

# Environmental signatures associated with cholera epidemics

Guillaume Constantin de Magny<sup>a</sup>, Raghu Murtugudde<sup>b</sup>, Mathew R. P. Sapiano<sup>b</sup>, Azhar Nizam<sup>c</sup>, Christopher W. Brown<sup>b,d</sup>, Antonio J. Busalacchi<sup>b</sup>, Mohammad Yunus<sup>e</sup>, G. Balakrish Nair<sup>f</sup>, Ana I. Gil<sup>g</sup>, Claudio F. Lanata<sup>g,h</sup>, John Calkins<sup>i</sup>, Byomkesh Manna<sup>f</sup>, Krishnan Rajendran<sup>f</sup>, Mihir Kumar Bhattacharya<sup>f</sup>, Anwar Huq<sup>j</sup>, R. Bradley Sack<sup>k</sup>, and Rita R. Colwell<sup>a,j,l,1</sup>

<sup>a</sup>Institute for Advanced Computer Studies, <sup>b</sup>Earth System Science Interdisciplinary Center, and <sup>l</sup>Maryland Pathogen Research Institute, College of Chemical and Life Sciences, University of Maryland, College Park, MD 20742; <sup>c</sup>Department of Biostatistics, Emory University Rollins School of Public Health, Atlanta, GA 30329; <sup>d</sup>Satellite Climate Studies Branch, National Oceanic and Atmospheric Administration, College Park, MD 20742; <sup>e</sup>International Centre for Diarrhoeal Disease Research, Bangladesh, Dhaka, 1000, Bangladesh; <sup>f</sup>National Institute of Cholera and Enteric Diseases, Beliaghata, Kolkata 700010, India; <sup>g</sup>Instituto de Investigacion Nutricional, Lima 12, Peru; <sup>h</sup>School of Medicine, Universidad Peruana de Ciencias Aplicadas, Lima 12, Peru; <sup>i</sup>Environmental Systems Research Institute, Inc., Vienna, VA 22182; and <sup>j</sup>Department of International Health and <sup>k</sup>Department of Environmental Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD 21205

Contributed by Rita R. Colwell, September 26, 2008 (sent for review August 1, 2008)

The causative agent of cholera, *Vibrio cholerae*, has been shown to be autochthonous to riverine, estuarine, and coastal waters along with its host, the copepod, a significant member of the zooplankton community. Temperature, salinity, rainfall and plankton have proven to be important factors in the ecology of *V. cholerae*, influencing the transmission of the disease in those regions of the world where the human population relies on untreated water as a source of drinking water. In this study, the pattern of cholera outbreaks during 1998–2006 in Kolkata, India, and Matlab, Bangladesh, and the earth observation data were analyzed with the objective of developing a prediction model for cholera. Satellite sensors were used to measure chlorophyll a concentration (CHL) and sea surface temperature (SST). In addition, rainfall data were obtained from both satellite and *in situ* gauge measurements. From the analyses, a statistically significant relationship between the time series for cholera in Kolkata, India, and CHL and rainfall anomalies was determined. A statistically significant one month lag was observed between CHL anomaly and number of cholera cases in Matlab, Bangladesh. From the results of the study, it is concluded that ocean and climate patterns are useful predictors of cholera epidemics, with the dynamics of endemic cholera being related to climate and/or changes in the aquatic ecosystem. When the ecology of *V. cholerae* is considered in predictive models, a robust early warning system for cholera in endemic regions of the world can be developed for public health planning and decision making.

ecology | epidemiology | microbiology | remote sensing

*Vibrio cholerae*, a serious pathogen for humans, has been the subject of intense study for more than a century, yet the discovery that this bacterium is a natural inhabitant of riverine, estuarine, and coastal waters throughout both temperate and tropical regions of the world was made only relatively recently (1). Before the late 1970s, transmission of cholera was believed to occur exclusively by person-to-person contact, with epidemics initiated by contaminated water and food. In 1855, Sir John Snow hypothesized that devastating epidemics of cholera are caused by contaminated drinking water (2). In his time, the germ theory of disease had not been proven nor accepted and, obviously, it was not understood that the epidemic strain of *V. cholerae* was a bacterium naturally occurring in the aquatic environment (1, 3).

It is now recognized that *V. cholerae* is a component of coastal and estuarine microbial ecosystems, with the copepod species of zooplankton that comprise the aquatic fauna of rivers, bays, estuaries and the open ocean serving as host for the bacterium (4–7). *V. cholerae* can be found attached to the carapace and in the gut of copepods in large numbers, the copepod essentially serving as a vector for this human pathogen (1, 8, 9). A single copepod, for example, can contain as many as  $10^3$ – $10^5$  *V. cholerae*

cells (10). Because a concentration of  $10^9$  ml<sup>-1</sup> *V. cholerae* comprises an infective dose, ingestion of untreated water containing a relatively small number of copepods carrying *V. cholerae* can initiate the disease (11). Therefore, conditions favorable for multiplication of copepods and related chitinous zooplankton species for which *V. cholerae* is commensal or symbiotic will result in an increase in the number of *V. cholerae*. The importance of copepods in cholera transmission was demonstrated in a study showing that the number of cholera cases in Bangladeshi villages was significantly reduced when a simple filtration method that effectively removed the plankton and particulate matter was used to treat drinking water (12, 13).

