

RESEARCH

Open Access



# Seasonality of cholera in Kolkata and the influence of climate

Debbie Shackleton<sup>1\*</sup>, Theo Economou<sup>2,3</sup>, Fayyaz Ali Memon<sup>1</sup>, Albert Chen<sup>1</sup>, Shanta Dutta<sup>4</sup>, Suman Kanungo<sup>4</sup> and Alok Deb<sup>4</sup>

## Abstract

**Background** Cholera in Kolkata remains endemic and the Indian city is burdened with a high number of annual cases. Climate change is widely considered to exacerbate cholera, however the precise relationship between climate and cholera is highly heterogeneous in space and considerable variation can be observed even within the Indian subcontinent. To date, relatively few studies have been conducted regarding the influence of climate on cholera in Kolkata.

**Methods** We considered 21 years of confirmed cholera cases from the Infectious Disease Hospital in Kolkata during the period of 1999–2019. We used Generalised Additive Modelling (GAM) to extract the non-linear relationship between cholera and different climatic factors; temperature, rainfall and sea surface temperature (SST). Peak associated lag times were identified using cross-correlation lag analysis.

**Results** Our findings revealed a bi-annual pattern of cholera cases with two peaks coinciding with the increase in temperature in summer and the onset of monsoon rains. Variables selected as explanatory variables in the GAM model were temperature and rainfall. Temperature was the only significant factor associated with summer cholera (mean temperature of 30.3 °C associated with RR of 3.8) while rainfall was found to be the main driver of monsoon cholera (550 mm total monthly rainfall associated with RR of 3.38). Lag time analysis revealed that the association between temperature and cholera cases in the summer had a longer peak lag time compared to that between rainfall and cholera during the monsoon. We propose several mechanisms by which these relationships are mediated.

**Conclusions** Kolkata exhibits a dual-peak phenomenon with independent mediating factors. We suggest that the summer peak is due to increased bacterial concentration in urban water bodies, while the monsoon peak is driven by contaminated flood waters. Our results underscore the potential utility of preventative strategies tailored to these seasonal and climatic patterns, including efforts to reduce direct contact with urban water bodies in summer and to protect residents from flood waters during monsoon.

**Keywords** Generalized additive modelling, Climate, Cholera, Cross-correlation lag analysis, India

\*Correspondence:

Debbie Shackleton  
dmshackleton@gmail.com

<sup>1</sup> College of Engineering, Mathematics, and Physical Sciences, University of Exeter, Exeter EX4 4QF, UK

<sup>2</sup> Department of Mathematics, University of Exeter, Exeter, UK

<sup>3</sup> Climate and Atmosphere Research Centre, The Cyprus Institute, Nicosia, Cyprus

<sup>4</sup> National Institute of Cholera and Enteric Diseases, Kolkata, India

## Background

Cholera is an infectious diarrheal disease caused by contamination with pathogenic strains (O1 or O139) of the bacteria *Vibrio cholerae*. While the disease has been successfully eradicated in many parts of the world thanks to major improvements to drinking water and sanitation infrastructure, the seventh global cholera pandemic persists in many others such as India. Bordering the Bay of



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Bengal, often considered to be the ‘homeland’ of cholera, the Indian State of West Bengal experiences the highest cholera burden in the country [1], much of which is concentrated in the densely populated state capital Kolkata.

Cholera is highly climate-sensitive and is broadly considered to be exacerbated by the effects of climate change [2–4]. This is particularly true with regards to the endemic cholera which persists across the Bengal Delta, a vast river delta surrounding the Bay of Bengal consisting of Bangladesh in the East, and West Bengal in the West. Transmission in this region occurs predominately due to contact with contaminated water sources in the environment [5]. Within Bangladesh, the relationship between cholera and climate has been studied intensively [6]. An interesting, and potentially unique phenomenon with endemic cholera in this region is the dual-peak seasonality of cases. This is well documented in Bangladesh which experiences peaks both pre- and post-monsoon with a marked abated occurrence of cases during the monsoon itself [7–9]. Within the context of Bangladesh, association between climate and cholera appears to differ between peaks. With regards to rainfall, while the post-monsoon peak has been demonstrated to hold a strong positive association with rainfall [7, 9, 10], during the summer the inverse is true with drier years tending to lead to stronger early peaks [7, 11]. An even more complex relationship appears to exist between cholera and sea surface temperature (SST) in the Bay of Bengal. While evidence gathered from historical cholera mortality records (1891–1940) across the Bengal Delta found coastal SST was positively associated with pre-monsoon cholera, no relationship was found with post-monsoon cholera [8]. In contrast a more recent study in Bangladesh [10] found a weak association coastal SST pre-monsoon, but a much stronger association post-monsoon. The effects of air temperature in this region have been less well considered, but a 2018 study found a significant increase in cholera risk following a heatwave on wet days, but not dry days [12]. Research in other regions has suggested that non-linear relationships may be present between cholera and climate variables. For example, non-linear relationships have been found between rainfall and cholera in Yemen [13], and between *Vibrio Cholerae* abundance and SST in the North Atlantic [14].

Compared with the rich epidemiological research background of cholera in Bangladesh however, research into the environmental influencers of Kolkata is much more limited despite the presence of endemicity in the region. Geographically, the city of Kolkata is similar to the Bangladesh capital Dhaka where much of the research has been focussed. At a distance of 250 km apart, both are dense urban centres north of the Bay of Bengal with large slum populations who share the same language and many

similar cultural practices. However, a study by de Magny et al. [15] has shown that despite their similarities and physical proximity, the influence of climate on cholera in Bangladesh and Kolkata is markedly different.

To date, research conducted into the relationship between cholera and climate in Kolkata has largely been limited to case studies describing particular outbreaks [16], or communities [17]. De Magny et al. [15] used a generalized linear model to assess the effects of rainfall and ocean temperature and chlorophyll levels on cholera in Kolkata. They found a significant influence of rainfall and ocean chlorophyll however only five years of data were used due to limited data available at the time. Significant research gaps remain in clarifying this relationship. Within the Bengal region, the potential presence of nonlinearity in cholera-climate associations has not been ascertained. Within the more specific context of Kolkata, no studies have yet considered the influence of ambient temperature, nor have individual relationships by season have not been considered.

