Predictive Time Series Analysis Linking Bengal Cholera with Terrestrial Water Storage Measured from Gravity Recovery and Climate Experiment Sensors

Antarpreet Jutla,* Ali Akanda, Avinash Unnikrishnan, Anwar Huq, and Rita Colwell*

Department of Civil and Environmental Engineering, West Virginia University, Morgantown, West Virginia; Department of Civil and Environmental Engineering, University of Rhode Island, Kingston, Rhode Island; Department of Civil and Environmental Engineering, Portland State University, Portland, Oregon; Maryland Pathogen Research Institute, University of Maryland, College Park, Maryland; Center for Bioinformatics and Computational Biology, University of Maryland, College Park, Maryland; Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland

Abstract. Outbreaks of diarrheal diseases, including cholera, are related to floods and droughts in regions where water and sanitation infrastructure are inadequate or insufficient. However, availability of data on water scarcity and abundance in transnational basins, are a prerequisite for developing cholera forecasting systems. With more than a decade of terrestrial water storage (TWS) data from the Gravity Recovery and Climate Experiment, conditions favorable for predicting cholera occurrence may now be determined. We explored lead–lag relationships between TWS in the Ganges–Brahmaputra–Meghna basin and endemic cholera in Bangladesh. Since bimodal seasonal peaks in cholera in Bangladesh occur during spring and autumn seasons, two separate logistical models between TWS and disease time series (2002–2010) were developed. TWS representing water availability showed an asymmetrical, strong association with cholera prevalence in the spring ($\tau = -0.53$; $P < 0.001$) and autumn ($\tau = 0.45$; $P < 0.001$) up to 6 months in advance. One unit (centimeter of water) decrease in water availability in the basin increased odds of above normal cholera by 24% (confidence interval [CI] = 20–31%; $P < 0.05$) in the spring, while an increase in regional water by 1 unit, through floods, increased odds of above average cholera in the autumn by 29% (CI = 22–33%; $P < 0.05$).

INTRODUCTION

Cholera remains a global threat, notably in coastal regions where the disease is endemic.¹ Cholera in the Bengal Delta, Indus River basin, and Congo basin, has been strongly associated with variability in large-scale processes, for example, availability of water in rivers.² Annual bimodal peaks of cholera³ in the Ganges–Brahmaputra–Meghna (GBM) basin are related to low flow during months of minimal rainfall when seawater intrusion occurs, carrying cholera bacteria associated with zooplankton⁴ from the coast to inland water bodies. In contrast, high river flow after heavy rains inundates large terrestrial landscapes, creating conditions favorable for cholera with cross-contamination, and the resulting mixing of bacteria-laden water spread over vast spatial scales.^{1–3,5–11} Growth of *Vibrio cholerae*, the causative agent for the disease, in riverine, estuarine, and coastal waters has been shown to be strongly linked to a set of biological parameters namely nutrients in water, temperature, pH, salinity, plankton, and zooplankton^{4,12} that are further associated with large-scale climatic processes.^{13,14} Because cholera bacteria are autochthonous to the aquatic environment and eradication of the pathogen is not possible, forecasting conditions favorable for growth of the cholera bacteria in the environment is achievable if relevant environmental conditions can be linked to large-scale geophysical phenomena.

With reliable estimates of river discharge, appropriate intervention strategies can be devised and implemented, especially when water and sanitation infrastructure is fragile and/or damaged by a natural disaster. Cycles of frequent and unexpected flooding and drought occur in the third largest freshwater basin formed by the two rivers, the Ganges and the Brahmaputra. These cycles create a pattern of hydroclimatic processes that are related to the occurrence of cholera in the Bengal Delta.

Because the GBM basin is extremely complex with water shared among three major countries: India, China, and Bangladesh, developing a holistic understanding of hydroclimatic processes is difficult, especially with the lack of data sharing at upstream locations in the basin. Although most of the Ganges basin falls within the jurisdiction of India, the Brahmaputra River flows through Chinese territory and both rivers enter Bangladesh at Hardinge Bridge and Bahadurabad, respectively. Therefore, without the required data, long-term forecasting of river discharge, particularly low flows, remains a challenge. A similar lack of upstream data poses a serious challenge to understand flows in the basin and, therefore, an even greater challenge to develop a forecasting system for advance warning of cholera.