Another advance in understanding cholera epidemics was made with the discovery that the epidemics are significantly influenced by climatic factors (1, 13–17). Several environmental drivers of the seasonal cycle of cholera in India and Bangladesh where cholera is endemic have been identified (1, 18). Ocean surveillance by satellite remote sensing was used to monitor changes in sea surface temperature (SST) and sea surface height (SSH) in the Bay of Bengal. The pattern of changes in these parameters were shown to be related to *V. cholerae* dynamics in coastal, estuarine, and riverine waters of the Bay of Bengal and to the cholera epidemics caused by these bacteria in that region of the world (14). That is, the distinct seasonal pattern of cholera, in countries where the disease is endemic has been correlated with environmental factors and climate that drive both copepod population dynamics and the seasonal peaks in abundance of *V. cholerae* in the aquatic ecosystem of the Bay of Bengal. Given this relationship, it is concluded that the variables related to copepod population dynamics can serve as a proxy for the estimation of *V. cholerae* abundance in the environment.

Lobitz *et al.* (14) were the first to explore the now well recognized relationship between SST, SSH and cholera incidence using remote sensing, describing changes in coastal ecosystems significantly related to the annual cyclic pattern of cholera epidemics. Significant progress has since been made in

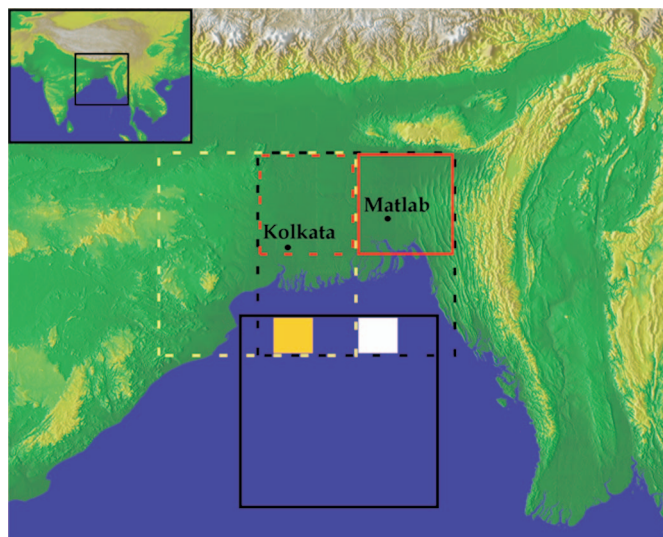
Author contributions: G.C.d.M., R.M., A.J.B., and R.R.C. designed research; G.C.d.M., R.M., M.R.P.S., and C.W.B. performed research; G.C.d.M., R.M., M.R.P.S., A.N., C.W.B., M.Y., G.B.N., A.I.G., C.F.L., J.C., B.M., K.R., M.K.B., A.H., R.B.S., and R.R.C. analyzed data; and G.C.d.M., R.M., M.R.P.S., A.N., C.W.B., A.J.B., G.B.N., A.I.G., C.F.L., J.C., K.R., A.H., R.B.S., and R.R.C. wrote the paper.

The authors declare no conflict of interest.

<sup>1</sup>To whom correspondence should be addressed at: Institute for Advanced Computer Studies, Biomolecular Sciences Building 296, Room 3103, University of Maryland, College Park, MD 20742. E-mail: rcolwell@umiacs.umd.edu.

This article contains supporting information online at [www.pnas.org/cgi/content/full/0809654105/DCSupplemental](http://www.pnas.org/cgi/content/full/0809654105/DCSupplemental).

