Empirical dynamic modeling reveals climatic drivers in dynamics of bacillary dysentery epidemics in China

Haisheng Wu¹,², Zhenjun Li²,², Xiaolin Yu¹, Qinghui Zeng¹, Jiumin Lin¹, Yuliang Chen¹, Siqi Ai¹, Pi Guo¹,⁵ and Hualiang Lin⁴

¹ Department of Preventive Medicine, Shantou University Medical College, Shantou 515041, People’s Republic of China
² State Key Laboratory for Infectious Disease Prevention and Control, National Institute for Communicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention, Beijing 102206, People’s Republic of China
³ Department of Hepatology and Infectious Diseases, the Second Affiliated Hospital, Shantou University Medical College, Shantou 515041, People’s Republic of China
⁴ Department of Epidemiology, School of Public Health, Sun Yat-sen University, Guangzhou, People’s Republic of China
⁵ Guangdong Provincial Key Laboratory of Infectious Diseases and Molecular Immunopathology, Shantou 515041, People’s Republic of China
⁶ These authors contributed equally to this work.

E-mail: pguo@stu.edu.cn and linhualiang@mail.sysu.edu.cn

Keywords: bacillary dysentery, climatic drivers, empirical dynamic modeling, nonlinear dynamics

Supplementary material for this article is available online

Abstract

At present, there is still a lack of studies to address the dynamics underlying epidemics of bacillary dysentery (BD), with particular concern on the role of climatic drivers across different regions of a country or the world. The variability of climate domains, non-linear interactions, and covariations of climatic variables pose challenges for explaining the correlation between environment and BD and identifying causal climatic drivers. In this nationwide study involving 31 provincial capital cities in China, we used the empirical dynamic modeling (EDM), which is a framework for nonlinear time series analysis, to explore climate-driven patterns of BD. We first identified possible temperature (i.e. mostly via its seasonality) and relative humidity driving BD dynamics nationally. Then, we used the EDM to estimate the causal intensity of temperature and relative humidity in different latitudes. The results reveal that the combined nonlinear effect of them on BD may be nationwide, but this effect is concealed due to their high correlation in northern regions. We also found an approximately S-shaped relationship between temperature and BD at the population level; while the effects of relative humidity on BD are strongly dependent on environmental details, especially at temperatures above 0 °C. Temperature may potentially mediate and affect the effects of relative humidity on BD. This nationwide study provides a unified explanation for causal climate drivers of BD, regardless of the different climatic domains and epidemic patterns of BD in diverse cities.

1. Introduction

Epidemics of bacillary dysentery (BD) have been a great scourge of human populations for centuries (Njamkepo et al 2016). The most important BD-associated pathogens worldwide are Shigella, Salmonella, and Campylobacter (Wu et al 2020). Notably, a majority of BD and related manifestations caused by the bacteria belonging to the genus Shigella, which is of considerable global public health concern (Kotloff et al 1999, von Seidlein et al 2006). As a type of Shigellosis primarily transmitted by the via contaminated food or water, BD is still a major public issue in developing countries and regions where hygiene and sanitation are poor (Cheng et al 2017). Globally, the burden of BD, and causes considerable morbidity and mortality per year (Mara and Horan 2003). According to a recent study of the global burden of diseases, Shigella was the second leading cause of diarrheal mortality in 2016 across all age groups, and responsible for 212 438 deaths and about 13.2% of all diarrhea deaths (Khalil et al 2018).

In Asian countries, a multicentre study showed that Shigella was isolated from 2927 of 56 958 diarrhea
episodes, accounting for 5% of the total diarrhea episodes, and *Shigella* appears to be more ubiquitous in Asian’s impoverished populations (von Seidlein *et al*. 2006). In China, *Shigella flexneri* was identified as the most prevalent type that caused BD, accounting for about 64% BD cases (Chang *et al*. 2016). Although the mortality of BD has shown a downward trend in the past decade (Koloff *et al*. 2018), there is still a considerable burden of BD diseases in the global community (Khalil *et al*. 2018). For example, a recent epidemiological analysis shows that the incidence of BD was 11.24 cases per 100 000 person in 2014 in China (Chang *et al*. 2016).

In fact, BD is a disease that has been widely distributed throughout the world, and it has a spatially heterogeneous distribution across different countries or different regions of a country (Anderson *et al*. 2016). Although BD epidemics can be affected by some factors such as industrialization, hygiene deterioration, predisposition because of malnutrition, and shortage of specific therapies (Mather *et al*. 2014), these factors are not enough to explain the geographical heterogeneity in the current epidemiological distribution. Epidemiological evidence indicates that climate changes with anticipation of more frequent, intense and prolonged extreme temperature events are posed a major threat to the occurrence of infectious diseases (McMichael 2013, Altizer *et al*. 2013). It is, therefore, generally recognized that ambient temperature is a possible environmental determinant, and contributes to the dynamics of infectious diseases (Altizer *et al*. 2013, Xiao *et al*. 2017, Zhao *et al*. 2018). This also occurs in the dynamics of BD. Previous studies suggest that meteorological factors promote the emergence of BD (Chen *et al*. 2012, Lee *et al*. 2017). In China, an increasing number of studies used mathematical models to examine and quantify associations between climate and BD, including time series Poisson regression models (Hu *et al*. 2018), negative binomial regression models (Lee *et al*. 2017), distributed lag non-linear models (Cheng *et al*. 2017), and generalized additive models (Li *et al*. 2015). However, the conclusions of these correlation-based studies are inconsistent, indicating that there is biogeographical heterogeneity of the effects of climate on BD (Wu *et al*. 2020). For example, some previous studies have shown that the incidence of BD is positively linked to ambient temperature or relative humidity (Zhang *et al*. 2008, Li *et al*. 2013, Zhao *et al*. 2016), while another study revealed that the incidence of BD is negatively linked to relative humidity (Li *et al*. 2013).

