Developing a forecasting model for cholera incidence in Dhaka megacity through time series climate data

Salima Sultana Daisy, A. K. M. Saiful Islam, Ali Shafqat Akanda,
Abu Syed Golam Faruque, Nuhu Amin and Peter Kjær Mackie Jensen

ABSTRACT

Cholera, an acute diarrheal disease spread by lack of hygiene and contaminated water, is a major public health risk in many countries. As cholera is triggered by environmental conditions influenced by climatic variables, establishing a correlation between cholera incidence and climatic variables would provide an opportunity to develop a cholera forecasting model. Considering the auto-regressive nature and the seasonal behavioral patterns of cholera, a seasonal-auto-regressive-integrated-moving-average (SARIMA) model was used for time-series analysis during 2000–2013. As both rainfall ($r = 0.43$) and maximum temperature ($r = 0.56$) have the strongest influence on the occurrence of cholera incidence, single-variable (SVMs) and multi-variable SARIMA models (MVMs) were developed, compared and tested for evaluating their relationship with cholera incidence. A low relationship was found with relative humidity ($r = 0.28$), ENSO ($r = 0.21$) and SOI ($r = -0.23$). Using SVM for a 1 °C increase in maximum temperature at one-month lead time showed a 7% increase of cholera incidence ($p < 0.001$). However, MVM (AIC = 15, BIC = 36) showed better performance than SVM (AIC = 21, BIC = 39). An MVM using rainfall and monthly mean daily maximum temperature with a one-month lead time showed a better fit (RMSE = 14.7, MAE = 11) than the MVM with no lead time (RMSE = 16.2, MAE = 13.2) in forecasting. This result will assist in predicting cholera risks and better preparedness for public health management in the future.

Key words | cholera, climatic variables, Dhaka, forecasting, SARIMA

Salima Sultana Daisy
A. K. M. Saiful Islam (corresponding author)
Institute of Water and Flood Management (IWFM),
Bangladesh University of Engineering and
    Technology (BUET),
Dhaka 1000,
Bangladesh
E-mail: *saiful3@gmail.com*

Ali Shafqat Akanda
Department of Civil and Environmental
    Engineering,
University of Rhode Island,
Kingston, RI 02881,
USA

Abu Syed Golam Faruque
Centre for Nutrition and Food Security,
International Centre for Diarrhoeal Disease
    Research,
Bangladesh

Nuhu Amin
Environmental Intervention Unit, Enteric and
    Respiratory Disease Program, Infectious Disease
    Division,
International Centre for Diarrhoeal Disease
    Research,
Bangladesh

Peter Kjær Mackie Jensen
Copenhagen Center for Disaster Research,
    Department of Public Health,
University of Copenhagen,
Copenhagen,
Denmark

INTRODUCTION

Cholera is an infection of the small intestine; marked by profuse, watery, secretory diarrhea with or without vomiting; caused by the bacterium *Vibrio cholerae*. This can result in acute dehydration and, without treatment, it can even cause death within a few hours (McElroy & Townsend 2014). Cholera bacteria transmission occurs via the fecal-oral route primarily by drinking water or eating food that has been contaminated. Worldwide, about 1.3 billion people are at risk of cholera in endemic countries; an estimated burden of 2.9 million (uncertainty range: 1.3–4.0 million) cases; and 95,000 (uncertainty range: 21,000–143,000) deaths per year in endemic countries as of 2015 (Ali *et al.* 2015). After seven pandemics (spread over a continent) in the last 200 years, cholera remains endemic (≤1% mortality rate) in many developing countries in Asia, Africa and Latin America. The seventh cholera pandemic began in Indonesia in 1961, but the disease has reemerged as a global killer since the 1990s (Kotar & Gessler 2014).

Recently, the mortality rate for epidemic cholera (>3% mortality rate) was recorded as high as 6.4% in 2010 in Haiti, 6% in 2000 in Madagascar, 4.3% in 2008–2009 in Zimbabwe, 4% in 2006–2007 in Angola, 3.8% in 2010 in Nigeria, and 3.3% in 2006–2007 in Sudan (Enserink 2010).

A thorough review on the relationship between cholera incidence and the climatic variables by using different statistical methods in different locations of the world has been summarized in Table 1. In the review, climatic variables rainfall, maximum temperature, minimum temperature, relative humidity, El Niño southern oscillation (ENSO) and southern oscillation index (SOI) were found statistically significant relationship with cholera incidence in different countries of the world. In Africa, rainfall, temperature and seas surface temperature (SST) play significant roles on cholera outbreaks, e.g. in Ghana, rainfall and SOI (De Magny et al. 2006); in Nigeria, mean temperature, rainfall and relative humidity (Leckebusch & Abdussalam 2015); in Senegal, rainfall (de Magny et al. 2012); in Zambia, maximum temperature and rainfall (Luque Fernández et al. 2009); in Zanzibar, minimum temperature and rainfall (Reyburn et al. 2011); and in southeastern Africa (Uganda, Keya, Rwanda, Burundi, Tanzania, Malawi, Zambia, and Mozambique) (Jutla et al. 2015; Mendelsohn & Dawson 2008; Trærup et al. 2010), mean temperature and SST (Paz 2009). In America (Haiti) rainfall plays a vital role (Eisenberg et al. 2013; Righetto et al. 2013). In South America, sea surface temperature and ambient temperature have an effect on cholera outbreaks e.g., in Peru (Checkley et al. 2000; Colwell 1996). In Asia, rainfall, temperature (mean, minimum and maximum), relative humidity, SST, sea surface height, and river discharge influence increasing cholera outbreaks, e.g. in India (Rajendran et al. 2011; Sebastian et al. 2015) and in Bangladesh (Bouma & Pascual 2001; Pascual et al. 2002; Akanda et al. 2009; Islam et al. 2009; Hashizume et al. 2010). Moreover, social risk factors, e.g. poverty, sanitation conditions, and untreated drinking water, play important roles in the transmission and outbreak of cholera (Ali et al. 2002a, 2002b; Charles & Ryan 2011; Reiner et al. 2012).

Cholera outbreak is observed in two seasons in Bangladesh, namely, pre-monsoon (March–May) and post-monsoon (September–November) season (Lipp et al. 2002; Akanda et al. 2009). Generally, cholera outbreaks show dual peaks annually in some parts of Bangladesh (e.g. in Dhaka and Matlab), while single seasonal peaks in other parts (e.g. pre-monsoon peak in Mathbaria in the southwest coast and post-monsoon peak in Chhatak in the northeast flood-prone area) (Akanda et al. 2009; Akanda et al. 2011; Alam et al. 2011; Bertuzzo et al. 2012). Pre-monsoon cholera outbreak in coastal areas of Bangladesh and the capital region of Dhaka is associated with salinity intrusion caused by low flow situations in regional rivers, a surrogate for dry season water scarcity that provides an optimum environment for growth and increased abundance of V. cholerae pathogens (Louis et al. 2003; Vital et al. 2007); while post-monsoon outbreaks in Dhaka and other inland regions have shown strong links to water abundance in flood-prone areas causing seasonal floods all over Bangladesh (Akanda et al. 2009; Jutla et al. 2011). Moreover, the pre-monsoon triggering cause of cholera outbreak includes scarcity of seasonal safe drinking water due to many drinking water sources going dry because of seasonal declination of the groundwater table all over Bangladesh; while in Dhaka city, this water scarcity is due to over-exploitation of groundwater and low river water availability for surface water treatment plants. In laboratory tests, it has been shown that salinity and temperature are important factors for influencing the growth of V. cholerae (Batabyal et al. 2014), and V. cholerae can survive more when aided by copepods (Huq et al. 1983). On the other hand, monsoon floods inundating large inland areas with stagnant water and rain-flushed nutrients provide a growth environment for pathogens (Islam et al. 2007). With the recession of seasonal flood water, available water-borne pathogens including cholera in combination with scarcity of safe drinking water when many drinking water sources are contaminated with flood water cause a second 'post-monsoon' outbreak (Akanda et al. 2009, 2011). Some environmental indicators such as water temperature and water depth in some water bodies in Bangladesh showed a significant lagged correlation with cholera outbreaks (Huq et al. 2005). Moreover, climate variability, for example extreme dry conditions and high temperature leading to droughts, or heavy rainfall leading to floods that occurred caused by ENSO, may lead to enhance cholera outbreaks in the future (Field et al. 2014). Recently Martinez et al. (2017) evaluated the effect of climate covariate ENSO on cholera incidence in Dhaka using two

Table 1 | Studies of relations of cholera incidences and climatic variables with different statistical methods

