Technical Report

Relationship between climate variability and occurrence of diarrhoea and cholera

A pilot study using retrospective data from Kolkata, India

World Health Organization
REGIONAL OFFICE FOR South-East Asia

World Health Organization
Centre for Health Development

2011

Submitted by:

National Institute of Cholera & Enteric Diseases (NICED)

Kolkata, India

INDIAN COUNCIL OF MEDICAL RESEARCH
NEW DELHI

This report has been prepared based on a technical services agreement between NICED and WHO Kobe Centre, Japan. The following persons contributed in preparation of different aspects of this report:

Dr G.B. Nair
Project Team Leader
Director, National Institute of Cholera & Enteric Diseases (NICED)
P-33, CIT Road, Scheme-XM, Beliaghata, Kolkata 700 010
E-mail: nairgb@icmr.org.in; gbnair_2000@yahoo.com
Tel: +91-33-2363 3373

Dr Alok Kumar Deb
Scientist 'D', NICED

Dr Suman Kanungo
Scientist 'B', NICED

Dr Anup Palit
Scientist 'E', NICED

Dr Susmita Chatterjee
Project Assistant

Mitali Sen
Project Assistant

Acknowledgements

This pilot study was carried out based on a generic research protocol developed to assess retrospectively the negative health impact of climate change on diarrhoeal diseases with emphasis on cholera and capacity of health system to cope with the consequences. The generic protocol was prepared for WHO South-East Asia Regional Office (SEARO) as an agreement for performance of work with the National Institute of Cholera and Enteric Diseases, Kolkata (NICED) and extensively reviewed through an informal scientific consultation carried out in Kolkata (India). The present study was developed by the same team, with collaboration and funding from WHO Kobe Centre. NICED also gratefully acknowledge the support of the Department of Communicable Diseases, SEARO.

© World Health Organization 2011

All rights reserved. Requests for permission to reproduce or translate WHO publications – whether for sale or for noncommercial distribution – should be addressed to the WHO Centre for Health Development, I.H.D. Centre Building, 9th Floor, 5-1, 1-chome, Wakinohama-Kaigandori, Chuo-ku, Kobe City, Hyogo Prefecture, 651-0073, Japan (fax: +81 78 230 3178; email: wkc@wkc.who.int).

The designations employed and the presentation of the material in this publication do not imply the expression of any opinion whatsoever on the part of the World Health Organization concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. Dotted lines on maps represent approximate border lines for which there may not yet be full agreement.

The mention of specific companies or of certain manufacturers' products does not imply that they are endorsed or recommended by the World Health Organization in preference to others of a similar nature that are not mentioned. Errors and omissions excepted, the names of proprietary products are distinguished by initial capital letters.

All reasonable precautions have been taken by WHO to verify the information contained in this publication. However, the published material is being distributed without warranty of any kind, either express or implied. The responsibility for the interpretation and use of the material lies with the reader. In no event shall the World Health Organization be liable for damages arising from its use.

This publication contains the collective commissioned research team and does not necessarily represent the decisions or the stated policy of the World Health Organization.

Table of contents

List of tables and figures

List of Abbreviations:

ACF	Auto-Correlation Function
AR	Auto Regression
ARIMA	Auto Regressive Integrated Moving Average
CCF	Cross-Correlation Function
LOWESS	Locally Weighted Scatterplot Smoothing
NICED	National Institute of Cholera & Enteric Diseases
PACF	Partial Auto-Correlation Function
RRSC	Regional Remote Sensing Centre
SD	Standard Deviation
SEARO	WHO South-East Asia Regional Office
SSH	Sea Surface Height
SST	Sea Surface Temperature
WHO	World Health Organization

Introduction

Infectious diseases, once expected to be eliminated as a significant public health problem, remain one of the leading causes of death in the world. Many factors have contributed to the persistence and increase in the occurrence of infectious diseases, such as environmental changes, societal changes, deteriorating health care, mass food production, human behavior, public health infrastructure and microbial adaptation. Climate change, at the levels projected by current global climate models, may have important and far-reaching effects on infectious diseases, especially those transmitted by arthropod vectors, as well as water-borne diseases. Although most scientists agree that global climate change influences the transmission dynamics of infectious diseases, the exact nature and extent of such influence still remain uncertain.

The poorest countries are likely to be disproportionately affected by climate change impacts on health, due to structural shortcomings involving inadequate public health infrastructure, poor disease surveillance, weak health systems, lack of human and financial resources and inadequate emergency preparedness and management systems, among others. High population density may compound climate change vulnerability due to poverty, further displacing the burden of health outcomes towards the most disadvantaged regions and countries. The WHO South East Asia Region (comprising 11 countries; see http://www.who.int/about/regions/searo) is home to 26% of the world's population and 30% of the world's poor (1). Because of this combination of factors, the consequences of climate change could be disastrous for the region.

Acute diarrhoeal disease is one of the most important health-related impacts linked to short term and long term changes in the climate. The frequency and intensity of extreme climate events such as droughts, floods and cyclone directly impact the prevalence of diarrhoeal diseases. Currently, 1.3 billion episodes of diarrhoea occur worldwide each year, mostly in children, with an average of 2-3 episodes per child per year (2). According to WHO's global diarrhoeal disease burden estimate (3), the majority of the 62 451 000 DALYs lost in 2001 were from developing country people, where children suffer from as many as 12 episodes of diarrhoea each year.

Cholera infections vary greatly in frequency, severity, and duration in different parts of the world. In some south Asian countries it is endemic, whereas in some parts of Africa and South America sporadic outbreaks occur, which is also a common feature in the endemic areas. Past studies have shown that cholera varies with seasons (4). Environmental and climatic factors are believed to be the key players for this temporal variation (5-7). These climatic conditions, however, are intermingled with other environmental and socio-demographic conditions, which also play a profound role in variation in the incidence of diarrhoeal disease.

Several researchers have postulated a link between heavy rainfall and flooding – whether resulting from El Niño-associated events or from other meteorological impacts – and subsequent outbreaks of diarrhoeal diseases (8). Recent analysis between cholera time series and El Nino Southern Oscillation (ENSO) with specific time intervals has found that this climate phenomenon may account for over 70% of the disease variance (9). Besides rainfall, two other remote consequences of inter-annual climate variability – sea surface temperatures (SSTs) and chlorophyll-a in the Bay of Bengal – are proposed to influence cholera in Bangladesh (8). In Peru, studies showed that a 5^0C rise in temperature caused a 200% increase in diarrhoea admission during 1997–1998 and each 1^0C rise in temperature caused 8% increase in the risk of getting severe diarrhoea in children (10).

Studies have also demonstrated a relationship between climate variability and non-cholera diarrhoea. Studies from Bangladesh showed that the number of non-cholera diarrhoea cases increased with heavy rainfall and higher temperature, particularly in lower socioeconomic settings with poor sanitation (11). Influence of climate variation on diarrhoeal diseases has been evident in Pacific Islands also. Evidence shows that the higher the temperature and the lower the potential water availability, the higher the occurrence of diarrhoea is (12).

Notwithstanding a plethora of information on climate-diarrhoea and climate-cholera relationships generated through several studies, the exact nature and magnitude of such relationships are still not well understood. This is mostly due to the failure to account for the potentially confounding effects of important non-climate factors such as nutritional status (13–16), personal hygiene (13,17,18), immunization (especially measles) (13), infant feeding practices (13,15,19), safe drinking water and environmental sanitation (17–21), education level (15,22), socioeconomic condition and household overcrowding (23). We embarked on the current study in an attempt to obtain data on these variables and adjust for their effects during evaluation of the climate-diarrhoea relationship.

Objectives

General: To determine the effects of climate variability on changes in the occurrence of diarrhoeal diseases over time

Specific: 1) to clarify the relationship between climatic factors and occurrence of diarrhoea and cholera (specifically through the evaluation of association of the following variables through collection and analysis of 10 years (1999–2008) retrospective data) and 2) to identify and test a range of non-climatic factors that can potentially influence the above relationship.

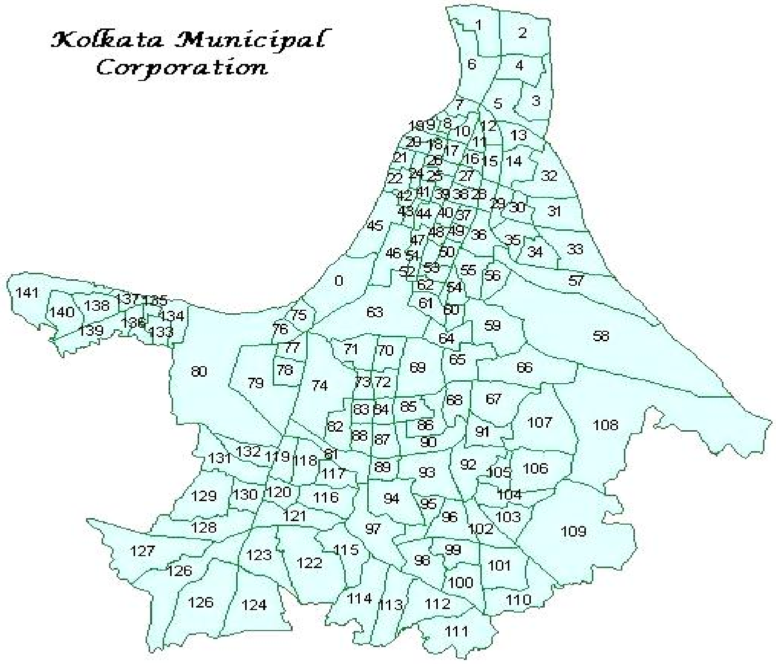
Methods

Study design

This was a retrospective longitudinal study involving analysis of ten-year secondary data for the period 1999–2008.

Study area

Data pertaining to the Kolkata Municipal Corporation area only were considered for this study. The rationale being – (a) the climate data collected from local weather stations were relevant for this municipal area and (b) diarrhoea data used in this study were obtained from the Infectious Diseases Hospital, Kolkata, where most (86%) of the diarrhoea admissions were from Kolkata Municipal Corporation area.



The Kolkata Municipal Corporation (KMC) area, comprising of 141 civic administrative units (called municipal wards), has an area of 185 km^2. Kolkata, a city more than 300 years old, is currently the capital of the state of West Bengal in eastern India. Geographically, it is located at 22°33′N and 88°20′E, just beneath the Tropic of Cancer. The Bay of Bengal, which lies to the south of Kolkata, is approximately 120 km away. Being close to the sea, Kolkata is little more than 5 meters above sea level and the city's weather is also greatly influenced by it. Physically, Kolkata is connected to the Bay of Bengal by the Hooghly River that flows on the western side of the city in a north-south direction. The river provides the city with a gentle slope towards the east and south-east. In the eastern part of the city, large stretches of wetlands and marshy areas of swamps, spreading over an area of 12 500 hectares, provide a unique urban ecosystem in an environmentally sensitive area. Much of the city was originally a vast wetland, reclaimed over the decades to accommodate the city's burgeoning population. Kolkata has a tropical wet-and-dry climate (Köppen climate classification *Aw*). There are three major seasons, summer, monsoon and winter, with a brief intervening period of "autumn". The annual mean temperature is 26.8 C (80.2 F); monthly mean temperatures range from 19 C to 35 C. Summers are hot and humid with temperatures in the low 30s and during dry spells the maximum temperatures often exceed 40 C (104°F) during May and June. Winter tends to last for only about two and a half months, with seasonal lows dipping to 9–11 C (54 F–57°F) between December and January. The highest ever recorded temperature is 43.9 C (111.0 F) and the lowest is 5 C (41.0°F).[1] On average, May is the hottest month with daily temperatures ranging from a low of 27 C (80.6°F) to a maximum of 37°C (98.6 F), while January, the coldest month has temperatures varying from a low of 12 C (53.6 F) to a maximum of 23 C (73.4°F). Often during early summer, dusty squalls followed by thunderstorms or hailstorms and heavy rains with hail lash the city, bringing relief from the humid and heat wave conditions. All these climatic conditions yield a favorable environment for considerably high incidence and persistence of diarrhoea and cholera; in fact, historically this area is often called the "Cradle of Cholera". Thus, Kolkata offered us a unique opportunity to study the relationship between diarrhoea (and cholera) and a number of climate factors.

Data collection

To determine the associations as stated before, the data requirement and possible sources of the required data are outlined below.

Climate data

(a) Meteorological data:

> We collected ten-year retrospective data on maximum and minimum temperature (°C), rainfall (mm), and relative humidity (%) in the morning and evening. These data were obtained from Alipore Weather Station, Kolkata (Indian Meteorological Dept., Govt. of India).

(b) Remote-sensing data:

> These data were obtained from the Regional Remote Sensing Centre (RRSC), Kharagpur, West Bengal. We planned to obtain the ten-year data on sea surface temperature (SST), sea surface height (SSH) and chlorophyll-a. However, data on SSH could not be supplied by RRSC and the supplied data on chlorophyll-a were incomplete (contained missing data for several months) and unusable for our purposes. Hence, only SST data were used for the present study.

Disease data

The Infectious Diseases (I.D.) Hospital, Kolkata caters to all kinds of infectious disease (including diarrhoea) cases arising from Kolkata and its adjoining districts. This institute (NICED) is conducting surveillance of diarrhoea cases admitted to this hospital for almost the past two decades. Every fifth diarrhoea case admitted to this hospital is enrolled in this surveillance system on two randomly selected days per week. Demographic, socioeconomic and clinical history are obtained from these enrolled cases and stool specimen are also collected to identify the causative organism, especially cholera. These surveillance data for the relevant time period were used as a source of diarrhoea and cholera data in the present study.

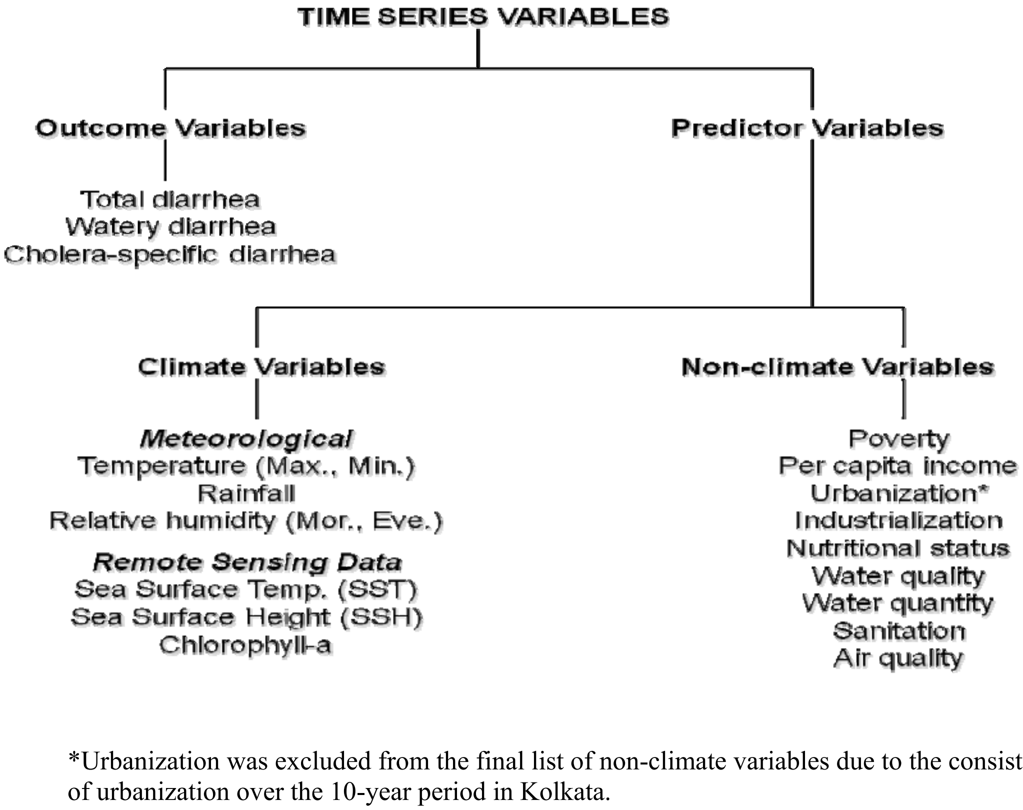
Non-climate data – The following steps were followed for identification of relevant non-climate factors:

1. Lists of possible factors were individually prepared through an online search of relevant publications and reports. These lists were compared, discussed, and finally combined to prepare the common list of non-climate factors that could affect occurrence of diarrhoea and/or cholera

2. The factors that were possibly not associated with various climate variables were excluded from the final list, so they could not act as possible confounders

3. The final list of possible non-climate factors that might need adjustment during evaluation of diarrhoea-climate association was identified.

