



# Climate and socioeconomic influences on interannual variability of cholera in Nigeria

Gregor C. Leckebusch\*, Auwal F. Abdussalam

School of Geography, Earth and Environmental Sciences, University of Birmingham, UK



## ARTICLE INFO

### Article history:

Received 9 November 2014

Received in revised form

20 April 2015

Accepted 26 April 2015

Available online 18 May 2015

### Keywords:

Cholera

Climate variability

Infectious diseases

Health

Statistical modelling

## ABSTRACT

Cholera is one of the most important climate sensitive diseases in Nigeria that pose a threat to public health because of its fatality and endemic nature. This study aims to investigate the influences of meteorological and socioeconomic factors on the spatiotemporal variability of cholera morbidity and mortality in Nigeria. Stepwise multiple regression and generalised additive models were fitted for individual states as well as for three groups of the states based on annual precipitation. Different meteorological variables were analysed, taking into account socioeconomic factors that are potentially enhancing vulnerability (e.g. absolute poverty, adult literacy, access to pipe borne water). Results quantify the influence of both climate and socioeconomic variables in explaining the spatial and temporal variability of the disease incidence and mortality. Regional importance of different factors is revealed, which will allow further insight into the disease dynamics. Additionally, cross validated models suggest a strong possibility of disease prediction, which will help authorities to put effective control measures in place which depend on prevention, and or efficient response.

© 2015 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## 1. Introduction

The link between climate and specific infectious diseases has been observed for over a century. One of these meteorologically-sensitive infectious diseases that remain a major health burden in Nigeria for several decades is cholera. The influence of climate on cholera dynamic has been well documented in literature, for example, in Asia (e.g., Bouma and Pascual, 2001; Pascual et al., 2000), South America (e.g., Colwell, 1996; Speelman et al., 2000), and in Africa (e.g., De Magny et al., 2012; Fernandez et al., 2009; Traerup, 2010). On seasonal time scales cholera, as other diseases in Nigeria (e.g., Olouko et al., 2014), shows a seasonality with increased burdens from May to August. Additionally, social risk factors are playing an important role in transmission and outbreak of cholera, for example, the disease has been termed the 'disease of poverty' (Charles and Ryan, 2011; Snowden, 2008) and was associated with inadequate environmental sanitation conditions and untreated drinking water (e.g., Ali et al., 2002a, 2002b; Hashizume et al., 2007; Penrose et al., 2010; Rajendran et al., 2011; Reiner et al., 2012; Talavera and Perez, 2009). Previous studies have demonstrated the possibility of predicting cholera epidemics (e.g., Reyburn et al., 2011), however, the importance of combining the effects of social risk factors in addition to meteorological conditions

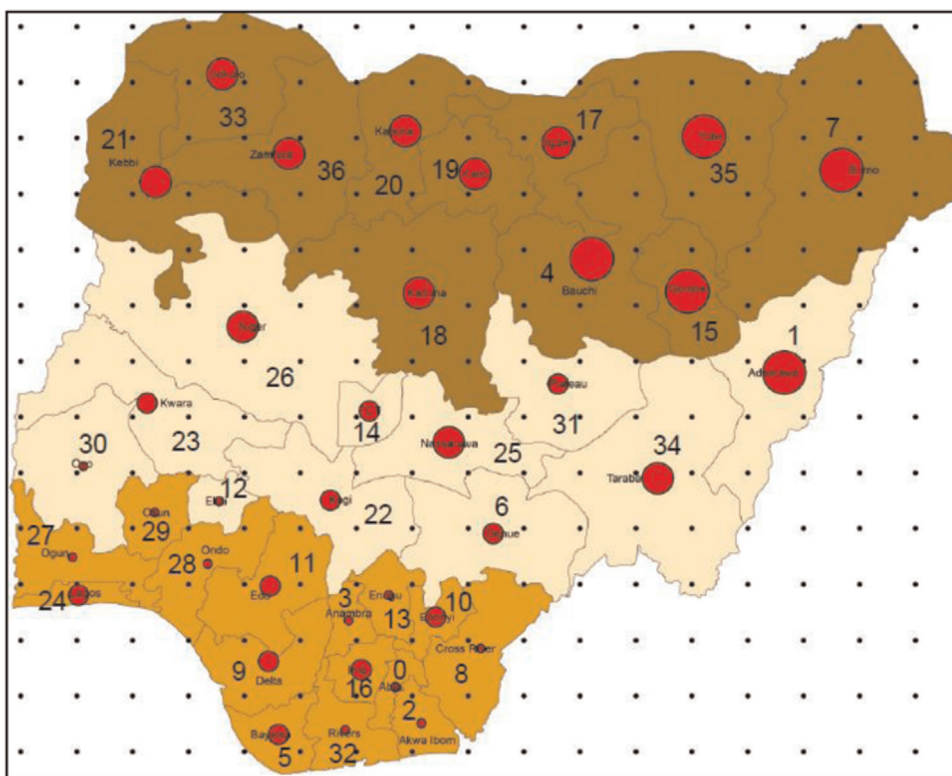
in studying the dynamics of the disease has been pointed out (e.g., Pascual et al., 2000; Chou et al., 2010).

In Nigeria, cholera is one of the primary causes of morbidity and mortality, with incidence occurring in both small outbreaks and large epidemics. The transmission of cholera in Nigeria might be facilitated by numerous factors such as lack of access to safe drinking water, unhygienic environment, environmental disasters, literacy level, population congestion, and internal conflicts which may results to population displacement to Internally Displaced Persons (IDP) camps. Provision of safe drinking water remains a serious issue of concern and this necessitate people even in cities to buy street vended water which has the high risk of being contaminated. Typical areas at risk might include population living in urban and peri-urban slums, these areas are mostly densely populated by low income earners and basic infrastructures are not readily available. Despite the availability of the oral cholera vaccines, anecdotal evidence reveals that this effective control method is not yet commonly used in Nigeria. The main control method is mainly treatment through rehydration with oral salts after infection (WHO, 2011).

The current study aims to statistically model the influences of meteorological and socioeconomic factors on the interannual and spatiotemporal variability of cholera disease in Nigeria. Despite the fact Nigeria is reporting the largest number of cholera of cases and deaths to WHO (WHO, 2012), this study is the first to investigate this type of relationship. The model development and results are based on 22-years (1990–2011) of clinically diagnosed hospital-

\* Corresponding author.

E-mail address: [G.C.Leckebusch@bham.ac.uk](mailto:G.C.Leckebusch@bham.ac.uk) (G.C. Leckebusch).



**Fig. 1.** Map of Nigeria showing: (a) ERA interim grid points covering the country (black dots), (b) grouping of states based on annual rainfall totals into region R1, R2, and R3 (from South to North, coloured; details confer text), and (c) annual mean of cholera incidence rate (IR) for individual states between 2000 and 2011 (red circles). Additional information on states name is added in the appendix (Appendix Table A1). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

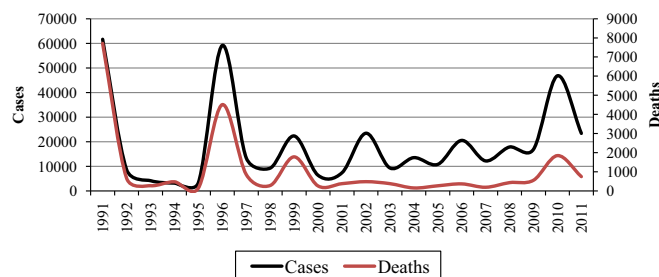
reported cases of cholera, and also 12-years (2000–2011) states level reported cholera cases and deaths (cf. Fig. 1). The study will provide more understanding of the meteorological and socio-economic drivers of cholera outbreak in Nigeria, which could help to a larger degree in the epidemic prediction, thereby allowing authorities to effectively prepare and respond in good time to prevent outbreaks through measures such as vaccinating the vulnerable population. Our paper is the first to report a relationship between meteorological and socioeconomic conditions and cholera in Nigeria.