Variables	Type of statistical analysis	Location; time, reference	Findings
Rainfall; Southern Oscillation Index (SOI)	Cross-correlation analysis	Africa – Ghana; 1975–1995 (De Magny et al. 2006)	Strong statistical association between cholera outbreak and climatic variables under scrutiny
Mean temperature; rainfall; relative humidity	Generalized Additive Modeling (GAM) and Multiple Linear Regression (MLR)	Africa – Nigeria; 1990–2011 (Leckebusch & Abdussalam 2015)	Climatic variables, most especially temperature and rainfall, play an important role in explaining the cholera dynamics
Rainfall	Cross-correlation analysis	Africa – Senegal; May 10–December 31, 2005 (de Magny et al. 2012)	The influence on cholera transmission of the intense rainfall over a densely populated and crowded region was detectable for both Dakar and Thiès, Senegal
Maximum temperature; rainfall	Poisson autoregressive model	Africa – Lusaka, Zambia; 2003–2006 (Luque Fernández et al. 2009)	1 °C rise in temperature 6 weeks before the onset of the outbreak explained 5.2% of the increase in the cholera cases and a 50 mm increase in rainfall 3 weeks before explained an increase of 2.5%
Minimum temperature; rainfall;	SARIMA model	Africa – Zanzibar; 2002–2008 (Reyburn et al. 2011)	1 °C rise in temperature at four months lag resulted in a 2-fold increase of cholera cases, and an increase of 200 mm of rainfall at two months lag resulted in a 1.6-fold increase of cholera cases
Mean temperature; SST	Poisson regression model	Africa – Southeastern Africa; 1971–2006 (Paz 2009)	Annual mean temperature and SST had significant impact on cholera incidence during the studied period
Rainfall	Multivariate Poisson	America – Haiti; October 2010 – December 2011 (Righetto et al. 2013)	A clear correlation between rainfall events and cholera outbreaks
Rainfall	Quantitative analysis using a combination of statistical and dynamic models	America – Haiti; 2010 (Eisenberg et al. 2013)	Increased rainfall was significantly correlated with increased cholera incidence 4–7 days later
Maximum temperature; rainfall; relative humidity	SARIMA model	Asia – Vellore, India; 2000–2010 (Sebastian et al. 2015)	50% decrease of cholera cases from 2000–2004 to 2005–2010. During 2000–2004, there was a positive significant association between rainfall and cholera cases ($r = 0.51$, $p < 0.001$) and this was not observed in 2005–2010
Mean temperature; relative humidity; rainfall	SARIMA and GLM	Asia – Kolkata, India; 1996–2008 (Rajendran et al. 2011)	Cholera was associated higher RH (>80%) with 29 °C temperature with intermittent average (10 cm) rainfall
Rainfall; mean temperature	Poisson regression model	Asia – Dhaka, Bangladesh; 1996–2002 (Hashizume et al. 2008)	Weekly cholera cases increased and decreased by 14 and 24% respectively for 45 ± 10-mm of rainfall over 0–8 and 0–16 weeks lag
ENSO (El Niño/ Southern Oscillation)	Scale-dependent correlation (SDC) analysis, Singular spectrum analysis (SSA)	Asia – Bangladesh; 1980–2001 (Rodó et al. 2002)	A strong and consistent association between cholera levels and ENSO is apparent in the last two decades
Minimum temperature; maximum temperature; rainfall and SST	SARIMA model	Asia – Matlab, Bangladesh; 1988–2001 (Ali et al. 2013)	6% increase in cholera incidence with a minimum temperature increase of 1 °C

models (mechanistic temporal model and statistical spatio-temporal model).

Considering all the studies relevant to Bangladesh and other countries as summarized above, the following research questions are yet to be addressed for the densely populated megacity Dhaka: (1) which climatic variable, single- or multi-variable can predict cholera incidence better? (2) how to develop a cholera forecast model based on this correlation? and (3) is there any location-dependent correlation between cholera incidence and climatic variables, or not? Hence, this study is aimed at addressing the above research questions for better predicting cholera incidence; so that preparedness and emergency response plans can be taken into consideration in a more comprehensive way than at present. For doing so, a regression model seasonal autoregressive integrated moving average (SARIMA) was found suitable for this study which refers a relevant method for time series analysis due to its forecasting capability and better information on time-related changes (Helfenstein 1991).

## DATA AND METHODS

Study area and related data

Dhaka cholera incidence data has been used for this study because: (i) Dhaka is at high risk of endemic cholera because of the high population density as well as seasonal flooding and proximity to the Bay of Bengal; (ii) Dhaka exhibited a dual peak of cholera incidence annually like Matlab, for which a similar study was conducted by Ali *et al.* (2013) where a single variable SARIMA model with minimum temperature and SST was found to be triggering cholera outbreaks (Table 1) and rainfall did not influence cholera; and (iii) long-time series (2000–2013) continuous cholera incidence data is available from the International Centre for Diarrheal Disease Research, Bangladesh (icddr,b), which receives most cholera patients in and around Dhaka megacity (Figure 1).

Three-hourly local meteorological data of Dhaka station, such as rainfall, maximum temperature, minimum temperature and relative humidity, by the Bangladesh Meteorological Department (BMD) was summarized as

monthly data for the last 14 years from January 1, 2000 to December 31, 2013. The weather station is located at latitude 23°46′N, longitude 90°23′E. There was no missing data during the study period. In Dhaka district there are two meteorological stations by BMD, one at Tejgaon which is at the center of Dhaka city and another at Dhaka international airport located in the northern part of Dhaka. The Tejgaon station's data is used in this study as this is more reliable than the airport station's data as suggested by BMD. Satellite-derived ENSO (source: www.cpc.ncep.noaa.gov/data/indices/sstoi.indices) and SOI (source: www.cpc.ncep.noaa.gov/data/indices/soi) at the index Niño 3.4 were also used in this study (Niño 3.4 (5N–5S, 170 W–120 W) anomalies may be thought of as representing the average equatorial sea surface temperatures (SSTs) across the Pacific from about the dateline to the South American coast. The Niño 3.4 index typically uses a five-month running mean, and El Niño or La Niña events are defined when the Niño 3.4 SSTs exceed ±0.4 °C for a period of six months or more. For more information please refer to https://climatedataguide.ucar.edu/climate-data/nino-sst-indices-nino-12-3-34-4-oni-and-tni).

Dhaka is a densely-populated megacity with a population of 14.2 million in 2011; the 2001 population was 9.7 million, therefore the decadal growth rate was 46% (BBS 2014). Dhaka has a tropical wet and dry climate which has a distinct monsoonal season with an annual average temperature of 26.1 °C and rainfall of 2,149 mm based on the last 30 years of data (WB 2017). During January 2000 to December 2013, the mean (± SD) monthly rainfall, maximum temperature, minimum temperature, and relative humidity of 164 (±177) mm, 30.7 (±3.3) °C, 22.1 (±4.5) °C, and 82 (±6.4)%, respectively indicate that Dhaka generally has warm and humid weather.

