertuzzo, E., Mari, L., Gatto, M., Rodriguez-Iturbe, I., Rinaldo, A., 2012. The role of aquatic reservoir fluctuations in long-term cholera patterns. Epidemics 4 (1), 33–42.

[41] Rinaldo, A., Bertuzzo, E., Mari, L., Righetto, L., Blokesch, M., Gatto, M., Casagrandi, R., Murray, M., Vesenbeckh, S. M., Rodriguez-Iturbe, I., 2012. Reassessment of the 20102011 Haiti cholera outbreak and rainfall-driven multiseason projections. Proceedings of the National Academy of Sciences 109 (17), 6602–6607.

[42] Rodó, X., Pascual, M., Fuchs, G., Faruque, A. S. G., 2002. ENSO and cholera: A nonstationary link related to climate change? Proceedings of the national Academy of Sciences 99 (20), 12901–12906.

[43] Ruiz-Moreno, D., Pascual, M., Bouma, M., Dobson, A., Cash, B., 2007. Cholera Seasonality in Madras (19011940): Dual Role for Rainfall in Endemic and Epidemic Regions. EcoHealth 4 (1), 52–62.

[44] Sperna Weiland, F. C., Tisseuil, C., Durr, H. H., Vrac, M., van Beek, L. P. H., 2012. Selecting the optimal method to calculate daily global reference potential evaporation from CFSR reanalysis data for application in a hydrological model study. Hydrology and Earth System Sciences 16 (3), 983–1000.

[45] Weiss, M., Menzel, L., 2008. A global comparison of four potential evapotranspiration equations and their relevance to stream flow modelling in semi-arid environments. Advances in Geosciences 18.

[46] WHO, 2015. Cholera. World Health Organisation. Accessed on: 2014-04-08.
URL http://www.who.int/mediacentre/factsheets/fs107/en