## 2. Investigation methods and data collection

Two statistical modelling approaches were adopted for this study in order to take advantage of the two different set of disease data obtained. The choices of explanatory variables were based on previous studies that have already documented the importance of these variables. These include meteorological variables such as maximum and minimum temperatures, rainfall, and relative humidity (e.g., Hashizume et al., 2007; Rajendran et al., 2011; Reyburn et al., 2011), absolute poverty (e.g., Traerup, 2010) adult literacy (e.g., Hashizume et al., 2007) access to safe drinking water (e.g., Penrose et al., 2010) and population density (Ali et al., 2002a, 2002b).

### 2.1. Epidemiological data

An overview of the annual variability of reported cholera cases in Nigeria (from WHO sources, cf. Fig. 2) for the last 30 years reveals a high level of interannual variability. To investigate spatio-temporal variability, two different sets of epidemiological records of suspected cholera cases were obtained. Firstly, monthly counts



**Fig. 2.** Reported number of annual cholera cases and deaths to the World Health Organisation (WHO) from Nigeria between 1991 and 2011. Data sourced from WHO: Global health observatory data repository (<http://apps.who.int/gho/data/node.main.175>).

of clinically diagnosed cholera cases reported between 1990 and 2011 were collected from selected hospitals in Northwest Nigeria: Kano, Sokoto and Gusau. The selection of hospitals was based on proximity to meteorological stations with long-term records of measurements, and consistency in reporting cases of infectious disease. Data from these hospitals were successfully used to study meningitis in this region (Abdussalam et al., 2014a, 2014b), which corroborates the quality of data from these sources. Secondly, annual counts of cholera cases and deaths for the entire country at individual state level were obtained from the Nigerian National Centre for Disease Control (NCDC)-a unit of the Federal Ministry of Health: disease surveillance data across the country covering the time period 2000 and 2011 (cf. Fig. 1).

### 2.2. Meteorological data

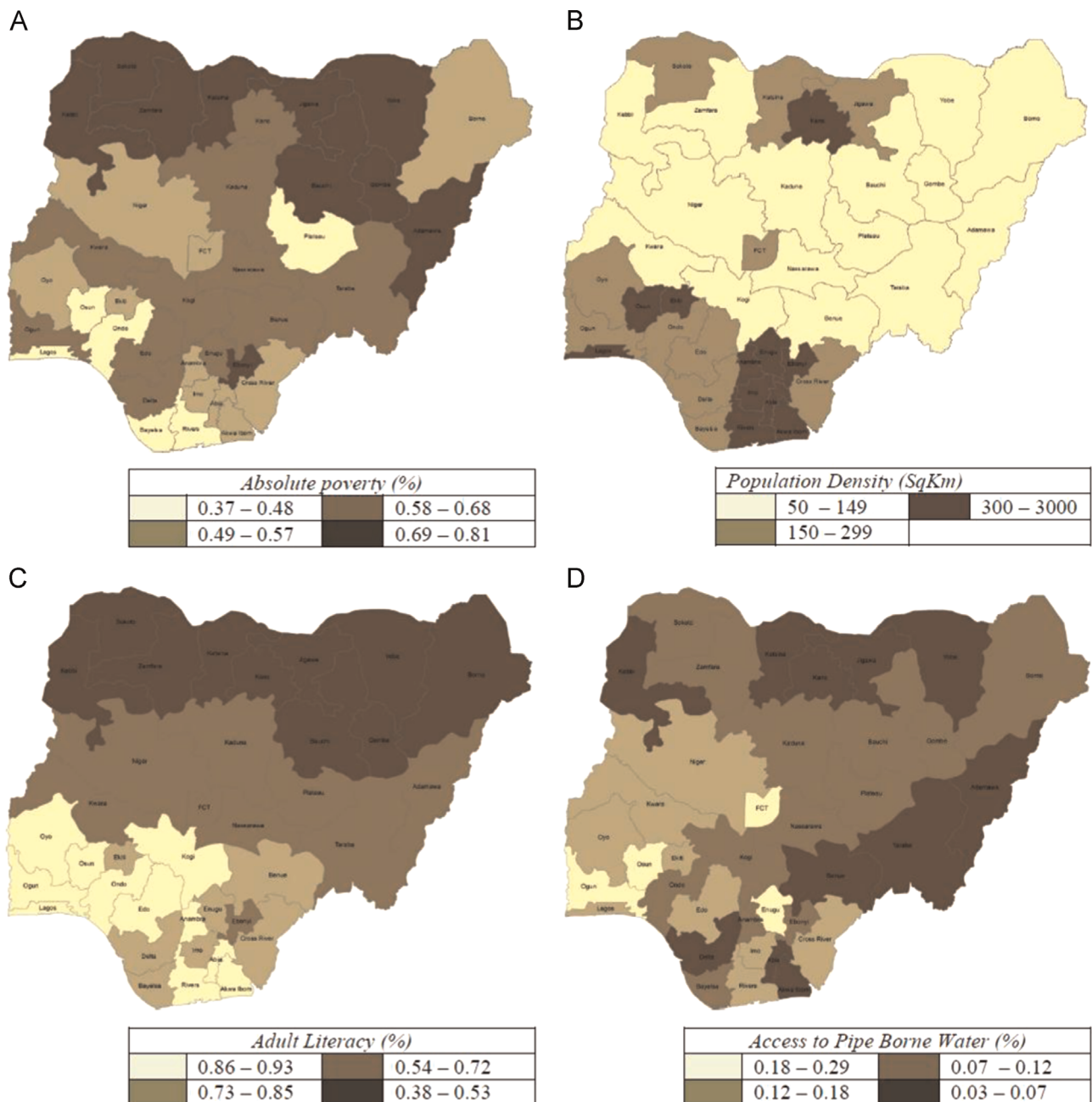
Based on the epidemiological data obtained, two sets of meteorological data were used. Firstly, digital records of four variables

from airport-based meteorological stations located in each of the three cities were obtained from the Nigerian Meteorological Agency (NMA): quality controlled daily precipitation sums, maximum and minimum temperatures, and relative humidity. Secondly, due to non-availability of meteorological station data to represent individual states of Nigeria, for this study we employ the use of ERA-interim reanalysis data (Dee et al., 2011). Almost each state is represented by related grid-cell information with only some exceptions because of their size (the location of each grid-cell within Nigeria is indicated in Fig. 1). Daily time series of surface values for maximum and minimum temperatures alongside precipitation were extracted from each grid-cell between 2000 and 2011. These variables were validated using the available station data from the three stations. Seasonal averages of maximum

and minimum temperatures for the hottest months (from March to June) and annual rainfall totals were computed from the extracted daily time series.

### 2.3. Socioeconomic and demographic data

Annual state level socioeconomic data between 2000 and 2011 were obtained from the Nigerian National Bureau of Statistics (NBS). Data obtained includes: a) percentages of population having access to pipe borne water; b) adult literacy; c) absolute poverty, and d) population density (cf. Fig. 3). A clear spatial heterogeneity in these parameters is revealed with differences between the individual states in Nigeria, but showing also some regional coherency. State's population census (from 2006) was obtained from the



**Fig. 3.** Socio-economic conditions and its spatial distribution in Nigeria: (A) mean percentage of population living in *absolute poverty* per state between 2000 and 2011. Absolute poverty is measured as the percentage of population with income less than some fixed proportion of median income; (B) *population density* in population per square kilometre (Sqkm); (C) mean percentage of population with *adult literacy* per state between 2000 and 2011. Adult literacy is assessed on the ability to read and write with understanding, in English or in any of the Nigerian native languages; (D) mean percentage of people with access to *pipe born water*.



Nigerian Population Commission (NPC), Abuja, Nigeria. Annual population estimate for each state was calculated forward and backward using Nigerian population growth rate index provided by World Bank (<http://data.worldbank.org/country/nigeria>). Population density for each state was computed by dividing each state's population with its aerial cover.

