



OPEN

Meteorological and social conditions contribute to infectious diarrhea in China

Xiang Yang¹, Weifeng Xiong¹, Tianyao Huang² & Juan He¹✉

Infectious diarrhea in China showed a significant pattern. Many researchers have tried to reveal the drivers, yet usually only meteorological factors were taken into consideration. Furthermore, the diarrheal data they analyzed were incomplete and the algorithms they exploited were inefficient of adapting realistic relationships. Here, we investigate the impacts of meteorological and social factors on the number of infectious diarrhea cases in China. A machine learning algorithm called the Random Forest is utilized. Our results demonstrate that nearly half of infectious diarrhea occurred among children under 5 years old. Generally speaking, increasing temperature or relative humidity leads to increased cases of infectious diarrhea in China. Nevertheless, people from different age groups or different regions own different sensitivities to meteorological factors. The weight of feces that are harmfully treated could be a possible reason for infectious diarrhea of the elderly as well as children under 5 years old. These findings indicate that infectious diarrhea prevention for children under 5 years old remains a primary task in China. Personalized prevention countermeasures ought to be provided to different age groups and different regions. It is essential to bring the weight of feces that are harmfully treated to the forefront when considering infectious diarrhea prevention.

Infectious diarrhea is a preventable disease yet remains a major public health problem around the world¹. According to World Health Organization, diarrheal disease accounts for approximately 1.8 million deaths each year and is the second leading cause of death in children under five years old, which is responsible for killing around 0.525 million children every year. Developing countries suffer much more badly from infectious diarrhea^{2,3}. China is a middle-income country with vast territory and varied climates. The Chinese government provides heavily subsidized healthcare as well as wide public health education, yet still endures significant diarrhea burden. China ranks among the top 15 countries on the number of diarrheal morbidity⁴. In China, the infectious diarrhea cases account for nearly 20% of those of 45 notifiable infectious diseases during the last 10 years⁵. Despite the high burden, our ability to predict and prepare for diarrheal case surges remains limited^{6,7}.

Researches have found out that the incidence of infectious diarrhea has a very close relationship with seasonal change for decades^{8–12}, in particular with meteorological conditions^{13,14}. In general, various studies done in different parts of the world suggest that increasing temperature or relative humidity leads to increasing infectious diarrhea cases^{15–24}, only with a few exceptions^{25,26}. Extreme weather events such as floods or El Nino warm events greatly trigger infectious diarrhea in various localities^{14,21,27–31}. For instance, higher maximum temperature and more sunshine in the pre-monsoon period are found to have a tendency to enhance the first peak of the diarrhea occurrence in tropical area³². Climate was considered to dominate high- and low-bacterial enteric diseases periods in Vietnam³³. Higher ambient temperature decreases and higher rainfall increases diarrheal risk in rural sites of south India³⁴. Etiological and meteorological factors had age-specific effects on the prevalence of infectious diarrhea in Jiangsu province of China³⁵. La Nina conditions lagged 0–5 months are associated with higher than average incidence of under-age-5 diarrhea in the early rainy season in Botswana³⁶.

In addition to the meteorological conditions, other studies illustrate that some social conditions also contribute to incidence of infectious diarrhea^{24,37–40}. Those social factors can be classified into 4 categories. The first one is the resource of public health, like medical and technical personnel⁴¹, access of hygienic conditions^{42–44} and nutrition status^{45–49}. The second one is the safety of water, in particular, the quality and quantity of clean water available for cooking, drinking and washing^{42,50–53}. The third one is the fecal contamination, like the septic system densities⁵¹, access to private sanitation⁵⁴. The fourth one is the local forest coverage^{24,55}. For instance, it has been found that medical and technical personnel per thousand persons were significantly negatively related to the risk of bacterial dysentery in Sichuan province of China⁴¹. Use of river water for cooking and cleaning

¹Beijing University of Chinese Medicine, No. 11, Bei San Huan Dong Lu, Chaoyang District, Beijing 100029, China. ²Tsinghua University, Haidian District, Beijing 100084, China. ✉email: hejuan6428@sina.com