Previously published studies have focused on flooding in Bangladesh caused by the two rivers, since 80–90% of the flow in both rivers occurs during monsoon season (June–October). Droughts or low flows in November through May have not been analyzed previously. Yet, within the context of conditions predictive of cholera in the region, low flow water availability has been associated with limiting bacterial intrusion into the inland water system.³ With the availability of more than a decade of data from new satellite sensor, Gravity Recovery and Climate Experiment (GRACE) Forecasting useful information on estimates on changes in terrestrial water storage (TWS), which represents an integrated assessment of water stored in different hydrological reservoirs, is possible. TWS is the sum of storage changes in wetlands, rivers, groundwater, and soil moisture.^{15,16} It provides a valuable indicator of changes that occur in hydrological conditions at the level of large basin scales.^{1,7,8} TWS can be used to estimate river discharge,^{18–20} groundwater abundance and depletion,¹⁷ hydrological droughts,²¹ and flooding.²² However, linking TWS with climate-influenced water-borne diarrheal diseases has not yet been done, especially to establishing association of TWS with cholera, a disease that is in turn strongly linked with low water availability in the basin. If advance warning of conditions associated with outbreaks can be

*Address correspondence to Antarpreet Jutla or Rita Colwell, Department of Civil and Environmental Engineering, West Virginia University, P.O. 6013, Morgantown, WV 26506. E-mails: asjutla@mail.wvu.edu or rcolwell@umiacs.umd.edu

provided, it would represent a significant public health benefit and if satellite observations provide the data, the need for ground-based monitoring would be eliminated.

The risk of a cholera outbreak is high if both “hydroclimatological risk” (defined as a combination of large-scale geophysical processes and conditions optimum for bacterial growth) and “societal risk” (defined as unavailability of safe drinking water and proper sanitation) coincide. Societal risk is key to spread of disease in a population, but difficult to characterize primarily due to nonavailability of data to quantify at various spatial scales. Hydroclimatological risks can be determined from the association of geophysical conditions and cholera. The hypothesis, therefore, is that downstream water availability, in the form of river flow, is a complex function of upstream total water, including water stored in soil, wetlands, below surface as groundwater, and rivers, and can be estimated from data on upstream water conditions. The goal of this study was to determine hydroclimatological risk of cholera in Bangladesh from the association of water availability and cholera prevalence time series and to define maximum lead time from TWS for prediction of conditions favorable for cholera.

DATA AND METHODS

River discharge data. Monthly river discharge data from April 2002 through December 2010 for the Ganges and Brahmaputra rivers were obtained from the Institute of Water Modeling, Dhaka, Bangladesh. The gages were located at the Hardinge Bridge (Ganges) and Bahadurabad (Brahmaputra) where the two rivers enter Bangladesh from India. Following methodology proposed by Akanda and others,³ combined river flows were calculated as the sum of the flows in the two rivers.

TWS data. GRACE sensors provided TWS data²³ that were obtained from National Aeronautics and Space Administration's Jet Propulsion Laboratory servers (ftp://podaac-ftp.jpl.nasa.gov/allData/tellus/L3/land_mass/RL05/netcdf/; accessed September 16, 2014) for the same period as the river discharge data (April 2002 to December 2010). TWS represents total water available, which is the sum of groundwater, surface water, and soil moisture. The data were available on approximately 100-km grids over entire landmass and was processed for the GBM basin covering, source and travel of two rivers,²⁴ between longitudes of 78.00–94.00° E and latitudes of 23.75–30.50° N. The monthly time series was obtained after calculating the average of all pixels of TWS in the domain.

Disease data. Cholera prevalence data, on a monthly scale, were obtained from the International Center for Diarrheal Disease Research, Bangladesh (ICDDR,B) hospitals in Dhaka. Cholera data from ICDDR,B are perhaps one of the longest datasets available.²⁵ Cholera prevalence is recorded as the percent of new cholera-infected patients among a statistical subset of patients (every 50th patient enrolled in the hospital maintained by ICDDR,B) selected from a total pool of patients visiting the hospital for treatment during a given week.²⁵ Although more than 30 years of data were available, only a subset of data that overlapped with satellite data availability was used in the study.