#### 2.4. Model overview

Two modelling approaches were adopted: the Generalised Additive Modelling (GAM) approach and the Multiple Linear Regression (MLR) approach. The choices for these approaches were informed by the different spatial and temporal characteristics exhibited in the two sets of the disease data. The hospital data has the advantage of having higher resolution (monthly) and longer time series of cases (1990–2011), but socioeconomic data are not available at this spatial and temporal scale. While on the other hand, the state level data is on annual timescale and span for only 12 years (2000–2011), but socioeconomic data are available at states level and for the period of the study. GAM was used to model the monthly hospital reported cases, while MLR was used to model the state level data.

##### 2.4.1. Generalised additive models (GAM) for the analysis of monthly hospital data from NW Nigeria (22 years)

Several studies have used the Poisson regression model to investigate the relationships between meteorological and environmental variables with cholera (e.g., Chou et al., 2010; Paz, 2009). The flexibility and less restrictive modelling environment provided by GAM makes it suitable for this study. This is because the model will allow attuning and accounting for the many confounding factors that may have additional effects to the disease, such as social and behavioural factors, population dynamics, migration, and sanitation. Another good reason for choosing GAM is because the models' smoothing function has the benefit of automatically dealing with both non-linear and non-monotonic associations between the outcome variable and the predictors without necessarily the use of variable transformation or polynomial terms. This ultimately avoids using the assumption of normal distribution in the data. The explanatory variables include meteorological variables aggregated monthly between the three selected stations, as well as the previous incidence of cholera. Additionally, all other unobserved seasonally-varying climatic and non-climatic factors that may influence the disease were represented in GAMs by a smooth function of time,  $s(t)$ , which was modelled as a low-degree cubic spline that changes monthly over the course of the annual cycle (Dukic et al., 2012). We assume that cholera counts,  $y_{it}$  follow independent Poisson distributions, in that case a log-link function was used (Cameron, 1998), with mean  $\mu_{it}$ , where  $i=1, \dots, 22$  denotes the years, and  $t=1, \dots, 12$  denotes the month within each year. The GAM formulation is thus:

$$\log(\mu_{it}) = s(t) + X_{it}\beta$$

The expected cholera counts ( $E(y_{it}) = \mu_{it}$ ) in year  $i$  in month  $t$  therefore depend upon the vector of coefficients  $\beta$ , which contains the effects of climate variables collected in the covariate matrix  $X_{it}$ , and upon the effects of the unobserved seasonally-varying factors,  $s(t)$ .

##### 2.4.2. Multiple linear regression (MLR) analysis for annual state-level cholera data (12 years)

Stepwise MLR was applied (e.g., Stocco et al., 2010; Thomson et al., 2006; Yaka et al., 2008) to model the meteorological and socioeconomic influences on the spatiotemporal variability of

cholera cases and deaths in Nigeria between 2000 and 2011. MLR is a powerful statistical technique which uses the equation of a straight line to predict the outcome of a dependant variable from a linear combination of independent predictor variables:

$$Y_i = b_0 + b_1X_1 + b_2X_2 + \dots + b_nX_n + \varepsilon_i$$

where  $Y_i$  is the outcome in the dependent variable,  $b_0$  is a constant,  $b_1$  is the coefficient associated with the first predictor variable ( $X_1$ ) and so on and  $\varepsilon_i$  is the term which calculates the difference between the observed and predicted value of  $Y_i$ .

Since the annual cholera cases and deaths are counts, transformation was carried out in two stages. First, because the annual sum cases have a natural trend with respect to population, incidence rate (IR) were calculated, defined by the number of case per 100,000 of hospital intake, also Case Fatality Rate (CFR) was computed for the death counts, defined as the proportion of fatal cases in relation to the total cases within a specified time. Secondly, considering the skewed nature of both the IR and CFR, their distribution was normalised by a log transformation as LogIR and LogCFR, respectively. Finally, to avoid trend effects from influencing the models' output, all time series included in the models were linear de-trended.

#### 2.5. Model fitting

For the GAM models development, monthly cholera counts for the three hospitals in Kano, Sokoto, and Gusau were aggregated, and variables of the corresponding three meteorological stations averaged. Monthly cholera counts were aggregated in order to derive a regional perspective, which will be linked to regional climate variability. We do not have any indication about specific biases in the individual hospital counts for cholera, in fact we could confirm that for meningitis the local hospital data are highly representative of district level data from WHO for a shorter time period (Abdussalam et al., 2014a). Fig. 4 shows the individual contribution of each hospital per year.

Three GAMs were fitted (denoted as models A, B, C): model A was fitted with the combination of non-lagged climatic variables and cases from the previous month; model B, with only the 1-month lagged explanatory climatic variables and cases from the previous month; and model C is the same as GAM A, except previous cases were excluded. All three GAMs were tested for a variety of degrees-of-freedom (DOF) for the fit of  $s(t)$ , but it was found that those models in which  $s(t)$  has 4 DOF have the best fit. The estimated effect of confounding factors represented in the GAM as a smooth function of time is similar and follows the seasonality of the disease, with the months of April–August having the highest values of  $s(t)$  (cf. Fig. 5). Models A and B are designed to be explanatory and predictive models respectively, while C is for climate change assessment, since the number of cases in the previous month is unknown in the future. The best models were selected by minimising the Akaike Information Criteria (AIC). The retained variables include mean monthly maximum and minimum temperatures, precipitation totals, average relative humidity, and previous cases.

In the second approach, a stepwise MLR was used to fit the best model for three groups of states (regions 1, 2, and 3) as shown in Fig. 1. The 36 states and the Federal Capital Territory (FCT) were grouped into terciles based on their annual rainfall totals, grouping was made in order: (a) to investigate the spatiotemporal differences of the influences of both meteorological and socioeconomic factors on cholera incidence and deaths across these three regions, and (b) to have enough data samples (Knofczynski and Mundfrom, 2008) which will allow for more robust statistics during model fit.

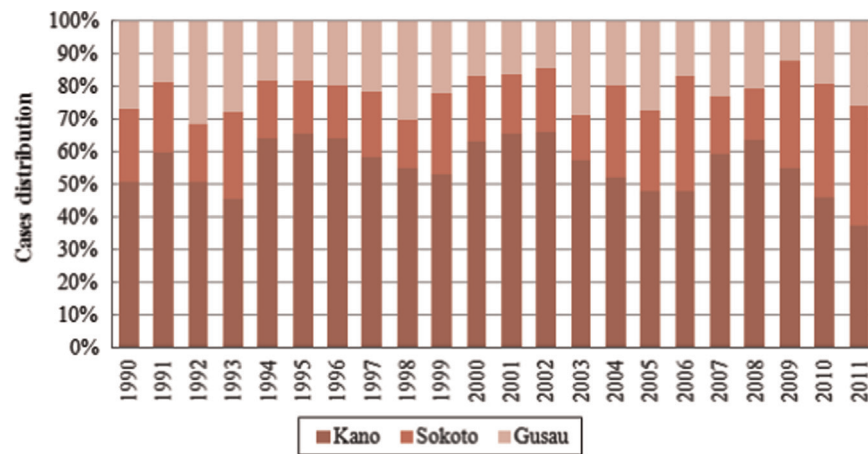


Fig. 4. Distribution of cholera cases from the three selected hospitals (Kano, Sokoto, and Gusau) in northwest Nigeria between 1990 and 2011.