In this study, we aim to fill these gaps by using Generalized Additive Modelling (GAM) to analyse the potentially nonlinear relationships between cholera and three environmental variables: temperature, rainfall, and sea surface temperature (SST) during the period 1999–2019. We also take a closer look at lag times with cross-correlational lag analysis. An additional novel aspect in this study is comparing the use of rainfall runoff with rainfall as a variable. By considering runoff explicitly, we aim to consider the role of contamination of water sources via rainfall runoff in transmission. Finally, we will consider the influence of climate variables on both seasons separately to capture any season-specific relationships.

## Methods

### Study area

The city of Kolkata lies on the east bank of the Hooghly River, a major Ganges distributary, around 160 km North of its origin into the Bay of Bengal (Fig. 1). It has a total population of around 45 million residents, including many who live in tightly crowded urban slums. The city has an average population density of 24,306 persons per km<sup>2</sup>.

### Epidemiological data

A dataset of stool samples from diarrhoeal patients reporting to the Infectious Disease Hospital (IDH) in Kolkata under their diarrhoeal surveillance system during the 21-year period 1999–2019 was obtained from the Indian Council of Medical Research - National Institute of Cholera and Enteric Diseases (ICMR-NICED). In the surveillance system, every fifth patient on two randomly selected days of the week (representing around 6% of



**Fig. 1** Map of study area showing regions where environmental data were obtained. Bounding box 1 (BB1) shows the 0.5 × 0.5 degree grid cell centred on (88.25E, 22.75 N), bounding box 2 (BB2) denotes the 1 × 1 degree grid cell centred on (88.5E, 21.5 N)

total patients) was tested for several pathogens including O1 and O139 *Vibrio cholerae*. We extracted the number of samples which tested positive for either O1 or O139 *Vibrio cholerae* where the patient was registered as residing within the Kolkata Municipal Corporation (KMC) region. The dataset from 1999 to 2007 was un-digitized and pre-aggregated at a monthly resolution. We therefore considered two datasets: a 21-year monthly dataset from 1999 to 2019, and a 12-year dataset (2008–2019) aggregated weekly. The monthly dataset was normalised to a 30-day month to account for differences in month length (Eq. 1).

$$case\ S_{normalized,month} = case\ S_{raw,month} \cdot \frac{30}{days.in.month} \quad (1)$$

The decadal census populations of Kolkata were recorded as 4,399,819, 4,572,876, and 4,496,694 during the years 1991, 2001, and 2011 respectively (data were unavailable for 2021) [18]. We therefore considered the population of Kolkata to be relatively stable, and

therefore ignored any marginal fluctuations in population in our analysis.

**Climate data**

Monthly mean daytime temperature (°C) and total rainfall (mm/month) from were obtained from satellite estimations produced by the Climate Research Unit (CRU) at the University of East Anglia from the CRU TS 4.04 dataset [19] (available at <http://www.cru.uea.ac.uk/cru/data/hrg/>). Rainfall runoff ( $kg\ m^{-2}\ s^{-1}$ ) was estimated using the mean monthly output of the following five land-surface models: DLEM [20], ISAM Trendy [21], JULES 1.0 [22], LPX-BERN [23], and ORCHIDEE [24]. We further estimated weekly mean daytime temperature (°C) and total precipitation (mm/day) by extracting daily estimations provided by the CRU Japanese Reanalysis (CRU JRA – available at <https://catalogue.ceda.ac.uk/>) [25] and aggregating weekly. Each dataset was available at a 0.5 × 0.5 degree spatial resolution and the cell with centroid longitude = 88.25°E, latitude = 22.75°N was chosen since it covered the majority of the KMC area (BB1 in

Fig. 1). Monthly SST data (°C) at a 1°x1° resolution were extracted from the Hadley Centre Global Sea Ice and Sea Surface Temperature (HadISST) dataset [26] which combines in-situ and adjusted satellite measurements. We use the grid cell centred at 88.5°E, 21.5°N which refers to the area where the Bay of Bengal meets the mouth of the Hooghly River (BB2 in Fig. 1).

**Statistical analysis**

**Identification of seasonality**

A box plot of monthly cholera cases, temperature, and rainfall was produced for visual inspection of seasonality within the monthly dataset. Two distinct ‘cholera seasons’ are demarcated through visual inspection, where the season period is defined according to patterns in cholera cases rather than meteorological patterns.

**Nonlinear climate relationships**

To allow for the potential presence of non-linear associations in the data, we explored the association between cholera cases and environmental variables using Generalized Additive Models (GAMs) [27] applied to the 1999–2019 monthly dataset [27]. The GAM analysis was conducted using the R package *mgcv* [28]. A more in-depth explanation of GAM models is given in the [Supplementary materials](#) and is briefly summarized here. GAMs can be considered as an extension to generalized linear models (GLMs) which relaxes the linearity assumption and allows for significantly more flexible model fitting. Here, we model the cholera case data as sampled from a negative binomial distribution (due to the count nature and presence of over-dispersion) where the mean (count) is characterized as the sum of smooth functions of the environmental variables. We considered the variables rainfall, modelled rainfall-runoff, temperature and coastal SST as environmental inputs to the model.

The smooth functions in our model were constructed using penalized cubic regression splines. The flexibility of the smooths can be tuned by increasing the number of knots (k) where a greater number of knots permits a more flexible smooth. Since overfitting is prevented by a smoothing penalization term, we set k to the minimum value above which the influence on the results was negligible to allow sufficient degrees of freedom to describe the true relationship within a manageable computational cost.

In our model, we include the season as a factor within each smooth of an environmental variable which permits

the smooths to ‘interact’ with the season factors and produce a particular smooth for each environmental variable in each cholera season. Here, to prevent any potential misinterpretation of environmental correlations as incidental seasonal patterns such as those that may be caused by seasonal migration or major annual festivals, we include a smooth for month using a cyclic penalized cubic regression spline. Similarly, to account for long-term trends not caused by environmental changes (e.g., due to improvements to sanitation or changes in health seeking behaviour) we include a smooth function of date.

To reduce bias linked with assumptions about lags in the connection between environmental variables and cholera, we analysed five different lag configurations for each of the four climate variables, as well as the option of omitting each variable. These configurations were as follows: variable omitted, in sync with the cholera occurrence month, delayed by one month, delayed by two months, mean average of concurrent and one-month lag, mean average of a one-month and a two-month lag. This examination led to a total of  $N = 6^4 - 1 = 1296$  model variations. The general formula for each GAM model is given in Eq. 2.

$$\begin{aligned} \log[E(Y_m)] = & \beta_0 + s_{season}(temperature_{m,lag}) + s_{season}(rainfall_{m,lag}) \\ & + s_{season}(runoff_{m,lag}) + s_{season}(SST_{m,lag}) + s(date) \\ & + s(month) + season \end{aligned} \tag{2}$$

Where  $E(Y_m)$  represents the expected monthly confirmed cholera cases in month  $m$ ,  $\beta_0$  represents the intercept, and  $s_{season}()$  represents the smooth functions within a particular cholera season.