Thus, the final list of all possible variables on which data needed to be collected was prepared as follows.



*Urbanization was excluded from the final list of non-climate variables due to the consistent (100%) level of urbanization over the 10-year period in Kolkata.

Identification of possible sources for relevant non-climate data:

1. Health on the March, Dept. of Health &Family Welfare, Govt. of West Bengal
2. Bureau of Applied Economics & Statistics, Govt. of West Bengal
3. Directorate of Census Operations, West Bengal, Kolkata – 700 069
4. Kolkata Municipal Corporation
5. Other published reports

The research team tried to collect data on the identified non-climate variables (as stated above) from these sources. However, some of these data were not available for the study area and data on other variables were either incomplete or not measured uniformly. For example, after a long discussion, it was considered that "poverty" could be measured by a surrogate variable "people living *below the poverty line*" – popularly known as "BPL", which is an indicator set by the government to determine the percentage of "poor" people in an area that are eligible for different kinds of government assistance. However, this is based on gross monthly income of the families; in addition, the income cut-off to determine whether a family belongs to "BPL" is changed from time to time. Moreover, the Kolkata Municipal Corporation (KMC) has maintained formal lists of BPL families only for the past 2–3 years. Thus, it was difficult to use poverty data for this population. Data on "per capita income", "industrialization" and "nutritional status of children" were only available from the Census data and/or data from the National Family Health Survey (NFHS). Census data only indicates decadal status and NFHS data have been collected only twice (NFHS-2 in 1998–99 and NFHS-3 in 2005–06) during the period of concern for this study. Thus, these data too were inadequate for use in the analysis. Data on water quality could be retrieved for most of the study period from government sources. However, these data were not reported uniformly for different years under the study and it was not possible to make them uniform and comparable over the entire measurement period – leading us to drop these variables from our analysis also. Other variables of potentially important explanatory power had to be dropped from the analysis for similar reasons. Thus, in our final analysis, we only considered the disease data and data on climatic factors including remotely sensed data as described below.

Data entry, cleaning and editing:

Software used

For data entry, cleaning, editing, merging: MS-Excel

For data analysis: SYSTAT 12.0, SPSS Statistics 17.0

Data extrapolation

We obtained diarrhoea/cholera data as daily values from the ongoing diarrhoea surveillance system where every fifth admitted cases were captured on two randomly selected days in every week. From average number of cases captured per surveillance day in any month and using the preceding information, we estimated the total number of diarrhoea admissions in that month. For estimating monthly number of cholera cases, we additionally used information on type of diarrhoea (only watery or loose diarrhoea was considered) and whether appropriate laboratory test was done to identify cholera. For climate variables, monthly averages were calculated from the collected daily values of each variable.

Handling of missing values

There were only a few (less than 1%) missing values in the climate variables – these were replaced by weekly averages surrounding that value. In a few places, rainfall was marked as "trace", which were replaced by 0.05 (since the lowest detectable rainfall was 0.1mm).

Validation of estimations

Yearly values of estimated numbers for diarrhoea have been matched with actual admission data from I.D. Hospital, Kolkata (where the surveillance is going on) and have been found satisfactory. We also collected actual monthly admission data (for randomly selected six months) from the I.D. Hospital (especially for cases coming from Kolkata area) to check whether they also match with the estimated monthly values, which was satisfactory too.

Data analysis:

Initially, we conducted univariate descriptive analysis of the collected data, followed by univariate and bivariate time series analysis. Results from these preliminary analyses were used for appropriate time series models and conducting further explanatory analysis.

DESCRIPTIVE ANALYSIS

This involved plotting of original (observed) data for each variable to visually check any fluctuations or increasing/decreasing trend over the selected ten-year period. We also tried to determine whether we had proportionately more (or less) days per year that were abnormally hot (or cool), more (or less) humid or more (or less) rainy.

UNIVARIATE TIME SERIES ANALYSIS

Univariate time series analysis consisted of three main parts.

1. *Running of sequence plots*

Initially, we constructed sequence plots for each time series. These plots helped us visually check two main aspects related to each series – stationarity and the model form (additive or multiplicative).

2. *Identification of different components of a time series*

Any time series usually consists of three different components – trend, seasonal, and irregular. This was done by *decomposing* the time series (using moving averages).

3. *Checking stationarity and lag period (for autocorrelation) for each series*

This was done using two approaches – plots of autocorrelation function (ACF) and partial autocorrelation function (PACF).

BIVARIATE TIME SERIES ANALYSIS

Scatterplots

Scatterplots were used to identify correlations between two time series variables on interval scale and to produce the best-fit curve. In our analyses, we used a LOWESS curve that possibly better demonstrated the relationships.

Cross-correlation

Cross correlation was used as a standard method of estimating the degree to which two series were correlated (how values in one series affected the values in the other) at different time lags. The cross correlation coefficients range between -1 and +1. We plotted cross-correlation function (CCF) of disease data (diarrhoea and cholera) with the climate and remote-sensing data. These plots indicated the time lags (in months) for which diarrhoea (or cholera) occurrences had strongest correlation with each predictor.

TIME SERIES MODELING

We used the ARIMA model to check whether and how the predictors in our dataset influence the outcomes (diarrhoea or cholera). The model assumptions were based upon the findings of the univariate and bivariate analysis as mentioned above.

The results obtained from these analyses are described in the following section.

Results

DESCRIPTIVE ANALYSIS

Table 1: Yearly distribution of max., mean, and min. values of variables

YEAR		TEMP_MAX	TEMP_MIN	RAINFALL	RH_MOR	RH_EVE	DIAR_MTH	CHOL_TOT	SST
1999	Max.	36.3	27.1	24.9	89.7	89.2	3701	704	30.5
	Mean	31.682	22.995	7.077	80.231	72.097	1857.62	284.60	27.463
	Min.	26.0	14.0	.0	69.2	48.8	775	0	24.8
2000	Max.	35.1	26.7	9.3	86.4	85.0	2411	448	29.7
	Mean	31.430	22.776	4.232	80.212	72.358	1491.81	145.97	27.828
	Min.	26.7	14.8	.0	69.1	56.8	620	0	24.7
2001	Max.	35.5	26.7	13.5	88.1	84.0	2294	553	30.2
	Mean	31.663	22.755	4.209	81.066	71.602	1427.15	187.90	27.584
	Min.	25.9	13.0	.0	73.5	53.9	844	0	23.3
2002	Max.	35.3	27.6	14.8	87.6	84.3	2062	452	30.7
	Mean	31.556	22.956	5.080	79.453	70.548	1465.11	99.19	27.575
	Min.	26.2	15.6	.0	71.2	51.4	698	0	24.6
2003	Max.	36.3	27.3	11.4	85.2	85.9	2601	220	29.9
	Mean	31.382	22.827	4.612	81.156	74.257	1433.86	106.56	27.788
	Min.	24.2	12.5	.0	76.5	59.1	568	22	24.9
2004	Max.	36.9	27.7	12.6	87.7	86.7	2983	1044	29.6
	Mean	31.561	22.805	4.534	79.182	71.434	1856.54	355.06	27.588
	Min.	24.7	14.5	.0	72.6	53.0	756	0	24.7
2005	Max.	35.9	28.0	14.5	86.9	85.2	2175	320	29.9
	Mean	31.670	22.958	5.165	80.058	70.878	1443.01	146.02	27.964
	Min.	25.3	15.2	.0	73.4	55.3	809	0	24.6
2006	Max.	35.3	27.1	19.0	86.5	84.3	2381	473	29.9
	Mean	32.069	23.073	5.268	78.275	67.588	1445.79	156.78	27.538
	Min.	26.9	14.0	.0	64.6	43.5	775	0	25.1
2007	Max.	35.5	27.2	23.1	86.0	82.5	2635	741	30.3
	Mean	31.372	22.791	6.440	79.987	69.673	1696.41	276.29	27.781
	Min.	26.1	14.0	.0	72.0	51.1	878	0	24.8
2008	Max.	36.3	28.8	18.6	86.9	85.7	2497	741	30.1
	Mean	31.400	22.985	5.243	79.918	70.247	1424.82	334.76	27.666
	Min.	26.1	14.6	.0	71.5	55.5	852	0	24.0
Total	Max.	36.9	28.8	24.9	89.7	89.2	3701	1044	30.7
	Mean	31.578	22.892	5.186	79.954	71.068	1554.21	209.31	27.677
	Min.	24.2	12.5	.0	64.6	43.5	568	0	23.3

The initial data description consisted of summarizing all the variables under study. The yearly maximum, minimum and mean values of each of the variables were computed for the ten years (1999–2008) under the study (Table 1). Apparently there were no significant changes in maximum, minimum temperatures, relative humidity (morning and evening) and sea surface temperature over the ten years. However, maximum rainfall showed some fluctuations. Maximum monthly diarrhoea cases showed a declining trend from 1999 to 2002, from 2003 the number started increasing, with a peak in 2004. From 2005–2008, there were fluctuations in the maximum number of diarrhoea cases. Total cholera cases declined continuously up to 2003, with a sudden rise in 2004. A decline in 2005 was followed by increasing cases again from 2006. From this data, it seems that there was an outbreak of diarrhoea and cholera in 2004.

We also plotted monthly averages of original (observed) data for each variable – namely, temperature (maximum and minimum), rainfall, relative humidity (morning and evening), SST, diarrhoea admissions (total and watery) and cholera cases. Since all results from analysis of watery diarrhoea resembled the results obtained from analysis of total diarrhoea admissions, the graphs and plots of watery diarrhoea cases are not separately presented separately in this report. All variables contained monthly values for 120 months (from January 1999 to December 2008).

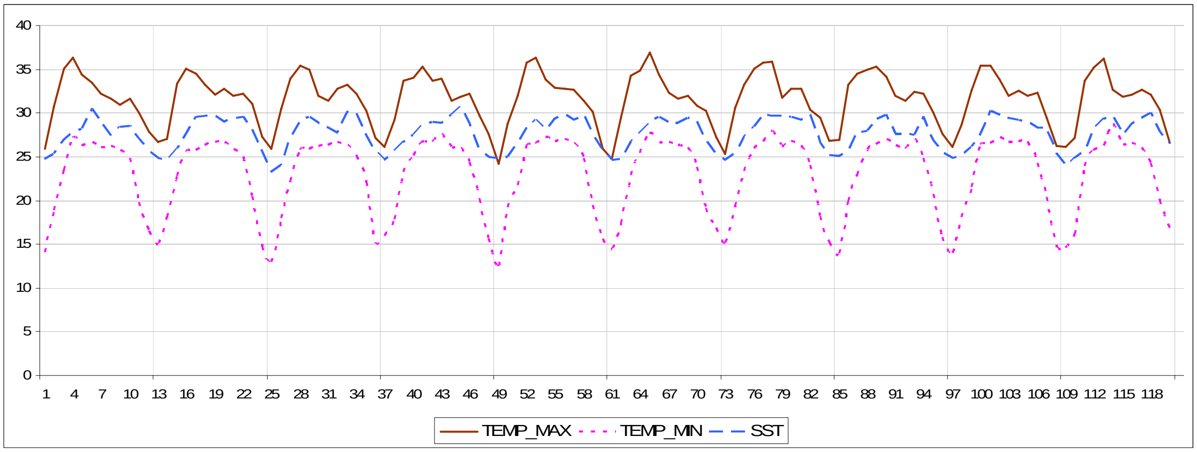


Fig. 1(a). **Monthly averages of temperatures (°C, max. & min.) and SST (°C) in Kolkata, 1999–2008**

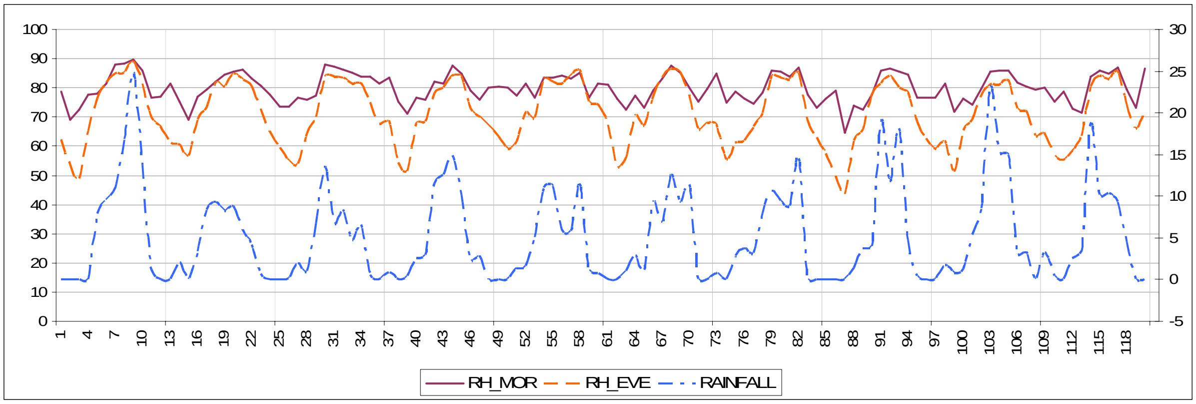


Fig. 1(b). Monthly averages of relative humidity (%, morning & evening) and rainfall (mm, plotted on secondary Y-axis) in Kolkata, 1999–2008

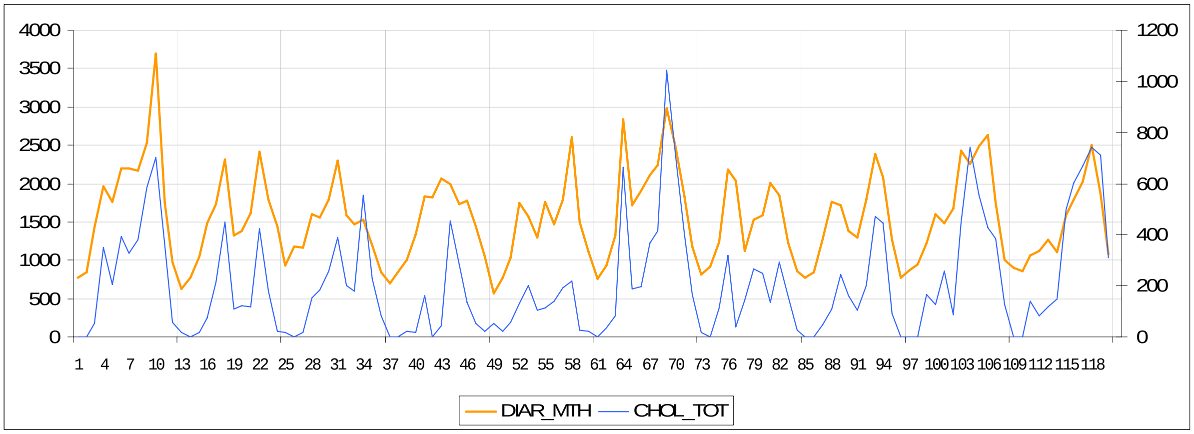


Fig. 1(c). Monthly averages of number of diarrhoea and cholera cases (plotted on secondary Y-axis) in Kolkata, 1999–2008

From Fig. 1 (a–c), it appeared that there were not substantial changes in maximum, minimum and the sea surface temperatures over the period. However, relative humidity seemed in a decreasing trend, whereas rainfall showed a fluctuating pattern. Since the above plots showed only average values for each variable in each month, they did not give us any idea of whether there had been extreme values and if there was any increase or decrease in occurrence of such extreme values over the specified time period. For example, it would be interesting to know whether the number of hotter days or days without rains increased over the years. To get a grasp of these trends, we calculated the mean and standard deviation of the daily values for the whole 10-year period for each of these variables and plotted yearly distributions along with trend lines for values exceeding one SD of the mean values (Fig. 2 a–d).