Three time series were generated by joining the time series of each state for both cholera cases and deaths, and their respective meteorological and socioeconomic variables. For each of the regions (R1, R2, and R3), 3 models were fitted: the first (MLR-A) and second (MLR-B) sets of models consist of only meteorological and socioeconomic predictors respectively, while the third model (MLR-C) comprises of the combination of both. The stepwise MLR models were separately developed for IR and CFR with correlated meteorological and socioeconomic variables ( $r > 0.4$  and  $p$ -value  $< 0.05$ ). During model development, collinearity diagnostics and autocorrelation checks were performed. Depending on the model, the explanatory variables includes seasonal maximum and minimum temperatures, rainfall totals, population density, and percentages of population with access to pipe borne water, absolute poverty, and adult literacy.

## 2.6. Model evaluation

Models were assessed using the cross validation correlation (CVC) (Kohavi, 1995) and the skill score (Murphy, 1988) metrics. Validation was carried out between observed versus simulated values for each model. Three fold cross validation was performed by partitioning the data into three consecutive subsets of equal length, followed by successively excluding one of these subsets, and fitting the model to the remaining data to predict the values of the excluded subset. The fitted values for the three subsets were

then combined to obtain a single time series in order to compare with the observed time series

To further determine the prediction accuracy of the models, we computed the skill score of each model by comparing the models' predicted RMSE,  $E_{pre}$ , with that of a reference model  $E_{ref}$ .

$$\text{Skill score} = 1 - (E_{pre}/E_{ref})$$

here  $E_{pre}$  and  $E_{ref}$  represent the RMSE of the model-predicted cases compared to the observed cases, and that of the long term monthly mean of the observed cases

In order estimate the average comparative importance of each predictor variable in all models, the relative influence (RI) of each covariate with respect to all the covariates in the models were computed based on the long-term monthly means of each covariate as well as the monthly value of  $s(t)$  for GAMs, and based on the long-term annual means of each covariate over the 12 year period of data for the MLR models (e.g. Abdussalam et al., 2014a).

## 3. Results for modelling cholera

In recent years, Nigeria has experienced increase in cholera cases and deaths, example, in 2010 alone, the country reported about 41,784 cases and 1716 deaths (CFR 4.1%) from 222 districts in 18 states (WHO, 2013), in which most of the cases comes from the northern part of the country (cf. Fig. 1). Generally, both the individual and aggregated counts of the monthly hospital-reported

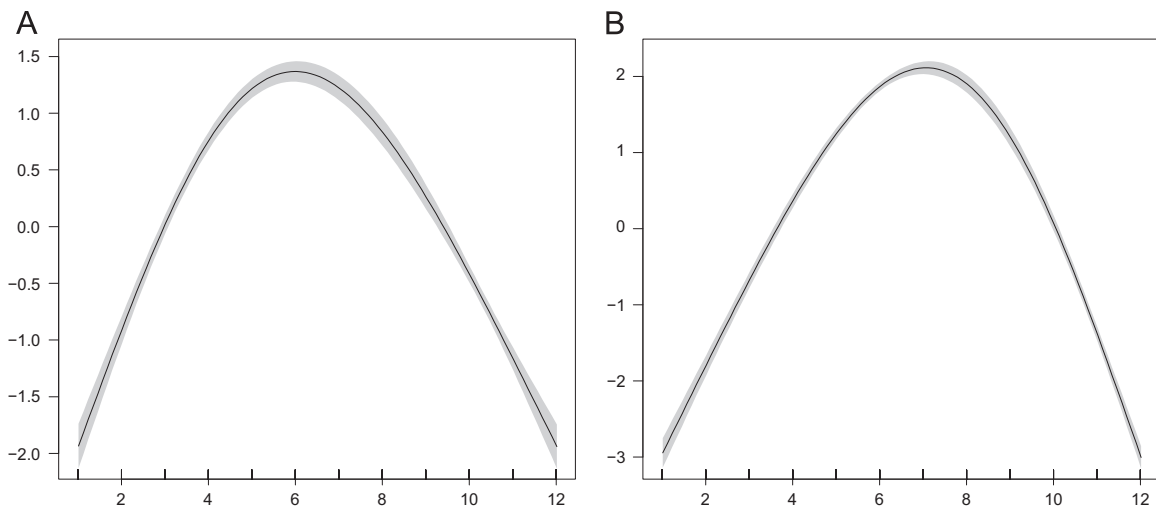


Fig. 5. Estimated smooth function of time ( $s(t)$ ) for 12 months; model A (left) and model B (right). The shaded area shows the 95% confidence interval about  $s(t)$ .

**Table 1**  
Estimates for GAM A, B and C. Model A is fitted non-lagged climate variables, B is fitted with only 1-month lagged climate covariates, while C has the same composition with A, but previous cases are not included. Abbreviations: coef.=coefficient; std. error=standard error.

Predictors	Model A			Model B			Model C		
	Coef.	Std.error	p-Value	Coef.	Std. error	p-Value	Coef.	Std. error	p-Value
Monthly mean $T_{max}$ (°C)	0.321	0.037	0.001	0.246	0.006	0.001	0.384	0.023	0.000
Monthly mean $T_{min}$ (°C)	0.153	0.018	0.101	0.117	0.008	0.007	0.216	0.039	0.018
Rain (mm)	0.206	0.241	0.221	0.305	0.021	0.101	0.281	0.173	0.131
Relative humidity (%)	0.036	0.007	0.001	0.112	0.103	0.001	0.162	0.102	0.041
Previous cases	0.047	0.002	0.201	0.102	0.012	0.104	Na	Na	Na

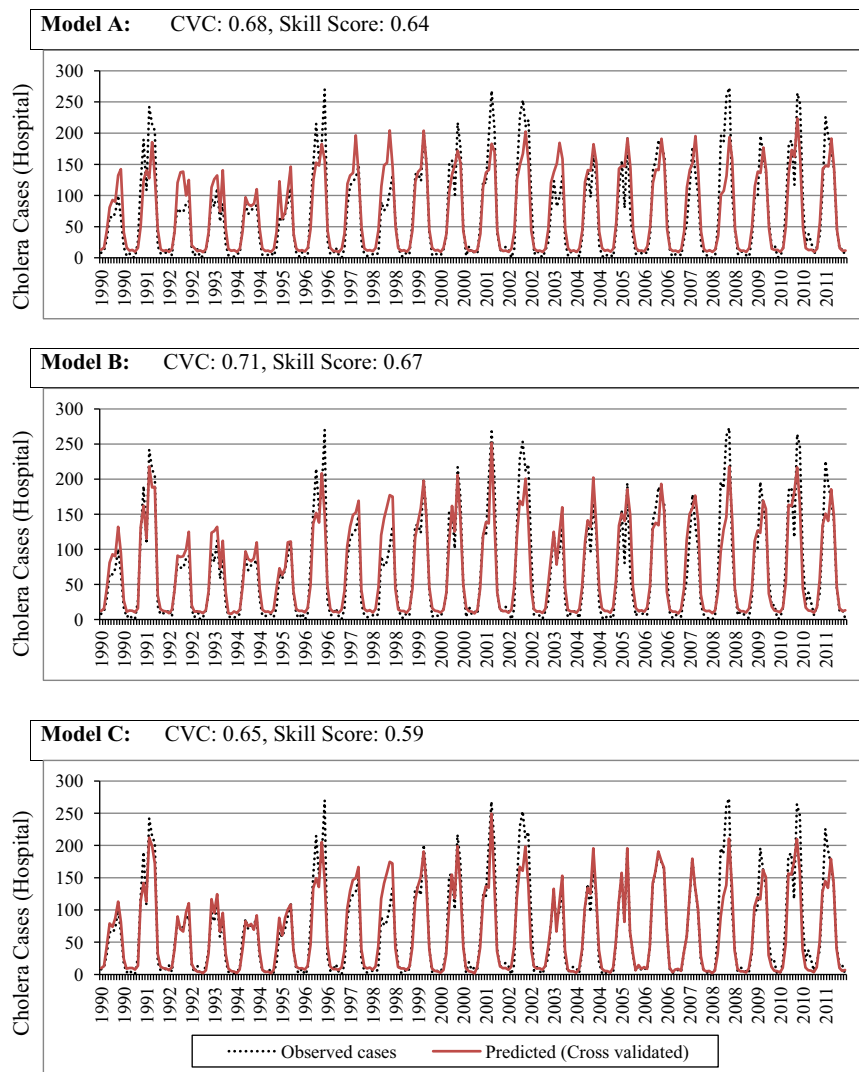
cholera cases exhibit a marked annual cycle, with yearly disease maxima occurring between the month of April and August.