© 2008 by The National Academy of Sciences of the USA



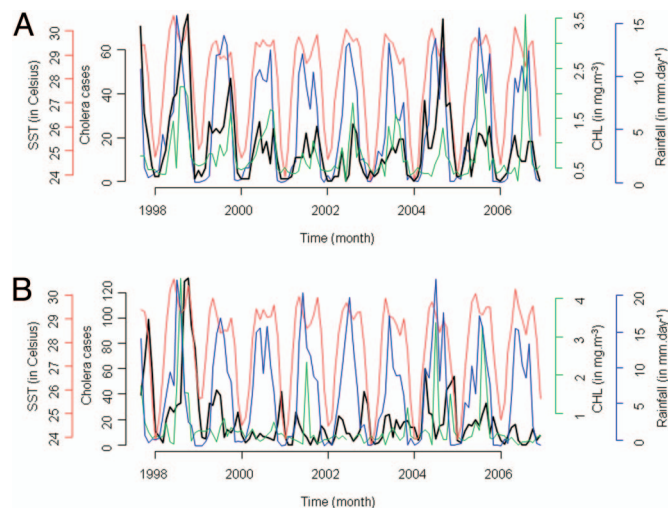
**Fig. 1.** Map showing the region from which the environmental and epidemiological data were obtained. Locations for extraction of the monthly mean values for CHL and SST are indicated by (i) one degree resolution squares, orange for Kolkata (centered in 20.5N/88.5E) and white for Matlab (centered in 20.5N/90.5E), and for (ii) the five degree box average for both sites by the black solid line square (centered in 18.5N/89.5E). Locations where monthly mean rainfall values were obtained are indicated as follows: (i) two and half degree resolution points for Kolkata are delimited by a red dashed line square (centered in 23.75N/88.75E) and for Matlab by a red solid line square (centered in 23.75N/91.25E), and (ii) for the five degree box average for Kolkata by the yellow dashed line square (centered in 22.5N/87.5E) and for Matlab by the black dashed square (centered in 22.5N/90.0E). Background image represents the ground elevation coded from dark green, the lowest value, to white, the highest [The Global Land One-kilometer Base Elevation (GLOBE) NOAA, NGDC; (<http://www.ngdc.noaa.gov/mgg/topo/globe.html>); (48)].

modeling interannual variability of cholera epidemics in Bangladesh, employing mathematical models to reproduce the seasonal cycle of cholera (19, 20). In one such study, a significant correlation was found between selected environmental variables, namely precipitation, SST, and CHL, with cholera cases in a single epidemic during 2000–2001 in KwaZulu-Natal, South Africa (21). In that analysis, SST and CHL data were estimated from satellite sensor measurements, and a conclusive finding was that robust predictive associations require multiyear analyses (21).

Here, we report on the association between selected environmental variables and confirmed cholera cases in two separate human populations resident in Kolkata, India, and Matlab, Bangladesh, in the northern Bay of Bengal, a marginal sea of the Indian Ocean (Fig. 1). Different satellite-derived data (CHL and SST) for the nearest coastal environment for each of the two geographical areas, and local rainfall during a nine year period (Fig. 2), September 1997–December 2006, for which all three environmental datasets were available, were analyzed. To determine which environmental signatures were associated with cholera epidemics in Kolkata and in Matlab, and to demonstrate the capacity to predict cholera, a historical approach was taken in modeling the relationship between the dynamics of the cholera epidemics and related environmental factors.

## Results

For Kolkata, a positive relationship was observed between the number of cholera cases and CHL anomaly ( $CHL_{ano}$ ) for both the single grid-point and the box-averaged values (see *Materials and Methods* and Fig. 1 for details). A  $1 \text{ mg/m}^3$  increase in monthly mean  $CHL_{ano}$  for the single grid-point value was



**Fig. 2.** Epidemiological and environmental dynamics for (A) Kolkata, India, and (B) Matlab, Bangladesh. The cholera cases, the CHL, the SST and the rainfall are shown in black, green, red and blue lines, respectively.

associated with an increase of 32.5% [95% confidence interval (CI) = 8.3%–62.0%] in the number of cholera cases, after controlling for the annual cycle and the “persistence” effect arising from density dependence in cholera transmission.

The relationship described above for  $CHL_{ano}$  was also observed for the single grid-point value of rainfall anomaly ( $PRE_{ano}$ ) and the number of cholera cases in Kolkata. A  $1 \text{ mm/day}$  increase in monthly mean  $PRE_{ano}$  was associated with an increase of 6.5% [95% CI = 1.6%–11.7%] in the number of cholera cases, after controlling for the annual cycle and the “persistence” effect.

The final model for Kolkata is

$$y|\eta \sim \text{Po}(\mu, \sigma^2)$$

$$\log(\mu) = \eta = \beta_0 + \beta_1 I_{MAM} + \beta_2 I_{JJA} + \beta_3 I_{SON} + \beta_4 \log(1 + chol_{t-1}) + \beta_5 CHL_t + \beta_6 Rain_t \quad [1]$$

where coefficients are detailed in Table 1. Observed cholera case time-series and the prediction of both the fitted model and the fitted cross-validation model are displayed in Fig. 3A. The comparison of observed cholera cases against the prediction of the fitted models (Fig. 3B) shows that the cross-validation model tends to underestimate more of the highest values of cholera cases than the fitted model. The pseudo  $R^2$  value for the fitted model is equal to 88.17%.