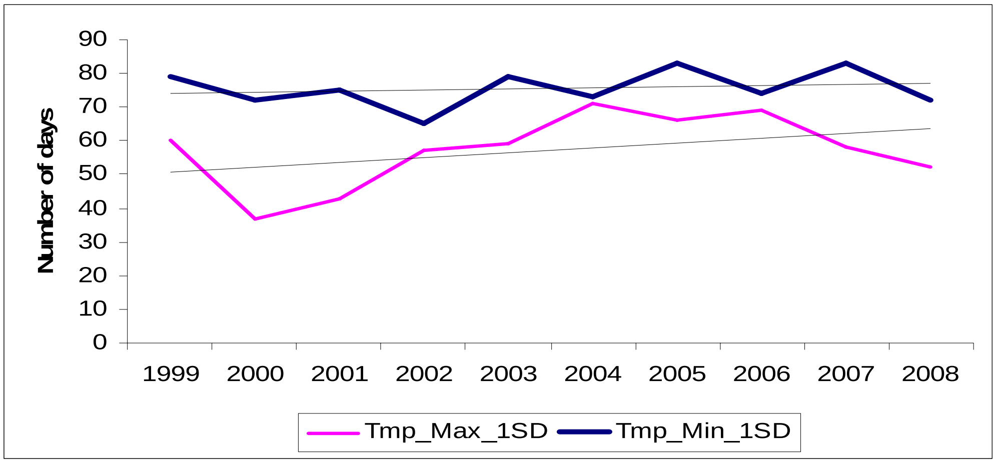


Fig. 2(a). **Yearly distribution of no. of days when maximum temperature remained 1 SD above 10-year mean and minimum temperature remained 1 SD below 10-year mean, Kolkata, 1999–2008**

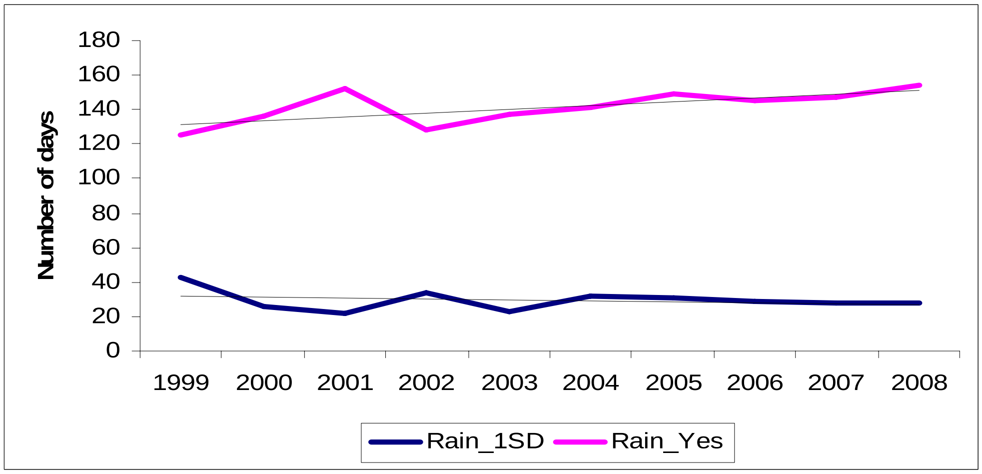


Fig. 2(b). **Yearly distribution of no. of days when rainfall was 1 SD above 10-year mean and no. of rainy days was 1 SD above 10-year mean, Kolkata, 1999–2008**

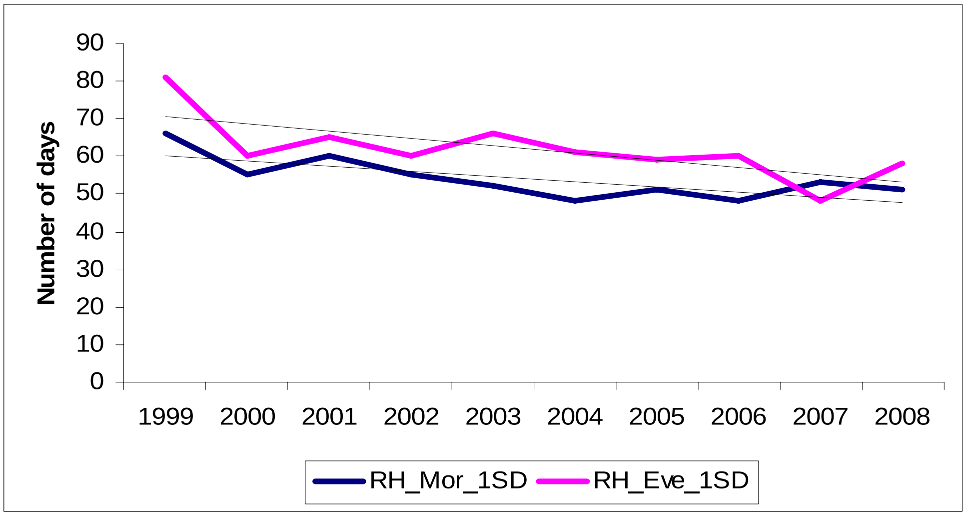


Fig. 2(c). **Yearly distribution of no. of days when relative humidity (morning and evening) remained 1 SD above 10-year mean, Kolkata, 1999–2008**

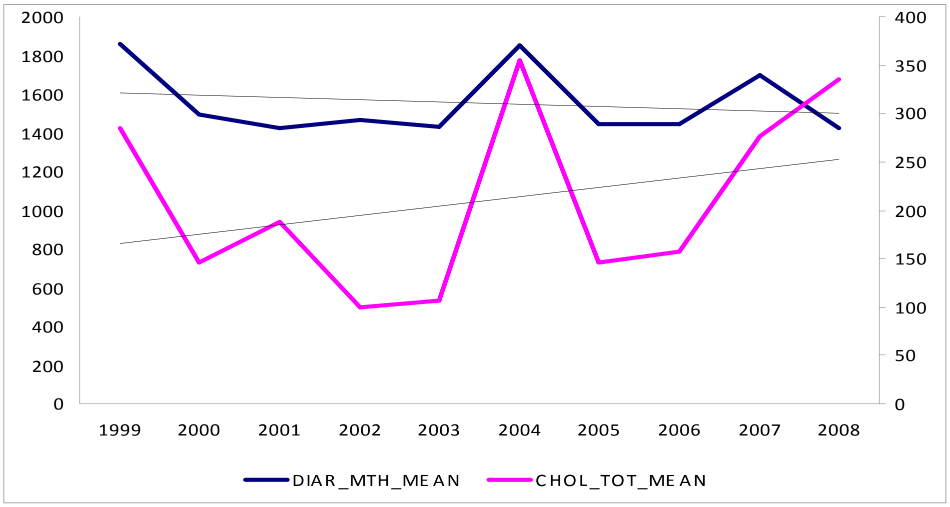


Fig. 2(d). **Yearly distribution of mean number of diarrhoea and cholera (plotted on secondary Y-axis) cases, Kolkata, 1999–2008**

Figures 2 (a–d) demonstrated that over the ten-year period under study, there had been an increase in the number of hotter days and cooler nights; although the number of rainy days increased during that period, the number of days with higher rainfall decreased. As we could guess from these findings, the number of more humid days also showed a decreasing trend. Interestingly, although the occurrence of diarrhoea decreased over the period, the number of cholera cases actually had an increasing trend.

Thus, these descriptive plots of our data provided us with some idea about the nature of changes in the study variables over the specified 10-year time period. However, since these data actually consisted of time series of the respective variables, the characteristics of these variables and their changes over time would be more clearly elaborated by doing time series analysis of these variables, as described in the following sections.

UNIVARIATE TIME SERIES ANALYSIS

The first step in the time series analysis was constructing run sequence plots for each variable over time. From these plots, it seemed that an additive model would be appropriate for the observed time series of individual variables, since the magnitude of the seasonal fluctuation did not vary much over time.

$$Y_t = S_t + T_t + E_t$$

The decomposition for the additive model was performed using the following steps:

a) Determine the duration of one season. One season was considered as 12 months in this study.

b) Compute the centered moving average of the series, with a moving average window width of one season. All seasonal variability is eliminated from the moving average series; therefore, subtracting the moving average from the observed series isolates the seasonal component plus irregular component.

c) Assume the seasonal component is constant from season to season. The seasonal component is then computed from the isolated series as the average of each point in one season. The resulting values represent the average seasonal component of the series.

d) The original series was modified by subtracting the seasonal component. The resulting series was the seasonally adjusted series (i.e. the one from which the seasonal component had been removed).

e) The combined trend and cyclical component was approximated by applying to the seasonally adjusted series.

f) Finally, the irregular component was determined by subtracting the trend-cycle component from the seasonally adjusted series.

The sequence plots along with the trend of respective variables are presented below (Fig. 3 a–h).