### 3.1. Monthly cholera counts for 3 hospitals in NW Nigeria with generalised additive models (GAM)

All meteorological variables correlated positively with disease cases in all three models, and *mean monthly maximum temperature* and *monthly rainfall totals* appear to be the most important predictors. The model estimates and standard errors for GAMs are presented in Table 1. Fig. 6 shows the 22-years models fit for time

series of the observed number of cases, and that of the predicted cases for GAMs A, B and C. Model with meteorological variables lagged by one month appears to capture the monthly and inter-annual variability of the cholera cases more accurately.

Cross validation statistics are presented in Table 2, revealing high levels of correlations and skill scores in all the three models. As expected from the model fit (Fig. 6), the lagged model has a better statistics values if compared with non-lagged model A as measured by CVC and skill score (0.71 and 0.67). This demonstrates a good indication for the possibility of potentially predicting cholera cases with a month time lead in this region.



**Fig. 6.** The fit for GA models A, B and C (“observed”; dashed-black “predicted”; red). Model A is fitted non-lagged climate variables, B is fitted with only 1-month lagged climate covariates, while C has the same composition with A, but previous cases are not included. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

**Table 2**

Model validation results given CVC and skill score. 'Full' and 'CV' stands for full and 3-fold cross validated models respectively.

Models	Correlation		Skill score	
	Full	CV	Full	CV
A	0.703	0.685	0.659	0.641
B	0.734	0.713	0.704	0.672
C	0.681	0.652	0.623	0.592

The relative influence (RI) for the four months in which cholera cases is generally high is shown in Fig. 7. In both models shown, mean monthly maximum temperature and monthly rainfall totals show a comparatively important influence across the four months with the highest RI of about 35% and 28% across the months. The influence of mean monthly minimum temperature and that of average relative humidity remains almost the same across the months. The function  $s(t)$  which accounts for unobserved explanatory variables, varies in influence from about 8–15%, while previous cases also remains relatively the same, with the highest influence in the month of May.

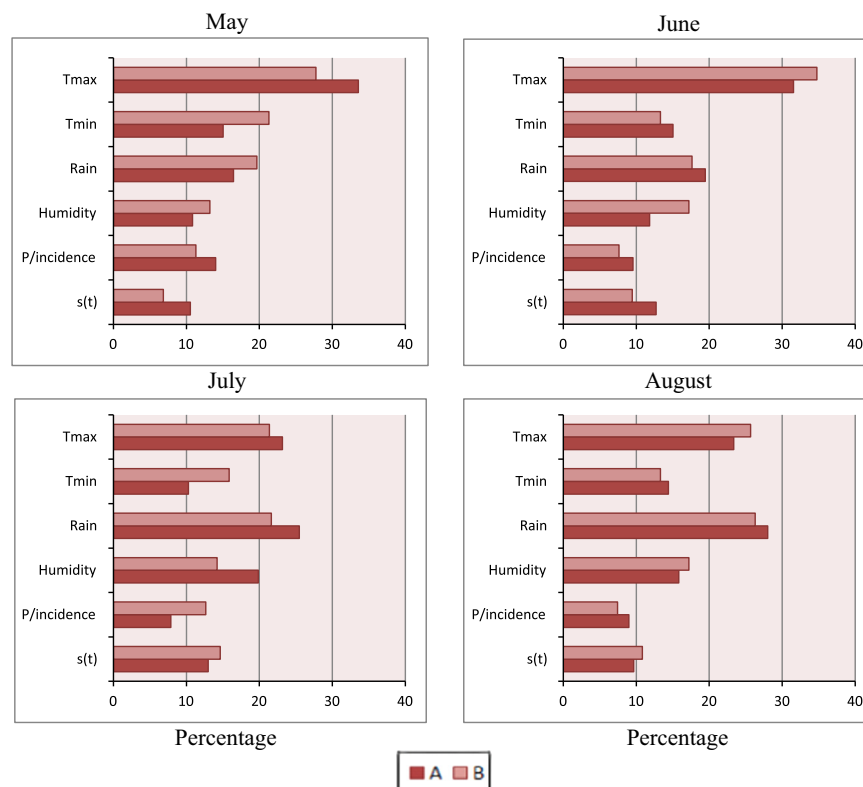
### 3.2. Incorporation of socio-economic influences on the link between climate variability and cholera on state level with Multiple Linear Regression (MLR)

Information about potential socio-economic factors is only available for twelve years on an annual basis per state, which is too short to robustly identify any statistical relation between socio-economic factors, climate and cholera on a single state basis. Nevertheless, for all 36 states and FCT in Nigeria this information is available which allows in principle for an analysis of the spatial distribution of the relationship between climate, social factors and

cholera incidence. To gain an estimate of socio-economic factors in steering cholera incidence and mortality rate, we grouped available state-based information (meteorological data and socio-economic conditions) into 3 larger regions with similar members depending on annual rainfall totals (terciles).

A positive significant relationship between cholera IR and *annual seasonal maximum and minimum temperatures, rainfall, absolute poverty*, and *population density* was observed, while a negative but significant relationship with *access to pipe borne water* and *adult literacy* ( $R^2$  ranges from 0.20–0.60,  $p < 0.05$ ) is found (without figure). Table 3 presents the regression coefficient of both IR and CFR models. Regardless of region and model, individually, maximum temperature, rainfall, and water source appear to be the most important variables in explaining the variability of the disease. Adult literacy is the least important predictor with respect to IR, whilst population density explains the lowest proportion of variability in the CFR. With the exception of water source, socio-economic variables explain more variability in the CFR models than the IR. A higher proportion of variability is consistently explained in both IR and CFR models with the combination of meteorological and socioeconomic explanatory variables. Fig. 8 shows the models' fit for time series of the observed IR, and that of the predicted cases for the best MLR model (with the combination of meteorological and socioeconomic variables, MLR-C). Models with only meteorological or socioeconomic variables do not accurately capture the interannual variability like that of the combination of both. This emphasises the need why socioeconomic variables should be considered in modelling frame work of this nature.

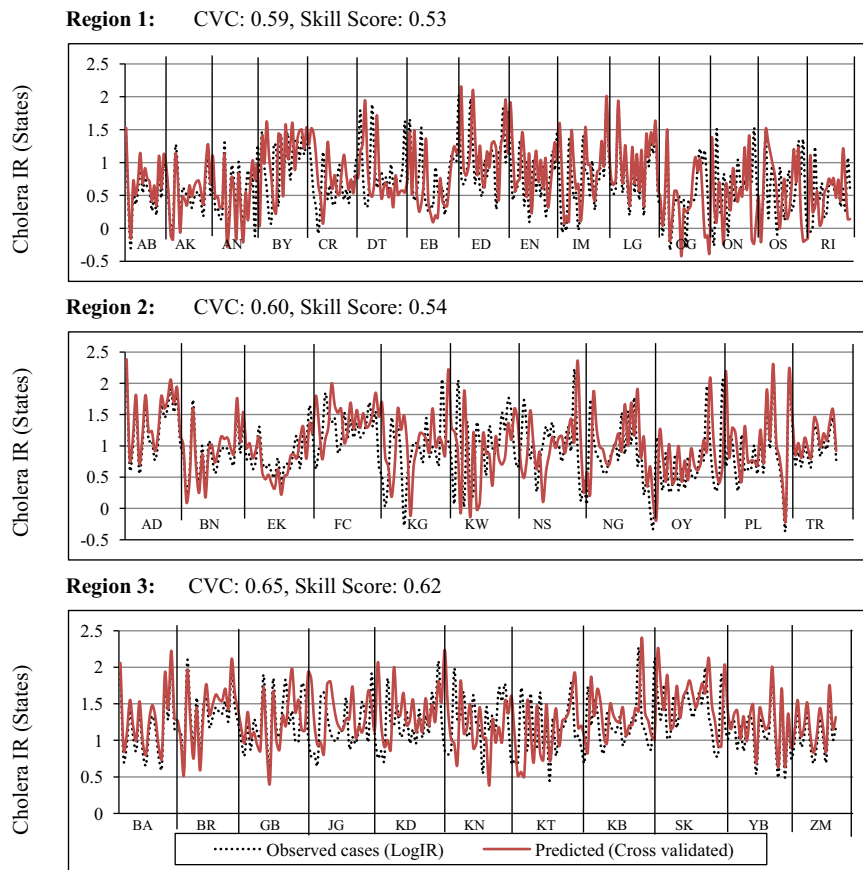
Cross validation statistics for all models and for all the three regions are presented in Table 3, all the models have a relatively good statistics, but as expected, models with the combination of both meteorological and socioeconomic variables show better statistic values compared with other models that are either climate-based or socioeconomic-based only. Climate-based models