The strength of each model was assessed by calculating the Akaike Information Criterion (AIC) [29] which measures goodness-of-fit while penalizing additional model complexity. The model with the lowest AIC score was selected as the final model.

We anticipated temporal autocorrelation in the counts, and after exploratory analysis of the autocorrelation function (ACF) and the partial ACF (PACF) of the residuals of the chosen model (Figure S1), an AR(1) process was indicated. We therefore included an additional one-month lagged term of the response variable (as a linear covariate with its own associated Beta coefficient) in the final model (Eq. 3). This inclusion of an autoregressive term to deal with autocorrelation in the data is a well-established approach [30]. An ACF plot of the final model (Figure S2) confirmed this to be a reasonable assumption.

---


$$\begin{aligned} \log[E(Y_m)] = & \beta_0 + s_{season}(temperature_{m,lag}) + s_{season}(rainfall_{m,lag}) + s_{season}(runoff_{m,lag}) \\ & + s_{season}(SST_{m,lag}) + s(month) + s(month - of - year) + season + \log[y_{m-1}] \end{aligned} \tag{3}$$


---



### Lag analysis

We next used cross-correlation analysis to identify the lags associated with the strongest relationship between climate variables and cholera cases. Specifically, we measured the Pearson's correlation between climate and cholera time series at lag periods from 0 to 25 weeks stratified by season. We utilised the 12-year weekly aggregate datasets from 2008 to 2019 to assess the effect of lag times at a finer temporal resolution.

### Limitations

In our study we conclude that the relationship observed between rainfall and cholera during the monsoon season is due to urban flooding. However, measuring rainfall-induced flooding using monthly averages does not fully capture tropical cyclones and short periods of highly intense rainfalls responsible for the flash flooding which frequently occurs in Kolkata [31]. It is therefore possible that the associations with rainfall found in this study are underestimated. Another issue in this study is the uncer-

with an increase in temperature as summer approaches (Fig. 2B), and the second after the onset of Monsoon rains (Fig. 2C). We therefore demarcated summer cholera as cases reported during the four-month period March–June, and monsoon cholera as cases reported during the six-month period July–December. January and February are considered as ‘non-cholera’ seasons. Summary statistics for cholera cases, temperature and rainfall are displayed in Table 1.

### GAM analysis

The model variation with the lowest AIC score (AIC = 1312.3) and selected for analysis retained temperature and rainfall as explanatory variables but omitted SST and rainfall runoff (Eq. 4). The selected lag for temperature was the mean average of the concurrent and previous month, no lag was chosen for rainfall and its concurrent value was selected. The model was able to explain 55.5% of the deviance in the data and residual analysis (Figure S3) suggests model assumptions are reasonable.

$$\log[E(t)] = \beta_0 + s_{\text{season}}(\text{temperature}_{m,\text{lag}=\text{mean}(0,1)}) + s_{\text{season}}(\text{rainfall}_{m,\text{lag}=0}) + s(\text{month}) + s(\text{month} - \text{of} - \text{year}) + \text{season} + \log[y_{t-1}] \quad (4)$$

tainty in the cholera case data. Only cholera patients who report to the ID hospital in Eastern Kolkata may be included in the confirmed cases. Therefore, reported case numbers are highly sensitive to changes in health-seeking behaviour. It is possible, for example, that during periods of heavy rains typically associated with high cholera risk, residents may be more encouraged to seek hospital treatment. Further, as only around 6% of ID hospital patients are selected for testing, the total number of confirmed cases is low, with an average of only around 10 per month. This makes the data highly susceptible to random noise which could potentially mask statistically significant relationships.