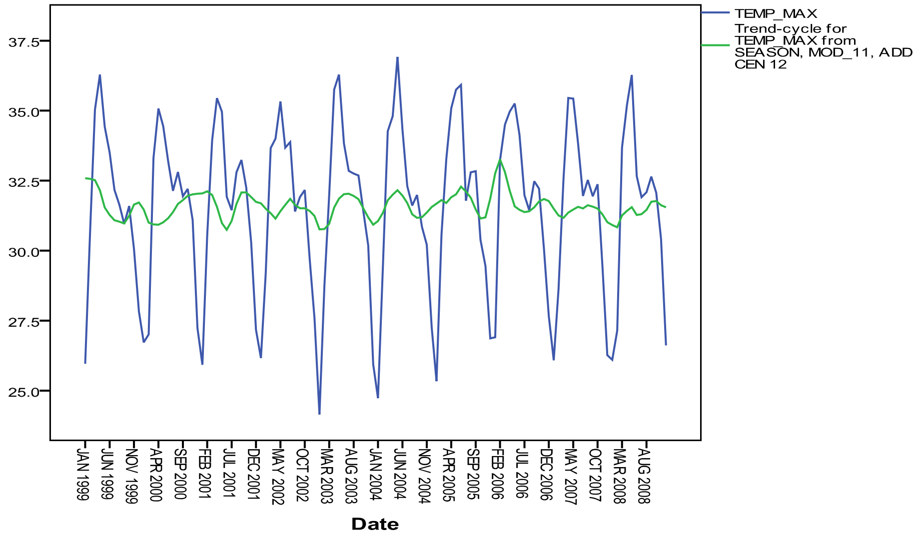


Fig. 3(a). Sequence plot and trend of maximum temperature, Kolkata, 1999–2008

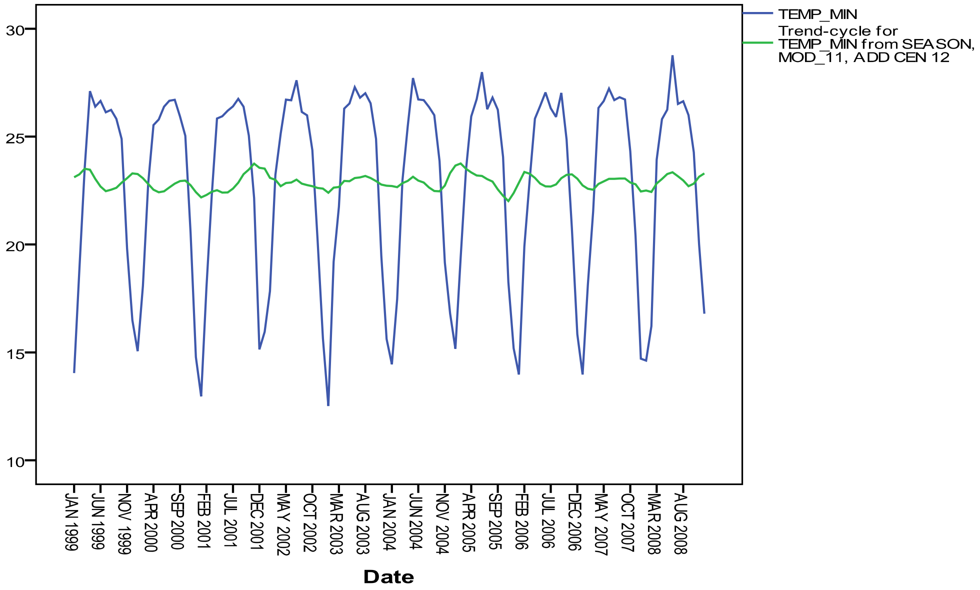


Fig. 3(b). Sequence plot and trend of minimum temperature, Kolkata, 1999–2008

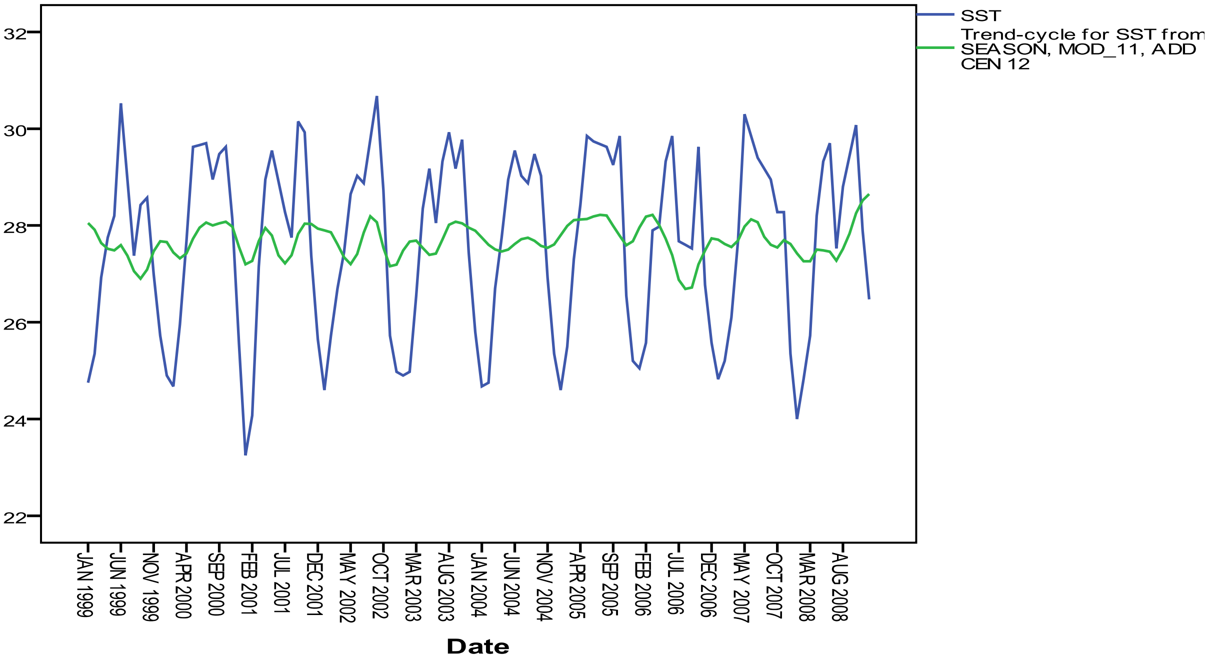


Fig. 3(c). Sequence plot and trend of SST, Kolkata, 1999–2008

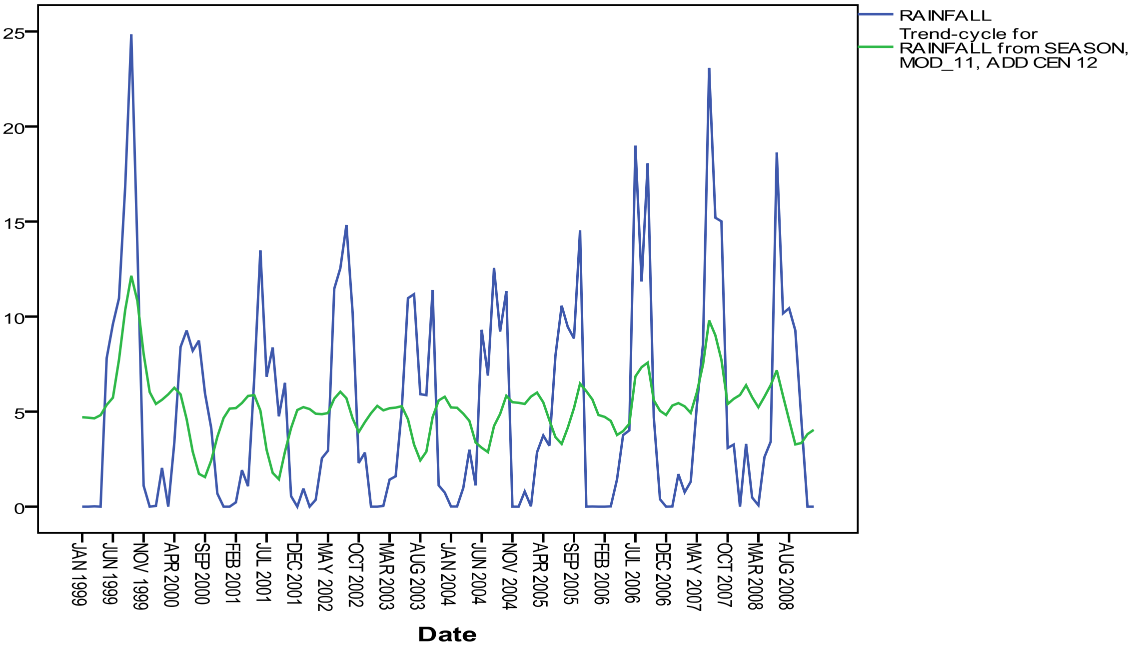


Fig. 3(d). Sequence plot and trend of rainfall, Kolkata, 1999–2008

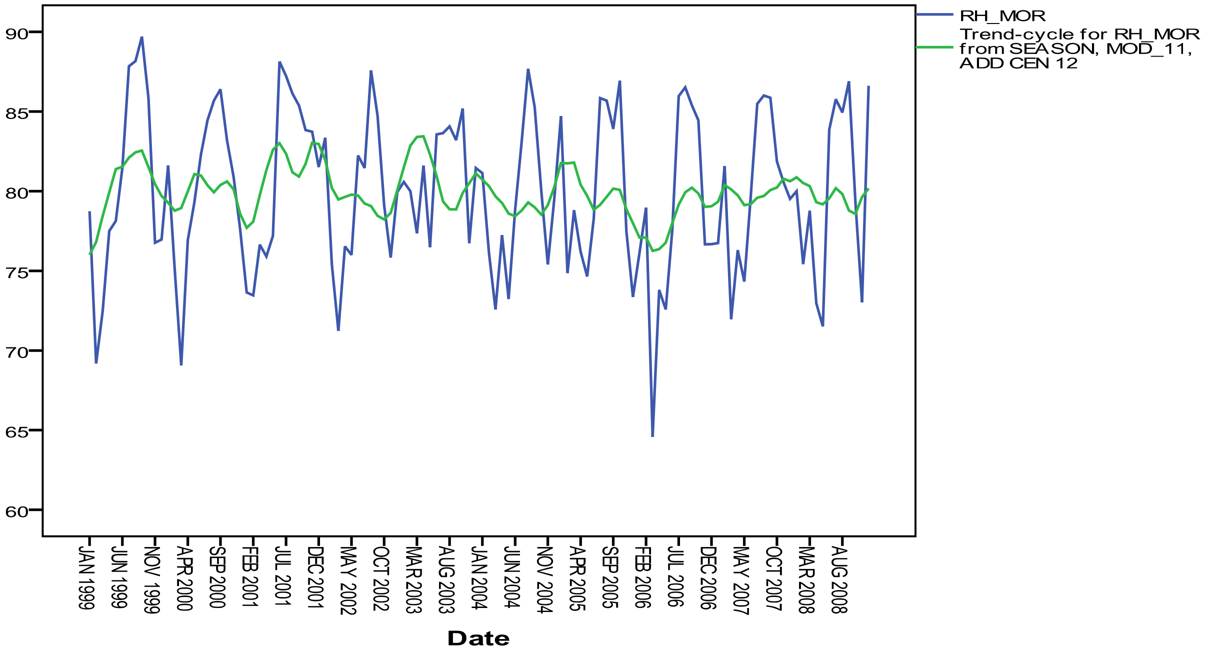


Fig. 3(e). Sequence plot and trend of rel. humidity (morning), Kolkata, 1999–2008

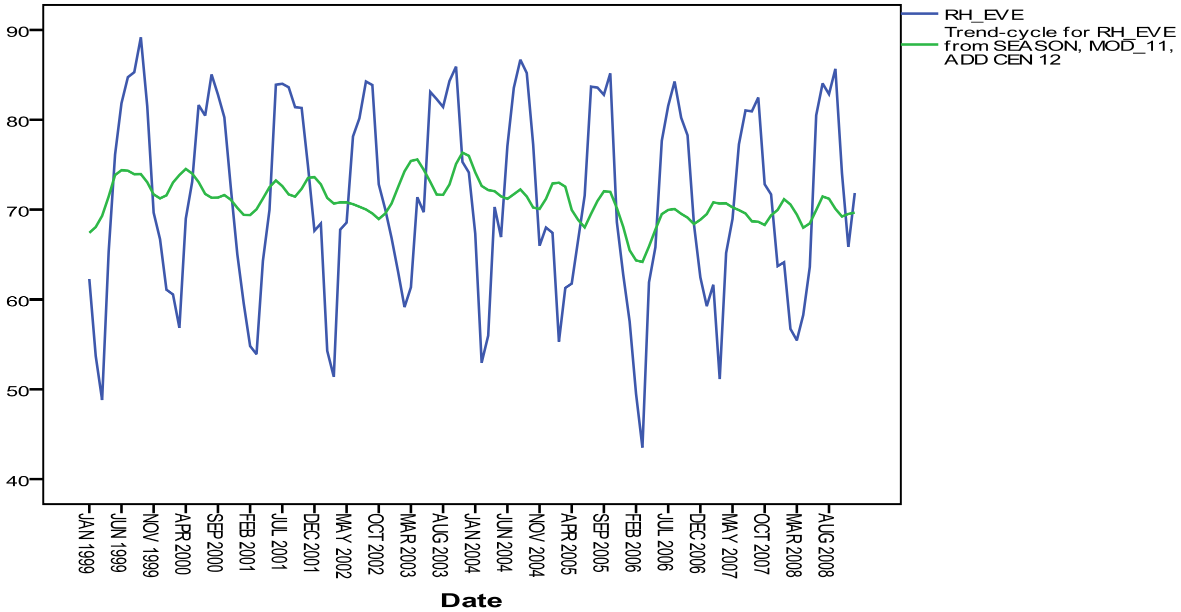


Fig. 3(f). Sequence plot and trend of rel. humidity (evening), Kolkata, 1999–2008

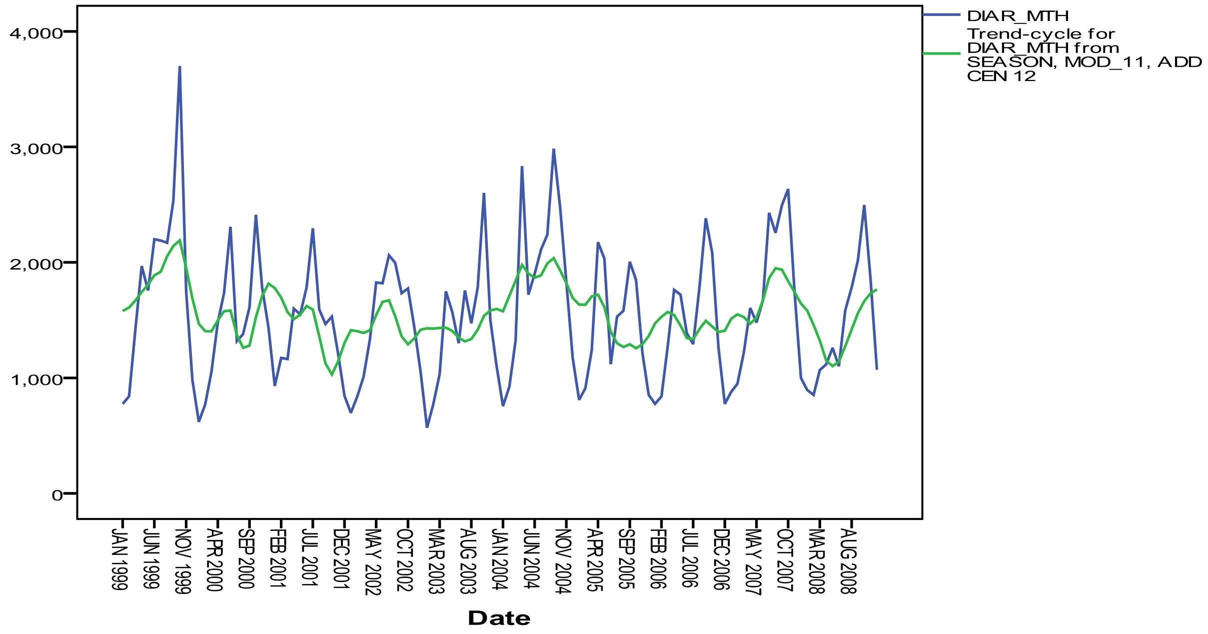


Fig. 3(g). Sequence plot and trend of diarrhoea cases, Kolkata, 1999–2008

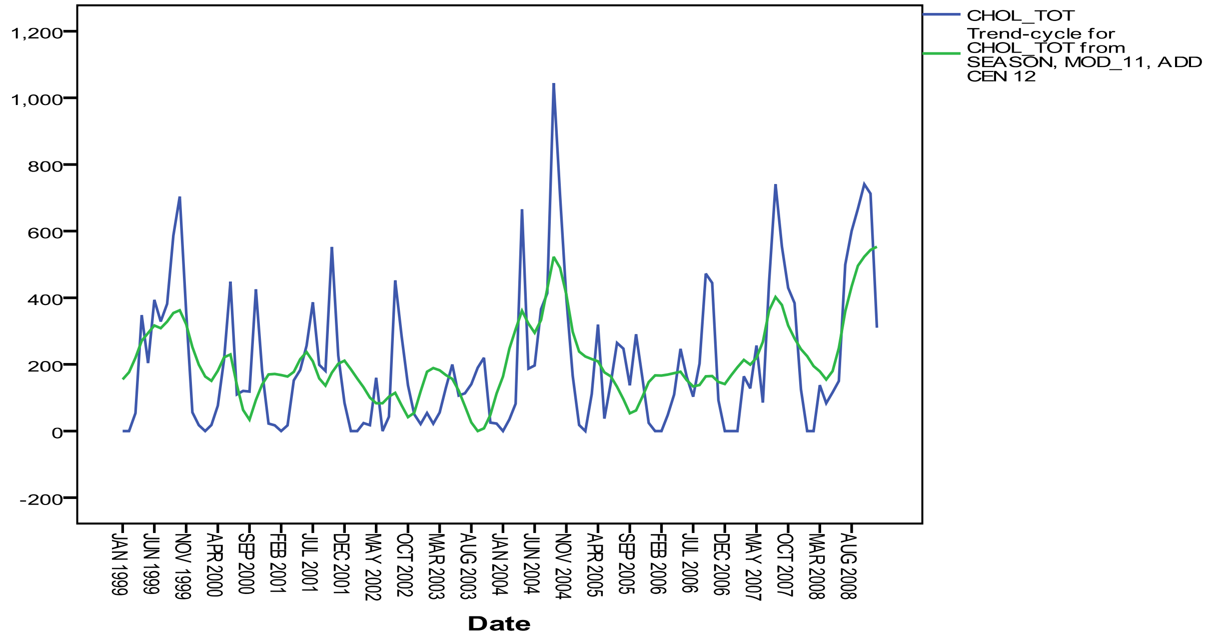


Fig. 3(h). Sequence plot and trend of cholera cases, Kolkata, 1999–2008

The above plots (Fig. 3 a–h) revealed that while the maximum and minimum temperatures in Kolkata remained almost at the similar level over the 10-year period, the sea surface temperature had a slightly increasing trend. The rainfall and relative humidities (morning and evening), however, showed a decreasing trend. The occurrence of diarrhoea and cholera showed opposite trends – while diarrhoea cases decreased over time, the number of cholera cases increased in recent years.

The decomposition of each time series also provided seasonal factors for each variable. The seasonal patterns identified from our data are presented below in two ways – (a) overall seasonal pattern as identified from monthly distribution of seasonal components of whole data, and (b) comparison of seasonal patterns within each year for the 10-year period under study.