Statistical analysis. We examined the exploratory relationships using correlation analysis supported by models for each

data variables. Correlation values were calculated using non-parametric Kendall tau (τ).²⁶

Model between cholera and TWS. Binomial logistical regression models were developed to estimate the probabilistic likelihood of above- and below-average cholera prevalence. On the basis of previous studies on hydroclimatological association of cholera with river discharge,^{3,11} separate models were developed for the spring (February, March, April, and May) and autumn (July, August, September, and October) seasons. Output was compared (detailed description is available in reference²⁷) using three measures of association: percentage of concordant pairs, Somer's D, and Goodman–Kruskal Gamma, as well as goodness of fit (Hosmer–Lemeshow and Pearson χ^2). Measure of association established relationships between response variable (above-average cholera incidence) and predicted probabilities of the variables, showing the effect of predictors (TWS) on seasonal peaks. Goodman–Kruskal Gamma is a nonparametric rank correlation and ranges between -1 and 1 , with 1 representing perfect association. Somer's D is an ordinal measure of association through rank statistics and is an extension of the Kendall τ rank correlation,²⁸ with a value of 1 indicating perfect association. Hosmer–Lemeshow is a χ^2 test that assesses whether observed events match the expected output event in subgroups of a model population.²⁹

Model between river discharge and TWS. Bayesian linear regression^{30,31} models were developed between river discharge and TWS at different lead times. Bayesian analysis^{30,32,33} was chosen over the classical regression since it provided probabilistic estimate of parameters, as well as credible confidence intervals (CIs)³¹ based on sample simulations. The general form of the model is $Q_1 = \alpha + \beta \text{ TWS} + \epsilon$ where Q_1 is the natural log of river discharge time series, α and β are the model parameters assumed to be normally distributed and ϵ is the normally distributed error in the model. The model was run 10,000 times using the Markov chain Monte-Carlo approach implemented in the Gibbs sampler.^{31,33} The first 2,000 samples were discarded and performance statistics (mean, Pearson correlation [R^2], bias, root mean square error, and Nash–Sutcliffe efficiency [NSE]¹¹) between observed and simulated river discharges were calculated.

RESULTS

Figure 1A shows the seasonality of cholera, computed using average of prevalence values for each month, indicating that the disease has a bimodal pattern in the region. TWS, on the other hand, revealed a unimodal pattern, with one seasonal peak in the month of June. Negative values of TWS represent low water availability than average and vice versa. October through March yielded negative TWS values whereas April through August showed positive values, indicating more available water mass than average. Figure 1B shows the visual representation of cholera and TWS time series, normalized between 0 and 1. Kendall τ correlation between monthly time series of cholera and TWS was not statistically significant ($\tau = 0.14$), as anticipated. Therefore, we divided the time series into two seasons, spring and autumn, based on observed seasonality of cholera (Figure 1A).

Spring cholera. Our previous study indicated 2-month lead time between river discharge and spring cholera.³ Therefore, we expect that the lead time between seasonal cholera and

FIGURE 1. (A) Annual seasonality of cholera in Dhaka, Bangladesh, and terrestrial water storage (TWS) anomaly for the Bengal delta and (B) time series of cholera prevalence (%; black line) and TWS (millimeter water; gray line) from April 2002 to December 2009. The data were normalized between 0 and 1 for comparison.

TWS to be more than 2 months. Table 1 shows a statistically significant correlation ($\tau = -0.533$; $P < 0.001$) between TWS and cholera with a 6-month lead time (model ST6). The correlation value declined as the lead time decreased to 2 months. We observed a positive correlation between TWS and cholera at 1-month lead time (model ST1 $\tau = 0.410$; $P < 0.05$). Logistical regression models for 6-month lead time indicated that 1 unit (centimeter) decrease in available water increased the odds for above-average cholera in the spring season by 26% (95% CI = 20–31) and by 30% (95% CI = 25–36) for a 5-month lead time. Goodness of fit metrics, Pearson χ^2 , and Hosmer–Lemeshow tests showed values within the range of 0.474–0.602

and 0.609–0.695, for 6- and 5-month models, respectively. Similarly, the measures of association metrics—Somers' D and Goodman–Kruskal Gamma—were high (greater than 0.75) for ST6 and ST5 models, with an overall concordant pairs ranging from 85% to 89%. Figure 2A shows observed cholera and fitted probabilities obtained using regression models, indicating the regression models adequately capture the variability in the time series.