**Fig. 7.** The relative influence of explanatory variables on *cholera* in northwest Nigeria for GA models A and B for the months in which the disease burden is typically highest: May, June, July, and August.

**Table 3**  
Model estimates and validation results, given by regression coefficient ( $R^2$ ), cross validation correlation (CVC), and skill score. 'Full' and 'CV' stands for full and 3-fold cross validated models, respectively. MLR-A: climate only based model; MLR-B: socioeconomic only based model, and MLR-C: combined model.

Region	Incidence rate models					Mortality rate models				
	$R^2$	CVC		Skill score		$R^2$	CVC		Skill score	
		Full	CV	Full	CV		Full	CV	Full	CV
Region 1										
MLR-A	0.349	0.591	0.532	0.494	0.451	0.203	0.451	0.448	0.422	0.443
MLR-B	0.149	0.387	0.362	0.303	0.278	0.123	0.359	0.334	0.252	0.232
MLR-C	0.365	0.604	0.593	0.594	0.534	0.316	0.562	0.542	0.571	0.567
Region 2										
MLR-A	0.194	0.441	0.433	0.395	0.378	0.118	0.343	0.331	0.235	0.259
MLR-B	0.231	0.481	0.443	0.445	0.423	0.168	0.404	0.396	0.327	0.325
MLR-C	0.403	0.634	0.601	0.611	0.541	0.261	0.511	0.477	0.481	0.467
Region 3										
MLR-A	0.119	0.345	0.324	0.314	0.293	0.124	0.351	0.332	0.288	0.249
MLR-B	0.350	0.592	0.589	0.541	0.534	0.301	0.544	0.535	0.420	0.414
MLR-C	0.429	0.655	0.652	0.642	0.623	0.375	0.615	0.601	0.594	0.562



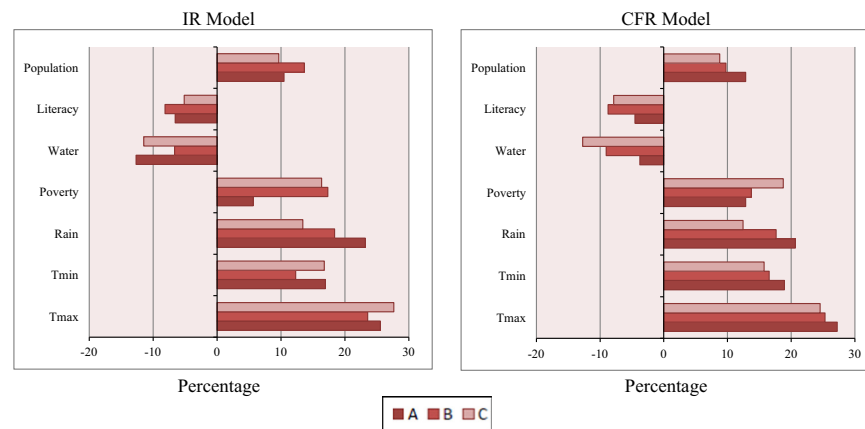
**Fig. 8.** Model fit for IR models in the three regions (R1, R2, and R3) ("observed" black, "predicted" red) from 2000 to 2011. Models contain the combination of both meteorological and socioeconomic explanatory variables (seasonal maximum and minimum temperatures, annual rainfall totals, percentage of population in absolute poverty, population having access to pipe borne water, adult literacy, and population density per square kilometre). Region 1, 2, and 3 consist of time series from 15, 11, 11 states, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

have a better statistics in the southern part of the country (region R1) if compared with the north (region R3) and vice-versa when looking at the socioeconomic-based ones. Overall the best model in all the regions is the model for IR which uses the combination of both, meteorological and socioeconomic explanatory variables from the northern region (R3) as measured by CVC and skill score (0.65 and 0.62).

The relative influence of the explanatory variables for models

of IR and CFR across the study period is shown in Fig. 9. In all models seasonal average of maximum temperature and annual rainfall totals shows a comparatively important influence with the highest RI of about 27% and 26%, respectively. The influence of seasonal average minimum temperature, adult literacy and that of population density remains almost the same across the regions and models. The RI of absolute poverty and water source is higher in the northern part of the county (region R3).





**Fig. 9.** The relative influence of explanatory variables (seasonal maximum and minimum temperatures, annual rainfall totals, percentages of population in absolute poverty, having access to pipe borne water, adult literacy, and population density per square kilometre) on cholera incidence rate in Nigeria for models MLR-A, MLR-B, and MLR-C, across the study period.

#### 4. Summary, discussion and conclusion

In this study, both GAM and MLR statistical techniques were employed to model the influences of meteorological and socio-economic conditions on the interannual variability of cholera in Nigeria. GAMs were used to model the monthly aggregate counts of clinically diagnosed hospital-reported cholera cases from 1990 to 2011 in northwest Nigeria, explanatory variables in this models includes mean monthly maximum and minimum temperatures, monthly rainfall totals, monthly average relative humidity, and 1-month previous incidence and  $s(t)$ . While stepwise MLR was used to investigate the spatiotemporal variability of cholera incidence and mortality rate using annual state level data between 2000 and 2011. Here, explanatory variables include seasonal maximum and minimum temperatures, annual rainfall totals, population density, and percentages of population with absolute poverty, adult literacy, and access to pipe borne water. These approaches were adopted because of the differences in the spatial and temporal characteristics of the disease data, and also based on the availability of socioeconomic data. The hospital case data exhibit a marked annual cycle, with yearly disease maxima occurring between the months of April and August, while the state level annual data indicates increase in cases of cholera with most of the cases being reported from the northern part of the country.

All models in both the two approaches pointed out to the importance of meteorological variables in explaining the disease dynamics, most especially temperature and rainfall. The general positive relationship observed between cholera, temperature and rainfall has previously been reported by studies all over the world. Here we manage for the first time to identify and quantify this for Nigeria, especially with respect to the spatio-temporal distribution of the link responsible for interannual variability of Cholera. Temperature is related to food contamination, which may consequently serve as a vehicle for cholera transmission (Rabbani and Greenough, 1999) depending on the physio-chemical properties of the contaminated food. Similarly, rainfall is well documented to have a positive relationship with cholera (Reyburn et al., 2011; Hashizume et al., 2011). The seasonal trend observed in the monthly hospital time series corresponds with the rainy season; this is consistent with what has been found in other regions (e.g. Richard et al., 1999). The link between rainfall and cholera was explained to be as a result of flooding which exposes population to the bacterium (Emch et al., 2010; Hashizume et al., 2008), this explanation might also be applicable to Nigeria. Because in the northern part of Nigeria, the beginning of rain season is usually associated with heavy downpours which consequently bring about

flooding, thereby increasing the risk of contaminating sources of drinking water through sewage collapse. In slums areas, where most of the residents are using a local toilet system (pit latrines), during heavy downpours sewage water can overflow or seep through the ground into local sources of drinking water like wells, or by pressure into roasted pipes that leaks due to aging or lack of maintenance. On the other hand during the peak of the dry and hot season people tend to source drinking and cooking water from sources with higher risk of contamination, which includes stagnant waters and wells with lower depths. A typical example is the cholera outbreak in March 1999 in Kano city where scores of lives were lost, this outbreak was directly associated with the interruption of domestic water supply (WHO, 2013), which necessitate residents to source water from contaminated sources. Another problem that might facilitate cholera outbreak during the rainy season is the poor drainage system which is a peculiar characteristic of major cities in Nigeria, after heavy rainfall houses are usually flooded in some cases sub-merged with dirty water from open gutters.