In contrast to  $CHL_{ano}$ , the  $PRE_{ano}$  five degree grid box average was not retained in the model, suggesting that the major effects of precipitation occur locally rather than at a large scale, reflecting localized precipitation, even in a large scale monsoon system (22). SST was not retained in the model, either with the single grid point anomaly or with the five degree grid box average. It could be because of collinearity between the variables, most of the variance likely being already explained by  $CHL_{ano}$  and  $PRE_{ano}$ .

For Matlab, a positive relationship was observed between number of cholera cases and the single grid-point data for  $CHL_{ano}$ , with a one month lag. A  $1 \text{ mg/m}^3$  increase in the monthly mean  $CHL_{ano}$  was related to the number of cholera cases increasing by 31.4% [95% CI = 13.0%–52.7%], after controlling for the annual cycle and the “persistence” arising from density dependence in cholera transmission. When the model was run using a five degree box for the averaged data, a positive

**Table 1. Summary of the model obtained for Kolkata, India, and Matlab, Bangladesh**

	Model parameters								Df
	Intercept	$I_{MAM}^*$	$I_{JJA}^*$	$I_{SON}^*$	$\text{Log}(\text{chol}_{t-1})^*$	$CHL_t$	$\text{Rain}_t$	$CHL_{t-1}$	
<b>Kolkata</b>									
Coefficient	-0.264	1.597	1.524	1.476	0.578	0.281	0.063		103
P value	0.39	$1.10^{-7}$	$4.10^{-7}$	$1.10^{-6}$	$9.10^{-12}$	$7.10^{-3}$	$9.10^{-3}$		
SE	0.31	0.28	0.28	0.29	0.07	0.10	0.02		
<b>Matlab</b>									
Coefficient	0.676	0.674	0.168	0.659	0.637			0.273	103
P value	$2.10^{-2}$	$2.10^{-3}$	0.44	$9.10^{-4}$	$2.10^{-14}$			$5.10^{-4}$	
SE	0.29	0.22	0.22	0.19	0.07			0.08	

\*As described in *Materials and Methods*, this parameter was maintained in the model without considering the *P* value.

relationship was observed between  $CHL_{ano}$  and the number of cholera cases. Upwelling occurs mainly along the coast of the Bay of Bengal, thus the five degree box yielded a significantly weakened coastal upwelling signal. As for Kolkata, SST was not retained in the model presumably because of the same reasons. The final model for Matlab is

$$y|\eta \sim \text{Po}(\mu, \sigma^2)$$

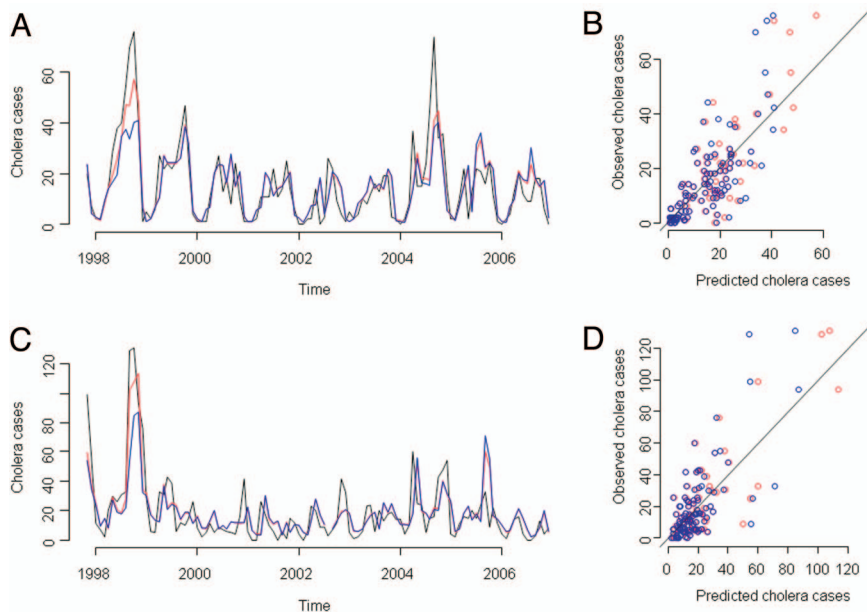
$$\log(\mu) = \eta = \beta_0 + \beta_1 I_{MAM} + \beta_2 I_{JJA} + \beta_3 I_{SON} + \beta_4 \log(1 + \text{chol}_{t-1}) + \beta_5 CHL_{t-1} \quad [2]$$

where coefficients are detailed in Table 1. Observed cholera case time series and the prediction of both the fitted model, and the fitted cross-validation model are displayed in Fig. 3C. The contrast of observed cholera cases against the prediction of the fitted models (Fig. 3D) shows that the cross-validation model tends to underestimate more of the peak cholera case values than the fitted model, with both models overestimating the low number of cholera cases. The pseudo  $R^2$  value for the fitted model is equal to 71.46%.