Autumn cholera. A statistically significant positive correlation was observed for the time series of TWS and autumn cholera at 4- ($\tau = 0.311$; $P < 0.05$), 5- ($\tau = 0.510$; $P < 0.05$), and 6- ($\tau = 0.450$; $P < 0.05$) month lead times. A negative

TABLE 1
Performance of spring cholera model with TWS

Model	τ	α	β	Goodness of fit		Measure of association		
				Pearson	HL	Concordant (%)	Somers' D	GKG
ST6	-0.533*	-0.455*	-0.304*	0.474	0.603	89	0.770	0.770
ST5	-0.401*	-2.017*	-0.357*	0.602	0.695	85	0.750	0.770
ST4	-0.260*	-1.794*	-0.227*	0.432	0.107	70	0.380	0.390
ST3	-0.177†	-1.306†	-0.144†	0.370	0.257	60	0.270	0.260
ST2	0.015†	1.678†	0.173†	0.340	0.183	66	0.320	0.210
ST1	0.410*	1.875*	0.274*	0.590	0.168	80	0.650	0.640

α = intercept; β = TWS coefficient; OR = odds ratio; GKG = Goodman–Kruskal Gamma; HL = Hosmer–Lemeshow; total number of data point = 28; TWS = terrestrial water storage; ST6, ST5, ST4, ST3, ST2, and ST1 = spring models at 6-, 5-, 4-, 3-, 2-, and 1-month lead time, respectively.

*Statistically significant at $P < 0.05$.

†Statistically not significant.

FIGURE 2. Radar plot for observed and fitted probabilities using binominal logistic regression including 95% upper (UCI) and lower (LCI) confidence intervals for (A) spring and (B) autumn cholera.

but statistically relevant correlation was obtained at 1-month lead time ($\tau = -0.352$; $P < 0.05$). Logistical regression, when assessing above-average cholera during the autumn season, indicated P values greater than 0.6 for 5- and 6-month lead-time models to determine goodness of fit Table 2. Concordant pairs (greater than 70%) as well as Somers' D and Goodman–Kruskal Gamma (P values > 0.5) were obtained as measure of association between the modeled and predicted cholera prevalence. One unit (centimeter) increase in available water increased the odds of above-average cholera in the autumn season by 29% (95% CI = 20–31) for 6-month and 12% (95% CI = 10–19) for 5-month lead times. Figure 2B shows probabilities for above- and below-average cholera incidence with CIs using logistical regression model were satisfactorily simulated when compared with observed data.

TWS and river discharge. Nonparametric Kendall τ correlation coefficients (Figure 3) were found to be statistically significant up to a 3-month lead time between a real averaged TWS and combined discharge with highest value obtained at 2-month lead time ($\tau = 0.791$; $P < 0.001$), after which the correlation dropped significantly. A strong correlation for combined discharge was not unexpected since coarse resolution of GRACE provides TWS anomalies that are estimated more reliably on regional scales than on localized small scales. Model statistics (Table 3) between observed and simulated river discharge, obtained from the Bayesian linear regression, showed strongest correlation between observed and simulated flow values at 2-months lag ($R^2 = 0.733$, NSE = 0.700). Furthermore, statistically significant values were also obtained with the 3-month lag model. Figure 4 shows the upper (97.5%) and lower (2.5%) quartiles as well as mean simulated (50% quartile) and observed river flow time series for the combined river discharge model at 2-month lead time. It should be noted that calculation of quartiles and interpretation are different from the classical approach. The dashed lines in Figure 4 indicate probability of simulated discharge in the 95% credible probabilistic interval, provided TWS and observed discharge follow normal distribution. Figure 5A shows results for 10,000 samples (after the 2,000 burn-in was removed), indicating samples were randomly drawn from distribution. Figure 5B presents probability density functions (PDFs) for parameter β for combined discharge model (RD2). Contribution of regression coefficients in the Bayesian analysis can be approximated by the probability value from the PDFs such that the ratio of samples lie to the left or right of the zero of the total number of iterations.³³ All sample values were to the right of zero (Figure 5B) indicating TWS

TABLE 2
Performance of autumn cholera model with TWS

Model	τ	α	β	Goodness of fit		Measure of association		
				Pearson	HL	Concordant (%)	Somers' D	GKG
AT6	0.451*	1.688*	0.167*	0.655	0.601	80	0.600	0.600
AT5	0.510*	0.510*	0.088*	0.509	0.602	77	0.550	0.550
AT4	0.311*	-0.184*	0.088*	0.380	0.461	74	0.500	0.500
AT3	0.039‡	-0.234‡	0.058‡	0.330	0.229	60	0.180	0.180
AT2	0.025‡	1.589‡	-0.074‡	0.293	0.369	66	0.330	0.330
AT1	-0.302†	1.330‡	-0.085‡	0.257	0.466	70	0.420	0.420

α = intercept; AT6, AT5, AT4, AT3, AT2, and AT1 = autumn models at 6-, 5-, 4-, 3-, 2-, and 1-month lead time, respectively; β = TWS coefficient; OR = odds ratio; GKG = Goodman–Kruskal Gamma; HL = Hosmer–Lemeshow; total number of data point = 28; TWS = terrestrial water storage.