## Results

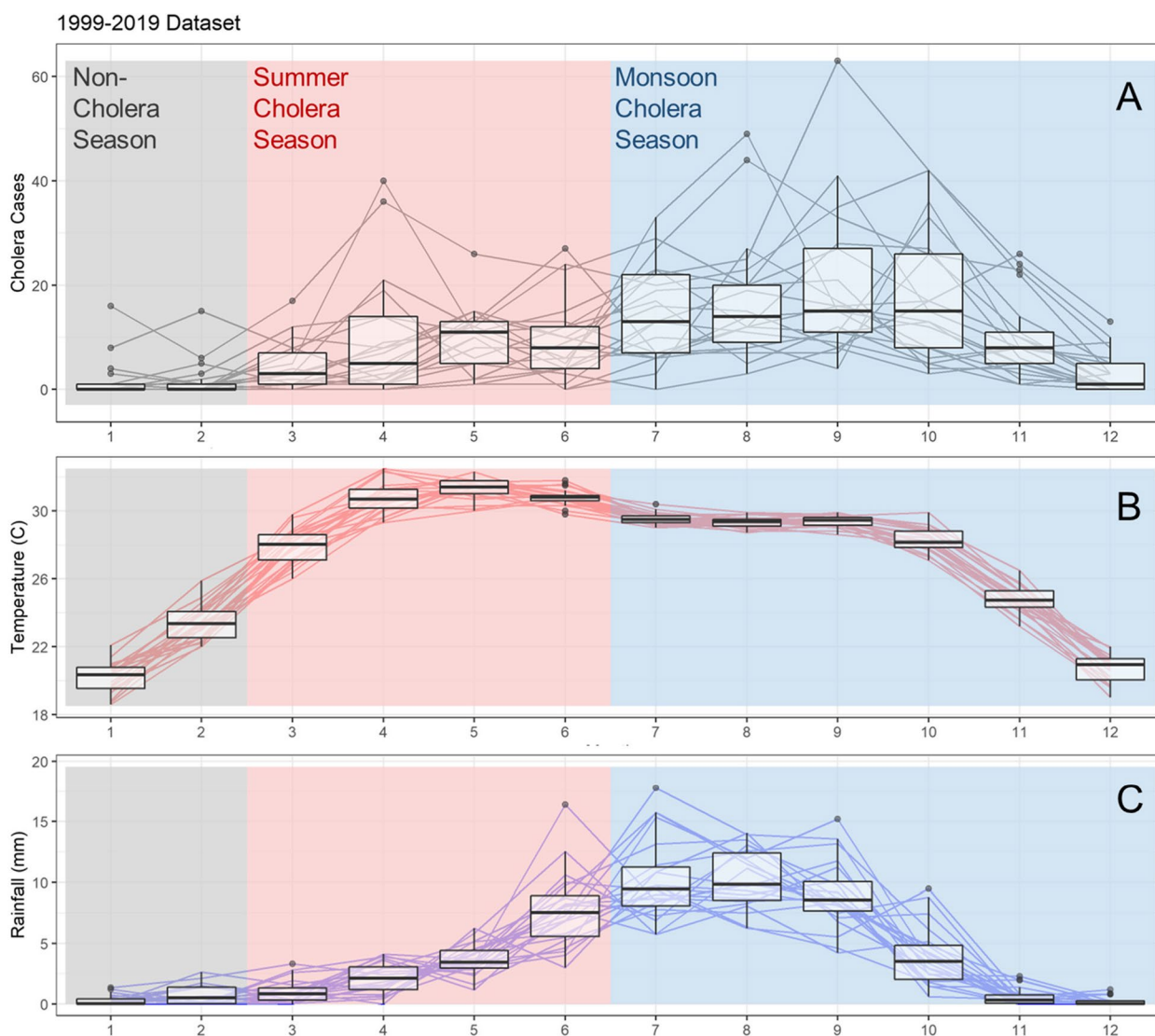
### Descriptive statistics

A total of 2479 confirmed cases of cholera were recorded between 1999 and 2019, the time series is given in Figure S4. To detect the presence of seasonality within cholera cases, we plotted a boxplot of total cholera cases per month from 1999 to 2019 (Fig. 2A). A subtle but distinct bi-annual pattern can be witnessed in the monthly data, the first peaking in April/May with a slight lull in June, followed by a larger peak around September. It can further be seen from that the first peak coincides

The partial effect curves for each variable-season combination are shown in Fig. 3. These represent the component effect of each environmental term in the model which, when combined with the long-term and seasonal partial effects (Figure S5), autoregressive terms and intercept  $\beta_0$ , sum to the overall prediction of the model. The y-axis denotes the logged relative risk and can be interpreted as the logged relative expected cholera case count with respect to the seasonal mean.

During the summer season the influence of increased temperature appears greatest at mean temperatures above 29 °C, with a maximum relative risk (RR) of 3.8 (95%CI 2.35–6.13) occurring at 30.3 °C. Temperatures below 25 °C were also associated with reduced risk where a mean temperature of 24.4 °C has an estimated RR of 0.16 (95%CI 0.06–0.45). From 3 °C it appears there is a slightly negative association between monthly rainfall and cholera cases, however this relationship is not statistically significant.

During monsoon season, the influence of temperature was considerably less pronounced with only temperatures below 22.7 °C being significantly associated with reduced RR, where RR is estimated at 0.37 (95%CI 0.16–0.82) at 21.4 °C. However, the effect of rainfall was much more influential. A total monthly rainfall of 550 mm was



**Fig. 2** Variation in monthly values between 1999–2019 for **A** confirmed cholera cases in Kolkata, **B** mean daytime temperature in the Kolkata area, **C** total precipitation in the Kolkata area. Each line represents values for one year. Boxplots are overlaid to display mean and interquartile ranges for each month. Background colour represents cholera season demarcation, non-cholera season is shown as grey, red as summer, and blue as monsoon

associated with an RR of 3.38 (95%CI 1.37–7.86). It can be seen from Fig. 3 that all three significant relationships are considerably non-linear.

An overall linearly decreasing trend beginning from 2010 can be witnessed in the smooth for long-term trends (Figure S5A). Interestingly the model has estimated a null seasonal trend (Figure S5B) which suggests that the seasonal patterns observed in the data were fully explained by the environmental smooths.

**Lag times**

We computed the strength of the correlation between rainfall and temperature time series with the weekly cholera case time series at lags from 0 to 25 weeks (Fig. 4). We found that the rainfall was significantly and positively correlated with cholera cases during the monsoon season at lags from 0 to 8 weeks with the strongest association occurring at 3 weeks lag. Interestingly a smaller but still significant negative correlation is witnessed from lags 16–21 weeks. Rainfall was

**Table 1** Descriptive summary of seasonal cholera cases, temperature, rainfall and runoff during the years 1999–2019 using monthly dataset

	Season	Mean	Min	Max	Standard Deviation
Monthly Cases	total	9.84	0	63	10.54
	monsoon	13.71	0	63	11.68
	summer	8.11	0	40	8.04
	non-cholera	1.69	0	16	3.61
Temperature (°C)	total	27.20	18.60	32.50	3.83
	monsoon	27.00	19.00	30.40	3.31
	summer	30.20	26.00	32.50	1.48
	non-cholera	21.80	18.60	25.90	1.89
SST (°C)	total	27.64	22.02	30.87	2.29
	monsoon	28.20	23.92	30.11	1.76
	summer	28.73	25.82	30.87	1.27
	non-cholera	23.76	22.02	25.63	0.81
Rainfall (mm/month)	total	126.8	0.00	551.6	132.6
	monsoon	175.3	0.00	551.6	150.2
	summer	110.2	0.00	491.8	93.7
	non-cholera	14.4	0.00	74.1	18.9
Runoff (kg m <sup>-2</sup> s <sup>-1</sup> )	total	16.92	0.66	119.04	23.66
	monsoon	29.11	1.81	119.04	27.86
	summer	6.19	0.66	59.09	8.12
	non-cholera	1.83	1.02	4.67	0.80

not correlated with cholera cases during summer at any of the lags considered. Conversely, temperature was positively associated with cholera cases during the summer season from lags 1–13 weeks, peaking at 7 weeks.