Laboratory-confirmed cholera patients' data of the icddr,b were used for the same period of January 2000 to December 2013. icddr,b is commonly known as the 'cholera hospital', where the most severely affected cholera patients in and around Dhaka megacity come for treatment. Therefore, the cholera patients received at icddr,b are assumed to be representative of all the cholera patients in Dhaka megacity. During this period, out of all diarrheal patients recorded in icddr,b, 19% were identified as cholera patients (5,939 out of 30,984). Of all the cholera patients,
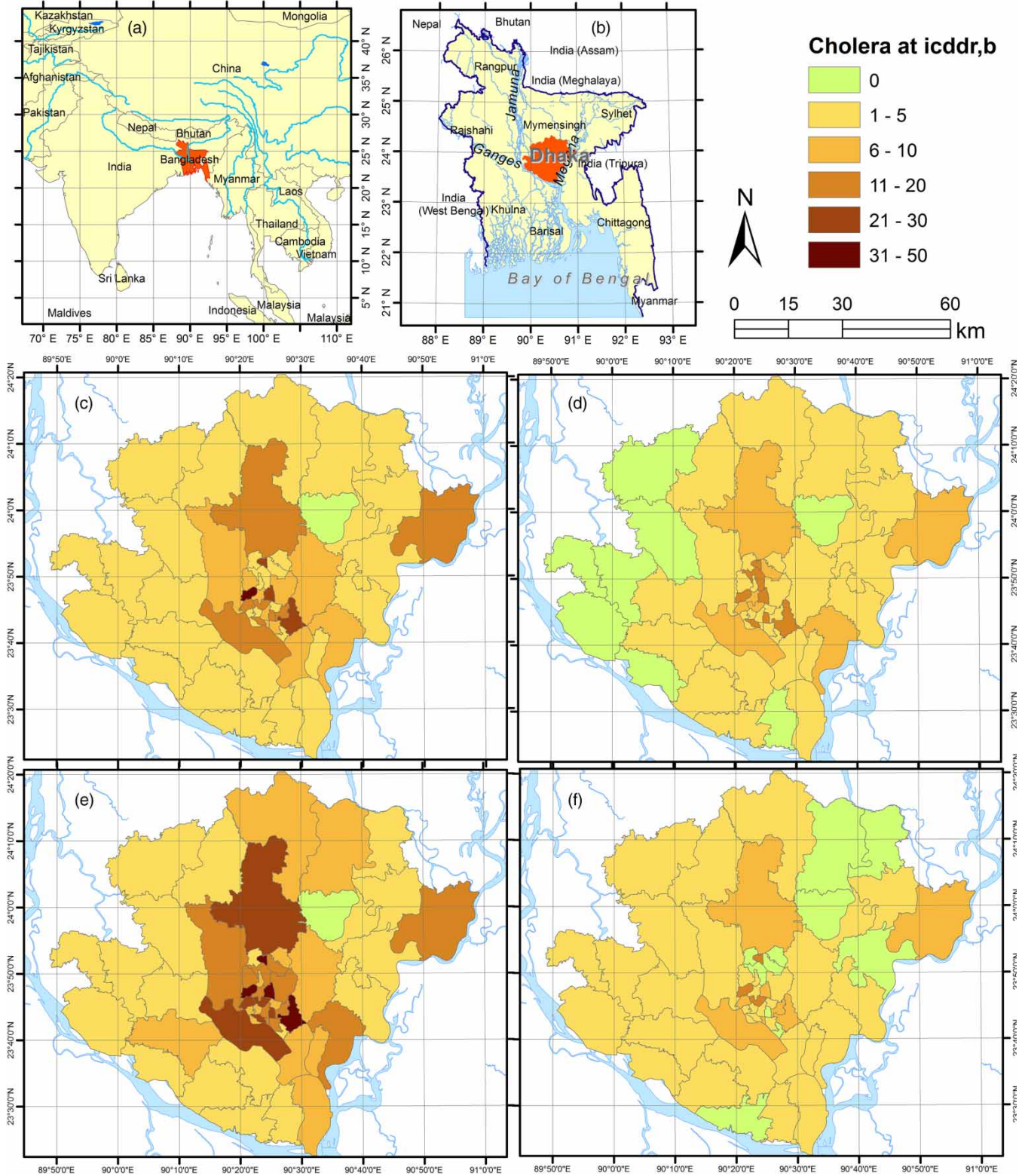
Figure 1 | (a) Location map of Bangladesh; (b) Location map of Dhaka; (c) Mean cholera incidence for 2000–2013 in and around Dhaka, admitted at icddr,b; (d) standard deviation of cholera incidence; (e) mean value with (plus) one standard deviation of cholera incidence; and (f) mean value without (minus) one standard deviation of cholera incidence.
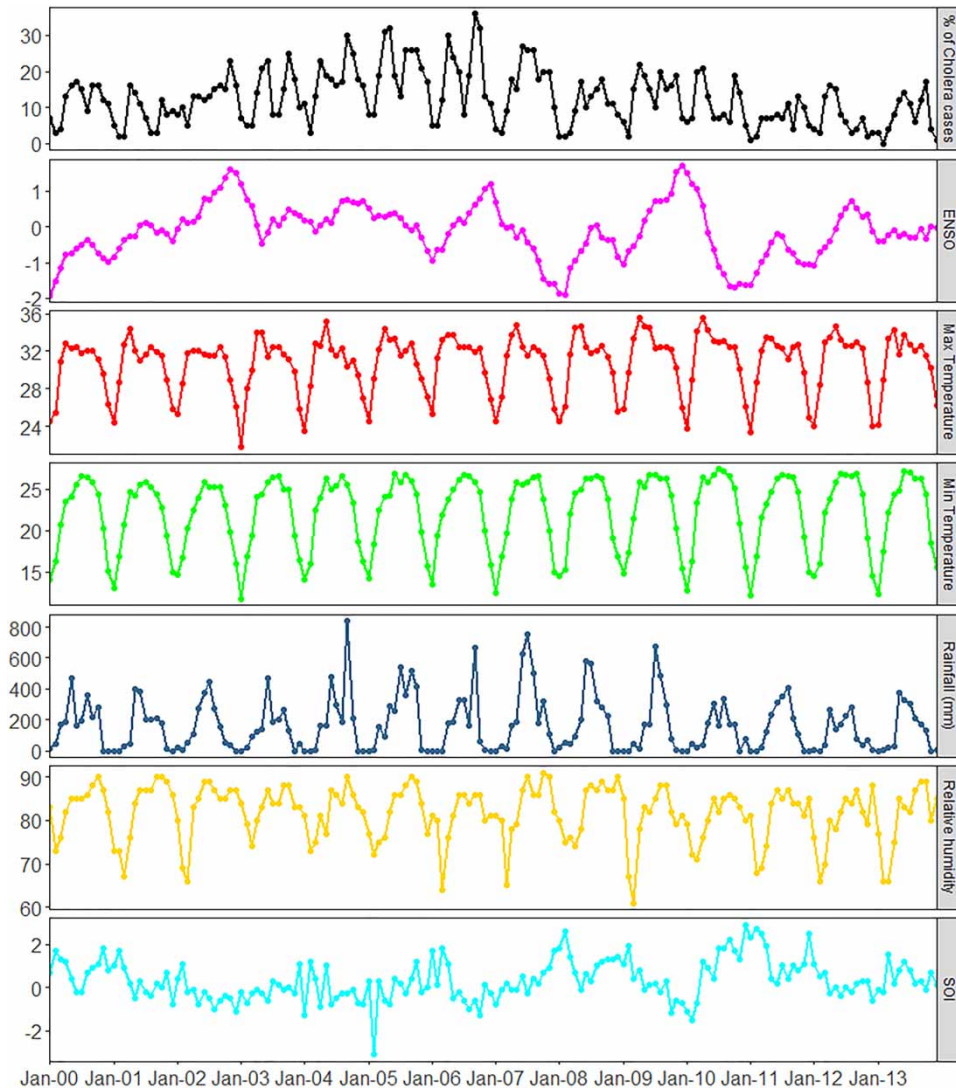
Figure 2 │ Time series plot of percentage (%) of cholera cases obtained from total patients visiting in icddr,b, Dhaka, ENSO, average monthly maximum temperature (°C), minimum temperature (°C), total rainfall (mm), relative humidity (%),and SOI during 2000–2013.

55% were less than five years of age, while 39% were greater than 15 years and the rest (6%) were between 5 and 15 years. A greater number of cholera patients was male (56%), while the rest (44%) were female. The average monthly cholera patients recorded in icddr,b and related climatic variables during 2000–2013 depicted the well-known bimodal distribution (Figure 2) pattern over an annual cycle: one occurred in the months of March–May (pre-monsoon) and the other in the months of September–November (post-monsoon). The pre-monsoon peak was higher than the post-monsoon peak during the studied period.

Evolution to SARIMA modelling

The overall flow chart of how SARIMA models were developed is shown in Figure 3. First, to investigate the delayed effects on cholera incidence, climate variables were temporally lagged by 0, 1, 2 and 3 months by cross correlation analysis. When one or more lagged associations of climatic variables with cholera incidence were found, it was then identified to be useful for SARIMA modeling. The Box–Jenkins modelling approach (Box *et al.* 2015) was used to carry out the time series analysis because Box and Jenkins first introduced the autoregressive integrated moving
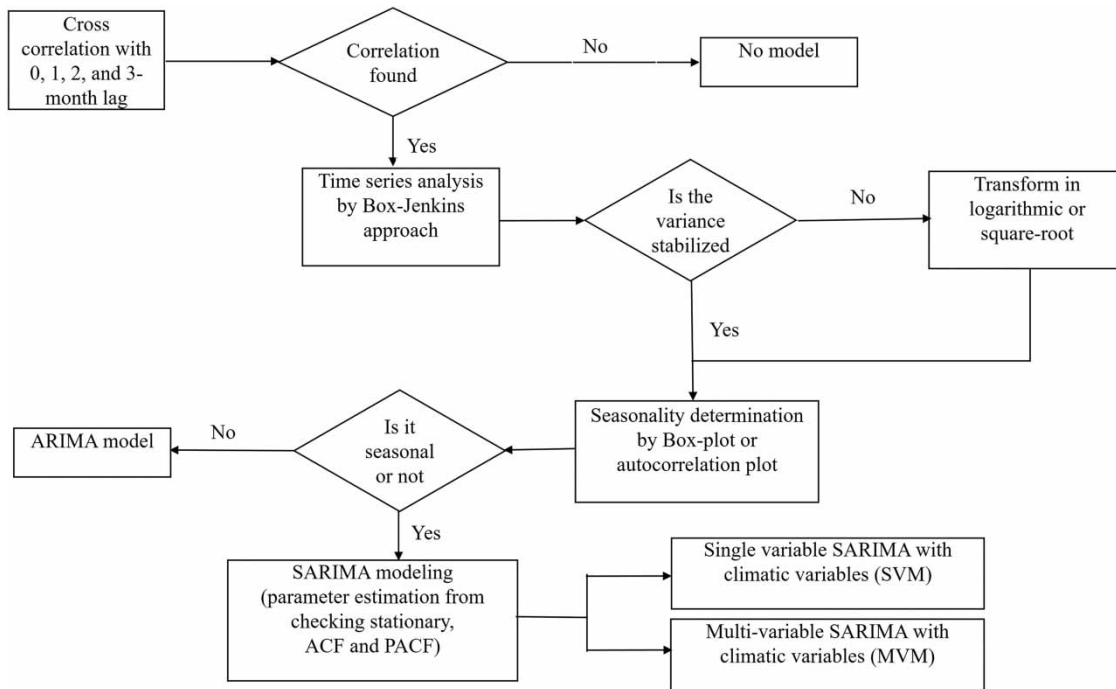
Figure 3 │ Flow chart of evolution of SARIMA modeling.

average (ARIMA) model in 1976 and this has now become the most popular method for time series forecasting. The mean-range plot (the range is plotted against the means for each seasonal period) of untransformed and logarithm or square root transformed series was carried out to stabilize the variance of cholera incidence as the logarithm transformation is required, when the mean-range plot shows a random scatter about a straight line (Helfenstein 1986). Seasonal patterns of a time series can be examined by box-plot or autocorrelation plot. As cholera incidence showed the seasonality, a seasonal-auto-regressive-integrated-moving-average (SARIMA) model was used as a time series analysis tool by fitting to time series data, either to better understand the data or to predict future points in the series (forecasting) when the data shows a seasonality. SARIMA can estimate the effects of climatic variables on cholera incidence which makes the appropriate model for forecasting cholera transmission as its integrated functions controlling seasonal variation, autocorrelation and long-term trends (Zhang *et al.* 2008; Lal *et al.* 2013). The SARIMA model is formed by including an additional *seasonal* term in the ARIMA model and is written as SARIMA $(p,d,q)(P,D,Q)_m$. In this model, m denotes the

number of periods per season, and p the autoregressive (AR) order, d the differencing order or integration term (I) and q the moving average (MA) order for non-seasonal parts of the model. P, D and Q denote the seasonal parts of AR, differencing by integration (I), and MA, respectively. The non-seasonal (p,d,q) and seasonal (P,D,Q) order of the model was determined by: (i) the differencing order determined by checking stationarity (that is, its mean, variance and autocorrelation should be approximately constant through time) from unit root test; (ii) the order of autoregressive by partial autocorrelation function (PACF); and (iii) the order of moving average by autocorrelation function (ACF).