*Statistically significant at $P < 0.05$.

†Statistically significant at $P < 0.10$.

‡Statistically not significant.

FIGURE 3. Kendall tau (τ) correlation between river discharge and terrestrial water storage (TWS) up to 5-month lead time.

FIGURE 4. Observed and simulated flow for combined rivers for 2-month lead time. (LCI and UCI are lower [2.5%] and upper [97.5%] quartiles).

had a true positive association with river discharge. Table 4 shows Bayesian analysis of the mean, standard deviation, and quartile ranges of the model parameters (α and β). It also shows Monte-Carlo error (MC error), an indication of variation in the mean of the parameter after simulation over 10,000 samples. Ideally, MC errors should be smaller than the standard deviation of the parameter suggesting model parameters are calculated with adequate precision and is less than 0.0005% of the standard deviations in all of the simulations.

DISCUSSION AND CONCLUSION

Results of this study indicate that the cholera prevalence time series for Dhaka is strongly associated with water availability in the upstream regions. The hypothesis that upstream water conditions in a large basin are related to cholera that occur downstream has been validated since bimodal peaks in cholera were found to be statistically related with lead time up to 5–6 months to TWS anomalies in the GBM basin. Clearly this correlation may not be directly linked to geophysical processes. Previous research linking cholera with hydroclimatic processes have convincingly shown that, to understand a cholera in the Bengal Delta, seasonality of the disease must be separately investigated.^{3,11,34} In summary, Figure 6 presents a schematic hydroclimatological cascade linking cholera with GRACE sensor data in which low and high flows can be interpreted as creating favorable conditions for growth of cholera bacteria but each for different reasons. Both high and low river flows were found to be statistically related with a lead time of at least 2 months for autumn and spring cholera peaks.³ Results of this study show that river discharge in

the basin is strongly influenced by TWS anomalies that occur 2–3 months in advance. Therefore, it is statistically conceivable that TWS is correlated with cholera for at least 5–6 months in advance, without having to determine river discharge in the Ganges and Brahmaputra rivers. This finding is informative since it eliminates the need to collect discharge data in a highly contested region where water is crucial and sharing water can be a contentious, if not volatile issue.

A negative correlation between spring cholera peak and water anomalies through GRACE at 5- to 6-month lead time is observed. Logistical models at both lead times appeared to produce statistically viable results from which average cholera in the spring season is likely to increase with a decrease in water availability. This indirect measure of water availability using GRACE is, in fact, the sum of all water sources, including groundwater and surface water. During low-flow months, the groundwater component of TWS is likely to play a role in the observed relationship. However, a negative anomaly in TWS may be due to decreased surface water availability upstream of the basin. In any case, growth of bacteria under low-flow conditions in rivers that have limited seawater intrusion has been confirmed previously.³ Also noted was a positive correlation between TWS and spring cholera at 1-month lead time. The previous studies^{3,8} used ca. 15 years of cholera data (1985–1999) and linked those data with river discharge, observing no relationship between spring outbreaks and low flows at 1-month lead time. Therefore, the positive correlation in model ST1, may either be circumstantial or a new relationship emerging in Bengal cholera. One explanation may be that the number of bacteria required to initiate an outbreak may reach the threshold owing to changes in climatic

TABLE 3
Model statistics for observed and simulated combined river discharge at different lead times

Model	Observed mean	Simulated mean	R^2	Bias	RMSE	NSE
RD0	30,340.433	21,982.035	0.076	8,358.398	26,638.401	0.029
RD1	30,550.059	26,171.555	0.372	4,378.504	21,568.640	0.327
RD2	30,714.627	29,683.108	0.733	1,031.519	10,832.200	0.700
RD3	30,756.599	26,426.267	0.580	4,330.332	17,721.008	0.553
RD4	30,447.377	21,855.908	0.136	8,591.469	26,080.425	0.028
RD5	30,064.952	20,181.398	0.010	9,883.555	28,032.432	0.136

NSE = Nash–Sutcliffe efficiency; RMSE = root mean square error.
Unit for mean, bias, and RMSE is cusecs; NSE range between -1 and 1.