## Discussion

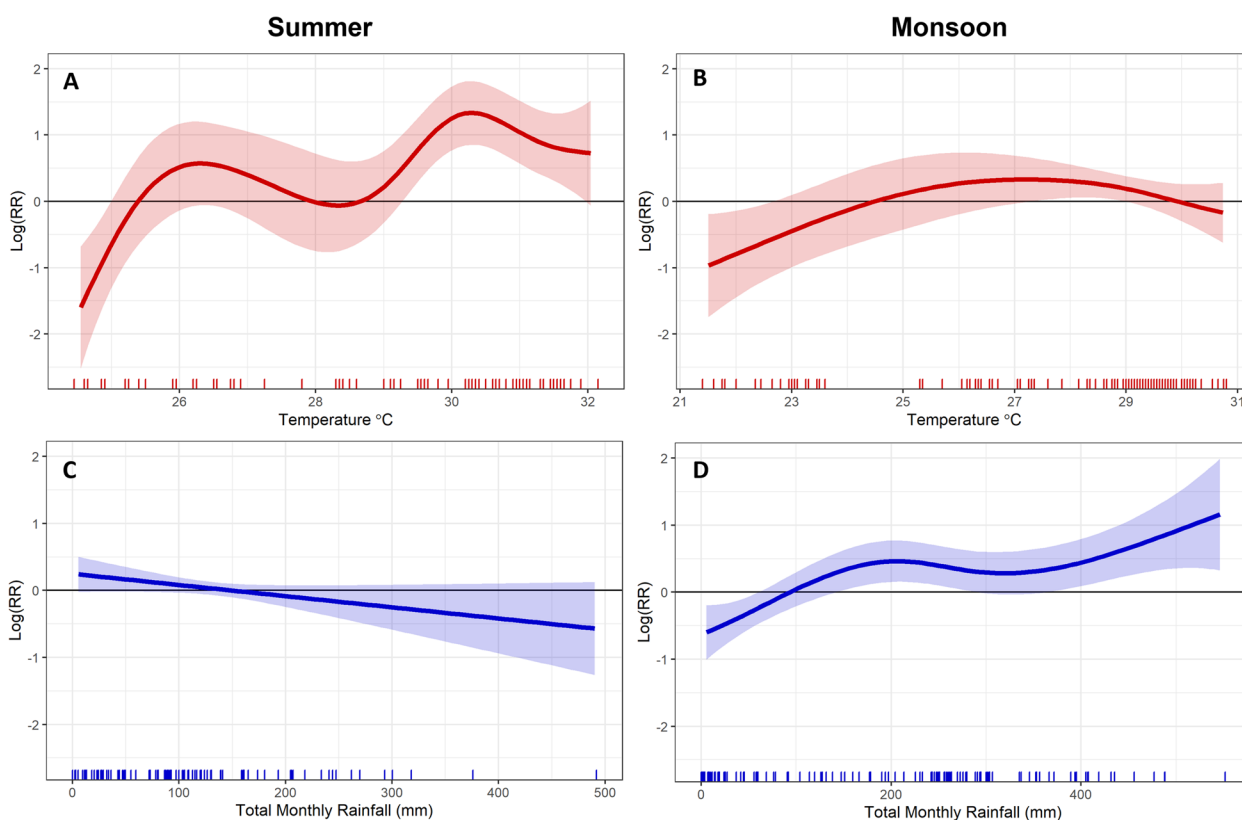
In this study, we explored the relationship between climate and cholera in Kolkata. We determined that cholera in Kolkata demonstrates bi-annual seasonality with distinct summer and monsoon peaks. Further, we found that there is a difference in the factors which influence the magnitude of summer and monsoon cholera peaks; while temperature is the main driver of summer cholera, rainfall dominates the relationship with monsoon cholera. The model which best fit the data did not include SST or rainfall runoff, suggesting that neither variable is helpful in explaining the relationship between climate and cholera. We also found the peak associated lag time between temperature and cholera cases in the summer was more than double that of peak lag time between rainfall and cholera during monsoon.

The seasonal cholera pattern we found in Kolkata is similar to that found in Bangladesh, though with a less pronounced inter-peak period with the secondary peak

beginning during the monsoon rather than following it [7]. However, the pattern is markedly different to that recorded in a historical dataset of the region (then known as Calcutta) during 1891–1940. During that period, the city experienced a single large peak occurring during the summer [8]. We suggest the following explanation for the change in single summer peak cholera in the early 1900s to the dual-peak pattern with maximum during the monsoon observed in our study. A reported 100% of residents of modern (2020) Kolkata have some form of access to a treated municipal water supply [32] meaning that their exposure to the multiplying *Vibrio cholerae* in urban water bodies is likely to be considerably reduced compared with the period 1890–1941, and thereby lessening the potency of this mechanism and reducing the magnitude of the summer peak. While access to sanitation in the region is likely to have increased significantly over the past 100 years, the proportion of households with an improved sanitation facility was only 60.9% and 48.4% in 2019 and 2015 respectively [32] indicating that residents remain vulnerable to the effects of flood water mixing with raw sewage during the monsoon. We propose that a combination of lower immunity levels (due to the lessened summer peak) and greater vulnerability to flood waters due to higher population density (around 4x greater in 1991 compared with 1911 [33]) could have introduced the presence of a monsoon peak.

The difference in associated climate factors between summer and monsoon cholera suggests distinct mechanisms mediating cholera transmission in each season. A potential explanation for the strong association between temperature and summer cholera witnessed in our study is that urban surface water is an important transmission route early in the year. Kolkata residents often come into contact with urban water sources such as ponds, rivers and lakes for the purposes of washing, bathing and swimming [34–36].

We consider that there are two potential explanations for the association with temperature. The first, as posited by Akanda et al. [7] among others, is that the preference of *V.cholerae* for warmer waters leads to proliferation of the bacteria during hot periods. This is in line with laboratory studies which found that *V.cholerae* O1 cells multiply most effectively when incubated at warmer temperatures up to 30 °C [37, 38]. Thus, increased temperatures could lead to an increase in concentration of the pathogenic bacteria and thereby, due to the dose dependent nature of cholera infections [39], increase the probability of cholera infection at a given interaction with urban surface water bodies. A second, alternative explanation for the association between temperature and summer cholera cases is increased contact with water bodies such as ponds, canals and the Hooghly River as one of the