It is well known that the annual climate periodicity is weaker in the tropics than in temperate regions, which may relate to the seasonally driven patterns of some infectious diseases (Deyle *et al*. 2016a). For instance, previous studies indicated that the diversity of seasonal influenza patterns observed across temperate, subtropical and tropical climates might be accounted by environmental factors (Lipsitch and Viboud 2009, Tamerius *et al*. 2013). Nonetheless, in the context of pandemic COVID-19, an important still unresolved question of other infectious disease is disease seasonality and environmental dynamics. At present, there is still a lack of studies to address the dynamics underlying epidemics of BD, with particular concern on the role of climatic drivers across different regions of a country or the world. The difference in climate periodicity between tropical and temperate regions makes it more difficult to explain the relationship between the epidemic of BD and environmental correlations. For national or global studies, both a general correlative approach and a two-stage modeling strategy merging the effects of multiple regions might lead to biased associations between climate and BD. Furthermore, climate variables follow certain seasonal patterns and usually covary (Deyle *et al*. 2016a). Therefore, in the non-linear ecosystem, the complicated interdependence among climatic factors brings about a problem, namely how to distinguish their effects and accurately describe the causal driving factors of the ecosystem.

Previous studies on infectious diseases, such as influenza (Yu *et al*. 2013, Deyle *et al*. 2016a) and hand, foot and mouth disease (Xing *et al*. 2014), have tended to focus on a disease trend along latitudes. Although disease trends along longitudes may potentially exist, it seems more important to study the latitudinal trends of diseases, since tropical, and temperate climates are distributed along latitude. As a continent-like country, China has an important latitudinal gradient from the north to the south, and this opportunity may help to study the influence of climatic variables along this trend. Such a context spurred us to develop a biogeographical and macroecological perspective on climatic drivers of human dysentery. In order to overcome the existing challenge, we adopt a well-established empirical dynamic modeling (EDM) (Sugihara *et al*. 2012) to investigate climate-driven patterns of BD in China. The EDM framework has the ability to determine causal drivers, quantify the intensity of causality across latitudes, and approximately estimate functional responses in complex, state-dependent systems. Here, we divided this study into three parts, based on related population data from the 31 provincial capital cities (figure 1) with a relatively long time span from 2010 to 2018. First, we used a quantitative seasonal model to reveal the seasonal diversity and regularity of BD across different latitudes, which suggests the necessity of the EDM methods. Second, we provided a unified explanation for causal climate drivers of BD in China, by examining alternative hypotheses for the national climate drivers of BD in the city-level time series epidemic. Finally, we explored the mechanistic effects of causal climate drivers.
2. Materials and methods

2.1. Study sites
We included a total of 31 provincial capital cities in China in the current study. The investigated cities involve a wide range of latitudes and five different climatic domains. We used Qinling Moutain–Huai River Boundary Line to distinguish northern and southern Chinese cities. The climatic domains of northern cities are mainly mid-temperate and warm-temperate, while the southern cities are mainly tropical and subtropical (figure 1).

2.2. National surveillance programme
The Chinese Center for Disease Control and Prevention (China CDC) receive all possible, clinical- or laboratory-confirmed BD cases, by the National Noticable Infectious Disease Reporting System (NIDR), which is a web-based infection diseases monitor information system (Wang et al 2006). The NIDR system involves all healthcare facilities at the village, town, county, and city levels. The China CDC collects and analyzes the data, and then issues annual reports based on the data (Guo et al 2016).

2.3. Case definitions
We refer to the previous literature (Wang et al 2006) for the definition of the suspected, clinically diagnosed, and laboratory diagnosed cases of BD. We adopted the definition for a suspected case of BD as an acute diarrhea patient with at least one of the following symptoms: abdominal pain, fever, tenesmus, tenderness in the left lower quadrant, and bloody or mucus stool. In addition to the symptoms mentioned above, medical workers confirm a clinically diagnosed case if the results of fecal microscope revealing more than 15 white blood cells and some red blood cells per high-power field ($\times 400$). And we define a laboratory diagnosed case if a Shigella species was isolated from fecal culture.

2.4. BD surveillance data
Based on the statutory procedures for surveillance of infectious diseases, medical worker should report any suspected or confirmed case of BD through this system to the China CDC within 24 h after diagnosis, by use of a standardized form. The standardized form includes basic demographic information (date of birth, gender, current address, and population classification), case classification (suspected...
or confirmed), date of symptoms onset, date of diagnosis, and date of death (if applicable). We collected the data of all suspected and confirmed cases of BD from 31 provincial capital cities in China, covering approximately 9 years (from 1 January 2010 to 31 July 2018). BD cases were aggregated by week, and used for analysis in this study. To calculate the city-specific incidence of BD, we collected the annual population size of each city from the CEIC Data platform (available at www.ceicdata.com/en) and the official website of the Bureau of Statistics in each city. For missing population data, we replace the city’s missing value of the population size with the average population size in its adjacent years. Then, we calculated the approximate week-incidence of BD through dividing the value of weekly cases by the population size of the year.

2.5. Meteorological data

We obtain the meteorological factors from the China Meteorological Data Service Center (CMDC, available at http://data.cma.cn/en), including mean daily temperature (°C), relative humidity (%), rainfall (mm), and sunshine-hour (h), from 1 January 2010 to 31 July 2018 in each provincial capital city in China. Previous studies have demonstrated that these four selected meteorological factors have potential impacts on the three key links in the spread of BD (pathogens, hosts/vectors, and transmission) (Xu et al 2014, Wu et al 2020). The CMDC adopts standard meteorological instruments to measure the meteorological data in each investigated city. Meteorological factors surveillance data covering 70 national baseline monitoring stations located in the local central area were collected from the capital cities. Then, we convert the daily measurements of meteorological factors to the weekly average data for further analysis.

2.6. Statistical analysis

This study included all suspected and confirmed cases of BD onset from 1 January 2010 to 31 July 2018. We used descriptive statistics to summarize the characteristics of BD in 31 provincial capital cities, and stratified data by gender, type of cases, age, population classification, year and season. We performed the Pearson’s correlation coefficient to measure the correlation among climate variables, stratified by region (nationwide, north and south).