FIGURE 5. (A) Plot of 10,000 samples of parameter β RD2 model and (B) estimated posterior probability density function (PDF) for parameter β combined river discharge model at 2-month lead time.

patterns. Population and urbanization in and around Dhaka city in Bangladesh is rapidly increasing, along with pollution of the regional riverine systems, putting stress on water availability. Using discharge data from 1985 to 2009, no relationship was found between low flows and spring cholera at 1-month lead time, which may indicate that the observed positive correlation in model ST1 may be spurious. As more data on TWS become available, the relationship should be investigated further. The positive relationship found with subsequent logistical model results indicate variability in TWS during seasons with high flows linked to the autumn cholera peak at 5–6 months lead time. Previous studies^{3,8} have also reported positive association of river discharge with autumn cholera at least 1 month in advance. GRACE TWS represents terrestrial water in the region, that is, the sum of groundwater, surface free-flowing water, and soil moisture storage. It cannot be assumed that variability of all water storage reservoirs will be essentially in the same phase, since the regional hydroclimatology indicated the months of June through October are dominated by heavy flow in the rivers.²⁴ Therefore, TWS during high river discharge may derive from surface water, while low-flow months represent a combination of river flow and other components of water storage in the region. TWS strongly influences river discharge, as evidenced by studies of several large basins.^{19,20} For the GBM basin, it has been shown that travel time for water from source (in the upper Himalayas) to gage locations where river discharge measurements are taken, is more than a month,²⁴ which confirms the strength of lagged correlations between combined discharge from the Ganges and Brahmaputra rivers. It must be noted that traditional forecasting of river flow in the GBM basin has been focused on determining links between large-scale teleconnections, such as El Niño Southern Oscillation (ENSO)^{35–37} and on 7- to 15-day lead times.^{24,38} Using Bayesian regression techniques, models with up to a 5-month lead time have been developed for combined discharge by incorporating all upstream pixels of the basin. Nevertheless, forecasting of river discharge can be physically improved by including sea surface

temperature (either localized or ENSO) affecting different modalities of the monsoon in the region, which is beyond the scope of this study. In addition, GRACE satellite data provide computed estimates on temporal variations in gravity fields of earth and with all remote sensing products, data from GRACE satellites may experience delays during acquisition, processing, and interpretation. In addition, TWS is an indirect measure that is linked with cholera incidence, and although it is the key contribution from results of this article, yet cautions should be exercised in physical interpretation of statistical relationships. The results reported cannot be interpreted as a direct causal mechanism, since actual counts of *V. cholerae* in the aquatic system in Dhaka region were not done, but the findings support a role for large-scale hydroclimatic processes in cholera outbreaks, indicating the potential for predicting conditions favoring cholera outbreaks. The disease burden can be significantly reduced if prevention measures are implemented proactively, actually well before outbreaks occur,³⁹ such as water purification to provide safe drinking water and improved sanitation to reduce cholera.⁷ For example, implementation of water, sanitation, and hygiene infrastructures in vulnerable localities will reduce cholera prevalence significantly, as shown in Haiti, bringing the overall case-fatality ratio to less than 2% from 12% during the initial weeks of the cholera epidemics. Large-scale hydroclimatological processes set appropriate stage for changes in the local to regional water environment that encompasses human access to safe water and sanitation, interaction with aquatic medium, and water availability—collectively categorized as societal risks for outbreak of cholera. These risks may be perceived as the optimal pathways for water contamination and pathogen transmission in a given human population. Data on access to safe water and sanitation access are not generally routinely collected (or made available) by health organization(s). World Health Organization/United Nations Children's Fund started a pilot program on collection of data on sanitation through its Joint Monitoring Program (<http://www.wssinfo.org/>), however, such data are available on aggregated levels of country scales, rendering it unusable for

TABLE 4
Bayesian analysis results for the discharge model at 2-month lead time (RD2)

	Model	Mean	Standard deviation	MC error	2.5% Percentile	Median	97.5% Percentile
α _RD2	RD2	4.308	0.010	< 0.000001	4.269	4.308	4.347
β _RD2		0.030	0.001	< 0.000001	0.027	0.030	0.033

α , β = model parameters; MC error = Monte-Carlo error.

FIGURE 6. Pathway to link terrestrial water storage (TWS) to predict conditions for cholera outbreaks 5–6 months in advance.

understanding local scale disease outbreaks. A dedicated effort should be made to recognize interaction of human population with water in regions that are susceptible to disease outbreaks, so that appropriate intervention strategies may be implemented.

The grand challenge that remains is a cholera prediction model for epidemic regions, whereby environmental conditions and transmission mechanisms are incorporated to provide accurate and actionable prediction. When either a major natural disaster strikes a region (town/province) or civil disorder damages living conditions of a population center, cholera tracking mechanism, employing satellite remote sensing of climate and environment parameters in the geographic area where a vulnerable population resides, must be the part of a public health early warning system and strategy preparation team armamentarium.