Goodness of fit

Akaike Information Criterion (AIC) (Akaike 1974) and Bayesian Information Criterion (BIC) (Schwarz 1978) were used to assist in evaluating the goodness-of-fit by penalizing the model parameters. The lower the AIC and BIC (Equations (1) and (2)) indicates a better model. Goodness of fit was also examined using both ACF and PACF (Dickey & Fuller 1979) of the residuals of the model and checking the Portmanteau test (Box & Pierce 1970; Ljung & Box 1978)

for white noise in residuals and a scatter plot of residuals versus predicted values:

$$AIC = -2 * \ln(maximum likelihood) + 2\kappa \qquad (1)$$

$$BIC = \ln(n)\kappa - 2 * \ln(maximum\ likelihood) \qquad (2)$$

where κ is the number of independently adjusted parameters within the model and n is the total number of observations.

Model forecasting

The cholera incidence data was divided into two: the data of 2000–2011 was used for the model parameter estimation, and the data of 2012 and 2013 for model forecasting. Root mean squared error (RMSE), mean absolute percentage error (MAPE) and mean absolute error (MAE) were used to verify the forecasting ability of the models (Shcherbakov *et al.* 2013). Lower value of RMSE, MAPE and MAE (Equations (3)–(5)) indicate a better fit of the data:

$$RMSE = \sqrt{\sum_{months=1}^{n} [(observed - forcasted)^2/number\ of\ months]} \qquad (3)$$

$$MAPE = \frac{100}{number\ of\ months} \sum_{months=1}^{n} |(observed - forcasted)/observed| \qquad (4)$$

$$MAE = \sum_{months=1}^{n} |(observed - forcasted)|/number\ of\ months \qquad (5)$$

Limitations

There are a few limitations of this study. For example, the cholera incidence of icddr,b data is assumed to be representative of the entire spatial extent of Dhaka megacity. This is due to the fact that detailed lab-tested cholera incidence data is only available at icddr,b, where cholera cases are diagnosed by the laboratory testing of stool. icddr,b receives most of the cholera patients for Dhaka and surrounding areas as it is renowned as the only cholera hospital in Dhaka, where a special response program is taken every year during the endemic outbreaks. The cholera data availability of only 14 years (January 2000–December 2013) may also be considered as a limitation.

RESULTS AND DISCUSSION

Results of evolution of models

By plotting the mean-range for each seasonal period (12 months), the logarithmic transformation was necessary to stabilize the variance of cholera incidence (Figure 4). All statistical analyses were performed on the logarithmically transformed cholera incidence. The seasonal pattern is quite evident in the box-plot (Figure 5) of the cholera incidence, where the incidence in April and May showed the pre-monsoon peak and in September and October the post-monsoon peak. The autocorrelation plot (Figure 6) showed the highest peak (0.54) at lag 12 which indicates annual seasonality. On the basis of the Augmented Dickey–Fuller (Fuller 2009) unit root test (test statistic = –6.77, whereas 0.01 = –3.49, 0.05 = –2.89, 0.1 = –2.58), the monthly cholera incidence was stationary, i.e. the differencing order for non-seasonal and seasonal (d, D) is zero. Finally, after checking the ACF and PACF plots, the SARIMA $(1,0,0)$ $(1,0,1)_{12}$ was the best fitted model based on the lowest AIC and BIC values. The plots of the ACF and PACF of the residuals of the chosen model showed no significant temporal correlation between residuals at different lags, and the scatter plot of the predicted values against the residuals showed no apparent pattern (Figure 7). Portmanteau Q statistics was 31.04 ($p = 0.84$), i.e. the regression model is quite acceptable.

Table 2 shows the cross-correlation of climatic variables with log-transformed monthly cholera incidence at 0, 1, 2, and 3 months' lag which provides the information on selecting the variables for detailed modelling with SARIMA. The results showed positive and high association with rainfall ($r = 0.43$ at 0-month lag), maximum temperature ($r = 0.61$ at 1-month lag) and minimum temperature ($r = 0.56$ at 0-month lag), while relative humidity ($r \leq 0.28$), ENSO ($r \leq 0.21$) and SOI ($r \leq -0.06$) showed low association with cholera incidence. Then, SARIMA models were run with all the variables individually as single variable SARIMA models (SVMs) as mentioned in Table 3 with lag of 0, 1, 2 and 3 months to check if any high association (low AIC and BIC values) can be made not depending on the cross-correlation values only, however, no high association was found for the latter three variables (relative
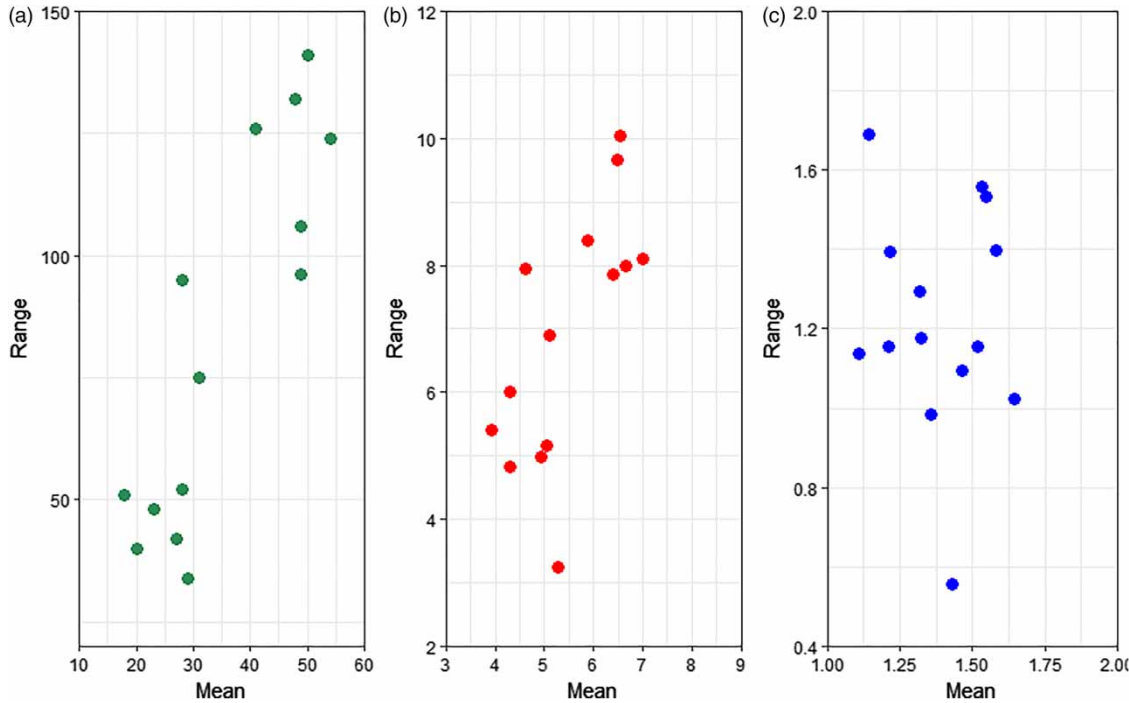
Figure 4 │ Mean-range plot of cholera incidence for (a) non-transformed data, (b) square-root transformation data, (c) log-transformed data.
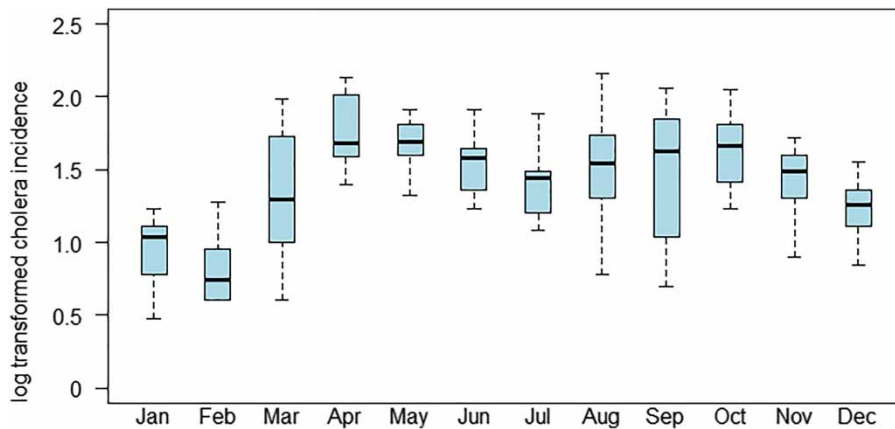

Figure 5 │ Box-plot of log-transformed cholera incidence recorded in icddr,b, Dhaka, 2000–2013.

humidity, ENSO and SOI) in Table 3. The relation of cholera incidence with maximum (AIC = 47, BIC = 66) and minimum temperature (AIC = 46, BIC = 65) showed better at the temporal lag 1 month and with rainfall (AIC = 52, BIC = 71) at the temporal lag 0 than other periods.