The positive relationship observed in the MLR models between cholera and socio economic variables, such as absolute poverty and population density is in agreement with previous findings (e.g., Matsuda et al., 2008; Snowden, 2008). Poverty is associated with environmental sanitation which can cause the disease to propagate (Rajendran et al., 2011; Reiner et al., 2012), while population densities allow the disease to spread quickly possibly via faecal contamination (Penrose et al., 2010; Rajendran et al., 2011). In a related study, Ali et al. (2002a), using a risk spatial model reported that areas associated with the risk of cholera morbidity and mortality are associated with population density, low educational status and proximity to surface waters.

In Nigeria, population living in urban and peri-urban slums are more at risk and vulnerable of contracting the disease, because these areas are mostly densely populated by low income earners and basic infrastructures are not readily available, which results in many people defecating in the open. In all the major cities of Nigeria public conveniences can hardly be found, according to United Nation Children's Fund (WHO and UNICEF, 2012) report over 34 million Nigeria are defecating outside. In areas like this, after a heavy downpour, surface outwash, collapsed sewages, and open drainage may lead to contamination of sources of drinking water like wells and rivers. Since cholera can spread via contaminated food and water, transmission of the disease will be made easy and rapid. The results also reveal negative relationships between IR and CFR with adult literacy, this shows that as more adults become literate, cases and deaths from cholera decreases, this finding is in

**Table A1**  
States name and codes.

SN	State	State code
0	Abia	AB
1	Adamawa	AD
2	Akwa Ibom	AK
3	Anambra	AN
4	Bauchi	BA
5	Bayelsa	BY
6	Benue	BN
7	Borno	BR
8	Cross River	CR
9	Delta	DT
10	Ebonyi	EB
11	Edo	ED
12	Ekiti	EK
13	Enugu	EN
14	FCT-Abuja	FC
15	Gombe	GB
16	Imo	IM
17	Jigawa	JG
18	Kaduna	KD
19	Kano	KN
20	Katsina	KT
21	Kebbi	KB
22	Kogi	KG
23	Kwara	KW
24	Lagos	LG
25	Nassarawa	NS
26	Niger	NG
27	Ogun	OG
28	Ondo	ON
29	Osun	OS
30	Oyo	OY
31	Plateau	PL
32	Rivers	RV
33	Sokoto	SK
34	Taraba	TR
35	Yobe	YB
36	Zamfara	ZM

line with the studies conducted in Latin America, Bangladesh, and Tanzania (Ackers et al., 1998; Ali et al., 2002b; Traerup 2010) respectively. Those with a higher level of education are expected to make more rational decisions by taking measures to avoid contracting the disease, and if infected they seek an immediate medical treatment for the disease before fatality occurs (Ali et al., 2002b). Water sources was also one of the important predictors in explaining the interannual variability in both IR and CFR, and was negatively associated, suggesting that as more people have access to safe drinking water there will be less contraction of the disease by individuals (e.g., Reiner et al., 2012; Rajendran et al., 2011).

MLR models show that socioeconomic variables contribute more in explaining the variability of both cholera IR and CFR in the northern part of the country (Table 3). Generally, socioeconomic data shows that this part of the country has lower level of adult literacy and higher level of poverty. Documented evidence has established this kind of spatial differences within countries whereby high cholera rates were attributed to poor socioeconomic status (e.g., Ali et al., 2002b; Penrose et al., 2010). Lower socioeconomic status in the northern part of the country may be the reason why IR and CFR are higher in this area and as a result why socioeconomic variables were able to explain a greater proportion of interannual variability in this region.

The study provided exploratory information on the influences of meteorological and socioeconomic explanatory variables on cholera interannual variability in Nigeria. Results from both modelling approaches highlighted the importance of both meteorological and socioeconomic variables in explaining and predicting the disease in Nigeria. It has been shown that increases in

temperature, rainfall, poverty, and population density may increase both cholera cases and deaths, while improvement of drinking water and adult literacy might reduce the risk of contracting the disease.

The results emphasises the importance of including socioeconomic factors in studies of this nature, this is because socioeconomic variables help in explaining a higher proportion of interannual variability in both the IR and CFR than using climate information alone. The new data provided by this study will serve as a strong basis for the potential prediction of cholera in Nigeria which could help authorities in controlling or even avoiding outbreaks. Also, with the growing concern of the potential impact of climate change on the dynamic of infectious diseases in the future, this study has provided a background for assessing the future impact, which is the next step of this study.

## Acknowledgements

The Nigerian Tertiary Education Trust Fund provided funding. Research of GC Lekebusch is supported by EU-FP7 Grant FP7-MC-CIG-322208 (EVE). We declare no actual or potential competing financial interest.

## Appendix A

See Appendix Table here [Table A1](#)

## References

- Abdussalam, A.F., Monaghan, A.J., Dukic, V.M., Hayden, M.H., Hopson, T.M., Lekebusch, G.C., Thorne, J., 2014a. Climate influences on interannual variability of meningitis incidence in northwest Nigeria. *Weather Clim. Soc.* 6 (1), 62–76. <http://dx.doi.org/10.1175/WCAS-D-13-00004.1>.
- Abdussalam, A.F., Monaghan, A.J., Steinhoff, D.F., Dukic, V.M., Hayden, M.H., Hopson, T.M., Thorne, J.E., Lekebusch, G.C., 2014b. The impact of climate change on meningitis in Northwest Nigeria: an assessment using CMIP5 climate model simulations. *Weather Clim. Soc.* 6 (3), 371–379. <http://dx.doi.org/10.1175/WCAS-D-13-00068.1>.
- Ackers, M.L., Quick, R.E., Drasbek, C.J., Hutwagner, L., Tauxe, R.V., 1998. Are there national risk factors for epidemic cholera? the correlation between socioeconomic and demographic indices and cholera incidence in Latin America. *Int. J. Epidemiol.* 27, 330–334.
- Ali, M., Emch, M., Donnay, J.P., Yunus, M., Sack, R.B., 2002a. Identifying environmental risk factors for endemic cholera: a raster GIS approach. *Health Place* 8, 201–210.
- Ali, M., Emch, M., Donnay, J.P., Yunus, M., Sack, R.B., 2002b. The spatial epidemiology of cholera in an endemic area of Bangladesh. *Soc. Sci. Med.* 55, 1015–1024.
- Bouma, M.J., Pascual, M., 2001. Seasonal and interannual cycles of endemic cholera in Bengal 1891–1940 in relation to climate and geography. *Hydrobiologia* 460, 147–156.
- Charles, R.C., Ryan, E.T., 2011. Cholera in the 21st century. *Curr. Opin. Infect. Dis.* 24, 472–477.
- Chou, W.C., Wu, J.L., Wang, Y.C., Huang, H., Sung, F.C., Chuang, C.Y., 2010. Modeling the impact of climate variability on diarrhea-associated diseases in Taiwan (1996–2007). *Sci. Total Environ.* 409, 43–51.
- Colwell, R.R., 1996. Global climate and infectious disease: the cholera paradigm. *Science* 274, 2025–2031.
- De Magny, G.C., Thiaw, W., Kumar, V., Manga, N.M., Diop, B.M., Gueye, L., Kamara, M., Roche, B., Murtugudde, R., Colwell, R.R., 2012. Cholera outbreak in senegal in 2005: was climate a factor? *PLoS One* 7, e44577.
- Dee, D.P., Uppala, S.M., Simmons, A.J., et al., 2011. The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Q. J. R. Meteorol. Soc.* 137, 553–597.
- Dukic, V., Coauthors, 2012. The role of weather in meningitis outbreaks in Navrongo, Ghana: a generalized additive modeling approach. *J. Agric. Biol. Environ. Stat.* 17, 422–460. <http://dx.doi.org/10.1007/s13253-012-0095-9>.
- Emch, M., Yunus, M., Escamilla, V., Feldacker, C., Ali, M., 2010. Local population and regional environmental drivers of cholera in Bangladesh. *Environ. Health* 9.
- Fernandez, M.A.L., Bauernfeind, A., Jimenez, J.D., Gil, C.L., El Omeiri, N., Guibert, D.H., 2009. Influence of temperature and rainfall on the evolution of cholera epidemics in Lusaka, Zambia, 2003–2006: analysis of a time series. *Trans. R. Soc. Trop. Med. Hyg.* 103, 137–143.