2.6.1. Seasonal multiple linear regression model

Because preliminary analyses using a wavelet approach did not well reveal changes in periodicity over time, we used Fourier decomposition to analyze detrended time series to characterize dysentery seasonality (Yu et al 2013). In this step, we standardized weekly time series in each city by dividing the number of cases per week by the annual case counts, to measure week-specific prevalence of the year. First, we created the heat maps based on standardized weekly time series to derive an empirical measure of seasonality and depict city-specific seasonal patterns of BD (Brouitin et al 2010, Yu et al 2013, Xing et al 2014). Then, to further quantify city-specific season characteristic of BD, we fitted a seasonal multiple linear regression model based on Fourier decomposition, which included time trends and harmonic terms on behalf of annual and semiannual periodicity, to the weekly time series of BD of each city (Yu et al 2013). We calculated the annual peaking timing and amplitude of both annual and semiannual periodicity after extracting the estimated model coefficients representing harmonic terms. On the basis that amplitude quantifies the difference between maximum and minimum of the fitted seasonal curves in different periodicity, we estimated the ratio of semiannual amplitude to the sum of annual amplitude and semiannual amplitude to assess relative importance of semiannual cycles. The closer the ratio gets to 0, the more it reveals the annual periodic dominance; otherwise, the closer the ratio gets to 1, the more it reveals the semiannual periodic dominance. Then we explored differences in the distribution of periodicity and annual peaking timing across latitude (Yu et al 2013, Xing et al 2014).

2.6.2. Empirical dynamic modeling (EDM)

The EDM is a quantitative, equation-free and mechanical modeling framework that utilizes observed time series to reconstruct the state space and underly- ing attractor of an unknown dynamic system (Dixon et al 1999, Sugihara et al 2012, Deyle et al 2013, 2016a, Tsonis et al 2015, Ye et al 2015). The process of building a manifold from time series is illustrated in a brief video animation (https://youtu.be/fevurdpiRYg). The state space in the EDM is the multivariate coordinate space in which each axis corresponds to a causally coupled system variable such as temperature, susceptible population size. Thus the state space transforms to a trajectory that traces out the system’s underlying attractor. This can be directly reconstructed from all required state variables; in addition, it can also be reconstructed by substituting lags of a time series. As a complete representation of the system, the dynamic attractor can replace the parametric equation to study complex dynamic systems.

With the use of R package ‘rEDM’, the EDM framework can be applied for several purposes, including detecting causation (Sugihara et al 2012), forecasting future states (Sugihara and May 1990, Sugihara 1994), tracking interactions (Deyle et al 2016b), investigating state-dependent functional responses (Deyle et al 2016a). In this article, we used the convergent cross-mapping (CCM) and the multivariate EDM to identify causal drivers across different latitudes. And we used the scenario exploration to assess the effect of small change in a climatic driver. We applied the Pearson’s correlation (ρ) to measure the EDM prediction skill.
2.6.3. Convergent cross-mapping (CCM)
The CCM is an approach to detect causality in nonlinear dynamic systems. The essential concept of CCM is that regarding prediction between variables as a check for causality (Dixon et al. 1999). For instance, if temperature has a causal effect on BD incidence, then information of temperature will be embedded in the dynamics of BD. Therefore, the attractor produced for the dynamics of BD allows us to predict the states of temperature.

In this study, we used the simplex projection for CCM, which only needs to determine one parameter $E$ (the embedding dimension). Referring to the previous study (Deyle et al. 2016b), we ascertained $E$ based on the optimal prediction of the cross-mapping lag of one week. Then, this $E$ value would be used to measure the unlagged cross-mapping skills. Our concern is to distinguish driving effects from mutual periodic fluctuations (i.e. seasonality) rather than to detecting convergence of cross-mapping skill. The method constructs a seasonal null model with surrogate time series, which can preserve the seasonal signal but randomize the interannual anomalies (Deyle et al. 2016a). We produced an ensemble of 500 null seasonal models by repeating shuffling procedure 500 times, and then compared model performance with the true time series. When the CCM prediction is significantly better for the real time series of climate factors than for the null surrogates, causal coercion is established (Tsonis et al. 2015).

We examined CCM prediction between raw BD incidence and climate drivers (temperature, relative humidity, rainfall and sunshine-hour). The Fisher’s method was applied to assess the meta-significance of the CCM tests of all cities. But it is worth noting that the Fisher’s method depends on the hypothesis that different $P$-values are independent, and the meta-significance may be conservative to a certain extent. Nevertheless, such an approach has been regarded as the best guide in previous work (Deyle et al. 2016a), due to the absence of practical ways to calculate the covariance between $P$-values.

2.6.4. Multivariate EDM: forecast improvement
Multivariate forecast improvement can be applied to detect causality in the case that a driving variable is stochastic (Dixon et al. 1999). This is based on the idea that including stochastic causal variables as a coordinate in the state space can improve the ability of nearest-neighbor prediction. For seasonal BD, climate factors cannot be considered as stochastic sequences, as information about them may already be included in the univariable-embedding (Sugihara et al. 2012, Deyle et al. 2016a). Therefore, we applied a modified method developed by a previous study (Deyle et al. 2016a).

Briefly, the raw BD incidence data was used in this step. We first select the optimal univariate embedded dimension $E'$ for each city; then we measured the improvement in forecasting skills, using simplex projection of the univariate embedding with $E = E' - 1$ and the same embedding, but with climate factor(s) as coordinate. This is because the univariate embedding with $E < E'$ does not contain complete information about the state system, but in this case adding information about a driver to multivariate embedding usually improves prediction skills. Finally, we used one-sample Wilcoxon tests to investigate the statistical significance of the forecast improvement, and paired two-sample Wilcoxon tests to detect the significant differences in improvement between climate factors.

2.6.5. Multivariate EDM: scenario exploration
A multivariate scenario exploration can evaluate the impact of a small perturbation in climate factors on BD incidence, across different system states. We applied the S-maps at each time step $t$ to forecast BD incidence one week later, with using a small increase ($+\Delta X/2$) and a small decrease ($-\Delta X/2$) of the observed value of climate driver $X(t)$. By examining the change in BD dynamics that results from a small change in driving variable, we can use scenario exploration to reveal the local effect that drivers have on the BD incidence. For climate driver $X$, we calculated the following formulas:

$$BD_+(t+1) = f(BD(t), BD(t - 1\ week), . . ., BD(t - (E - 2)\ week), X(t) + \Delta X/2)$$

$$BD_-(t+1) = f(BD(t), BD(t - 1\ week), . . ., BD(t - (E - 2)\ week), X(t) - \Delta X/2)$$

$$\Delta BD(t)/\Delta X = (BD_+(t+1) - BD_-(t+1))/\Delta X$$

In the above formulas, $BD(t+1)$ is a function of $X$ and all other state variables. We used $\Delta BD/\Delta X$ to approximately quantify the effect of $X$ at time $t$. We calculated 5% of the variance for variable $X$ in all cities as the value of $\Delta X$ (Deyle et al. 2016a).