### Discussion

Environmental factors were found to be statistically significant in two different geographical locations of the Indian continent in directly influencing the dynamics of cholera epidemics. A relationship between the short time lag (one month), coincident environmental conditions, and cholera epidemics was observed. Interestingly, despite the geographical proximity of Kolkata and Matlab, the effect of the environmental variables examined in this study was clearly different. These regional differences strongly indicate an important effect of the local environment and local zooplankton populations on the dynamics of cholera epidemics.

Indeed for Kolkata, higher values of  $CHL_{ano}$  reflected the more intense algal blooms than normal that occurred and led to larger zooplankton populations that were comprised mainly of crustacean copepods, a natural aquatic host for *V. cholerae* (1, 8, 9). From the analyses of the Dhaka, Bangladesh data, it was concluded that the tidal intrusion of coastal water carrying plankton into inland water could initiate increased human contact with the cholera vibrio (14), because water used for daily hygiene, personal consumption, and religious rites (e.g., ablu-



**Fig. 3.** Observed epidemiological data and prediction of fitted models. For Kolkata, (A) temporal dynamics of observed cholera cases, fitted model and cross-validation model shown in black, red and blue, respectively. (B) Scatterplot of observed cholera cases against (i) predicted cholera cases by fitted model in red circles, and (ii) predicted cholera cases by cross-validation model in blue circles. Black line represents perfect agreement between predicted and observed cases. For Matlab, C same as A. D same as B.

tions) in rural areas of Bangladesh and India is taken directly from local rivers or ponds essentially untreated (13, 23, 24). The relationship observed for  $PRE_{ano}$  similarly reflects the effect of unusually heavy precipitation influencing cholera transmission by feeding surface runoff into streams and rivers flooding the water supply (15, 16). It also can have a direct role in cholera dynamics by increasing nutrient runoff, inducing river enrichment responsible for local algal blooms (25, 26). During intense rainfall events, the tidal intrusion of coastal water carrying plankton into inland water influencing cholera transmission may be minimized by the increase of the river flows. We concluded that both factors are influencing cholera transmission independently.

For Matlab, the relationship between number of cholera cases and  $CHL_{ano}$  was as expected. Matlab comprises a densely populated cluster of villages in a flat, low altitude region representing a dense, natural hydrological system intersected by the Dhanagana River and having a strong tidal influence. The relative proximity of Matlab to the coastal region of the Bay of Bengal can explain both the significant relationship of  $CHL_{ano}$  with cholera and the one month lag for Matlab, because Matlab lies further inland than Kolkata.

It is evident that each location requires a different model because of local or regional specificity. While CHL was determined to be important in influencing the dynamics of cholera epidemics in both locations within the same month or with one month lag, a longer lead time was found to be necessary for the environmental factors to be useful in predicting epidemics of cholera and give health workers time to formulate plans, disseminate warnings and recommendations to the public. Nevertheless, even a short lead time, at the submonthly time scale, can provide sufficient warning so that precautionary measures can be taken to save lives. In this context, the robust results obtained in this study indicate that increasing the lead time for prediction in dynamic or statistical regional climate and/or ocean forecast models, will provide a valuable tool for public health purposes. A functional model for an early warning system will require finer temporal resolution of acquired data, particularly for Kolkata where the relationship appears to be submonthly although most likely a data-limitation issue at present. For practical applications of these models, environmental as well as epidemiological data need to be collected and compiled expediently to provide useful and reliable predictions of the onset, epidemics, and trends of cholera based on environmental variability. The environmental variables related to cholera dynamics in this study, rainfall and CHL, as well as predictable phenomena, such as the Madden-Julian Oscillation, El Niño-Southern Oscillation, and the more recently proposed Indian Ocean Dipole/Zonal mode (27–31), clearly point to the potential for prediction of cholera with a significant lead time. A simple statistical model presented here that incorporates only local environmental conditions, nevertheless, yielded significant statistical association between cholera and selected environmental factors.

Obviously, not all of the variance in the number of cholera cases in each region can be explained by this method alone, because cholera epidemics involve a complex and critical interplay of intrinsic dynamics with extrinsic drivers (20). For example, cholera is no longer a disease threat for developed countries, including the United States, even though the presence of *V. cholerae* O1 in the waters of the Chesapeake Bay and coastal states of the Gulf of Mexico has long been known (32–34). Because good sanitation is practiced and safe drinking water is available to the populations of the region, cholera epidemics no longer occur. However, *V. cholerae* does remain as a naturally occurring inhabitant year round in those aquatic environments, with the same spring and fall peaks in number of the cholera bacteria, and a similar relationship with its copepod host (32, 35).