- Hashizume, M., Armstrong, B., Hajat, S., Wagatsuma, Y., Faruque, A.S., Hayashi, T., Sack, D.A., 2007. Association between climate variability and hospital visits for non-cholera diarrhoea in Bangladesh: effects and vulnerable groups. *Int. J. Epidemiol.* 36, 1030–1037.
- Hashizume, M., Armstrong, B., Hajat, S., Wagatsuma, Y., Faruque, A.S.G., Hayashi, T., Sack, D.A., 2008. The effect of rainfall on the incidence of cholera in Bangladesh. *Epidemiology* 19, 103–110.
- Hashizume, M., Faruque, A.S.G., Terao, T., Yunus, M., Streatfield, K., Yamamoto, T., Moji, K., 2011. The Indian Ocean dipole and cholera incidence in Bangladesh: a time-series analysis. *Environ. Health Perspect.* 119, 239–244.
- Knofczynski, G.T., Mundfrom, D., 2008. Sample sizes when using multiple linear regression for prediction. *Educ. Psychol. Meas.* 68, 431–442.
- Kohavi, R., 1995. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: *Proceedings of the 14th International Joint Conference on Artificial Intelligence*, Lawrence Erlbaum Associates Ltd., pp. 1137–1145.
- Matsuda, F., Ishimura, S., Wagatsuma, Y., Higashi, T., Hayashi, T., Faruque, A.S.G., Sack, D.A., Nishibuchi, M., 2008. Prediction of epidemic cholera due to *Vibrio cholerae* O1 in children younger than 10 years using climate data in Bangladesh. *Epidemiol. Infect.* 136, 73–79.
- Murphy, A.H., 1988. Skill scores based on the mean square error and their relationships to the correlation coefficient. *Mon. Weather Rev.* 116, 2417–2424.
- Olouko, G., Umilla, B., Jaggernath, J., 2014. Perception and trends of associated health risks with seasonal climate variation in Oke-Ogun region, Nigeria. *Health Place* 25, 47–55.
- Pascual, M., Rodo, X., Ellner, S.P., Colwell, R., Bouma, M.J., 2000. Cholera dynamics and El Niño–Southern Oscillation. *Science* 289, 1766–1769.
- Paz, S., 2009. Impact of temperature variability on cholera incidence in South-eastern Africa, 1971–2006. *Ecohealth* 6, 340–345.
- Penrose, K., De Castro, M.C., Werema, J., Ryan, E.T., 2010. Informal urban settlements and cholera risk in Dar es Salaam, Tanzania. *PLoS Negl. Trop. Dis.* 4 (3) 10.1371/journal.pntd.0000631.
- Rabbani, G.H., Greenough, W.B., 1999. Food as a vehicle of transmission of cholera. *J. Diarrhoeal Dis. Res.* 17, 1–9.
- Rajendran, K., Sumi, A., Bhattachariya, M.K., Manna, B., Sur, D., Kobayashi, N., Ramamurthy, T., 2011. Influence of relative humidity in *Vibrio cholerae* infection: a time series model. *Indian J. Med. Res.* 133, 138–145.
- Reiner Jr., R.C., King, A.A., Emch, M., Yunus, M., Faruque, A.S.G., Pascual, M., 2012. Highly localized sensitivity to climate forcing drives endemic cholera in a megacity. *Proc. Natl. Acad. Sci. USA* 109, 2033–2036.
- Reyburn, R., Kim, D.R., Emch, M., Khatib, A., von Seidlein, L., Ali, M., 2011. Climate variability and the outbreaks of cholera in Zanzibar, East Africa: a time series analysis. *Am. J. Trop. Med. Hyg.* 84 (6), 862–869 10.4269/ajtmh.2011.10-0277.
- Richard, V., Tosi, C., Arzel, B., Kana, N., 1999. Endemic cholera in Chad: a real public health problem. *Med. trop.: rev. du Corps de sante colon.* 59, 169–172.
- Snowden, F.M., 2008. Emerging and reemerging diseases: a historical perspective. *Immunol. Rev.* 225, 9–26.
- Speelman, E.C., Checkley, W., Gilman, R.H., Patz, J., Calderon, M., Manga, S., 2000. Cholera incidence and El Niño-related higher ambient temperature. *J. Am. Med. Assoc.* 283, 3072–3074.
- Stocco, C., Leite, M.D., Labiak, V.B., Das Virgens, J.S., Nascimento, E., 2010. Influence of climatic variables on the incidence of meningitis and its space distribution in the city of Ponta Grossa-PR, 2001–2005. *Saude E Soc.* 19, 84–93.
- Talavera, A., Perez, E.M., 2009. Is cholera disease associated with poverty? *J. Infect. Dev. Ctries.* 3, 408–411.
- Thomson, M.C., Molesworth, A.M., Djingarey, M.H., Yameogo, K.R., Belanger, F., Cuevas, L.E., 2006. Potential of environmental models to predict meningitis epidemics in Africa. *Trop. Med. Int. Health* 11, 781–788.
- Traerup, S.L.M., 2010. The health impacts of climate change: A study of cholera in Tanzania. In: (BC3), B. C. F. C. C. (ed.) BC3 Working Paper Series 2010-01. Bilbao, Spain.
- WHO, 2011, August 11. Prevention and control of cholera outbreaks: World Health Organisation policy and recommendations. Retrieved from (<http://www.who.int/cholera/technical/prevention/control/en/>).
- WHO, 2013, June 18. Global task force on cholera control. Cholera country profile: Nigeria. Last updated January, 2012. Retrieved from (<http://www.who.int/cholera/countries/NigeriaCountryProfile2011.pdf>).
- WHO, 2012, June 17. Cholera fact sheet N107, 2012. Retrieved from (<http://www.who.int/mediacentre/factsheets/fs107/en/>).
- WHO and UNICEF, 2012: Progress on drinking water and sanitation: Joint Monitoring Programme update 2012. pp58. ISBN: 978 92 806 4632 0. Accessable Internet Source: [http://www.who.int/water\\_sanitation\\_health/publications/2012/jmp\\_report/en/](http://www.who.int/water_sanitation_health/publications/2012/jmp_report/en/).
- Yaka, P., Sultan, B., Broutin, H., Janicot, S., Philippon, S., Fourquet, N., 2008. Relationships between climate and year-to-year variability in meningitis outbreaks: a case study in Burkina Faso and Niger. *Int. J. Health Geog.* 7 (34) 10.1186/1476-072X-7-34 In this journal it is given: (page number not for citation purposes).