Notably, before scenario exploration, we first normalize the data on incidence, through dividing weekly incidence by the total reports per year in that city (averaged over all years reported). This is key for addressing the substantial differences in reporting rates between cities and putting the results of different cities on a roughly equal footing.

We performed all statistical analyses with R soft-ware (version 3.5.3), in which we applied the package ‘rEDM’ version 0.7.1 for all the EDM analyses. We adopted a $P$-value < 0.05 as statistically significant for all statistical tests.
3. Results

3.1. Demographic characteristics
During the investigated period, the China CDC surveillance system received 516 585 cases of BD from 31 provincial capital cities, of whom 508 324 (98.4%) cases were confirmed, 8261 (1.6%) were suspected, 54.0% were male, and 46.0% were female (table 1). For the age distribution, the high number of BD cases was in the age group 1–10 years 116 441 (22.5%), while the low number of cases was in the age group ≥76 years 20 160 (3.9%). Most BD cases (31.1%) were scattered children who did not attend school or kindergarten, of whom 59.3% were male. The annual number of reported cases has shown a downward trend, of which male cases were consistently higher than female cases in each observed year. Nationally, most cases (42.1%) occurred in the fall, while the number in spring (12.6%) is the lowest one throughout the investigated period.

3.2. Estimates of seasonal characteristics
Heatmaps in figures 2(A) and (B), reveal the empirical seasonal mode and diversity of seasonality across 31 provincial capitals. Nationally, the annual epidemic of BD was concentrated in the summer-fall with annual peak between June and August, which is consistent with the observed time series (figure 2(C)). In general, northern cities have a stronger and clearer annual periodicity than southern cities (figures 2(A) and (B)).

The median value of annual amplitude cycle was 74.0% in the northern cities at higher latitudes, which was significantly higher than the value of 53.1% in the southern cities (two-sample Wilcoxon test, \(T = 198, P\text{-value} < 0.001\)). Further, the estimating seasonality of BD displayed a strong latitudinal gradient, which increased with latitude in both annual amplitude (\(P\text{-value} < 0.001, \text{figure 3(A)}\)) and semi-annual amplitude (\(P\text{-value} < 0.001, \text{figure 3(B)}\)). There was no significant difference in the relative contribution of semi-annual component (\(P\text{-value} = 0.587, \text{figure 3(C)}\)) as well as annual peaking timing (\(P\text{-value} = 0.885, \text{figure 3(D)}\)) across different latitudes in China. The estimations of annual peaking timing were concentrated from the 28 weeks to the 33 week in the year across 31 provincial capitals, ranging from the 26.2 weeks to 35.5 weeks and with the median peak timing in the 30.5 week. The detailed results of the seasonal regression model for each city can be found in figures S1–S4 (available online at stacks.iop.org/ERL/15/124054/mmedia).

3.3. Identifying causal climate drivers
The box-and-whisker plots in figure 4 display the CCM forecast skill (\(\rho_{\text{CCM}}\)) of observed time series, and distributions of \(\rho_{\text{CCM}}\) for null surrogates in the 31 cities. The measured \(\rho_{\text{CCM}}\) of observed time series was drawn in the form of a red solid circle when the value was significantly higher than null distribution (\(P\text{-value} < 0.05\)); otherwise, it was drawn in the form of a red hollow circle. In the cities across different latitudes, relative humidity was a significant climate driver of BD incidence beyond seasonality, with a high meta-significance calculated by Fisher’s method (\(S = 129.3, P\text{-value} < 0.001\)). In addition, there was no meta-significance for temperature (\(S = 74.7, P\text{-value} = 0.13\)), rainfall (\(S = 58.1, P\text{-value} = 0.62\)) and sunshine-hour (\(S = 58.5, P\text{-values} = 0.60\)). These results reveal the only significant causal factor of relative humidity on BD, which is beyond seasonality and not obscured by confounders with periodic signals. However, paired Wilcoxon tests in \(\rho_{\text{CCM}}\) implied that the magnitude of cross-map predictive skill on average temperature was relatively high, compared with the results of relative humidity (\(T = 495, P\text{-value} < 0.001\)), rainfall (\(T = 496, P\text{-value} < 0.001\)) and sunshine-hour (\(T = 496, P\text{-value} < 0.001\)). These findings indicate that the temperature seasonality in the null model is predictive, further supporting the notion that seasonal temperature may be a climatic driver of BD dynamics.

The magnitude of \(\rho_{\text{CCM}}\) can represent the intensity of causal effect when there is no difference in all other conditions. Although the absolute levels of \(\rho_{\text{CCM}}\) for temperature and relative humidity in northern cities were higher comparing with southern cities (two-sample Wilcoxon test; temperature: \(T = 190, P\text{-values} < 0.05\); relative humidity: \(T = 207, P\text{-values} < 0.001\)), we cannot conclude that the causal effects of them were stronger in northern cities. The reason is that different seasonality across different latitudes makes the baseline predictability in these systems different, and ‘no difference in other conditions’ does not apply to different latitudes. On this basis, comparing \(\rho_{\text{CCM}}\) across different latitudes does not apply to measure the causal intensity in this study.