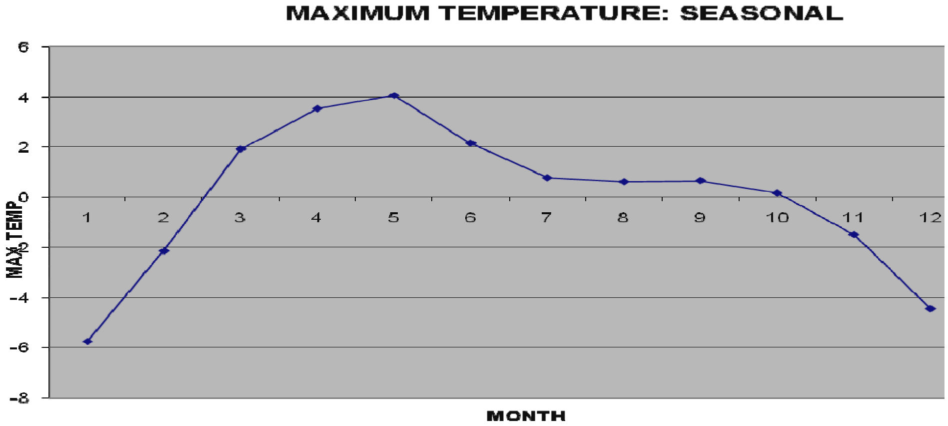


Fig. 4(a). Overall seasonal pattern of max. temperature, Kolkata, 1999–2008

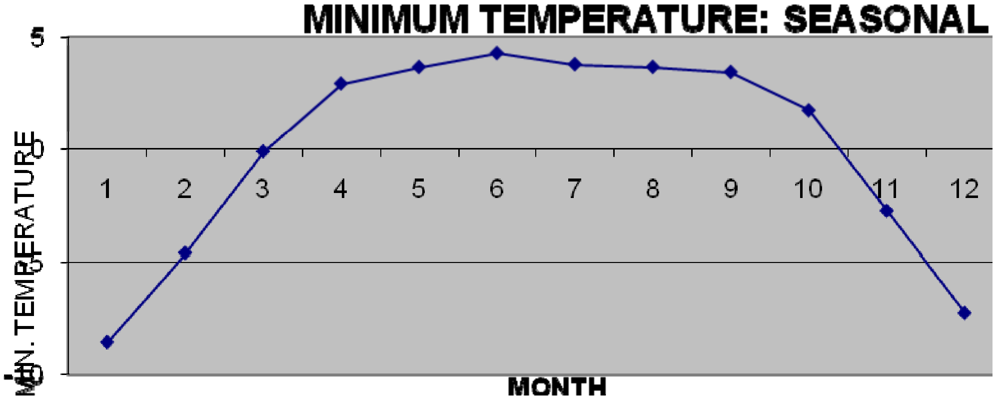


Fig. 4(b). Overall seasonal pattern of min. temperature, Kolkata, 1999–2008

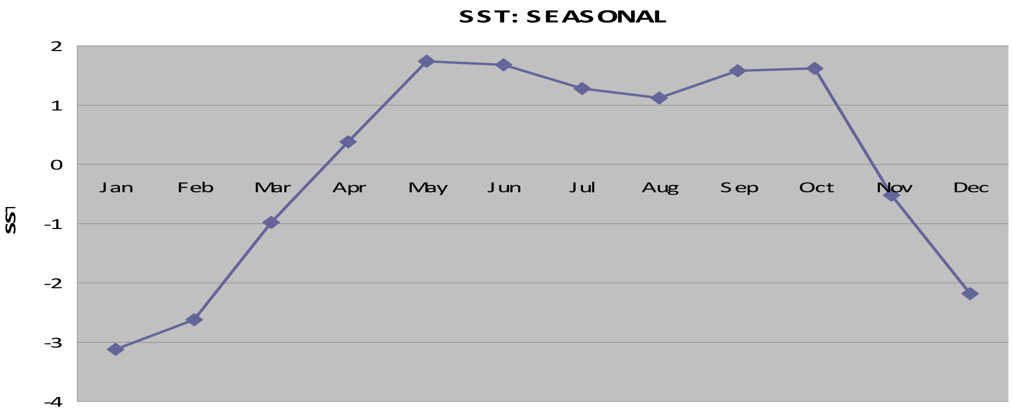


Fig. 4(c). Overall seasonal pattern of SST, Kolkata, 1999–2008

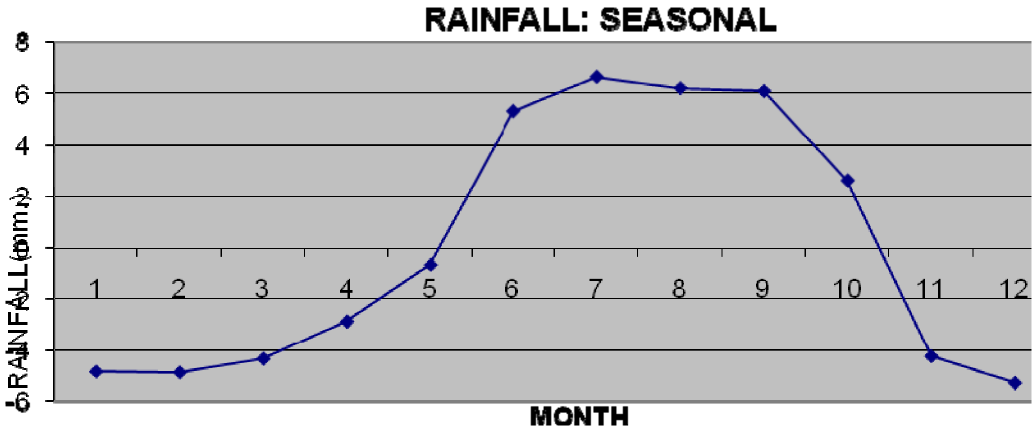


Fig. 4(d). Overall seasonal pattern of rainfall, Kolkata, 1999–2008

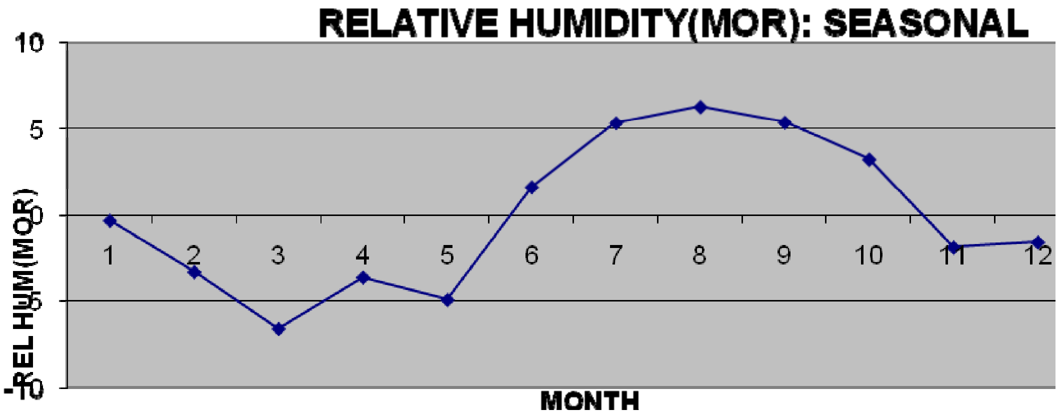


Fig. 4(e). Overall seasonal pattern of relative humidity (morning), Kolkata, 1999–2008

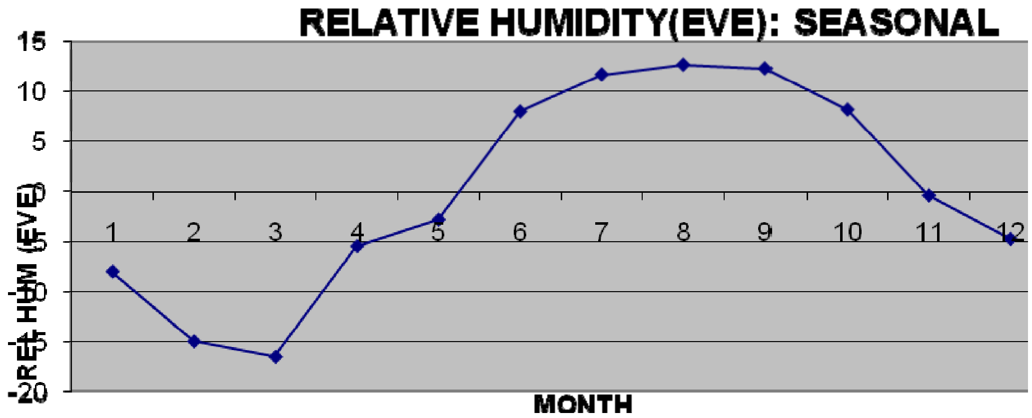


Fig. 4(f). Overall seasonal pattern of relative humidity (evening), Kolkata, 1999–2008

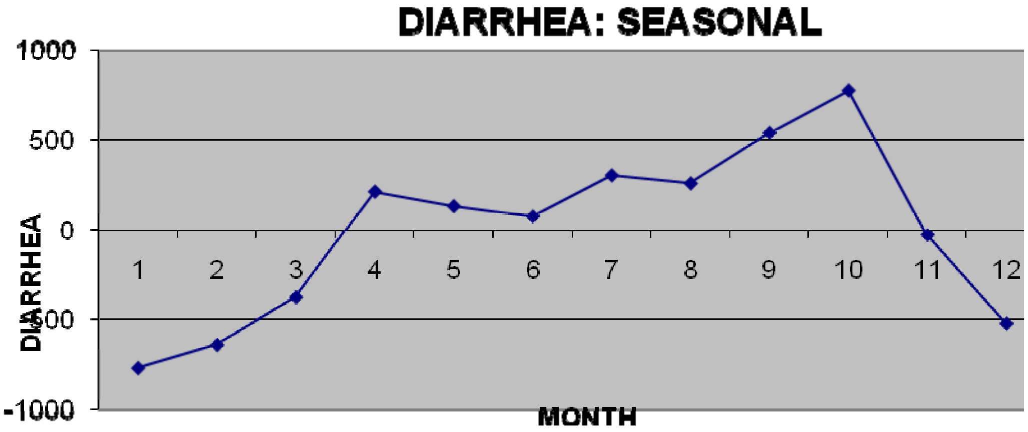


Fig. 4(g). Overall seasonal pattern of diarrhoea cases, Kolkata, 1999–2008

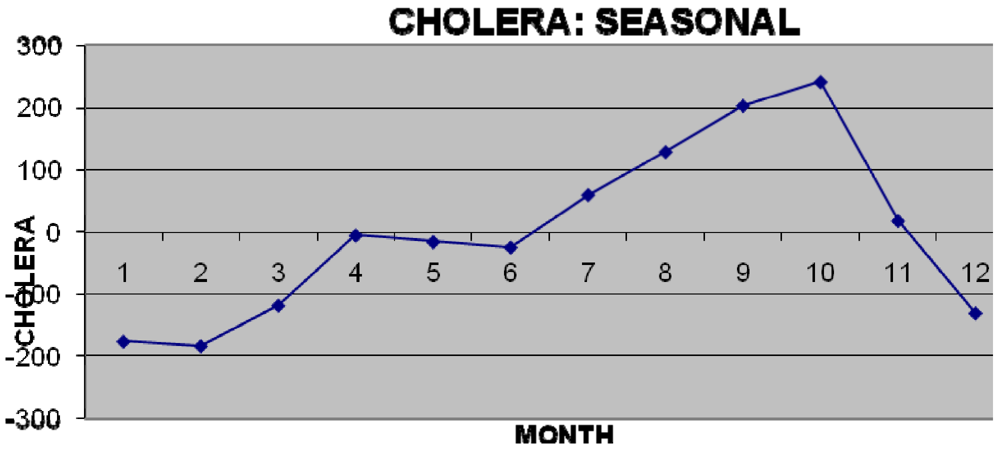


Fig. 4(h). Overall seasonal pattern of relative cholera cases, Kolkata, 1999–2008

From the 10-year period data, the overall seasonality demonstrated by each of the study variabes (as shown in Fig. 4 a–h) were summarized in the following table (Table 2).

Table 2. Overall seasonal patterns for the study variables during 1999–2008, Kolkata

Variable	Month(s) for		Other comments
	Highest value	Lowest value	
Maximum temperature	May	January	Overall, Mar–Jun had higher values
Minimum temperature	June	January	Overall, Apr–Sep had higher values
Sea surface temperature	May, October	January	Overall, May–Oct had higher values
Rainfall	July	December	Overall, Jun–Sep had higher values
Rel. humidity (morning)	August	March	Overall, Jul–Oct had higher values
Rel. humidity (evening)	August	March	Overall, Jun–Oct had higher values
Diarrhoea	October	January	Increasing trend after June
Cholera	October	February, Jan	Increasing trend after June

When seasonal patterns of different years (for the ten-year study period) were compared, none of the variables seemed to demonstrate a significant change over time, as shown in Fig. 5 a–c.

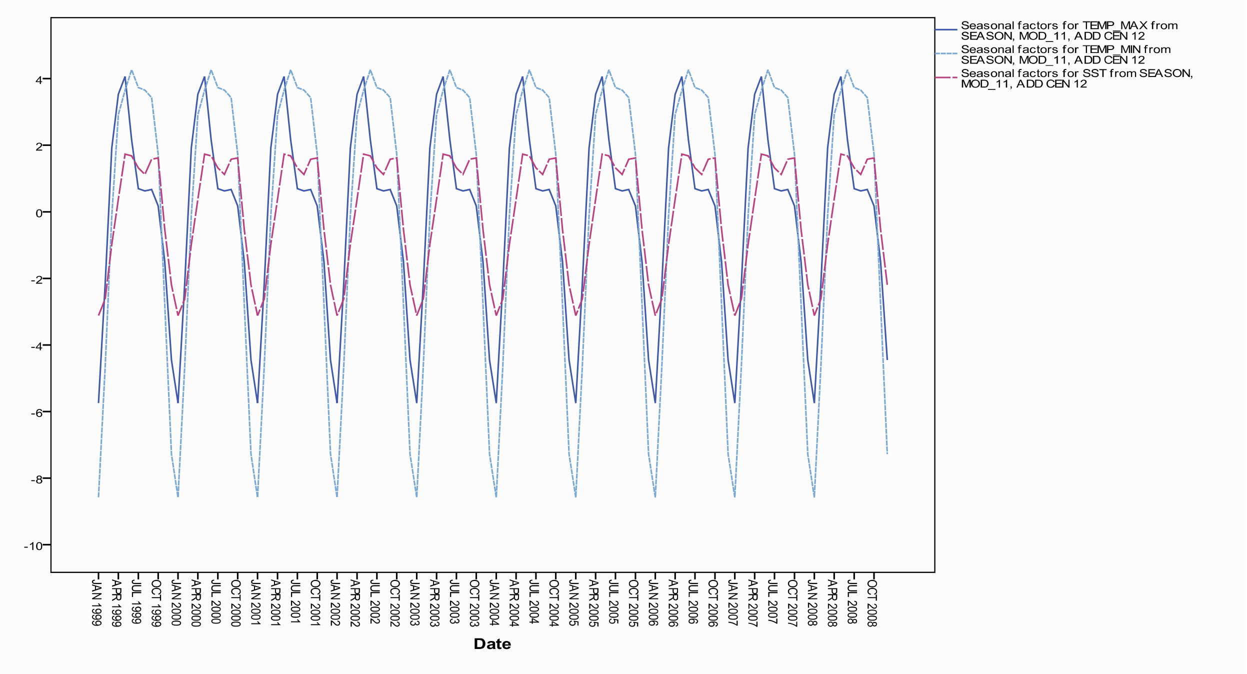


Fig. 5(a). **Comparison of seasonal patterns of maximum, minimum and sea surface temperatures, Kolkata, 1999–2008**

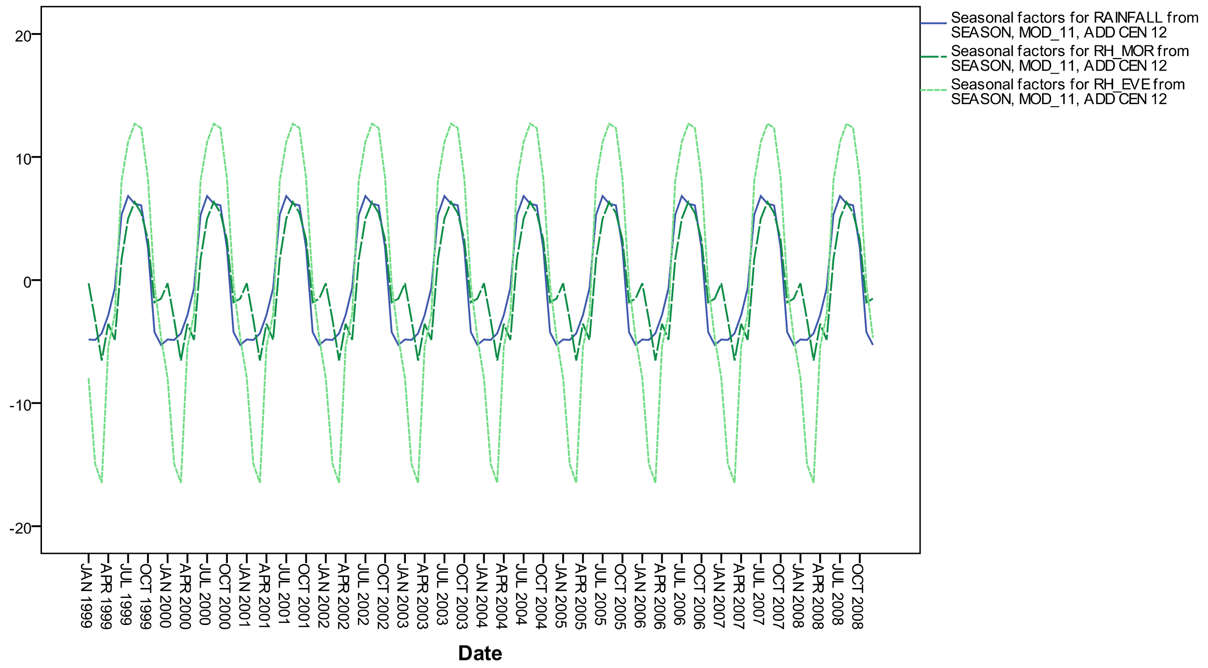


Fig. 5(b). **Comparison of seasonal patterns of rainfall and relative humidities (morning and evening), Kolkata, 1999–2008**