The single variable (SVM) SARIMA models (Table 3) show that an increase of the previous month (lag 1) 1 °C maximum temperature resulted in an increase of 7% cholera incidence ($p < 0.001$; AIC = 47, BIC = 66). At the temporal lag 0, an increase of 100 mm in rainfall resulted in a 4% increase of cholera incidence ($p = 0.04$; AIC = 52, BIC = 71) and an increase of 1 °C in minimum monthly temperature at 1-month lag resulted in a 5% increase of cholera incidence ($p < 0.001$; AIC = 46, BIC = 65). However, the multi-variable
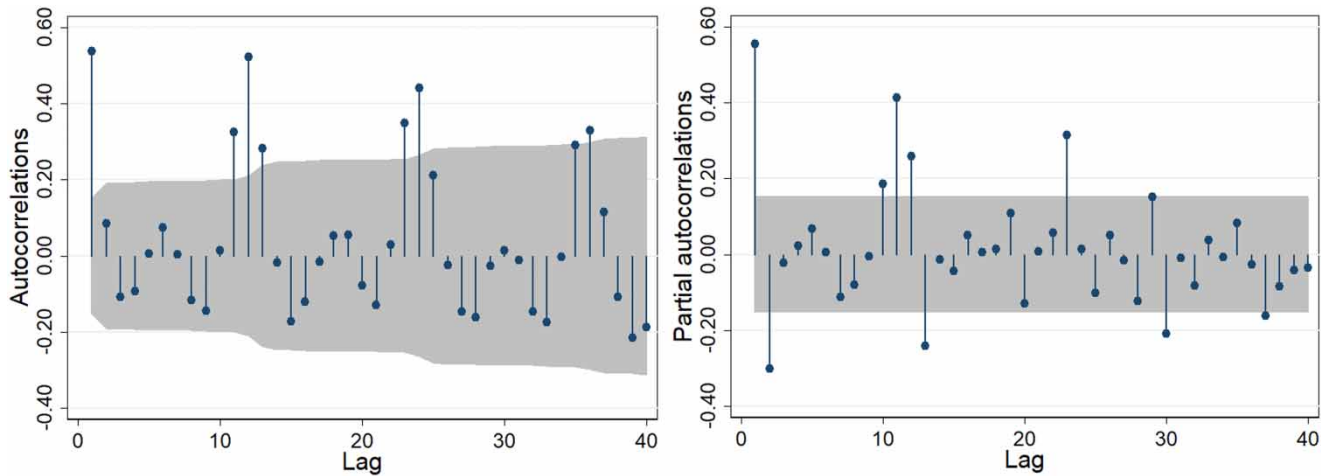
Figure 6 │ Autocorrelation function and partial autocorrelation plots with 5% significant limits of cholera incidence in Dhaka, 2000–2013.
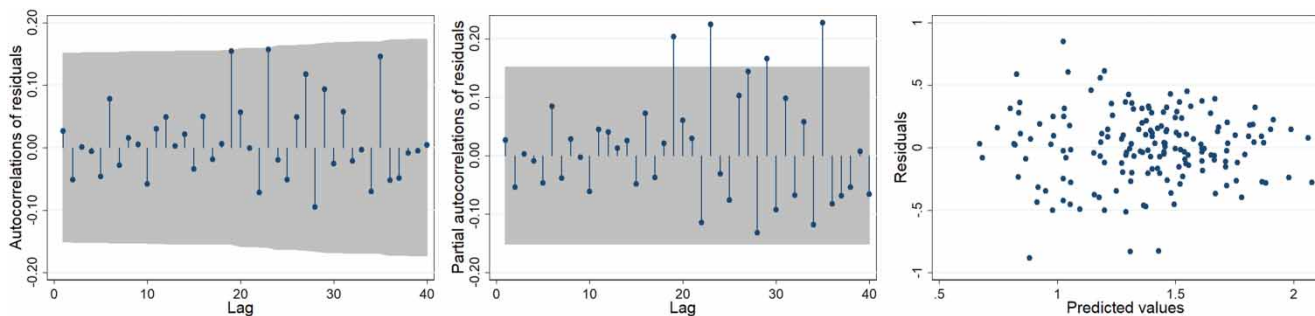
Figure 7 │ ACF and PACF plots of the residuals at different lags and scatter plot of residuals against predicted values of the SARIMA model.

Table 2 │ Cross-correlation analysis of climatic variables (Pearson's correlation coefficient) and log-transformed cholera incidence in Dhaka with a lag 0–3 months

Lag (month)	0	1	2	3
Rainfall	0.43	0.38	0.24	0.07
Minimum temperature	0.56	0.53	0.23	–0.12
Maximum temperature	0.56	0.61	0.31	–0.10
Relative humidity	0.28	–0.06	–0.25	–0.25
ENSO	0.21	0.16	0.09	0.03
SOI	–0.23	–0.17	–0.12	–0.06

SARIMA model (MVM) has been found better than the SVM in terms of error measurement of AIC and BIC as shown in Tables 3 and 4.

The data of monthly SST and minimum temperature showed the best result in studies by Ali *et al.* (2013) and monthly rainfall and minimum temperature by Reyburn *et al.* (2011); however, in this study, the combination of rainfall and maximum temperature (Table 4) fitted the best result. Therefore, finally, four combinations of MVMs A, B, C and D were fitted as listed in Table 5. The models with climatic variables were run with different time lags to check the effect of climatic variables on cholera incidence at lags of 0, 1, 2, and 3 months. Based on AIC and BIC (during 2000–2011), model B showed better fit (AIC = 15, BIC = 36) than the other three models (Table 6). This means that the interaction of rainfall ($p < 0.05$) and maximum temperature ($p < 0.001$) at 1-month lag yielded a significant association with cholera.

Evaluation of model forecast

The performance of models is shown in Figure 8 where the first 12-year period (1 January 2000–31 December 2011) was the model developing stage, while the later 2-years (1 January 2012–31 December 2013) was the model forecasting

Table 3 | Results of single variable SARIMA models (SVMs) for the period 2000–2013

Climatic variables	Lag (month)	*p*-value	AIC	BIC
Rainfall	0	0.04	52	71
	1	0.07	53	72
	2	0.787	58	77
	3	0.094	56	75
Maximum temperature	0	0.002	51	69
	1	<0.001	47	66
	2	0.074	55	74
	3	0.342	57	76
Minimum temperature	0	<0.001	47	66
	1	<0.001	46	65
	2	0.751	58	77
	3	0.901	58	77
Relative humidity	0	0.361	57	76
	1	0.940	58	77
	2	0.259	56	75
	3	0.495	56	76
ENSO	0	0.214	57	76
	1	0.183	56	75
	2	0.260	57	76
	3	0.918	58	77
SOI	0	0.089	54	73
	1	0.388	58	77
	2	0.441	58	77
	3	0.929	58	77

Table 4 | Results of multi-variable SARIMA models (MVMs) for 0-month lag

Climatic variables	*p*-value	AIC	BIC
Rainfall, and maximum temperature	0.008 (rainfall), <0.001 (maximum temperature)	42	64
Rainfall, and minimum temperature	0.063 (rainfall), 0.001 (minimum temperature)	44	66
Minimum temperature, and maximum temperature	0.088 (minimum temperature), 0.412 (maximum temperature)	49	70
Rainfall, minimum temperature, and maximum temperature	0.023 (rainfall), 0.542 (minimum temperature), 0.069 (maximum temperature)	44	69

Table 5 | Four combinations of multi-variable SARIMA models used in forecasting

Model name	Variables used in model
A	Cholera incidence with 0-month lagged rainfall and maximum temperature
B	Cholera incidence with 1-month lagged rainfall and maximum temperature
C	Cholera incidence with 2-month lagged rainfall and maximum temperature
D	Cholera incidence with 3-month lagged rainfall and maximum temperature

DISCUSSION

The results of this study illustrate that there is distinct seasonality (Figure 5) observed in *V. cholera* signatures throughout the world. Among various climatic variables, there is a significant association of rainfall and temperature (Table 2) with cholera incidence. These results are consistent with other studies (e.g. Colwell (2002) in Dhaka, Bangladesh; Reyburn *et al.* (2011) in Zanzibar, Tanzania). Although a very low positive effect of relative humidity ($r = 0.28$) was found at the current month (lag 0) and negative values at lags of 1, 2 and 3 months (Table 2), no significant effect of humidity could be found by time series analysis with the SARIMA model (Table 3). This result (effect of humidity on cholera transmission) is also consistent with other studies, e.g. Islam *et al.* (2009) in Matlab.