3.4. Causal intensity of suspected causal drivers
Figure 5 summarizes the results of the multivariate EDM method focusing on forecast improvement, which provides an alternative approach to determine the causal intensity of the suspected causal drivers (temperature and relative humidity) on BD. The results of one sample Wilcoxon test demonstrated that temperature was a driver of BD, since including temperature significantly improved predictive skills nationally (\(T = 472, P\text{-value} < 0.001\)). And relative humidity might be potential driver of BD with an approximately significance (\(T = 336, P\text{-value} = 0.09\)). According to the results of paired Wilcoxon test, in southern cities, using temperature and relative humidity together as embedded coordinates leads to a significant improvement compared with using either temperature (\(T = 388, P\text{-values} < 0.05\)) or relative humidity (\(T = 496, P\text{-values} < 0.001\)).
Table 1. Characteristics of bacillary dysentery from 31 provincial capital cities in China, from 1 January 2010 to 31 July 2018.

| Characteristics                          | Male          | Female         | Total         |
|------------------------------------------|---------------|----------------|---------------|
| Number of cases, number (%)              | 278 838 (54.0)| 237 747 (46.0) | 516 585       |
| Type of cases, number (%)                |               |                |               |
| Suspected cases                          | 4284 (51.9)   | 3977 (48.1)    | 8261          |
| Confirmed cases                          | 274 548 (54.0)| 278 838 (46.0) | 508 324       |
| Age groups, number (%)                   |               |                |               |
| 0–12 months                              | 49 032 (58.3) | 35 026 (41.7)  | 84 058        |
| 1–10 years                               | 70 053 (60.2) | 46 388 (39.8)  | 116 441       |
| 11–22 years                              | 34 753 (58.8) | 24 339 (41.2)  | 59 092        |
| 23–34 years                              | 42 868 (51.7) | 39 986 (48.3)  | 82 854        |
| 35–54 years                              | 40 470 (49.3) | 41 618 (50.7)  | 82 088        |
| 55–75 years                              | 31 507 (43.8) | 40 385 (56.2)  | 71 892        |
| 76 years and above                       | 10 155 (50.4) | 10 005 (49.6)  | 20 160        |
| Population classification, number (%)   |               |                |               |
| Students                                 | 39 837 (60.5) | 26 058 (39.5)  | 65 895        |
| Scattered children                       | 95 347 (59.3) | 65 353 (40.7)  | 160 700       |
| Infants                                  | 12 341 (59.4) | 8448 (40.6)    | 20 789        |
| Teachers                                 | 2060 (36.4)   | 3592 (63.6)    | 5652          |
| Medical workers                          | 505 (23.0)    | 1690 (77.0)    | 2195          |
| Cadres                                   | 19 835 (55.7) | 15 765 (44.3)  | 35 600        |
| Laborers and migrant workers             | 22 367 (67.3) | 10 849 (32.7)  | 33 216        |
| Farmers and herders                      | 23 110 (48.1) | 24 888 (51.9)  | 47 998        |
| Soldiers and police                      | 181 (85.0)    | 32 (15.0)      | 213           |
| Sailors and motorists                    | 372 (88.2)    | 80 (11.8)      | 452           |
| Catering industry staffs                 | 1389 (55.6)   | 1107 (44.4)    | 2496          |
| Public service workers                   | 459 (47.1)    | 515 (52.9)     | 974           |
| Business service workers                 | 8000 (50.2)   | 7933 (49.8)    | 15 933        |
| Housework and unemployment               | 19 973 (38.0) | 32 620 (62.0)  | 52 593        |
| Retirees                                 | 21 808 (42.9) | 29 027 (57.1)  | 50 835        |
| Others                                   | 6333 (55.8)   | 5010 (44.2)    | 11 343        |
| Unknown                                  | 4906 (51.4)   | 4642 (48.6)    | 9548          |
| Year of symptoms onset, number (%)       |               |                |               |
| 2010                                     | 49 533 (55.6) | 39 535 (44.4)  | 89 068        |
| 2011                                     | 43 745 (54.8) | 36 077 (45.2)  | 79 822        |
| 2012                                     | 35 814 (55.0) | 29 315 (45.0)  | 65 129        |
| 2013                                     | 31 767 (52.7) | 28 533 (47.3)  | 60 300        |
| 2014                                     | 27 725 (53.9) | 23 683 (46.1)  | 51 408        |
| 2015                                     | 25 171 (53.2) | 22 116 (46.8)  | 47 287        |
| 2016                                     | 23 834 (52.6) | 21 443 (47.4)  | 45 277        |
| 2017                                     | 22 948 (52.6) | 20 703 (47.4)  | 43 651        |
| 2018                                     | 18 301 (52.8) | 16 342 (47.2)  | 34 643        |
| Season of symptoms onset, number (%)     |               |                |               |
| Spring                                   | 34 922 (53.7) | 30 159 (46.3)  | 65 081        |
| Summer                                   | 77 002 (54.1) | 65 357 (45.9)  | 142 359       |
| Fall                                     | 116 603 (53.6)| 100 937 (46.4) | 217 540       |
| Winter                                   | 50 311 (54.9) | 41 294 (45.4)  | 91 605        |

alone. In northern cities, however, no significant difference in forecast improvement between including temperature alone and including both two factors ($T = 88$, $P$-value = 0.12). One possible explanation is that the average correlation coefficient between temperature and relative humidity in the north is higher than that in the south ($r = 0.21$ vs. $r = 0.10$, table 2), thus in the north including both temperature and relative humidity as embedding coordinates contains more identical information and causes a less improvement. Nationally, the multivariate-forecast improvement revealed the possibility of a nonlinear effect of temperature and relative humidity on BD, however, this effect might be concealed in northern region due to high correlation between temperature and relative humidity.

3.5. Driving and interactive mechanism of climatic drivers

The scenario exploration allows us to further investigate the driving and interactive mechanism for temperature and relative humidity, by calculating the rate of change of BD incidence as a function of them across system states. Figure S5 (available online at [https://stacks.iop.org/ERL/15/124054/mmedia](https://stacks.iop.org/ERL/15/124054/mmedia)) presents the city-by-city magnitude under the effect of temperature on BD. Overall, all the 31 provincial capital cities, across a large latitude range, showed
positive effects of temperature on BD. The magnitude, direction, and trajectory over the temperature range of this effect were further explored by plotting scenario exploration results for all cities together and then plotting quantile regression lines. A positive effect ($\Delta BD/\Delta T > 0$) was observed for the temperature nationally, but occasionally a slightly negative effect ($\Delta BD/\Delta T < 0$) is also observed (Figure 6A). The slightly negative effects occurred at high-temperature ranges, reinforced by a lower quantile regression with a small negative slope. Furthermore, an obvious positive median effect of temperature was observed nationally, with a slightly increasing slope between $-25^\circ$C and $16^\circ$C and...
decreasing slope between 16 °C and 35 °C. With the increase of temperature, the gradual increase and subsequent decrease in the rate of change of BD incidence revealed an S-shaped relationship between temperature and BD within the observed temperature range.