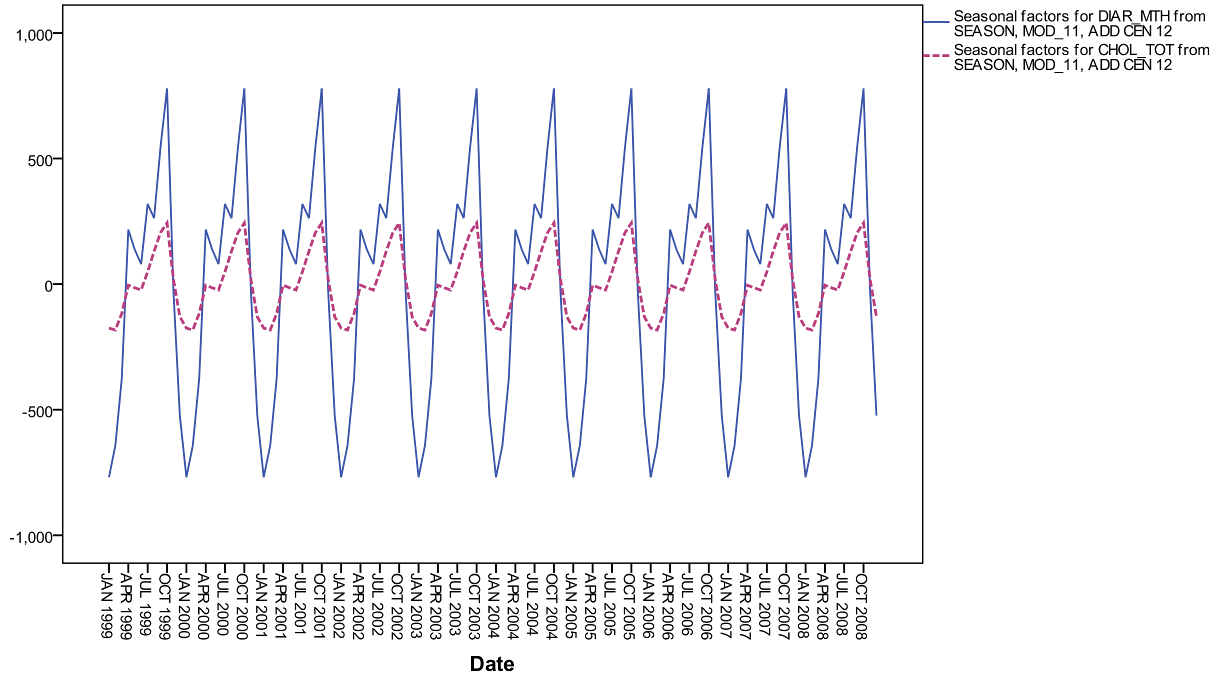


Fig. 5(c). **Comparison of seasonal patterns of occurrence of diarrhoea and cholera cases, Kolkata, 1999–2008**

2. Checking stationarity of the time series:

a) *Run sequence plots*

The initial run sequence plot of the data indicated that all the series seemed to have constant means when plotted, a visible sign of non-stationarity.

b) *Autocorrelation plots*

The autocorrelation plots (ACF plots) of estimated autocorrelations also died down slowly with increasing lag, indicating a second visual symptom of non-stationarity. Thus, *differencing* could be used to make the series stationary.

c) *Partial autocorrelation plots*

These plots demonstrated correlations at different lags, after adjusting for all correlations within the lag window. In the given plots (PACF plots) for all the variables under this study, the lag for strongest correlation for each time series had been demonstrated.

The autocorrelation and partial autocorrelation plots of each variable are shown in Fig. 6 (a–h).

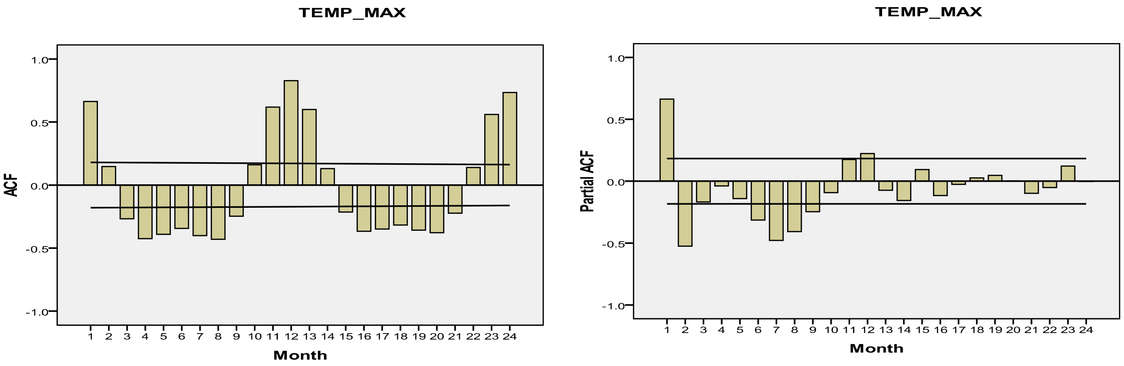


Fig. 6(a). **Autocorrelation and partial autocorrelation plots of maximum temperature, Kolkata, 1999–2008**

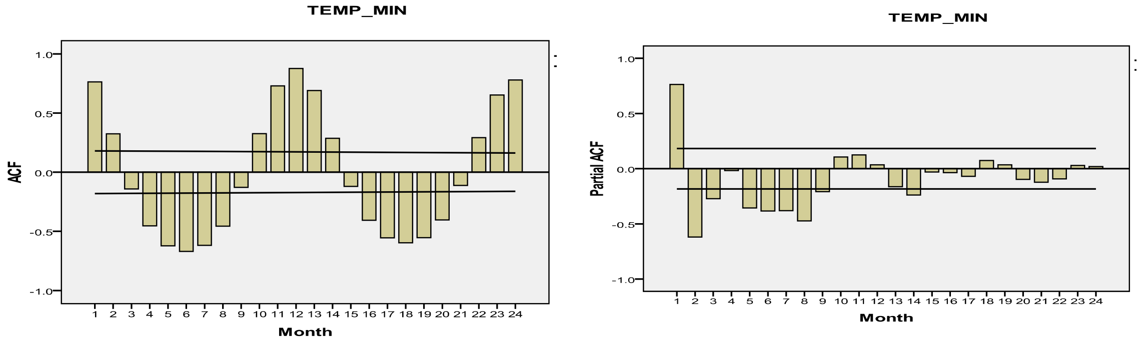


Fig. 6(b). Autocorrelation and partial autocorrelation plots of minimum temperature, Kolkata, 1999–2008

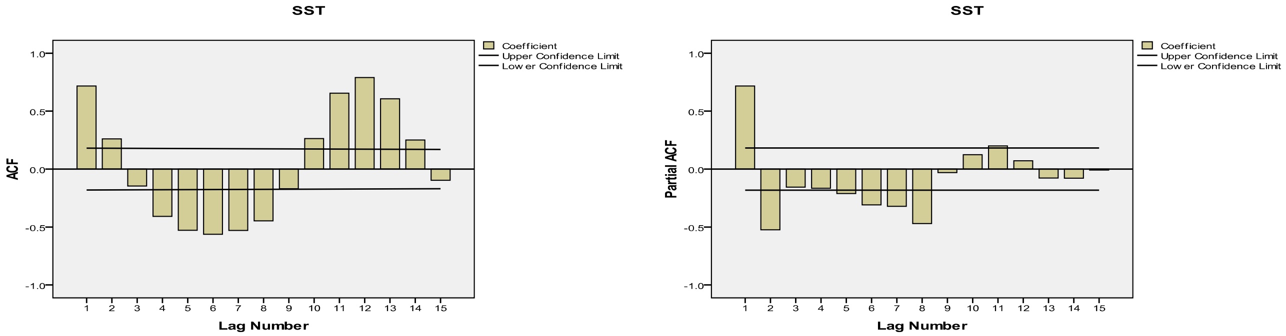


Fig. 6(c). Autocorrelation and partial autocorrelation plots of sea surface temperature, Kolkata, 1999–2008

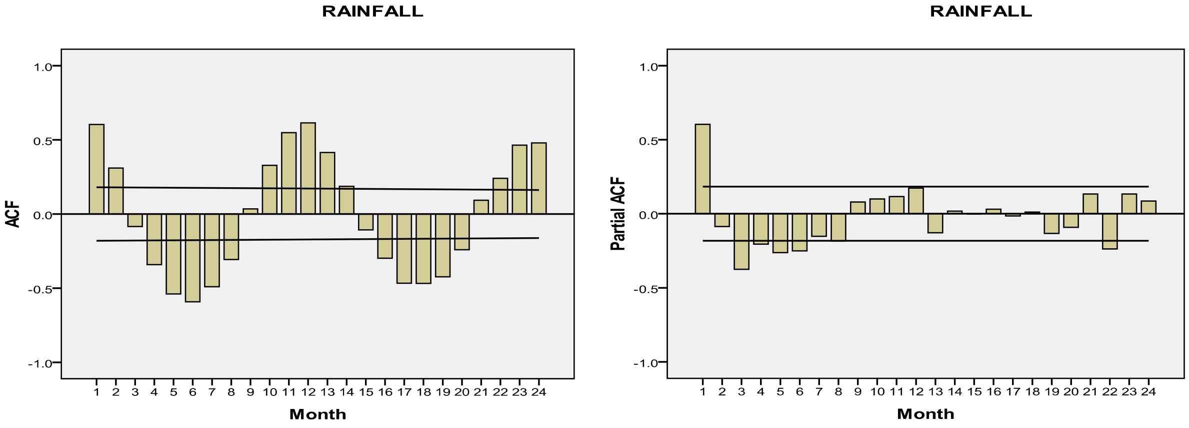


Fig. 6(d). Autocorrelation and partial autocorrelation plots of rainfall, Kolkata, 1999–2008

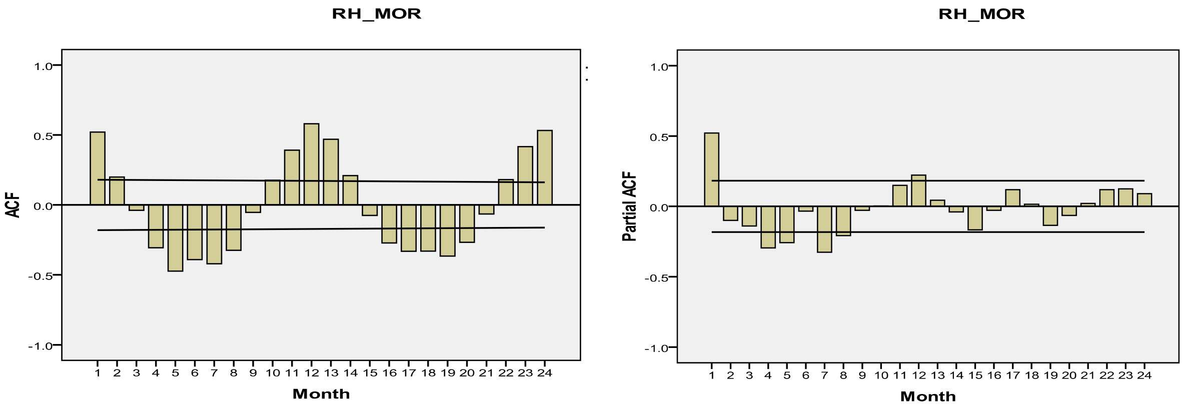


Fig. 6(e). **Autocorrelation and partial autocorrelation plots of relative humidity (morning), Kolkata, 1999–2008**

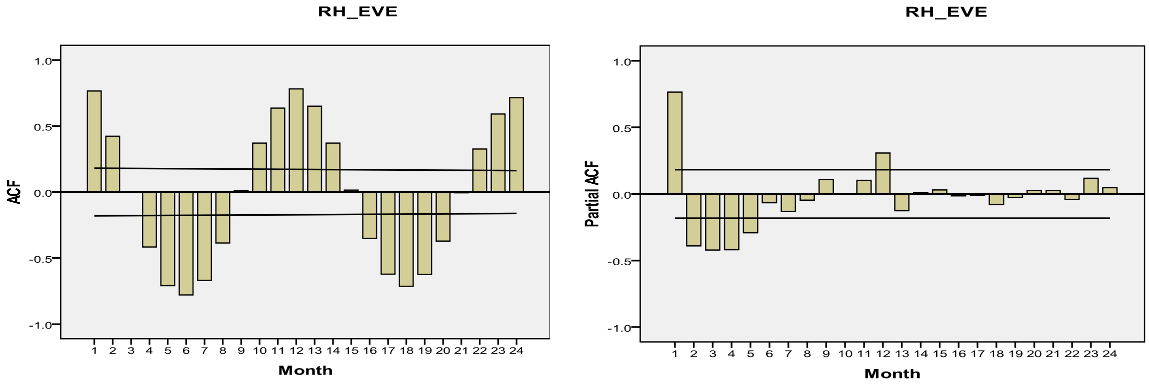


Fig. 6(f). **Autocorrelation and partial autocorrelation plots of relative humidity (evening), Kolkata, 1999–2008**

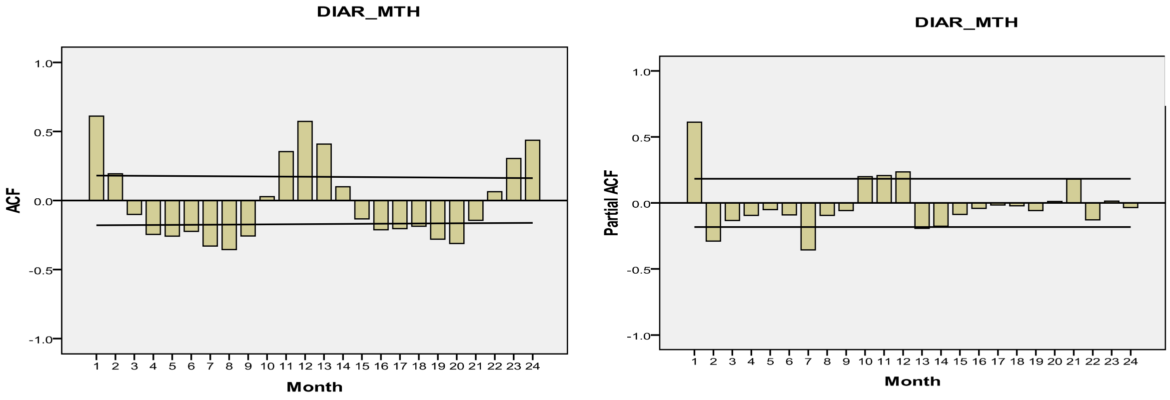


Fig. 6(g). **Autocorrelation and partial autocorrelation plots of diarrhoea, Kolkata, 1999–2008**

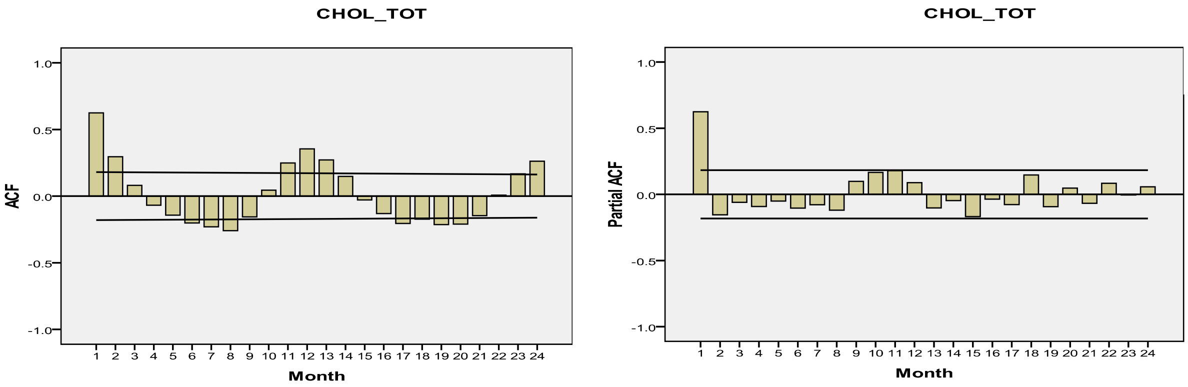


Fig. 6(h). **Autocorrelation and partial autocorrelation plots of cholera, Kolkata, 1999–2008**

After having the results of the univariate time series analyses, as described above, the next step was to carry out bivariate analysis to check relationships between the outcomes (diarrhoea and cholera) with each of the predictors. The results are shown below.