The probability of cholera incidence is high with high rainfall in Nha Trang, Vietnam and in Matlab, Bangladesh as documented by Emch *et al.* (2008); however, in Dhaka, Bangladesh, Hashizume *et al.* (2008) summarized that the risk of cholera may increase with both high and low rainfall. Although Hashizume *et al.* (2008) did not provide a quantitative analysis, they described a hypothetical pathway of increasing cholera cases due to high rainfall causing flooding conditions in Dhaka which may cause exposure to water contaminated with *V. cholerae*. However, flooding does not only depend on high rainfall but also the upstream river discharge as most areas of Bangladesh lie at the downstream of three large rivers in South Asia: the Ganges, Brahmaputra and Meghna. Low rainfall may also increase the incidence of cholera as hypothesized by Hashizume *et al.* (2008) where they argued that due to low rainfall

stage. Model B (1-month lag with rainfall and maximum temperature) fitted better between the simulated and observed cholera incidence than the other models (A, C and D) (Figure 8). The error measurements also indicated that model B (RMSE = 14.7, MAE = 11) showed more improved fitting compared to other models (Table 7).

Table 6 | Regression coefficients of the fitted SARIMA $(1,0,0)(1,0,1)_{12}$ (models A, B, C, D) during 2000–2011 in Dhaka, Bangladesh

Model parameter	Model A[a]			Model B[b]			Model C[c]			Model D[d]		
	β	SE	p-value	β	SE	p-value	β	SE	p-value	β	SE	p-value
AR	0.453	0.073	<0.001	0.464	0.072	<0.001	0.52	0.076	<0.001	0.553	0.075	<0.001
SAR	0.833	0.091	<0.001	0.728	0.165	<0.001	0.896	0.061	<0.001	0.880	0.061	<0.001
SMA	−0.53	0.140	<0.001	−0.462	0.198	0.02	−0.572	0.12	<0.001	−0.553	0.12	<0.001
Rainfall	0.00048	0.0002	0.016	0.00043	0.0002	0.03	0.00004	0.00018	0.788	−0.0002	0.00017	0.187
Maximum Temperature	0.062	0.016	<0.001	0.07	0.015	<0.001	0.026	0.022	0.242	−0.033	0.02	0.107

[a]log-likelihood = −2, AIC = 19, BIC = 39.
[b]log-likelihood = −1, AIC = 15, BIC = 36.
[c]log-likelihood = −10, AIC = 34, BIC = 55.
[d]log-likelihood = −10, AIC = 34, BIC = 55.

there may be water scarcity to a certain proportion of people of Dhaka city who rely on surface water for washing and bathing; therefore, the likelihood of multiple uses in water bodies may increase.

Temperature and increase of cholera incidence has a robust relationship, which is well documented in many studies (Lobitz *et al.* 2000; Speelmon *et al.* 2000; Sack *et al.* 2004; Huq *et al.* 2005). This relationship may be due to the multiplication of *V. cholerae*, which directly influences the abundance and toxicity of *V. cholerae* in aquatic environments (Hashizume *et al.* 2008); alternatively, high temperature may also have an indirect influence on pH levels or nutrients as an effect of increased growth of aquatic plants (Lipp *et al.* 2002), and it has already been documented in many studies (e.g. Lobitz *et al.* (2000) and Ali *et al.* (2013)) that the increase of sea surface temperature in the Bay of Bengal causes plankton bloom, which is a favorable condition for multiplication of *V. cholerae*.

The combined effect of rainfall and minimum temperature on cholera showed significant results at 1-month lag in Zanzibar, Tanzania, while the individual effect of 200 mm rainfall resulted in a 1.6% increase of cholera at 2-month lag and a 1 °C increase of minimum temperature at 4-month lag resulted in a 2% increase of cholera (Reyburn *et al.* 2011). In Matlab, Bangladesh, no combined effect was shown by Ali *et al.* (2013), however, an individually significant relationship was found with an increase of the minimum temperature of 1 °C at 0-month lag with a 6% increase of cholera. For Dhaka (this study), the correlation of cholera with maximum temperature (at 0 and 1-month lag), minimum temperature (0 and 1-month lag) and rainfall (0-month lag) individually showed better results (based on AIC and BIC in Table 3) than other climatic variables, i.e. relative humidity, ENSO and SOI. The model run with combined effect of rainfall and maximum temperature (Table 4) showed better results (low AIC and BIC) than individual effects of climatic variables (Table 3), that means, the performance of a multi-variable model (MVM) showed better results than a single variable model (SVM) which answers research question 1. This study also illustrates that previous month's rainfall and maximum temperature showed a better fit in forecasting (Table 7); that means cholera incidence can be forecasted one month earlier which answers research question 2. However, the rainfall and maximum
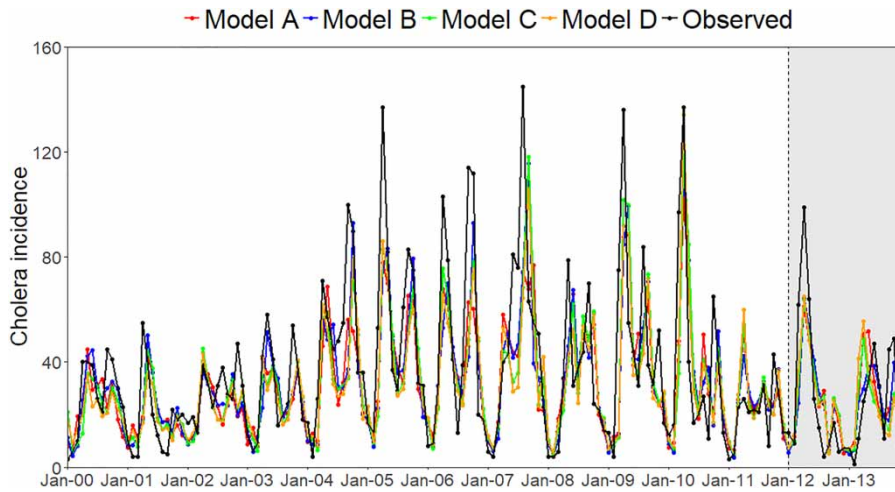
Figure 8 | SARIMA model of forecasting of cholera incidence in Dhaka, Bangladesh.

Table 7 | Forecasting accuracy of SARIMA models for cholera

Error measurement	Model A	Model B	Model C	Model D
RMSE	16.2	14.7	16.7	17.2
MAPE	1.22	1.04	1.13	1.18
MAE	13.2	11	13.2	13.4
n	24	24	24	24

RMSE: root mean squared error; MAPE: mean absolute percentage error; MAE: mean absolute error; n: number of observation.

temperature data should be measured accurately for obtaining an accurate forecasting of cholera outbreaks.

In Zanzibar, Tanzania and Matlab, Bangladesh, minimum temperature is a factor for cholera forecasting with both SVM and MVM. However, in Dhaka (this study), the combined effect of (i) rainfall and minimum temperature; or (ii) maximum and minimum temperature was found insignificant in MVMs (Table 4) while SVM with minimum temperature showed significant but higher AIC and BIC (Table 3) than MVMs, meaning that the effect of different climatic variables on cholera incidence is site or location specific, which answers research question 3. Therefore, for any specific area, individual cholera forecasting models should be developed and tested for better preparedness.

Martinez et al. (2017) developed a forecast model for Dhaka considering the influence of ENSO, however, the influence of ENSO on rainfall pattern is yet to be established over the Indian sub-continent region (Krishnamurthy &

Goswami 2000; Chowdhury 2003; Ihara et al. 2007; Izumo et al. 2010) while the rainfall affects positively on the cholera incidence in Dhaka (Hashizume et al. 2008). In this study, the cross-correlation with influence of ENSO has also been evaluated, however, low Pearson's correlation coefficient values (0.03–0.21, Table 2) were found in lags 0–3 months during the selection of climatic variables for forecasting cholera incidence, therefore, it was not selected for further analysis. Moreover, the model of this study is time series based forecasting using the SARIMA model, which is able to forecast every year (ENSO or non-ENSO years) with different lead time in months. Also, the model of this study is able to forecast cholera incidence of fall of 2012 that could not be fitted well by Martinez et al. (2017).

Climate change is likely to increase the frequency and intensity of drought as well as extreme rainfall leading to flood events in the future (IPCC 2014). Even if global warming is kept to 1.5 °C, the mountains of the Hindu Kush Himalaya region (upstream area of major rivers of Bangladesh) will likely be at least 0.3 °C higher (Wester et al. 2019). Fahad et al. (2018) found that at the end of this century the mean temperature increase over Bangladesh will vary from 3.2 to 5.8 °C where spatially southwest and south central parts of Bangladesh will experience a greater temperature rise than other parts. Such large warming from 3.2 to 5.8 °C could trigger a multitude of biophysical and socio-economic impacts such as increased glacial melting which may affect the annual water budget, i.e. less

predictive water availability during pre-monsoon and increasing frequency and severity of floods, therefore, the endemic cholera outbreaks both at pre- and post-monsoon may increase largely. Mohammed *et al.* (2017) showed that due to climate change high-end scenario (RCP8.5), the average timing of both floods and hydrological droughts is projected to shift earlier compared to the present hydrological regime, i.e. early onset of both flood and drought, therefore this time change may also adversely affect the dual peak cholera outbreaks annually that may prolong the cholera outbreaks.