However, different from the results of temperature, the same analysis for relative humidity did not present a clear effect on BD. The city-specific effects of relative humidity on BD manifested obvious diversity, suggesting the influence of relative humidity on BD is complex and strongly dependent on specific environmental details (figure S5). Although a small median effect of relative humidity on BD incidence was expected in China, the effect was occasionally much stronger, both positive and negative (figure 6(B)). Based on the scenario exploration of temperature and relative humidity, the influence's magnitude and direction of these climate drivers on the dynamics of BD rely on the system state (non-linear). To investigate whether there is mutual control or interaction between temperature and relative humidity, we focused on the effect of relative humidity on BD ($\Delta$BD/$\Delta$RH) as a function of temperature (figure 6(C)), and the effect of temperature on BD ($\Delta$BD/$\Delta$T) as a function of relative humidity (figure 6(D)). We found that BD incidence was insensitive to relative humidity when the temperature is below 0 °C (figure 6(E)). Conversely, BD incidence was sensitive to relative humidity when the temperature is above 0 °C, which had relatively large positive or negative effects (figure 6(F)).
4. Discussion

Previous studies, based on exposure-response relationships, have generally explained meteorological drivers of BD. However, it is a consensus that these correlation-based methods have limitation of understanding causal and interactive mechanisms in complex nonlinear systems at the population level, and exploring latitudinal trend in dysentery disease population dynamics and seasonal patterns. This nationwide study including 31 provincial capital cities that span a wide range of latitudes involving five climatic domains, may help to fill the gap.

A well-established EDM framework (Deyle et al 2016a) was adopted to address the general problem of identifying climate drivers in the dynamic system of BD. First, both experience-based seasonal patterns and quantitative seasonal models based on Fourier decomposition revealed the seasonal diversity and regularity of BD in the 31 provincial capitals in China. The periodicity and seasonality of BD manifest obvious effects of ‘the North is stronger, the South is weaker’. Second, we used the CCM to detect the effects of temperature (i.e. mostly via its seasonality) and relative humidity in driving BD dynamics, both of them have stronger causal effects than rainfall and sunshine-hour and play an important role at different latitudes throughout the country. Third, we used the multivariate forecast improvement to estimate the causal intensity of temperature, relative humidity, or both of them across different latitudes. And the results suggested that the combined nonlinear effect of temperature and relative humidity on BD may be nationwide. This joint effect was partly concealed by a higher correlation between temperature and relative humidity in northern region, compared with southern region. On this basis, a systematic understanding of the association among temperature, relative humidity and BD requires investigating the effects of their mutual interdependence rather than studying the different factors separately. Finally, the scenario exploration of the EDM indicated an approximately S-shaped relationship between temperature on BD. And the effects of relative humidity on BD are strongly dependent on environmental details, especially at temperatures above 0 °C. The
results indicated that temperature may potentially mediate the effects of relative humidity on BD.

The most common mechanism of dysentery pathogen transmission is the fecal-oral route, with affected hosts consuming contaminated food or water or coming into contact with the feces of infected humans or animals (Fearnley et al. 2011). Transmission is determined by many factors, including human immunity-based conditions (McMichael et al. 2006), extrinsic social, economic, climatic, ecological environment (Weiss and McMichael 2004), and their interaction mechanism. Spurred by previous studies, we inferred that temperature, relative humidity or both affect the incidence of BD by influencing one of the three aspects of disease (pathogens, hosts/vectors, and transmission). Our results are consistent with a previous laboratory study indicating a positive correlation between the proliferation rate of Salmonella and temperature within the range of 7.5 °C–37 °C (Baird-Parker 1994). Many correlation-based studies in the local areas also indicated that temperature is a critical climate factor of BD (Zhang et al. 2012, Li et al. 2013, Zhao et al. 2016, Lee et al. 2017, Wu et al. 2020), although the studies were not causal analyses and their results of exact effects are inconsistent. The seasonal modes of BD pathogens depend on temperature variation which has effects on gene transcription of Shigella (Wei et al. 2017). A warm environment helps improve the fitness of the Shigella pathogen and conducive to their survival (Wei et al. 2017). Besides, temperature variation is also likely to influence food production and human behaviors such as consumption and diet (Wu et al. 2020). Overall, these lead to an approximate S-shaped association between temperature and BD incidence across the whole temperature range in China.

In contrast, relative humidity did not manifest a distinct effect on the BD incidence as temperature. It has been reported that relative humidity influences the replication rates of bacterial and parasitic pathogens, and affects the survival rates of enteroviruses, hosts/vectors, and even human behavior (Huang et al. 2008, Xu et al. 2014). But the specific mechanisms of relative humidity on BD pathogens, hosts/vectors and transmission are far from conclusion. Our finding uncovered the distinct effects of relative humidity between different areas, fitting with a phenomenon in which some single-city studies found positive associations between relative humidity and BD while others found negative associations (Wu et al. 2020). This suggests that the effect of relative humidity on BD strongly dependent on other conditions. For instance, we found that temperature may potentially mediate the effects of relative humidity on BD, with a notable threshold as 0 °C. With the increase in temperature, the BD incidence becomes more and more sensitive to relative humidity. Presently, there are few qualitative investigations of the mechanisms involved in the effect of climate on BD, only two studies have been conducted from the perspective of the Chinese Medicine theory of Yunqi (Ma et al. 2013, Xu and Su 2016). According to this theory, BD occurs as a result of heat and humidity or heat and toxicity. One of the two

Figure 5. Forecast improvement of the multivariate EDM stratified by region. Causal effect is demonstrated if the EDM forecast skill ($\rho$) improves when a climatic variable is included in the EDM. Forecast improvement is quantified by $\Delta \rho = \rho_{\text{with driver}} - \rho_{\text{without driver}}$, where $\rho$ is the Pearson's correlation between observations and the EDM predictions. Including either temperature (T) or relative humidity (RH) or both T and RH leads to forecast improvement in forecast skill in all cities (A), northern cities (B) and southern cities (C).
analyses in Beijing city indicated that the pathogenic factors, including moisture, heat, and toxicity, more easily affect the human body in hot and humid conditions (Ma et al. 2013). However, such a viewpoint does not seem to be a national scale, as our results reveal that negative impacts of relative humidity on BD also occur in hot and humid conditions. The impacts of relative humidity on BD pathogens, hosts/vectors, and transmission, might rely on specific details in local area, such as temperature, socioeconomic factors (which is associated with population access to clean water and food), heterogeneity of population immunity, human dietary lifestyle, pathogen strains' susceptibility to drugs, and so on. The comprehensive influence of several different mechanisms might be responsible for the strongly state-dependent effect of relative humidity on BD. Further studies about this issue are necessary.