BIVARIATE TIME SERIES ANALYSIS

Scatterplots

Scatterplots were used to identify correlations between two time series variables on an interval scale and to produce the best-fit curve for these correlations. Here, we used LOWESS (locally weighted scatterplot smoothing) curve that could better demonstrate the relationships.

Cross-correlation

Cross-correlation was used as a standard method of estimating the degree to which two series were correlated (that is, how values in one series affected the values in the other) at different time lags. The cross correlation coefficients ranged between -1 and 1. The bound indicated maximum correlation and 0 indicated no correlation. A high negative correlation indicated a high correlation but of the inverse of one of the series.

Figures 7 (a–f) in the following pages showed the scatterplots and the cross-correlations.

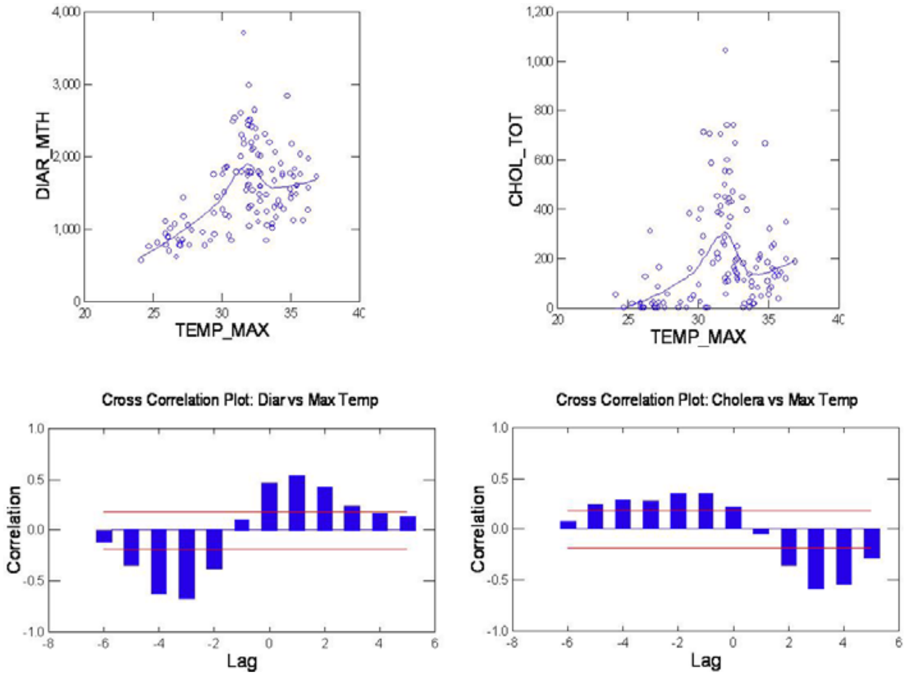


Fig. 7(a). **Scatterplots with LOWESS curve and cross-correlation plots: Maximum temperature vs. diarrhoea and cholera**

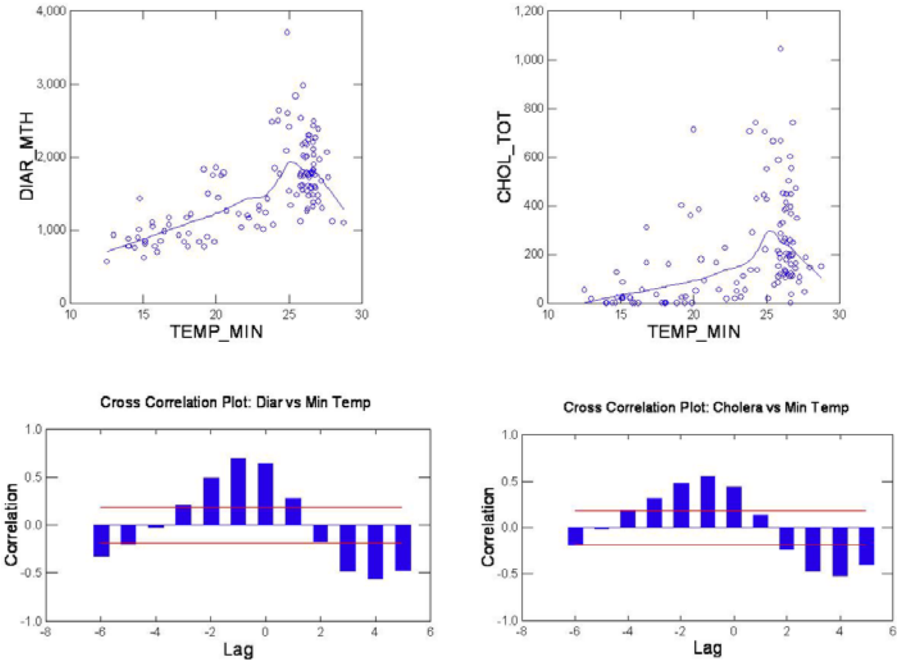


Fig. 7(b). **Scatterplots with LOWESS curve and cross-correlation plots: Minimum temperature vs. diarrhoea and cholera**

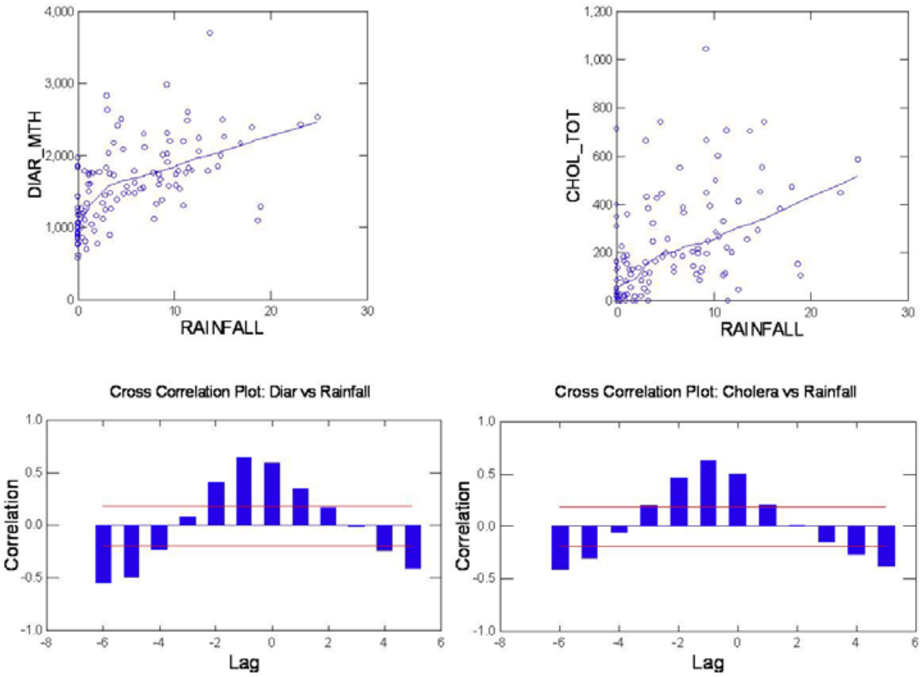


Fig. 7(c). **Scatterplots with LOWESS curve and cross-correlation plots: Rainfall vs. diarrhoea and cholera**

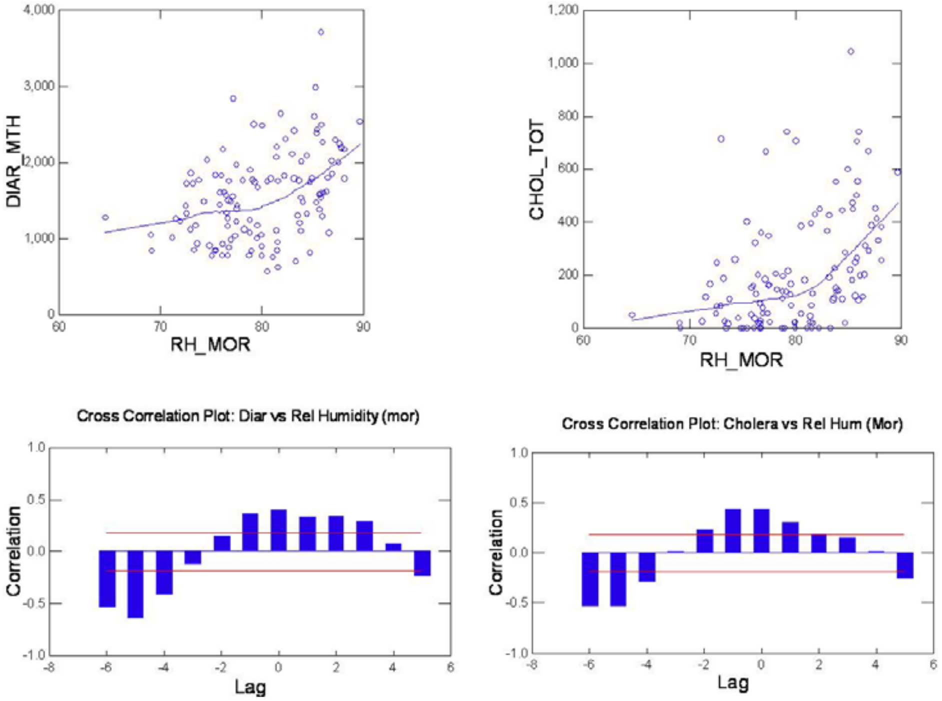


Fig. 7(d). **Scatterplots with LOWESS curve and cross-correlation plots: Relative humidity (morning) vs. diarrhoea and cholera**

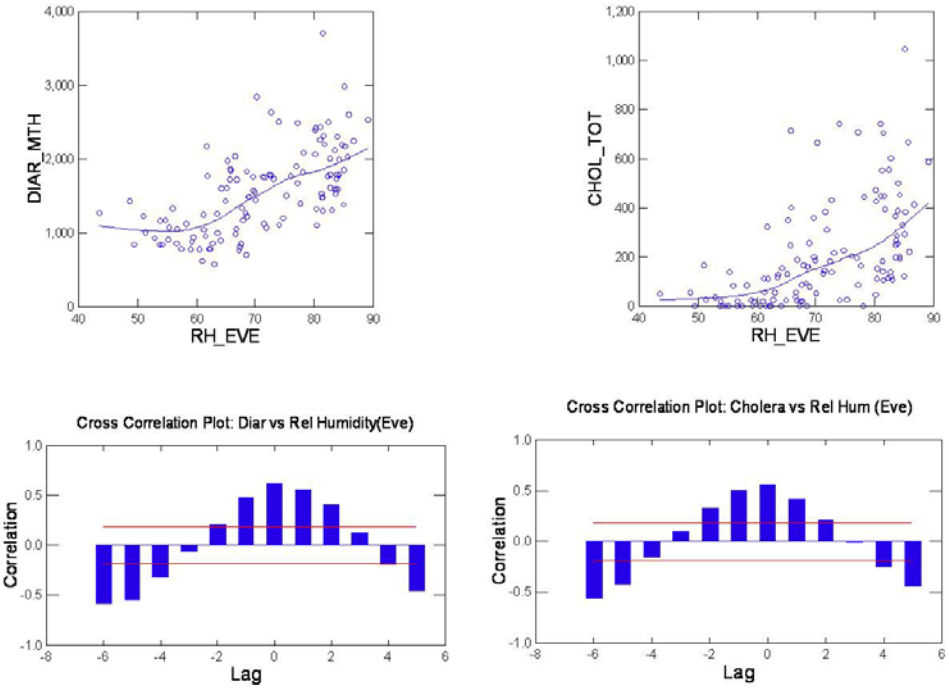


Fig. 7(e). **Scatterplots with LOWESS curve and cross-correlation plots: Relative humidity (evening) vs. diarrhoea and cholera**

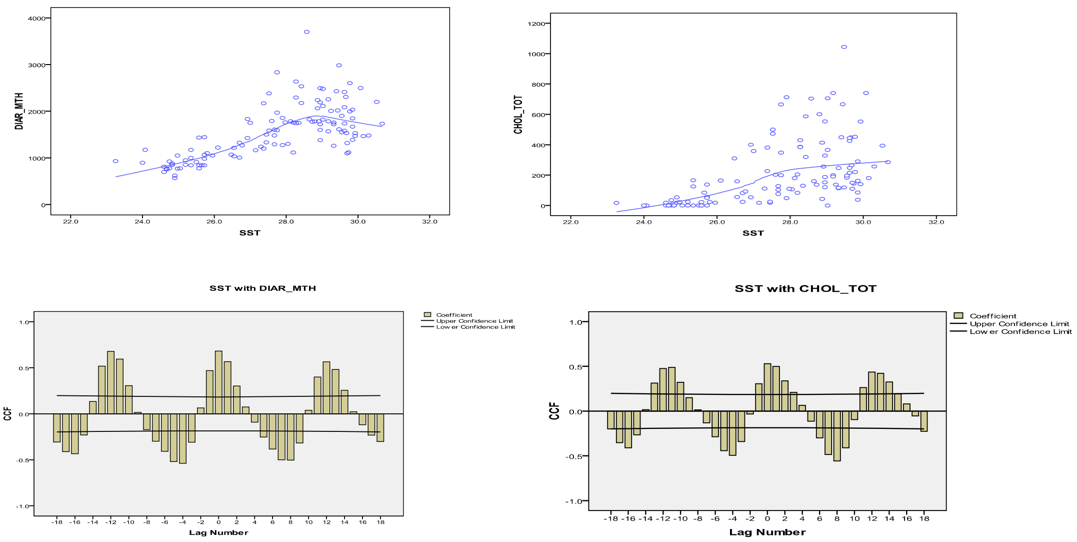


Fig. 7(f). **Scatterplots with LOWESS curve and cross-correlation plots: SST vs. diarrhoea and cholera**

The summary of the bivariate analysis findings are tabulated in Table 3.