CONCLUSIONS

This study is aimed at the developing and testing of a cholera forecasting model by establishing a relation between cholera incidence and climatic variables for Dhaka megacity in Bangladesh. The seasonal-auto-regressive-integrated-moving-average (SARIMA) model was found suitable as a forecasting cholera model because of its auto-regressive nature and seasonal behavior pattern. The SARIMA models showed a strong relation between cholera incidence and climatic variables in Dhaka, Bangladesh individually (rainfall, maximum temperature and minimum temperature) and also combined (rainfall and maximum temperature). For example, individual effect by single variable model showed that for a 1 °C monthly maximum temperature increase, cholera incidence increases by 7% ($p < 0.001$) at 1-month lag. That means the cholera incidence can be forecasted 1-month earlier with the temperature data, which is very promising for preparedness. However, the multi-variable model (Model B) with 1-month lag among all combinations of climatic variables and lags showed the best result with the lowest errors of AIC and BIC. This study also revealed that the relationship between cholera incidence and climatic variables varies with locations and climatic variables. Therefore, a forecasted cholera model is location-specific where climatic variables also vary with locations. Hence, one should analyze the location-specific climatic variables for forecasting cholera incidence.

The results of this study would be very important for a climatologist, an epidemiologist or a public health professional, who works with cholera incidence to develop preparedness and response plans. For a climatologist, it is important because climate change impact on cholera incidence may be predicted from this study as climatologists predict an increase of 1.4–5.8 °C in mean temperature over the next 100 years (Houghton *et al.* 2001). An epidemiologist would be helped by the new insights on environmental and climatic linkages of cholera outbreaks. A health professional may prepare for potential coping and adaptation strategies for potential climate change related health risks in Bangladesh. This study also contributes towards the development of a climate-based early warning system for cholera (Akanda *et al.* 2014).

ACKNOWLEDGEMENTS

REFERENCES

Akaike, H. 1974 A new look at the statistical model identification. *IEEE Trans. Automat. Contr.* **19** (6), 716–723.

Akanda, A. S., Jutla, A. S. & Islam, S. 2009 Dual peak cholera transmission in Bengal Delta: a hydroclimatological explanation. *Geophys. Res. Lett.* **36** (19), L19401.

Akanda, A. S., Jutla, A. S., 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. *Water Resour. Res.* **47** (3), W00H07.

Akanda, A. S., Jutla, A. S. & Colwell, R. R. 2014 Global diarrhoeal action plan needs integrated climate-based surveillance. *Lancet Glob. Health* **2** (2), e69–e70.

Alam, M., Islam, A., Bhuiyan, N., Rahim, N., Hossain, A., Khan, G. Y., Ahmed, D., Watanabe, H., Izumiya, H., Faruque, A. S. G., Akanda, A. S., Islam, S., Sack, R. B., Huq, A., Colwell, R. R. & Cravioto, A. 2011 Clonal transmission, dual peak, and off-season cholera in Bangladesh. *Infect. Ecol. Epidemiol.* **1** (1), 7273.

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

Ali, M., Emch, M., Donnay, J. P., Yunus, M. & Sack, R. B. 2002b The spatial epidemiology of cholera in an endemic area of Bangladesh. *Soc. Sci. Med.* **55** (6), 1015–1024.

Ali, M., Kim, D. R., Yunus, M. & Emch, M. 2013 Time series analysis of cholera in Matlab, Bangladesh, during 1988–2001. *J. Health Popul. Nutr.* **31** (1), 11–19.

Ali, M., Nelson, A. R., Lopez, A. L. & Sack, D. A. 2015 Updated global burden of cholera in endemic countries. *PLoS Negl. Trop. Dis.* **9** (6), e0003832.

Bangladesh Bureau of Statistics (BBS) 2014 *Population & Housing Census-2011: National Volume-3: Urban Area Report*. Bangladesh Bureau of Statistics, Statistics and Informatics Division, Ministry of Planning, Government of Bangladesh, Dhaka, Bangladesh.

Batabyal, P., Einsporn, M. H., Mookerjee, S., Palit, A., Neogi, S. B., Nair, G. B. & Lara, R. J. 2014 Influence of hydrologic and anthropogenic factors on the abundance variability of enteropathogens in the Ganges estuary, a cholera endemic region. *Sci. Total Environ.* **472**, 154–161.

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. *Geophys. Res. Lett.* **39** (5), L05403.

Bouma, M. J. & Pascual, M. 2001 Seasonal and interannual cycles of endemic cholera in Bengal 1891–1940 in relation to climate and geography. *Hydrobiologia* **460**, 147–156.

Box, G. E. & Pierce, D. A. 1970 Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *J. Am. Stat. Assoc.* **65** (332), 1509–1526.

Box, G. E., Jenkins, G. M., Reinsel, G. C. & Ljung, G. M. 2015 *Time Series Analysis: Forecasting and Control*. John Wiley & Sons, Inc, New York, USA.

Charles, R. C. & Ryan, E. T. 2011 Cholera in the 21st century. *Curr. Opin. Infect. Dis.* **24** (5), 472–477.

Checkley, W., Epstein, L. D., Gilman, R. H., Figueroa, D., Cama, R. I., Patz, J. A. & Black, R. E. 2000 Effects of El Niño and ambient temperature on hospital admissions for diarrhoeal diseases in Peruvian children. *Lancet* **355** (9202), 442–450.

Chowdhury, M. 2003 The el Niño-southern oscillation (ENSO) and seasonal flooding–Bangladesh. *Theor. Appl. Climatol.* **76** (1–2), 105–124.

Colwell, R. R. 1996 Global climate and infectious disease: the cholera paradigm. *Science* **274** (5295), 2025–2031.

Colwell, R. R. 2002 A voyage of discovery: cholera, climate and complexity. *Environ. Microbiol.* **4** (2), 67–69.

De Magny, G. C., Cazelles, B. & Guégan, J. F. 2006 Cholera threat to humans in Ghana is influenced by both global and regional climatic variability. *EcoHealth* **3** (4), 223–231.

de Magny, G. C., Thiaw, W., Kumar, V., Manga, N. M., Diop, B. M., Gueye, L., Kamara, M., Roche, B., Murtugudde, R. & Colwell, R. R. 2012 Cholera outbreak in Senegal in 2005: was climate a factor? *PLoS One* **7** (8), e44577.

Dickey, D. A. & Fuller, W. A. 1979 Distribution of the estimators for autoregressive time series with a unit root. *J. Am. Statist. Assoc.* **74** (366a), 427–431.

Eisenberg, M. C., Kujbida, G., Tuite, A. R., Fisman, D. N. & Tien, J. H. 2013 Examining rainfall and cholera dynamics in Haiti using statistical and dynamic modeling approaches. *Epidemics* **5** (4), 197–207.

Emch, M., Feldacker, C., Yunus, M., Streatfield, P. K., DinhThiem, V. & Ali, M. 2008 Local environmental predictors of cholera in Bangladesh and Vietnam. *Am. J. Trop. Med. Hyg.* **78** (5), 823–832.

Enserink, M. 2010 Haiti's outbreak is latest in cholera's new global assault. *Science* **330** (6005), 738–739.

Fahad, M. G. R., Islam, A. K. M. S., Nazari, R., Alfi Hasan, M., Islam, G. M. T. & Bala, S. K. 2018 Regional changes of precipitation and temperature over Bangladesh using bias-corrected multi-model ensemble projections considering high-emission pathways. *Int. J. Climatol.* **38** (4), 1634–1648.

Field, C. B., Barros, V. R., Dokken, D. J., Mach, K. J., Mastrandrea, M. D., Bilir, T. E., Chatterjee, M., Ebi, K. L., Estrada, Y. O., Genova, R. C. & Girma, B. 2014 *IPCC, 2014: Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, UK and New York, NY, USA.

Fuller, W. A. 2009 *Introduction to Statistical Time Series*. John Wiley & Sons, Inc, New York, USA.

Hashizume, M., Armstrong, B., Hajat, S., Wagatsuma, Y., Faruque, A. S., Hayashi, T. & Sack, D. A. 2008 The effect of rainfall on the incidence of cholera in Bangladesh. *Epidemiology* **19** (1), 103–110.

Hashizume, M., Faruque, A. S. G., Wagatsuma, Y., Hayashi, T. & Armstrong, B. 2010 Cholera in Bangladesh: 'climatic components of seasonal variation'. *Epidemiology* **21** (5), 706–710.

Helfenstein, U. 1986 Box-Jenkins modelling of some viral infectious diseases. *Stat. Med.* **5** (1), 37–47.

Helfenstein, U. 1991 The use of transfer function models, intervention analysis and related time series methods in epidemiology. *Int. J. Epidemiol.* **20** (3), 808–815.

Houghton, J. T., Ding, Y., Griggs, D. J., Noguer, M., Van Der Linden, P. J., Dai, X., Maskell, K. & Johnson, C. A. 2001 *Climate Change 2001: The Scientific Basis*. Cambridge University Press, UK and New York, NY, USA.

Huq, A., Small, E. B., West, P. A., Huq, M. I., Rahman, R. & Colwell, R. R. 1983 Ecological relationships between *Vibrio cholerae* and planktonic crustacean copepods. *Appl. Environ. Microbiol.* **45** (1), 275–283.