Received October 17, 2014. Accepted for publication September 8, 2015.

Published online November 2, 2015.

Acknowledgment: This research is funded in part from a NASA grant (NNX15AF71G) and WV NASA space grant program.

Authors' addresses: Antarpreet Jutla, Department of Civil and Environmental Engineering, West Virginia University, Morgantown, WV, E-mail: asjutla@mail.wvu.edu. Ali Shafqat Akanda, Department of Civil and Environmental Engineering, Tufts University, Medford, MA, E-mail: akanda@egr.uri.edu. Avinash Unnikrishnan, Department of Civil and Environmental Engineering, Portland State University, Portland, OR, E-mail: avinash.unnikrishnan@mail.wvu.edu. Anwar Huq, Maryland Pathogen Research Institute, University of Maryland, College Park, MD, E-mail: huqanwar@gmail.com. Rita Colwell, Maryland Pathogen Research Institute, University of Maryland, College Park, MD, Center for Bioinformatics and Computational Biology, University of Maryland, College Park, MD, and Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, E-mail: rcolwell@umiacs.umd.edu.

REFERENCES

1. Jutla A, Whitcombe E, Hasan N, Haley B, Akanda A, Huq A, Alam M, Sack RB, Colwell R, 2013. Environmental factors influencing epidemic cholera. *Am J Trop Med Hyg* 89: 597–607.
2. Jutla AS, Akanda AS, Islam S, 2010. Tracking cholera in coastal regions using satellite observations. *J Am Water Resour Assoc* 46: 651–662.
3. Akanda AS, Jutla AS, Islam S, 2009. Dual peak cholera transmission in Bengal Delta: a hydroclimatological explanation. *Geophys Res Lett* 36: L19401.
4. Huq A, Small EB, West PA, Huq MI, Rahman R, Colwell RR, 1983. Ecological relationships between *Vibrio cholerae* and planktonic crustacean copepods. *Appl Environ Microbiol* 45: 275–283.
5. Akanda AS, Jutla AS, Gute DM, Sack RB, Alam M, Huq A, Colwell RR, Islam S, 2013. Population vulnerability to biannual cholera outbreaks and associated macro-scale drivers in the Bengal Delta. *Am J Trop Med Hyg* 89: 950–959.
6. Akanda A, Jutla A, Gute D, Evans T, Islam S, 2012. Reinforcing cholera intervention through prediction-aided prevention. *Bull World Health Organ* 90: 243–244.
7. Akanda AS, Jutla AS, Colwell RR, 2014. Global diarrhoea action plan needs integrated climate-based surveillance. *Lancet Glob Health* 2: e69–e70.
8. Akanda AS, Jutla AS, Alam M, de Magny GC, Siddique AK, Sack RB, Huq A, Colwell RR, Islam S, 2011. Hydroclimatic influences on seasonal and spatial cholera transmission cycles: implications for public health intervention in the Bengal Delta. *Water Resour Res* 47: W00H07.
9. Jutla AS, Akanda AS, Griffiths JK, Colwell R, Islam S, 2011. Warming oceans, phytoplankton, and river discharge: implications for cholera outbreaks. *Am J Trop Med Hyg* 85: 303–308.
10. Jutla AS, Akanda AS, Islam S, 2012. Satellite remote sensing of space-time plankton variability in the Bay of Bengal: connections to cholera outbreaks. *Remote Sens Environ* 123: 196–206.
11. Jutla AS, Akanda AS, Islam S, 2013. A framework for predicting endemic cholera using satellite derived environmental determinants. *Environ Model Softw* 47: 148–158.