The findings of the present study are of practical public health implications. First, based on the quantitative seasonal model and EDM framework, our study revealed a remarkable latitudinal gradient...
in BD disease population dynamics in China. We found that temperature and relative humidity have causal effects on BD dynamics in population, and these climate drivers are important across geographical latitudes. Temperature has a consistently positive impact on BD dynamics, while the direct impact of relative humidity is indistinguishable, influenced by temperature and other environmental conditions. These findings lay potential foundation for laboratory studies that experimentally determine the threshold or range of significant climate drivers corresponding to the survivability and incubation period of BD pathogens. As we found that rising ambient temperature contributes to the spread of dysentery in population, some public health initiatives such as placing cooling equipment in schools and hospitals during hot, wet summers can be implemented and will help reduce the spread of the disease. Furthermore, our results suggest that the number of BD infections among children under 5 years of age is very high, which is consistent with previous evidence that children living in low- and middle-income settings have a large disease burden of BD (Kotloff et al 2018). The outbreaks of BD are usually linked to children in child-care facilities or schools (Scallan et al 2011). Because these sites might be reservoirs of Shigella that spread to the community, where attack rates within households can reach 33% (Mohle-Boetani et al 1995). The population-level findings in this study suggest that some measures such as strictly controlling the number of children’s daycare centers as well as improving environmental conditions during the annual summer vacation may be critical to reducing the incidence of BD.

**Data availability statement**

The data that support the findings of this study are available from the corresponding author upon reasonable request.

**Acknowledgments**

We thank the staff members at the hospitals, local health departments, and county-, district-, and prefecture-level Centers for Disease Control and Prevention for their great assistance in coordinating data collection. We thank the editor and the two anonymous reviewers for their professional suggestions which have greatly improved the manuscript.

**Conflict of interest**

The authors declare that they have no competing interests.

**Funding**

The study was funded by the National Key R&D Program of China (Grant No. 2018YFA0606200). The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**ORCID iD**

Hualiang Lin  [https://orcid.org/0000-0002-3643-9408](https://orcid.org/0000-0002-3643-9408)

**References**

Altizer S, Ostfeld R S, Johnson P T, Kurtz S and Harvell C D 2013 Climate change and infectious diseases: from evidence to a predictive framework. *Science* 341 514–9

Anderson M, Sansonetti P J and Marteyn B S 2016 Shigella diversity and changing landscape: insights for the twenty-first century. *Front. Cell. Infect. Microbiol.* 19 45

Baird-Parker A C 1994 Foods and microbiological risks. *Microbiology* 140 687–95

Brount H, Viboud C, Grenfell B T, Miller M A and Rohani P 2010 Impact of vaccination and birth rate on the epidemiology of pertussis: a comparative study in 64 countries. *Proc. Biol. Sci.* 277 3239–45

Chang Z et al 2016 The changing epidemiology of bacillary dysentery and characteristics of antimicrobial resistance of Shigella isolated in China from 2004–2014. *BMJ Infect. Dis.* 16 685

Chen M J, Lin C Y, Wu Y T, Wu P C, Lung S C and Su H J 2012 Effects of extreme precipitation to the distribution of infectious diseases in Taiwan, 1994–2008. *PloS One* 7 e34651

Cheng J et al 2017 Impacts of ambient temperature on the burden of bacillary dysentery in urban and rural Hefei, China. *Epidemiol. Infect.* 145 1567–76

Deyle E R, Fogarty M, Hsieh C H, Kaufman L, MacCall A D, Munch S B, Perretti C T, Ye H and Sugihara G 2013 Predicting climate effects on Pacific sardine. *Proc. Natl Acad. Sci. USA* 110 6430–5

Deyle E R, Maher M C, Hernandez R D, Basu S and Sugihara G 2016a Global environmental drivers of influenza. *Proc. Natl Acad. Sci. USA* 113 13081–6

Deyle E R, May R M, Munch S B and Sugihara G 2016b Tracking and forecasting ecosystem interactions in real time. *Proc. Biol. Sci.* 283 20152258

Dixon P A, Milicich M J and Sugihara G 1999 Episodic fluctuations in larval supply. *Science* 283 1528–30

Fearney E, Raupach J, Legala F and Cameron S 2011 Salmonella in chicken meat, eggs and humans. *Adelaide, South Australia, 2008 Int. J. Food Microbiol.* 146 219–27

Guo C et al 2016 Short-term effects of meteorological factors on pediatric hand, foot, and mouth disease in Guangdong, China: a multi-city time-series analysis. *BMJ Infect. Dis.* 16 524

Hu X, Ding G, Zhang Y, Liu Q, Jiang B and Ni W 2018 Assessment on the burden of bacillary dysentery associated with floods during 2005–2009 in Zhengzhou City, China, using a time-series analysis. *J. Infect. Public Health* 11 500–6