The results of this study provide a foundation on which to build a predictive capacity for cholera epidemics, hence, an early warning system for enhancing public health measures, especially for developing countries and areas of the world undergoing social disruption or climate change. Better spatial resolutions of the satellite datasets as well as finer temporal resolution, i.e., ground based network stations, will improve our ability to predict cholera outbreaks. It will also allow a continued environmental data validation as well as alleviate the problem of missing satellite data because of clouds or other potential technical issues. Selected environmental factors serve effectively as indicators for cholera, but an integrative, mechanistic and interdisciplinary model combining satellite and ground observations will be required. The Aquarius (<http://aquarius.gsfc.nasa.gov/index.php>) satellite mission, designed to measure global sea surface salinity, to be launched in 2009, will provide additional useful information because salinity has a critical influence on the ecology of *V. cholerae* (15, 36–38).

Subseasonal-to-interannual time-scale operational forecast systems that employ statistical forecasting tools and combine local ecosystem, coastal, and ocean data, including SST, SSH, CHL, and zooplankton concentration, from mechanistic hydrodynamic and ecosystem modeling to analysis of climate parameters from coupled climate forecast models (rainfall, river runoff, air temperature, and humidity) may prove more useful for Kolkata, where time-resolved environmental parameters to obtain submonthly lead times are needed, compared with Bangladesh, where the environmental lead time appears to be longer.

A fully integrated model for cholera prediction that takes into account the demographic components of human populations exposed to endemic cholera and the complexity of cholera epidemiology, as well as climate, and the environment, ecology and the genomics of *V. cholerae* will continue to be a focus of our research. Interannual variability of cholera (15, 20, 39, 40) and the strong El Niño that occurred in 1998 were found to be associated with the largest number of cholera cases for both Kolkata and Matlab and offer challenging examples of cholera and climate interaction.

## Materials and Methods

Data for Matlab, Bangladesh, and Kolkata, India, included in the analyses correspond to the monthly number of admitted patients for cholera symptoms individually confirmed by culture. The clinical specimens from which *V. cholerae* was isolated were from patients admitted either to the International Center for Diarrhoeal Diseases Research, Bangladesh (ICDDR, B) or to the Infectious Diseases Hospital of the National Institute of Cholera and Enteric Diseases (NICED) in Kolkata, India. Methods used for isolation and culture of *V. cholerae* have been reported elsewhere (41, 42). Additional computations detailed in [supporting information \(SI\) Text](#) showed no differences in using the monthly accumulation of cholera cases as a response variable in models instead of incidence.

Rainfall data used in the analyses were taken from the monthly merged satellite/gauge estimates of the Global Precipitation Climatology Project (GPCP) Version 2 (43). These data provide merged ocean/land estimates. Source of the monthly SSTs was Reynolds and Smith (44), generated from the NOAA Optimally Interpolated (v2) product provided by the National Climatic Data Center ([www.ncdc.noaa.gov](http://www.ncdc.noaa.gov)). CHL estimates were derived from the Sea-viewing Wide Field of View (SeaWiFS) sensor launched in 1997 (45), and provided by the NASA Goddard Distributed Active Archive Center (<http://daac.gsfc.nasa.gov>). CHL data are available by month from September, 1997 to the present. The spatial resolution of CHL and SST was a one degree grid and that for the rainfall data comprised a two-and-a-half degree grid. CHL and SST data were extracted from the measurements taken off the coast of the Bay of Bengal, whereas rainfall data (over land and ocean) were extracted coincident with the study locations (see Fig. 1). To evaluate potential differences at local or regional scale in the relationships between cholera cases and environmental variables, we also computed a five degree box-average for the environmental variables. For CHL and SST, it corresponds to the mean computed over a five by five square of data at one degree resolution, i.e., a mean of twenty five pixels (Fig. 1). For PRECIP, it corresponds to a parameter mean computed

over a five by five square of data at two-and-half degree grid resolution, i.e., a mean of four pixels (Fig. 1). Because the three environmental datasets show a strong seasonal pattern (Fig. S1), anomalies were computed to remove the effect of the annual cycle, which can result in spurious relationships between cotrending variables. The anomalies were constructed by calculating the climatological mean for each month and subtracting these means from each corresponding month from the original data.

Environmental factors for the months preceding the cholera outbreaks proved to be important. That is, for CHL, phytoplankton biomass increases followed by an increase in the zooplankton population, the latter harboring *V. cholerae*, both associated with an increase in SST. This process, however, is not instantaneous and a lag time of one to two months was used for an increase in the *V. cholerae* population after the increase in SST and CHL had occurred. A time lag after precipitation is mechanistically justified because precipitation during months preceding a cholera outbreak feeds streams and rivers with surface runoff, thereby playing its unique role in the initiation of cholera outbreaks, especially when flooding occurs (15, 16). The chain of these climate events and their lead times provide the basis for prediction of cholera outbreaks, and were exploited to demonstrate feasibility of prediction, hence an early warning system.