Table 3: Climate factors: The lag for strongest correlations with diarrhoea and cholera

Climate factor	Diarrhoea	Cholera
Temp_Max	-3 +1	-3 +1, +2
Temp_Min	-1	-1
Rainfall	-1	-1
RH_Mor	0	0, -1
RH_Eve	0	0
SST	0	0

Thus, all the time series as depicted above seemed to be non-stationary and followed an additive model for decomposed components. They also displayed seasonality – that is, periodic fluctuations in the data within any given year. The autocorrelation and partial autocorrelation plots demonstrated the lag times for which the series values had strongest correlations with other values in the same series. These correlation patterns helped develop appropriate prediction models for these series, as illustrated below.

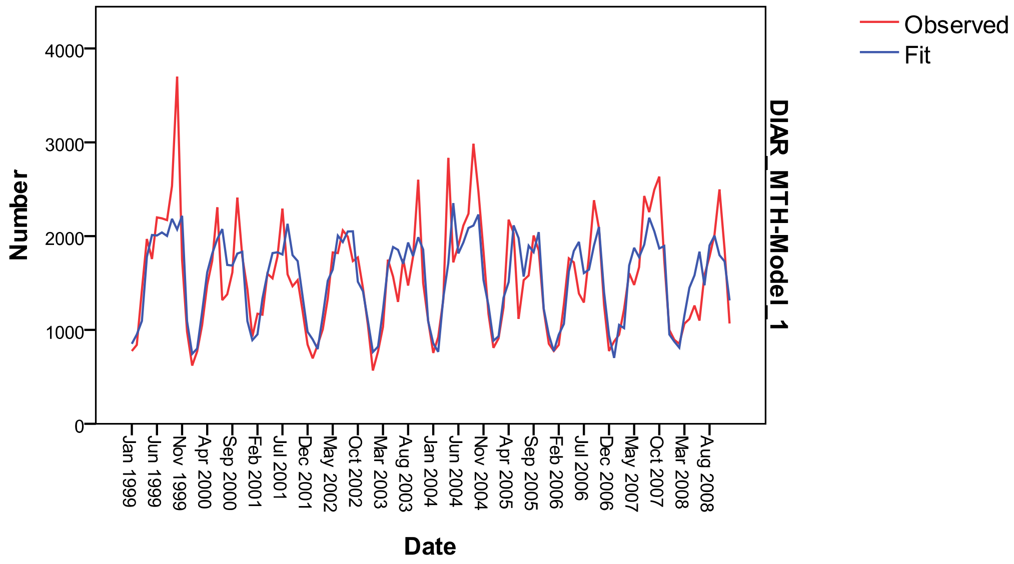
TIME SERIES MODELING

The explanatory analyses consisted of time series modeling using ARIMA models, where diarrhoea and cholera were treated as outcomes and the other variables (maximum and minimum temperatures, SST, rainfall and morning & evening relative humidities) were incorporated as the predictors. The assumptions were based on the findings from the univariate and bivariate time series analyses. The ACF plots indicated that probably an additive AR Model would be appropriate. We tested several assumptions in the models and checked model fitness in terms of fitness statistics (R^2 and seasonal R^2, as well as the Ljung-Box Q), correlogram of residuals and plots of observed data along with the model fit data. It appeared that for both outcomes (diarrhoea and cholera) the best fit models were ARIMA (2, 0, 0) (0, 0, 0). The model parameters and the plots of observed vs. fit data are presented below.

(a) Model for diarrhoea:

ARIMA Model Parameters

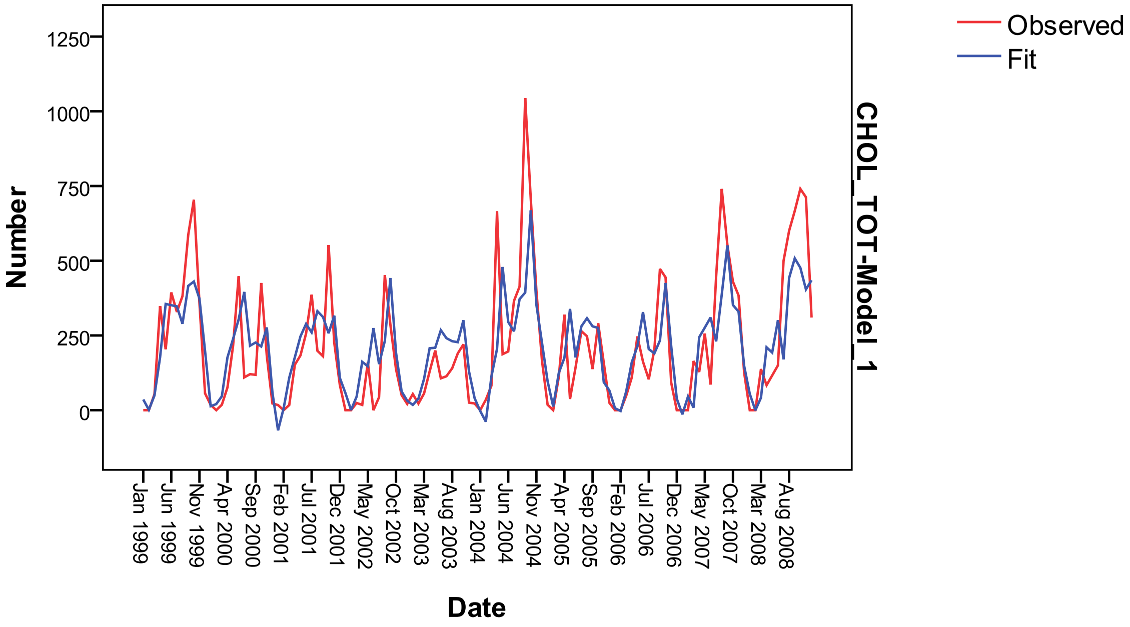
			Estimate	SE	t	Sig.
DIAR_MTH	Constant		-2922.441	1783.996	-1.638	.104
	AR	Lag 1	.481	.099	4.875	.000
		Lag 2	-.111	.099	-1.116	.267
TEMP_MAX	Numerator	Lag 0	49.985	54.715	.914	.363
TEMP_MIN	Numerator	Lag 0	.074	46.242	.002	.999
RAINFALL	Numerator	Lag 0	7.411	11.129	.666	.507
RH_MOR	Numerator	Lag 0	-6.576	14.247	-.462	.645
RH_EVE	Numerator	Lag 0	23.944	10.633	2.252	.026
SST	Numerator	Lag 0	60.732	44.217	1.374	.172



(b) Model for cholera:

ARIMA Model Parameters

			Estimate	SE	t	Sig.
CHOL_TOT	Constant		-1254.819	710.995	-1.765	.080
	AR	Lag 1	.534	.097	5.486	.000
		Lag 2	-.006	.104	-.060	.952
TEMP_MAX	Numerator	Lag 0	1.021	21.801	.047	.963
TEMP_MIN	Numerator	Lag 0	.326	18.268	.018	.986
RAINFALL	Numerator	Lag 0	1.878	4.615	.407	.685
RH_MOR	Numerator	Lag 0	.056	5.658	.010	.992
RH_EVE	Numerator	Lag 0	5.215	4.264	1.223	.224
SST	Numerator	Lag 0	37.601	17.738	2.120	.036



From the models illustrated above, it appeared that for both outcomes (diarrhoea and cholera), lag-1 autocorrelation was highly significantly predicted the outcome values. Additionally, the evening-time relative humidity and the SST significantly predicted values of diarrhoea and cholera, respectively.

Conclusions

This study involving retrospective 10-year hospital-based data on diarrhoea and cholera, as well as data on meteorological and remotely-sensed data (SST) highlighted some characteristics about climatic variability during this period and its possible relationship with occurrences of diarrhoea and cholera in Kolkata. The data indicated that over the past decade in Kolkata, there had been an increase in the number of hotter days, along with cooler nights – possibly an indication of extremes of temperatures in coming years. Simultaneously, there was a drop in relative humidity (morning and evening), resulting in less daily rainfall despite an increase in number of rainy days. Although the number of diarrhoea cases showed a decreasing pattern, the number of cholera cases increased in recent times. Moreover, both diarrhoea and cholera cases surged to an abnormally higher levels during 2004 – indicating occurrence of an outbreak at that time in Kolkata. The explanatory analysis using ARIMA modeling demonstrated that the previous month's number of diarrhoea and cholera cases were the best predictors for occurrences of diarrhoea and cholera in the following month; moreover, the relative humidity in the evening and the sea surface temperature also significantly predicted occurrences of diarrhoea and cholera respectively. However, these results should be interpreted very cautiously. No definitive positive/negative "trend" was noticed for any of the variables (climate or disease), except for "cholera", which showed an upward trend over last 10 years (probably indicating influence of factors other than those under consideration so far). Since these 10 years' data did not reflect any significant change over time, this time period (10 years) probably was not sufficient to assess changes in disease patterns due to climate change; at best these data could only indicate how disease occurrence changes with "season" (within each year).

Nevertheless, at least two important aspects were highlighted through these analyses. First, to prevent occurrence of cholera in this area of Gangetic West Bengal, sufficient attention should be given to possible factors other than climate variations – most notably breaches in sanitation system and water supply. In fact, Kolkata is one of the few cities in the world where pipelines for drinking water supply and drainage of waste water often run side-by-side, providing opportunities for mixing up of the two systems in case of leakages, which is not uncommon. Strengthening of these two systems, along with reconstruction of separate systems wherever possible, would prevent occurrences of diarrhoea and cholera at least to some extent. Second, changes in rainfall and temperature over the 10-year study period apparently give an indication of the potentially disastrous effects of climate change in Kolkata in the coming years. Perhaps this would be the best time to start preparing for adaptive measures (including legal steps and inter-departmental coordination) to prevent further detrimental changes in local climate and thereby preventing its ill-effects on human health.

Recommendations

There were indications that some relationships between the outcome variables and the predictor variables might exist. However, the exact nature of the relationships was not fully comprehensible or explained due to some important limitations of this retrospective study. For example, the chosen 10-year time period appeared too short to capture changes in climatic conditions in Kolkata; we could only observe effects of climate variability and seasonal changes on the disease data. Moreover, in this study we were especially interested in assessing the influence of selective non-climate factors in the climate-diarrhoea relationship. But the retrospective nature of the study did not allow us to have any control over measurement of any of the variables. Thus, although there were data available for many of the non-climate variables that could potentially act as confounders in the climate-diarrhoea (or cholera) relationships, they turned out to be unusable because they were not uniformly measured on the same time-scale or units or geographic locations. Data on some of the other variables were simply unavailable – they could not be even extrapolated for other available information. Most of these data gaps and inconsistencies could be overcome through a prospective study design, having the required data collected for sufficient time period. Hence, considering the importance of the subject matter, we strongly recommend conducting well-designed prospective studies that may give us better control over data collection and enhance the opportunity of examining the putative relationships between climate changes and occurrences of diarrhoea and cholera.

References

1. Economics and Development Resource Center. Basic statistics. Manila: Asian Development Bank; 2008. Available from: www.adb.org/statistics/pdf/basic-statistics-2008.pdf [accessed 28 May 2009].

2. Bhattacharya SK. Acute diarrhoeal disease in children. *Clin Immunol Immunopathol* 2003; 110: 1175- 81

3. World Health organization. World Health Report 2002. *Reducing Risk, Promoting Healthy Life*. WHO, Geneva 2002; 192

4. Colwell RR. Viable but non culturable bacteria: a survival strategy. *J Infect Chemother* 2000;6(2): 121-5

5. Bouma MJ, Pascual M. Seasonal and inter annual cycles of endemic cholera in Bengal 1891–1940 in relation to climate and geography. *Hydrobiologia* 2001; 460: 147-56

6. Alam M, Hasan NA, Sadique A, Bhuiyan NA, Ahmed KU, Nusrin S, *et al.* Seasonal cholera caused by Vibrio cholerae serogroups O1 and O139 in the coastal aquatic environment of Bangladesh. *Appl Environ Microbiol* 2006; 72(6): 4096-104

7. Heidelberg JF, Heidelberg KB, Colwell RR. Seasonality of Chesapeake Bay bacterioplankton species. *Appl Environ Microbiol* 2002; 68(11): 5488-97

8. Koelle K, Rodo X, Pascual M, Yunus M, Mostafa G. Refractory periods and climate forcing in cholera dynamics. *Nature* 2005; 436 [doi:10.1038/nature03820]

9. Rodo X, Pascual M, Fuchs G, Faruque ASG. ENSO and cholera: A nonstationary link related to climate change? *Proc Natl Acad Sci* 2002; 99(20): 12901-6

10. Checkley W, Epstein L, Gilman R, Figueroa D, Cama R, Patz J. Effect of El-Nino and ambient temperature on hospital admissions for diarrhoeal diseases in Peruvian children. *Lancet* 2000; 355: 442-50

11. Hazizume M, Armstrong B, Hajat S, Wagatsuma Y, Faruque ASG *et al*., Association between climate variability and hospital visits of non-cholera diarrhoea in Bangladesh: effects and vulnerable groups. *Int JEpidemiol*. [advance access published July 30, 2007]

12. Singh RBK, Hales S, de Wet N, Raj R, Hearnden M, Weinstein P. The Influence of Climate Variation and Change on Diarrhoeal Disease in the Pacific Islands. *Environ Health Perspect* 2001; 109(2):155-8

13. Zodpey SP, Deshpande SG, Ughade SN, Kulkarni SW, Shrikhande SN, Hinge AV. A prediction model for moderate or severe dehydration in children with diarrhoea. *J Diarrhoeal Dis Res* 1999; 17(1): 10-6

14. Rao S, Joshi SB, Kelkar RS. Changes in nutritional status and morbidity over time among pre-school children from slums in Pune, India. *Indian Pediatr* 2000; 37(10): 1060-71

15. Shah SM, Yousafzai M, Lakhani NB, Chotani RA, Nowshad G. Prevalence and correlates of diarrhoea. *Indian J Pediatr* 2003; 70(3): 207-11.

16. Ray SK, Roy P, Deysarkari S, Lahiri A, Mukhopadhaya BB. A cross sectional study of undernutrition in 0-5 yrs. age group in an urban community. *Indian J Matern Child Health* 1990; 1(2): 61-2

17. Bhattacharya SK. Progress in the prevention and control of diarrhoeal diseases since Independence. *Natl Med J India* 2003; 16(Suppl 2): 15-9

18. Fewtrell L, Kaufmann RB, Kay D, Enanoria W, Haller L, Colford JM Jr. Water, sanitation, and hygiene interventions to reduce diarrhoea in less developed countries: a systematic review and meta-analysis. *Lancet Infect Dis* 2005; 5(1): 42-52

19. Huttly SR, Morris SS, Pisani V. Prevention of diarrhoea in young children in developing countries. *Bull World Health Organ* 1997; 75(2): 163-74

20. Eisenberg JN, Scott JC, Porco T. Integrating disease control strategies: balancing water sanitation and hygiene interventions to reduce diarrhoeal disease burden. *Am J Public Health* 2007; 97(5): 846-52

21. Kumar Karn S, Harada H. Field survey on water supply, sanitation and associated health impacts in urban poor communities--a case from Mumbai City, India. *Water Sci Technol* 2002; 46(11-12): 269-75

22. 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(11): 1175-81

23. Mahendraker AG, Dutta PK, Urmil AC, Moorthy TS. A study of medico social profile of under five children suffering from diarrhoeal diseases. *Indian J Matern Child Health* 1991; 2(4): 127-30