Huang D, Guan P, Guo J, Wang P and Zhou B 2008 Investigating the effects of climate variations on bacillary dysentery incidence in northeast China using ridge regression and hierarchical cluster analysis. *BMJ Infect. Dis.* 8 130
Khalil I A et al 2018 Morbidity and mortality due to shigella and enterotoxigenic Escherichia coli diarrhoea: the global burden of disease study 1990–2016 Lancet Infect. Dis. 18 1229–40
Kotloff K L et al 1999 Global burden of Shigella infections: implications for vaccine development and implementation of control strategies Bull. World Health Organ. 77 651–66
Kotloff K L, Rudall M S, Platts-Mills J A, Pavlinac P and Zaidi A K M 2018 Shigellosis Lancet 391 801–12
Lee H S, Ha Hoang T T, Pham-Duc P, Lee M, Grace D, Phung D C, Thuc V M and Nguyen-Viet H 2017 Seasonal and geographical distribution of bacillary dysentery (shigellosis) and associated climate risk factors in Kon Tam Province in Vietnam from 1999 to 2013 Infect. Dis. Poverty 6 113
Li Z J, Zhang X J, Hou X X, Xu S, Zhang J S, Song H B and Lin H L. 2015 Nonlinear and threshold of the association between meteorological factors and bacillary dysentery in Beijing, China Epidemiol. Infect. 143 3510–9
Li Z et al 2013 Identifying high-risk areas of bacillary dysentery and associated meteorological factors in Wuhan, China Sci. Rep. 3 3239
Lipsitch M and Viboud C 2009 Influenza seasonality: lifting the fog Proc. Natl Acad. Sci. USA 106 3645–6
Ma S L, Tang Q L, Liu H W, He J and Gao S H 2013 Correlation analysis for the attack of bacillary dysentery and meteorological factors based on the Chinese medicine theory of Yunji and the medical-meteorological forecast model Chin. J. Integr. Med. 19 182–6
Mara D and Horan N J 2003 Handbook of Water and Wastewater Microbiology (Amsterdam: Elsevier)
Matther A E, Baker K S, McGregor H, Coupland P, Mather P L, Deheer-Graham A, Parkhill J, Bracegirdle P, Russell J E and Thomson N R 2014 Bacillary dysentery from World War 1 Med. Microbiol. (Amsterdam: Elsevier)
McMichael A J 2013 Globalization, climate change, and human health New Engl. J. Med. 369 96
McMichael A J, Woodruff R E and Hales S 2006 Climate change and human health: present and future risks Lancet 367 859–69
Mohle-Boetani J C, Stapleton M, Finger R, Bean N H, Poundstone J, Blake P A and Griffen P M 1995 Communitywide shigellosis: control of an outbreak and risk factors in child day-care centers Am. J. Public Health 85 812–6
Njamkepo E et al 2016 Global phylogeography and evolutionary history of Shigella dysenteriae type 1 Nat. Microbiol. 1 16027
Scallan E, Griffin P M, Angulo F J, Tauxe R V and Hoekstra R M 2011 Foodborne illness acquired in the United States—unspecified agents Emerg. Infect. Dis. 17 16–22
 Sugihara G 1994 Nonlinear forecasting for the classification of natural time series Phil. Trans. R. Soc. A 348 477–95
Sugihara G and May R M 1990 Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series Nature 344 734–41
Sugihara G, May R, Ye H, Hsieh C H, Deyle E, Fogarty M and Munch S 2012 Detecting causality in complex ecosystems Science 338 496–500
Tamerius J D, Shaman J, Alonso W J, Bloom-Feshbach K, Uejo C K, Comrie A and Viboud C 2013 Environmental predictors of seasonal influenza epidemics across temperate and tropical climates PLoS Pathog. 9 e1003194
Tsonis A A, Deyle E R, May R M, Sugihara G, Swanson K, Verbeni J D and Wang G 2015 Dynamical evidence for causality between galactic cosmic rays and interannual variation in global temperature Proc. Natl Acad. Sci. USA 112 3253–6
von Seidlein L et al 2006 A multicentre study of Shigella diarrhoea in six Asian countries: disease burden, clinical manifestations, and microbiology PLoS Med. 3 e353
Wang X Y et al 2006 Trend and disease burden of bacillary dysentery in China (1991–2000) Bull. World Health Organ. 84 561–8
Wei Y, Kouse A B and Murphy E R 2017 Transcriptional and posttranscriptional regulation of Shigella shut in response to host-associated iron availability and temperature Microbiologiy 6 e00942
Weiss R A and McMichael A J 2004 Social and environmental risk factors in the emergence of infectious diseases Nat. Med. 10 S70–6
Wu X, Liu J, Li C and Yin J 2020 Impact of climate change on dysentery: scientific evidences, uncertainty, modeling and projections Sci. Total Environ. 714 136702
Xiao X, Gasparrini A, Huang J, Liao Q, Liu F, Yin F, Yu H and Li X 2017 The exposure-response relationship between temperature and childhood hand, foot and mouth disease: a multi-site study from mainland China Environ. Int. 100 102–9
Xing W et al 2014 Hand, foot, and mouth disease in China, 2008–12: an epidemiological study Lancet Infect. Dis. 14 308–18
Xu F and Su Y 2016 Relationship research of the incidence of dysentery of Qian Gorlos area of Jilin Province with meteorological factors based on the five-Yun and six-qi theories Chin. J. Basic Med. Tradit. Chin. Med. 22 1635–7
Xu Z, Hu W, Zhang Y, Wang X, Tong S and Zhou M 2014 Spatiotemporal pattern of bacillary dysentery in China from 1990 to 2009: what is the driver behind? PLoS One 9 e104329
Ye H, Beamish R J, Glaser S M, Grant S C, Hsieh C H, Richards I J, Schnute J T and Sugihara G 2015 Equation-free mechanistic ecosystem forecasting using empirical dynamic modeling Proc. Natl Acad. Sci. USA 112 E1569–76
Yu H, Alonso W J, Feng L, Tan Y, Shu Y, Yang W and Viboud C 2013 Characterization of regional influenza seasonality patterns in China and implications for vaccination strategies: spatio-temporal modeling of surveillance data PLoS Med. 10 e1001552
Zhang Y, Bi P and Hillier J E 2008 Weather and the transmission of bacillary dysentery in Jilin, northern China: a time-series analysis Public Health Rep. 123 61–66
Zhang Y, Bi P, Sun Y and Hillier J E 2012 Projected years lost due to disabilities (YLDs) for bacillary dysentery related to increased temperature in temperate and subtropical cities of China J. Environ. Monit. 14 510–6
Zhao Q et al 2018 Modeling the present and future incidence of pediatric hand, foot, and mouth disease associated with ambient temperature in Mainland China Environ. Health Perspect. 126 047010
Zhao Y, Zhu Y, Zhu Z and Qu B 2016 Association between meteorological factors and bacillary dysentery incidence in Chaoyang city, China: an ecological study BMJ Open 6 e013376