There is a mechanistic basis for expecting up to two months lag for CHL, SST, and rainfall peaks in these drivers of cholera. A mechanistic approach allows translation of an empirical predictive understanding into quantitative prediction for cholera. To exploit the process empirically, variables included in the model were taken from both the coincident month and a one-month lag (t-1). CHL data also included a two-month lag (t-2). Indicator variables for quarterly means of cholera cases representing the annual cycle were included in the model as well as log number of cholera cases for the previous month. The latter is analogous to the differencing technique commonly used in an ordinary least squares linear regression (adapted for generalized linear models on a log scale), removing the "persistence" effect arising from density-dependence in cholera transmission (19). Both variables were included as fixed parameters in the models to remove residual autocorrelation.

A generalized linear model (GLM) with a Poisson distribution and a log link was used to model the data (46). As is common with such data, strong over-dispersion was apparent and was accommodated by using a quasi-Poisson GLM fitted using the R software package. This model estimates the degree of overdispersion and inflates standard errors accordingly. For the quasi-Poisson model, the constraint of equal mean and variance is relaxed so that the variance of the response is related to the mean  $\mu$  by

$$\text{Var}(Y) = \sigma^2 \mu \quad [3]$$

where  $\sigma^2$  is the (constant) dispersion parameter. Our model for counts in the presence of overdispersion can thus be written as

$$\begin{aligned} y|\eta &\sim \text{Po}(\mu, \sigma^2) \\ \log(\mu) &= \eta = \beta_0 + \beta_1 I_{\text{MAM}} + \beta_2 I_{\text{JJA}} + \beta_3 I_{\text{SON}} \\ &+ \beta_4 \log(1 + \text{chol}_{t-1}) + \beta_5 \text{CHL}_t \\ &+ \beta_6 \text{CHL}_{t-1} + \beta_7 \text{CHL}_{t-2} + \beta_8 \text{SST}_t \\ &+ \beta_9 \text{SST}_{t-1} + \beta_{10} \text{Rain}_t + \beta_{11} \text{Rain}_{t-1}, \end{aligned} \quad [4]$$

where  $y$  represents the observed counts,  $\eta$  is the fitted model, Po represents the Poisson distribution, the  $\beta_j$  are the model parameters, and  $I_{\text{MAM}}$ ,  $I_{\text{JJA}}$ , and  $I_{\text{SON}}$  represent the quarterly mean of cholera case index for March–April–May, June–July–August, and September–October–November, respectively and were used to remove any remaining annual cycle. Hypotheses of environmental factors driving cholera dynamics in the two locations was tested by using a 5% rejection range for significant variables. To evaluate the models computed on the whole dataset, cross-validations predictions were also made from the model by predicting cholera cases for each year using the final model structure, i.e., with only significant parameters, but estimating coefficients on the data set without data for the year under consideration. Then we assembled each predicted year to construct the cross-validation predictions for the whole period. The pseudo  $R^2$  proposed by (47) was used as a model diagnostic estimate adapted for quasi-Poisson model. Like the estimate of  $R^2$  commonly used for ordinary least squares linear regression, this statistical can be interpreted as the percentage of variability explained by the model.

SI. Further information is available in *SI Text* and *Fig. S2*.

**ACKNOWLEDGMENTS.** We thank the donor countries and agencies that provide unrestricted support to the International Centre for Diarrhoeal Disease Research, Bangladesh, for its operation and research; and the three referees for their constructive remarks and useful comments. This work was supported by National Oceanic and Atmospheric Administration Grant S0660009, Oceans and Human Health Initiative, and National Institutes of Health Grant 1 R01 AI39129.