Huq, A., Sack, R. B., Nizam, A., Longini, I. M., Nair, G. B., Ali, A., Morris, J. G., Khan, M. H., Siddique, A. K., Yunus, M. & Albert, M. J. 2005 Critical factors influencing the occurrence of *Vibrio cholerae* in the environment of Bangladesh. *Appl. Environ. Microbiol.* **71** (8), 4645–4654.

Ihara, C., Kushnir, Y., Cane, M. A. & Victor, H. 2007 Indian summer monsoon rainfall and its link with ENSO and Indian ocean climate indices. *Int. J. Climatol.* **27** (2), 179–187.

IPCC 2014 *Climate Change 2014 – Impacts, Adaptation and Vulnerability: Regional Aspects*. Cambridge University Press, United Kingdom and New York, NY, USA.

Islam, S., Brooks, A., Kabir, M. S., Jahid, I. K., Islam, M. S., Goswami, D., Nair, G. B., Larson, C., Yukiko, W. & Luby, S. 2007 Faecal contamination of drinking water sources of Dhaka city during the 2004 flood in Bangladesh and use of disinfectants for water treatment. *J. Appl. Microbiol.* **103** (1), 80–87.

Islam, M. S., Sharker, M. A. Y., Rheman, S., Hossain, S., Mahmud, Z. H., Islam, M. S., Uddin, A. M. K., Yunus, M., Osman, M. S., Ernst, R., Rector, I., Larson, C. P., Luby, S. P., Endtz, H. P. & Cravioto, A. 2009 Effects of local climate variability on transmission dynamics of cholera in Matlab, Bangladesh. *Trans. R. Soc. Trop. Med. Hyg.* **103** (11), 1165–1170.

Izumo, T., Vialard, J., Lengaigne, M., de Boyer Montegut, C., Behera, S. K., Luo, J.-J. & Yamagata, T. 2010 Influence of the state of the Indian Ocean Dipole on the following year's El Niño. *Nature Geosci.* **3** (3), 168–172.

Jutla, A. S., Akanda, A. S., Griffiths, J. K., Colwell, R. & Islam, S. 2011 Warming oceans, phytoplankton, and river discharge: implications for cholera outbreaks. *Am. J. Trop. Med. Hyg.* **85** (2), 303–308.

Jutla, A., Aldaach, H., Billian, H., Akanda, A., Huq, A. & Colwell, R. 2015 Satellite based assessment of hydroclimatic conditions related to cholera in Zimbabwe. *PloS One* **10** (9), 0137828.

Kotar, S. & Gessler, J. 2014 *Cholera: A Worldwide History*. McFarland & Co., Inc., Jefferson, NC, USA.

Krishnamurthy, V. & Goswami, B. N. 2000 Indian monsoon–ENSO relationship on interdecadal timescale. *J. Clim.* **13** (3), 579–595.

Lal, A., Ikeda, T., French, N., Baker, M. G. & Hales, S. 2013 Climate variability, weather and enteric disease incidence in New Zealand: time series analysis. *PloS One* **8** (12), e83484.

Leckebusch, G. C. & Abdussalam, A. F. 2015 Climate and socioeconomic influences on interannual variability of cholera in Nigeria. *Health Place* **34**, 107–117.

Lipp, E. K., Huq, A. & Colwell, R. R. 2002 Effects of global climate on infectious disease: the cholera model. *Clin. Microbiol. Rev.* **15** (4), 757–770.

Ljung, G. M. & Box, G. E. P. 1978 On a measure of lack of fit in time series models. *Biometrika* **65** (2), 297–303.

Lobitz, B., Beck, L., Huq, A., Wood, B., Fuchs, G., Faruque, A. S. G. & Colwell, R. 2000 Climate and infectious disease: use of remote sensing for detection of *Vibrio cholerae* by indirect measurement. *Proc. Natl. Acad. Sci.* **97** (4), 1438–1443.

Louis, V. R., Russek-Cohen, E., Choopun, N., Rivera, I., Gangle, B., Jiang, S. C., Rubin, A., Patz, J. A., Huq, A. & Colwell, R. 2003 Predictability of Vibrio cholerae in Chesapeake Bay. *Appl. Environ. Microbiol.* **69** (5), 2773–2785.

Luque Fernández, M. Á., Bauernfeind, A., Jiménez, J. D., Gil, C. L., Omeiri, N. E. & Guibert, D. H. 2009 Influence of temperature and rainfall on the evolution of cholera epidemics in Lusaka, Zambia, 2003–2006: analysis of a time series. *Trans. R. Soc. Trop. Med. Hyg.* **103** (2), 137–143.

Martinez, P. P., Reiner Jr, R. C., Cash, B. A., Rodó, X., Mondal, M. S., Roy, M. & King, A. A. 2017 Cholera forecast for Dhaka, Bangladesh, with the 2015-2016 El Niño: lessons learned. *PLoS One* **12** (3), e0172355.

McElroy, A. & Townsend, P. K. 2014 *Medical Anthropology in Ecological Perspective*. Westview Press, Colorado, USA.

Mendelsohn, J. & Dawson, T. 2008 Climate and cholera in KwaZulu-Natal, South Africa: the role of environmental factors and implications for epidemic preparedness. *Int. J. Hyg. Environ. Health* **211** (1–2), 156–162.

Mohammed, K., Islam, A. K. M. S., Islam, G. M. T., Alfieri, L., Bala, S. K. & Khan, M. J. U. 2017 Impact of high-end climate change on floods and low flows of the Brahmaputra river. *J. Hydrol. Eng.* **22** (10), 04017041.

Pascual, M., Bouma, M. J. & Dobson, A. P. 2002 Cholera and climate: revisiting the quantitative evidence. *Microbes Infect.* **4** (2), 237–245.

Paz, S. 2009 Impact of temperature variability on cholera incidence in southeastern Africa, 1971–2006. *Ecohealth* **6** (3), 340–345.

Rajendran, K., Sumi, A., Bhattachariya, M. K., Manna, B., Sur, D., Kobayashi, N. & Ramamurthy, T. 2011 Influence of relative humidity in *Vibrio cholerae* infection: a time series model. *Indian J. Med. Res.* **133** (2), 138–145.

Reiner, R. C., King, A. A., Emch, M., Yunus, M., Faruque, A. S. G. & Pascual, M. 2012 Highly localized sensitivity to climate forcing drives endemic cholera in a megacity. *Proc. Natl. Acad. Sci.* **109** (6), 2033–2036.

Reyburn, R., Kim, D. R., Emch, M., Khatiba, A., Von Seidlein, L. & Ali, M. 2011 Climate variability and the outbreaks of cholera in Zanzibar, East Africa: a time series analysis. *Am. J. Trop. Med. Hyg.* **84** (6), 862–869.

Righetto, L., Bertuzzo, E., Mari, L., Schild, E., Casagrandi, R., Gatto, M., Rodriguez-Iturbe, I. & Rinaldo, A. 2013 Rainfall mediations in the spreading of epidemic cholera. *Adv. Water Resour.* **60**, 34–46.

Rodó, X., Pascual, M., Fuchs, G. & Faruque, A. S. G. 2002 ENSO and cholera: a nonstationary link related to climate change? *Proc. Natl. Acad. Sci.* **99** (20), 12901–12906.

Sack, D. A., Sack, R. B., Nair, G. B. & Siddique, A. K. 2004 Cholera. *Lancet* **363** (9404), 223–233.

Schwarz, G. 1978 Estimating the dimension of a model. *Ann. Stat.* **6** (2), 461–464.

Sebastian, T., Anandan, S., Jeyaseelan, V., Jeyaseelan, L., Ramanathan, K. & Veeraraghavan, B. 2015 Role of seasonality and rainfall in *vibrio cholerae* infections: a time series model for 11 years surveillance data. *Clin. Epidemiol. Glob. Health* **3** (3), 144–148.

Shcherbakov, M. V., Brebels, A., Shcherbakova, N. L., Tyukov, A. P., Janovsky, T. A. & Kamaev, V. A. E. 2013 A survey of forecast error measures. *World Appl. Sci. J.* **24** (2013), 171–176.

Speelmon, E. C., Checkley, W., Gilman, R. H., Patz, J., Calderon, M. & Manga, S. 2000 Cholera incidence and El niño–related higher ambient temperature. *J. Am. Med. Assoc.* **283** (23), 3072–3074.

Trærup, S. L., Arigoni Ortiz, R. & Markandya, A. 2010 *The Health Impacts of Climate Change: A Study of Cholera in Tanzania.*

BC3 Working Paper Series 2010–01. Basque Centre for Climate Change (BC3), Bilbao, Spain.

Vital, M., Füchslin, H. P., Hammes, F. & Egli, T. 2007 Growth of vibrio cholerae O1 ogawa eltor in freshwater. *Microbiology* **153** (7), 1993–2001.

WB 2017 *Weatherbase: Historical Weather for Dhaka, Bangladesh.* Available from www.weatherbase.com (accessed 23 July 2017).

Wester, P., Mishra, A., Mukherji, A. & Shrestha, A. B. 2019 *The Hindu Kush Himalaya Assessment: Mountains, Climate Change, Sustainability and People.* Springer International Publishing, Basel, Switzerland.

Zhang, Y., Bi, P. & Hiller, J. 2008 Climate variations and salmonellosis transmission in Adelaide, South Australia: a comparison between regression models. *Int. J. Biometeorol.* **52** (3), 179–187.