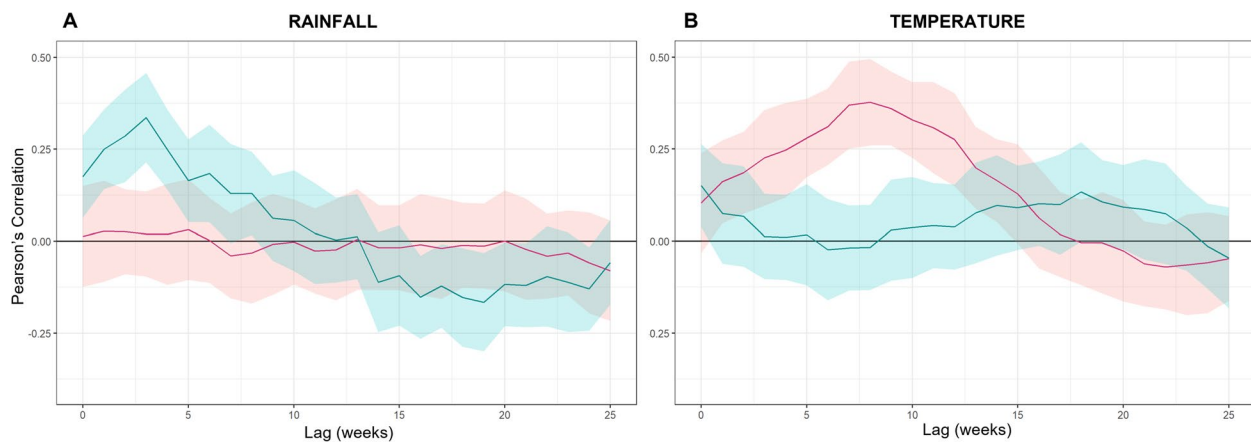
**Fig. 3** Partial effect curves for relationship between reported cholera cases with average temperature across 0- and 1-month lag (A, B) and rainfall in concurrent month (C, D) in summer(left) and monsoon (right) seasons after controlling for seasonal and long-term trends and the inclusion of a linear autoregressive covariate. The 95% confidence interval is shown with the lighter band. The null hypothesis of no effect is shown by the horizontal black line; significant relationships can be inferred in regions where the confidence interval does intersect the black line. The rug plot at the bottom of each plot displays the observed values for each covariate within each season. The y-axis represents the (logged) relative risk of cholera incidence with respect to the (logged) mean of the response variable

few available methods of ‘cooling off’ during heatwaves available to the urban poor. The results of our lag analysis support the former hypothesis as a longer lag time would be expected under such as scenario to allow for a build-up of bacteria in urban water bodies before reaching a critical threshold capable of causing cholera infection. The latter hypothesis would predict a much shorter associated lag time, namely the length the time between a given hot day, and the time taken for cholera to take hold enough for the patient to be admitted to hospital and the stool sample to be taken – a time likely to be no longer than 2–3 weeks.

A further interesting finding with regards to the relationship between cholera and temperature is its marked non-linearity, echoing findings previously reported in Azerbaijan where cholera cases peaked at an air temperature of 25 °C [40]. Similarly, our results show a linear correlation up until approximately 26 °C, after which the trend weakens.

The positive association with rainfall only witnessed in the monsoon season in our study is highly consistent with results from Bangladesh [12, 14]. One potential explanation for the positive association with rainfall, also suggested by Akanda et al. [10] among others, is that rainfall induced floods lead to mixing between flood and sewage waters, as well as an increased contact between residents and contaminated flood waters. An alternative explanation which has been suggested to explain cholera outbreaks in Central India [41] involves the role of rainfall runoff. In areas where open defecation remains an issue, including Kolkata [42], rainfall can wash human faecal matter from these open defecation sites into accessible water sources, leading to contamination and potential cholera outbreaks. However, this second explanation would suggest that modelled rainfall runoff would be a better predictor variable for cholera than rainfall, which was not supported in our study. Further a significant role





**Fig. 4** Correlation between aggregated seasonal cholera and concurrent/lagged seasonal rainfall (A) and temperature (B) during summer (red) and monsoon (blue) for lags 0–25 weeks. Lighter band represents the 95% confidence interval

of contaminated runoff would imply a positive association with rainfall during both dry and wet seasons, which is also inconsistent with our findings. A lack of positive association between cholera and rainfall in the summer remains consistent with the flooding hypothesis, as the rainfall which occurs during the summer generally does not lead to flood events [43]. We therefore suggest that the flooding hypothesis is likely the primary mechanism mediating the monsoon rainfall-cholera relationship.

Increases in SST in the Bay of Bengal have been suggested to be linked to cholera cases in the Bengal Delta by an associated increase in phytoplankton. This is suggested to increase both zooplankton populations and pH levels, both of which are considered to promote *Vibrio Cholerae* populations in local estuaries [6]. Coastal intrusion, especially during the summer, then leads to greater *Vibrio Cholerae* concentrations in the rivers upon which many urban areas are built upon. That our study found SST was not a useful indicator of cholera, highlights the spatial heterogeneity in SST-cholera relationships across the Bengal Delta. The reasons for these discrepancies remain unclear, especially given Kolkata's proximity to the Hooghly River, which is highly estuarine in Kolkata with a tidal oscillation of ~3 m [44], however this could indicate a decreased role of the river in cholera transmission in place of ponds and canals compared with Dhaka.

Our results suggest a mediating role of increased pathogenic *V.cholerae* in urban water bodies and as such efforts to reduce direct contact with urban water bodies, such as through public health campaigns or improved washing facilities preceding and during the summer season could be effective at reducing cholera cases. Conversely, our results suggest a mediating role of contact with contaminated flood waters during the monsoon cholera season, and therefore interventions focussing

on protecting residents from flood waters could be most effective in the short-term – and in the long-term, universal safe sanitation to prevent initial contamination of flood waters.

Our findings suggest that hotter summers and wetter monsoons are conducive to high numbers of cholera cases in Kolkata. This is concerning due to a widely projected increase in the volume of rain falling during the monsoon season in the South Asian region a result of greenhouse-gas forcing [45]. In addition summer heat-waves are predicted to become frequent, intense, and prolonged across South Asia with maximum exposure occurring in the Indo-Gangetic Plain [46]. This suggests that, from a climate standpoint, the vulnerability of Kolkata to cholera is likely to increase over time. It is therefore vital to remain vigilant in efforts to improve sanitation in the region.

## Conclusions

Our results find significant non-linear relationships between climate factors and cholera in Kolkata, with temperature driving summer cholera and rainfall driving monsoon cholera. We suggest the summer outbreak is mediated by increased pathogenic *Vibrio* concentration in urban water bodies, and the monsoon outbreak by contaminated rainfall-runoff. In this sense, we find that the relationship with climate in Kolkata is similar to that in Bangladesh. However, important differences were found including a lack of association with SST in Kolkata and a diminished early-monsoon lull. With regards to intervention strategies, we suggest that summer and monsoon peaks are considered separately with an increased emphasis on separation from urban water bodies during the summer season, and protection from flood waters in the monsoon.

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12879-023-08532-1>.

### Additional file 1.

### Acknowledgements

We thank the Indian National Institute of Cholera and Enteric Diseases (NICED) for making this work possible. We would also like to thank the Climate Research Unit and the University of East Anglia for making their meteorological estimations publicly available.

### Authors' contributions

SD and SK were instrumental in the epidemiological data collection. DS conducted the analysis and wrote the first draft of the manuscript. TE, FM, AC, and SK critically revised the manuscript. All authors have read and agreed to the published version of the manuscript.