- Colwell RR (1996) Global climate and infectious disease: The cholera paradigm. *Science* 274:2025–2031.
- Snow J (1855) *On the Mode of Communication of Cholera* (J. Churchill, London) 2nd Ed, pp 1–162.
- Johnson S (2006) *The Ghost Map: The Story of London's Most Terrifying Epidemic—And How It Changed Science, Cities, and the Modern World* (Riverhead Books, New York) p 320.
- Colwell RR, Kaper J, Joseph SW (1977) *Vibrio cholerae, Vibrio parahaemolyticus, and other vibrios: Occurrence and distribution in Chesapeake Bay* *Science* 198:394–396.
- Huq A, et al. (1983) Ecological relationships between *Vibrio cholerae* and planktonic crustacean copepods. *Appl Environ Microbiol* 45:275–283.
- Islam MS, Drasar BS, Sack RB (1994) The aquatic flora and fauna as reservoirs of *Vibrio cholerae*: A review. *J Diarr Dis Res* 12:87–96.
- Kaper JB, Morris JG, Jr, Levine MM (1995) Cholera. *Clin Microbiol Rev* 8:48–86.
- Nalin DR, Daya V, Reid A, Levine MM, Cisneros L (1979) Adsorption and growth of *Vibrio cholerae* on chitin. *Infect Immun* 25:768–770.
- Rawlings TK, Ruiz GM, Colwell RR (2007) Association of *Vibrio cholerae* O1 El Tor and O139 Bengal with the copepods *Acartia tonsa* and *Eurytemora affinis*. *Appl Environ Microbiol* 73:7926–7933.
- Colwell RR, Spira WM (1992) The ecology of *Vibrio cholera*. *Cholera*, eds Barua D, Greenough WB (Plenum Medical Book Co., New York), 3rd Ed, pp 107–127.
- Cash RA, et al. (1974) Response of man to infection with *Vibrio cholerae*. I. Clinical, serologic, and bacteriologic responses to a known inoculum. *J Infect Dis* 129:45–52.
- Huq A, et al. (1996) A simple filtration method to remove plankton-associated *Vibrio cholerae* in raw water supplies in developing countries. *Appl Environ Microbiol* 62:2508–2512.
- Colwell RR, et al. (2003) Reduction of cholera in Bangladeshi villages by simple filtration. *Proc Natl Acad Sci USA* 100:1051–1055.
- Lobitz B, et al. (2000) Climate and infectious disease: Use of remote sensing for detection of *Vibrio cholerae* by indirect measurement. *Proc Natl Acad Sci USA* 97:1438–1443.
- Lipp EK, Huq A, Colwell RR (2002) Effects of global climate on infectious disease: The cholera model. *Clin Microbiol Rev* 15:757–770.
- Pascual M, Bouma MJ, Dobson AP (2002) Cholera and climate: Revisiting the quantitative evidence. *Microbes Infect* 4:237–245.
- Ruiz-Moreno D, Pascual M, Bouma M, Dobson A, Cash B (2007) Cholera seasonality in Madras (1901–1940): Dual role for rainfall in endemic and epidemic regions. *Ecohealth* 4:52–62.
- Bouma MJ, Pascual M (2001) Seasonal and interannual cycles of endemic cholera in Bengal 1891–1940 in relation to climate and geography. *Hydrobiologia* 460:147–156.
- Pascual M, Rodo X, Ellner SP, Colwell R, Bouma MJ (2000) Cholera dynamics and El Niño–Southern Oscillation. *Science* 289:1766–1769.
- Koelle K, Rodo X, Pascual M, Yunus M, Mostafa G (2005) Refractory periods and climate forcing in cholera dynamics. *Nature* 436:696–700.
- Mendelsohn J, Dawson T (2007) Climate and cholera in KwaZulu-Natal, South Africa: The role of environmental factors and implications for epidemic preparedness. *Int J Hyg Environ Health* 211:156–162.
- Goswami BN, Venugopal V, Sengupta D, Madhusoodanan MS, Xavier PK (2006) Increasing trend of extreme rain events over India in a warming environment. *Science* 314:1442–1445.
- Huq A, et al. (2005) Critical factors influencing the occurrence of *Vibrio cholerae* in the environment of Bangladesh. *Appl Environ Microbiol* 71:4645–4654.
- Sack RB, et al. (2003) A 4-year study of the epidemiology of *Vibrio cholerae* in four rural areas of Bangladesh. *J Infect Dis* 187:96–101.
- Rao DVS (1973) Effects of environmental perturbations on short-term phytoplankton production off Lawsons-Bay, a tropical coastal embayment. *Hydrobiologia* 43:77–91.
- Valsaraj CP, Rao VNR (1994) Nitrogen limitation in the tropical waters of the Bay of Bengal. *Hydrobiologia* 286:139–148.
- Chen D, Cane MA (2008) El Niño prediction and predictability. *J Computational Phys* 227:3625–3640.
- Saji NH, Goswami BN, Vinayachandran PN, Yamagata T (1999) A dipole mode in the tropical Indian Ocean. *Nature* 401:360–363.
- Webster PJ, Moore AM, Loschnigg JP, Leben RR (1999) Coupled ocean-atmosphere dynamics in the Indian Ocean during 1997–98. *Nature* 401:356–360.
- Murtugudde R, McCreary JP, Busalacchi AJ (2000) Oceanic processes associated with anomalous events in the Indian Ocean with relevance to 1997–1998. *J Geophys Res Oceans* 105:3295–3306.
- Waliser D, et al. (2006) The experimental MJO prediction project. *Bull Amer Meteor Soc* 87:425–431.
- Heidelberg JF, Heidelberg KB, Colwell RR (2002) Seasonality of Chesapeake Bay bacterioplankton species. *Appl Environ Microbiol* 68:5488–5497.