12. Colwell RR, Kaper J, Joseph SW, 1977. *Vibrio cholerae*, *Vibrio parahaemolyticus*, and other vibrios: occurrence and distribution in Chesapeake Bay. *Science* 198: 394–396.
13. Huq A, Sack RB, Nizam A, Longini IM, Nair GB, Ali A, Morris JG, Khan MNH, Siddique AK, Yunus M, Albert MJ, Sack DA, Colwell RR, 2005. Critical factors influencing the occurrence of *Vibrio cholerae* in the environment of Bangladesh. *Appl Environ Microbiol* 71: 4645–4654.
14. Lobitz B, Beck L, Huq A, Wood B, Fuchs G, Faruque ASG, Colwell R, 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.
15. Lee H, Beighley RE, Alsdorf D, Jung HC, Shum CK, Duan J, Guo J, Yamazaki D, Andreadis K, 2011. Characterization of terrestrial water dynamics in the Congo Basin using GRACE and satellite radar altimetry. *Remote Sens Environ* 115: 3530–3538.
16. Swenson S, Wahr J, Milly PCD, 2003. Estimated accuracies of regional water storage variations inferred from the Gravity Recovery and Climate Experiment (GRACE). *Water Resour Res* 39: 1223.
17. Rodell M, Velicogna I, Famiglietti JS, 2009. Satellite-based estimates of groundwater depletion in India. *Nature* 460: 999–1002.
18. Reager JT, Thomas BF, Famiglietti JS, 2014. River basin flood potential inferred using GRACE gravity observations at several months lead time. *Nat Geosci* 7: 588–592.
19. Alsdorf D, Han SC, Bates P, Melack J, 2010. Seasonal water storage on the Amazon floodplain measured from satellites. *Remote Sens Environ* 114: 2448–2456.
20. Syed TH, Famiglietti JS, Chambers DP, 2009. GRACE-based estimates of terrestrial freshwater discharge from basin to continental scales. *J Hydrometeorol* 10: 22–40.
21. Awange JL, Forootan E, Kuhn M, Kusche J, Heck B, 2014. Water storage changes and climate variability within the Nile Basin between 2002 and 2011. *Adv Water Resour* 73: 1–15.
22. Crowley JW, Mitrovica JX, Bailey RC, Tamisiea ME, Davis JL, 2006. Land water storage within the Congo Basin inferred from GRACE satellite gravity data. *Geophys Res Lett* 33: L19402.
23. Landerer FW, Swenson SC, 2012. Accuracy of scaled GRACE terrestrial water storage estimates. *Water Resour Res* 48: W04531.
24. Jian J, Webster PJ, Hoyos CD, 2009. Large-scale controls on Ganges and Brahmaputra river discharge on intraseasonal and seasonal time-scales. *QJ Roy Meteorol Soc* 135: 353–370.
25. Longini IM, Yunus M, Zaman K, Siddique AK, Sack RB, Nizam A, 2002. Epidemic and endemic cholera trends over a 33-year period in Bangladesh. *J Infect Dis* 186: 246–251.
26. Jutla AS, Small D, Islam S, 2006. A precipitation dipole in eastern North America. *Geophys Res Lett* 33: L21703.
27. Hosmer DW, Hosmer T, Le Cessie S, Lemeshow S, 1997. A comparison of goodness-of-fit tests for the logistic regression model. *Stat Med* 16: 965–980.
28. Somers RH, 1962. A new asymmetric measure of association for ordinal variables. *Am Sociol Rev* 27: 799.
29. Hosmer D, Lemeshow S, 1989. *Applied Logistic Regression*. New York, NY: Wiley.
30. Clark JS, 2005. Why environmental scientists are becoming Bayesians. *Ecol Lett* 8: 2–14.
31. Ntzoufras I, 2009. *Bayesian Modeling Using WinBUGS*. Hoboken, NJ: Wiley.
32. Ahn J, Mukherjee B, Banerjee M, Cooney KA, 2009. Bayesian inference for the stereotype regression model: application to a case-control study of prostate cancer. *Stat Med* 28: 3139–3157.
33. Chu PS, Zhao X, 2007. A Bayesian regression approach for predicting seasonal tropical cyclone activity over the Central North Pacific. *J Clim* 20: 4002–4013.
34. Jutla A, Akanda AS, Huq A, Faruque ASG, Colwell R, Islam S, 2013. A water marker monitored by satellites to predict seasonal endemic cholera. *Remote Sens Lett* 4: 822–831.
35. Shukla J, Paolino DA, 1983. The southern oscillation and long-range forecasting of the summer monsoon rainfall over India. *Mon Weather Rev* 111: 1830–1837.
36. Torrence C, Webster PJ, 1999. Interdecadal changes in the ENSO–monsoon system. *J Clim* 12: 2679–2690.
37. Whitaker DW, Wasimi SA, Islam S, 2001. The El Niño southern oscillation and long-range forecasting of flows in the Ganges. *Int J Climatol* 21: 77–87.
38. Hirpa FA, Hopson TM, De Groeve T, Brakenridge GR, Gebremichael M, Restrepo PJ, 2013. Upstream satellite remote sensing for river discharge forecasting: application to major rivers in south Asia. *Remote Sens Environ* 131: 140–151.
39. Bartram J, 2008. Flowing away: water and health opportunities. *Bull World Health Organ* 86: 2.