### Funding

This work was supported by the EPSRC (WISE CDT; EP/L016214/1). There were no roles of the funding body in the design of the present study, collection, analysis, and interpretation of data, or in writing the manuscript.

### Availability of data and materials

Rainfall, temperature, and SST data are all publicly available, with details given in the methodology section. Monthly rainfall and temperature data can be accessed at <https://catalogue.ceda.ac.uk/> and weekly data at <https://crudata.uea.ac.uk/cru/data/hrg/>. Monthly SST data can be accessed at <https://catalogue.ceda.ac.uk/uuid/facafa2ae494597166217a9121a62d3c>. Epidemiological data is the property of the Indian Council of Medical Research (ICMR) and is not publicly available due to political, privacy and ethical concerns. Please contact the corresponding author to request the R code files.

### Declarations

#### Ethics approval and consent to participate

All ethical considerations and guidelines relating to the analysis of human epidemiological datasets were followed, namely full anonymization and aggregation of patient data beyond plausible identifiability. Data collection and analysis were approved by the Indian Council of Medical Research (ICMR).

#### Consent for publication

Not applicable.

#### Competing interests

The authors declare no competing interests.

Received: 1 February 2023 Accepted: 11 August 2023

Published online: 02 September 2023

### References

- Ali M, Sen Gupta S, Arora N, Khasnabis P, Venkatesh S, Sur D, et al. Identification of burden hotspots and risk factors for cholera in India: an observational study. *PLoS ONE*. 2017;12:e0183100. <https://doi.org/10.1371/journal.pone.0183100>.
- Usmani M, Brumfield KD, Jamal Y, Huq A, Colwell RR, Jutla A. A Review of the Environmental Trigger and Transmission Components for Prediction of Cholera. *Trop Med Infect Dis*. 2021;6. <https://doi.org/10.3390/tropicalmed6030147>.
- Christaki E, Dimitriou P, Pantavou K, Nikolopoulos GK. The impact of climate change on cholera: a review on the global status and future challenges. *Atmos (Basel)*. 2020;11(5):449. <https://doi.org/10.3390/ATMOS11050449>.
- Asadgol Z, Mohammadi H, Kermani M, Badirzadeh A, Gholami M. The effect of climate change on cholera disease: the road ahead using artificial neural network. *PLoS ONE*. 2019;14:e0224813. <https://doi.org/10.1371/JOURNAL.PONE.0224813>.
- Giebultowicz S, Ali M, Yunus M, Emch M. A comparison of spatial and social clustering of cholera in Matlab. *Bangladesh Health Place*. 2011;17:490–7. <https://doi.org/10.1016/j.healthplace.2010.12.004>.
- Shackleton D, Memon FA, Nichols G, Phalkey R, Chen AS. Mechanisms of cholera transmission via environment in India and Bangladesh: state of the science review. *Rev Environ Health*. 2023. <https://doi.org/10.1515/revh-2022-0201>.
- Akanda AS, Jutla AS, Islam MS. Dual peak cholera transmission in Bengal Delta: a hydroclimatological explanation. *Geophys Res Lett*. 2009;36:L19401. <https://doi.org/10.1029/2009GL039312>.
- Bouma MJ, Pascual M. Seasonal and interannual cycles of endemic cholera in Bengal 1891–1940 in relation to climate and geography. *Hydrobiologia*. Springer. 2001;460:147–56. <https://doi.org/10.1023/A:1013165215074>.
- Hashizume M, Faruque ASG, Wagatsuma Y, Hayashi T, Armstrong B, Hashizume M, et al. Cholera in Bangladesh: climatic components of seasonal variation. *Epidemiology*. 2010;21:706–10. <https://doi.org/10.1097/EDE.0b013e3181e5b053>.
- Akanda AS, Jutla AS, Alam M, De Magny GC, Siddique AK, Sack RB, et al. Hydroclimatic influences on seasonal and spatial cholera transmission cycles: implications for public health intervention in the Bengal Delta. *Water Resour Res*. 2011;47:1–11. <https://doi.org/10.1029/2010WR009914>.
- Koelle K, Rodó X, Pascual M, Yunus M, Mostafa G. Refractory periods and climate forcing in cholera dynamics. *Nature*. 2005;436:696–700. <https://doi.org/10.1038/nature03820>.
- Wu J, Yunus M, Ali M, Escamilla V, Emch M. Influences of heatwave, rainfall, and tree cover on cholera in Bangladesh. *Environ Int*. 2018;120:304–11. <https://doi.org/10.1016/j.envint.2018.08.012>.
- Zhao S, Musa SS, Qin J, He D. Associations between public awareness, local precipitation, and cholera in Yemen in 2017. *Am J Trop Med Hyg*. 2019;101:521–4. <https://doi.org/10.4269/ajtmh.18-1016>.
- Vezzulli L, Grande C, Reid PC, Hélaouët P, Edwards M, Höfle MG, et al. Climate influence on Vibrio and associated human diseases during the past half-century in the coastal North Atlantic. *Proc Natl Acad Sci U S A*. 2016;113:E5062–5071. <https://doi.org/10.1073/pnas.1609157113>.
- De Magny GC, Murtugudde R, Sapiano MRP, Nizam A, Brown CW, Busalacchi AJ, et al. Environmental signatures associated with cholera epidemics. *PNAS*. 2008;105:17676–81. <https://doi.org/10.1073/pnas.0809654105>.
- Mukhopadhyay AK, Deb AK, Chowdhury G, Debnath F, Samanta P, Saha RN, et al. Post-monsoon waterlogging-associated upsurge of cholera cases in and around Kolkata metropolis, 2015. *Epidemiol Infect*. 2019;147:e167. <https://doi.org/10.1017/S0950268819000529>.
- Sur D, Deen JL, Manna B, Niyogi SK, Deb AK, Kanungo S, et al. The burden of cholera in the slums of Kolkata, India: data from a prospective, community based study. *Arch Dis Child*. 2005;90:1175–81. <https://doi.org/10.1136/adc.2004.071316>.
- Bengal D. of COW. District Census Handbook, Kolkata. Kolkata: 2014. <https://doi.org/10.4324/9780203075906-24>.
- Harris I, Osborn TJ, Jones P, Lister D. Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset. *Sci Data*. 2020;7:1–18. <https://doi.org/10.1038/s41597-020-0453-3>.
- Tian H, Liu C, Zhang W, Tao B. The dynamic land ecosystem model (DLEM) for simulating terrestrial processes and interactions in the context of multifactor global change. *Acta Geogr Sin*. 2010;65:1027–47.
- Sitch S, Friedlingstein P, Gruber N, Jones SD, Murray-Tortarolo G, Ahlström A, et al. Recent trends and drivers of regional sources and sinks of carbon dioxide. *Biogeosciences*. 2015;12:653–79. <https://doi.org/10.5194/BG-12-653-2015>.
- Joint UK. Land Environment Simulator (JULES) n.d. <https://jules.jchmr.org/>. Accessed 25 Jan 2023.
- Stocker BD, Roth R, Joos F, Spahni R, Steinacher M, Zaehle S, et al. Multiple greenhouse-gas feedbacks from the land biosphere under future climate change scenarios. *Nat Clim Change*. 2013;2013(3):7. <https://doi.org/10.1038/nclimate1864>.

24. Sun Y, Goll DS, Chang J, Ciais P, Guenet B, Helfenstein J, et al. Global evaluation of the nutrient-enabled version of the land surface model ORCHIDEE-CNP v1.2 (r5986). *Geosci Model Dev*. 2021;14:1987–2010. <https://doi.org/10.5194/GMD-14-1987-2021>.
25. Kobayashi S, Ota Y, Harada Y, Ebata A, Moriya M, Onoda H, et al. The JRA-55 reanalysis: General specifications and basic characteristics. *J Meteorol Soc Jpn*. 2015;93:5–48. <https://doi.org/10.2151/jmsj.2015-001>.
26. Met Office, Hadley Centre. Global sea-ice coverage and SST (1870-Present), [Internet]. NCAS British Atmospheric Data Centre 2006. [http://badc.nerc.ac.uk/view/badc.nerc.ac.uk\\_\\_ATOM\\_\\_dataent\\_hadisst](http://badc.nerc.ac.uk/view/badc.nerc.ac.uk__ATOM__dataent_hadisst) (accessed 25 Jan 2023).
27. Wood SN. Generalized additive models: An introduction with R, second edition. CRC Press; 2017. <https://doi.org/10.1201/9781315370279/GENERALIZED-ADDITIVE-MODELS-SIMON-WOOD>.
28. Wood SN. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *J Royal Stat Soc (B)*. 2011;73:3–36.
29. Akaike H. Maximum likelihood identification of Gaussian Autoregressive moving average models. *Biometrika*. 1973;60:255–65.
30. Cameron AC, Trivedi PK. Time Series Data. Regression analysis of Count Data. Bloomington: Cambridge University Press; 2014.
31. Sen D. Flood hazards in India and management strategies. Natural and Anthropogenic Disasters, New Delhi: Springer; 2010.
32. Indian Ministry of Health and Family Welfare. National Family Health Survey 5 - Kolkata. Kolkata: International Institute for Population Sciences; 2020.
33. Government of Bengal. Reports of the Bengal Sanitary Board and the Public Health Department. vol. 1. Calcutta: Bengal Secretariat Press; 1912.
34. Sen Gupta S. Biomonitoring the status of aquatic bodies using zooplankton as surrogate species amidst urban landscape. *Holist Approach Environ*. 2022;12:144–54. <https://doi.org/10.33765/thate.12.4.2>.
35. Goswami SN, Trivedi R, Saha S, Mandal A. Seasonal variations of water characteristics in three urban ponds with different management practices at Kolkata of West Bengal. *India: J Entomol Zool Stud*; 2017.
36. Jha P, Banerjee S, Bhuyan P, Sudarshan M, Dewanji A. Elemental distribution in urban sediments of small waterbodies and its implications: a case study from Kolkata, India. *Environ Geochem Health*. 2020;42:461–82. <https://doi.org/10.1007/s10653-019-00377-5>.
37. Huq A, West PA, Small EB, Huq I, Colwell RRR. Influence of water temperature, salinity, and pH on survival and growth of toxigenic *Vibrio cholerae* serovar O1 associated with live copepods in laboratory microcosms. *Appl Environ Microbiol*. 1984;48:420–4. <https://doi.org/10.1128/aem.48.2.420-424.1984>.
38. Singleton FL, Attwell R, Jangi S, Colwell RR. Effects of temperature and salinity on *Vibrio cholerae* growth. *Appl Environ Microbiol*. 1982;44:1047–58.
39. Nelson EJ, Harris JB, Morris JG Jr, Calderwood SB, Camilli A. Cholera transmission: the host, pathogen and bacteriophage dynamic. *Nat Rev Microbiol*. 2009;7:693–702. <https://doi.org/10.1038/nrmicro2204>.
40. Gurbanov S, Akhmadov R, Shamkhalova G, Akhmadova S, Haley BJ, Colwell RR, et al. Occurrence of *Vibrio cholerae* in municipal and natural waters and incidence of cholera in Azerbaijan. *EcoHealth*. 2011;8:468–77. <https://doi.org/10.1007/s10393-012-0756-8>.
41. Dutta BP, Kumar N, Meshram KC, Yadav R, Sodha SV, Gupta S. Cholera outbreak associated with contaminated water sources in paddy fields, Mandla District, Madhya Pradesh, India. *Indian J Public Health*. 2021;65:46–50. [https://doi.org/10.4103/jiph.IJPH\\_1118\\_20](https://doi.org/10.4103/jiph.IJPH_1118_20).
42. Roy C, Sati VP, Biswas A, Kumar S. Status of drinking water, sanitation facilities, and hygiene in West Bengal: evidence from the National Family Health Survey of India (NFHS), 2019–2021. *J Water Sanitation Hygiene Dev*. 2023;13:50–62. <https://doi.org/10.2166/washdev.2023.228>.
43. Patra P, Satpati L. Precipitation Trends in the City of Kolkata and Its Implication on Urban Flooding. *Geogr Rev India*. 2017;4:335–51.
44. Prakash D, Tiwary CB, Kumar R. Ecosystem variability along the Estuarine Salinity Gradient: a case study of Hooghly River Estuary, West Bengal, India. *J Mar Sci Eng*. 2023;11:1–22. <https://doi.org/10.3390/jmse11010088>.
45. Turner AG, Annamalai H. Climate change and the South Asian summer monsoon. *Nature Climate Change*. 2012;2(8):587–95. <https://doi.org/10.1038/nclimate1495>.
46. Ullah S, Qinglong Y, Chen D, Sachindra DA, AghaKouchak A, Kang S, et al. Future Population exposure to Daytime and Nighttime Heat waves in South Asia. *Earths Future*. 2022;10(5):e2021EF002511. <https://doi.org/10.1029/2021EF002511>.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

### Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