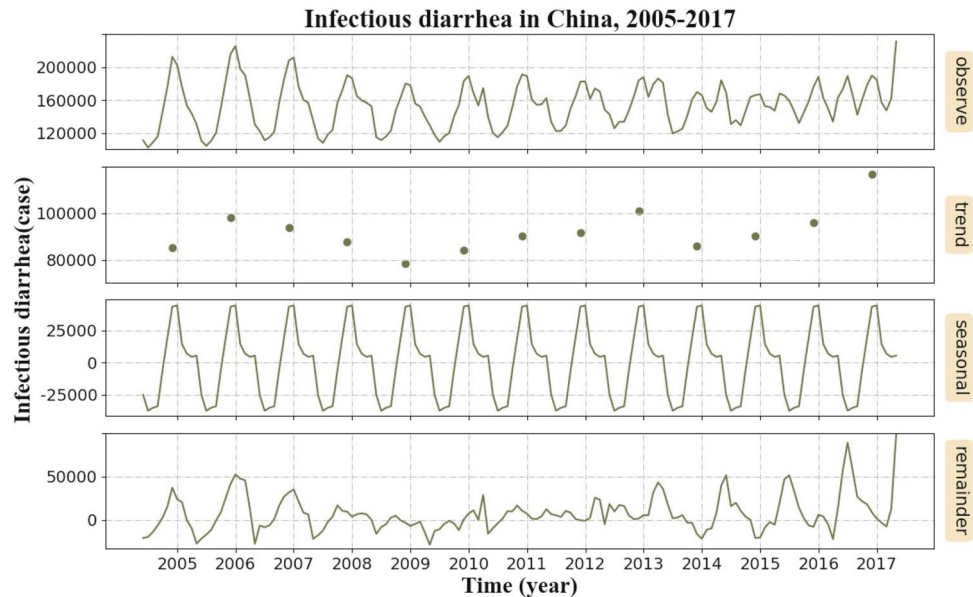


Figure 1. Monthly observed cases of infectious diarrhea in mainland of China, 2005–2017. Note: From top to bottom, the lines represent actually observed cases, trend components, seasonal components, and random components, respectively.

was highly associated with risk of typhoid fever in Blantyre, Malawi⁵³. Septic system densities were associated with infectious diarrhea in central Wisconsin⁵¹. Local forest coverage may mitigate the effects of extreme heat on cholera outbreaks in Bangladesh²⁴.

While the impacts of potential factors on diarrheal incidence have been well documented, existing researches generally suffer from three drawbacks. First, those researches mainly focused on only one kind of infectious diarrhea⁵⁶. Second, social factors that might contribute to statistical incidence of infectious diarrhea have not been put on enough attention^{24,38,39}. Third, with regard to the fact that the inter-relations among the predominant factors of infectious diarrhea are quite complicated in the real world, traditional statistics methods could be unsuitable for discovering these complex relationships^{22,57–60}.

In this study, we explore the link between monthly/yearly variability and all sorts of infectious diarrhea during 2005–2017 in China. The potential factors that we take into consideration are meteorological and social factors. The method for data analysis is a well established algorithm in the field of machine learning. The infectious diarrhea cases that have been reported to Chinese Center for Disease Control and Prevention (China CDC), are grouped into three classes: Class A (cholera); Class B (bacillary dysentery, typhoid and paratyphoid); and Class C (other infectious diarrhea). We apply some meteorological factors that are the most commonly used factors^{11,12}, including precipitation, wind speed, temperature, vapor pressure, and relative humidity.

Based on existing researches, we apply 4 categories of social factors to explore the linkage between social factors and infectious diarrhea. The factors we choose are the representative ones for each kind. In specific, we choose the quantity of centers for disease control and prevention (CDCP), the quantity of health supervision institute (HSI) and the quantity of health technicians (HT) to represent the resource of public health. As water pollution is usually the main threat to the safety of water⁶¹, the three major measuring elements of wastewater discharge are utilized to represent the safety of water. Those three elements are the amounts of wastewater discharge (WD), chemical oxygen demand emissions (CODE) and nitrogen output (NO). The weight of feces that are harmfully treated (FHT) is selected to represent the severity of fecal contamination. And the forest coverage rate (FCR) is used to represent the local forest coverage. The data we use in this study are all publicly available.

Results

Epidemiology of infectious diarrhea. A total of 14,396,560 cases of infectious diarrhea were reported during 2005–2017 in mainland of China. Cases aged under 5 years accounted for 46.78% (6,733,533/14,396,560) of all cases.

The overall situation of infectious diarrhea in mainland of China is shown in Fig. 1. The first pane in the figure is the monthly observed number of infected cases during 2005–2017. The second pane is the average number of infected diarrhea cases of the whole year, which represents the infectious diarrhea trend with time. The third pane considers the average number of infected cases of the specific month during the observed 13 years, which is applied to represent the seasonal pattern of infectious diarrhea. These monthly averages are then normalized by subtracting the mean of these averages. The fourth pane is the remained data, which are obtained by the observed number minus both the yearly and monthly averaged numbers. The corresponding month of each year labeled in the horizontal ordinate is July. The cases indicated a long-term increasing trend, which is shown in the second pane. In the third pane, a peak from May to August and a nadir from November to February are observed annually.

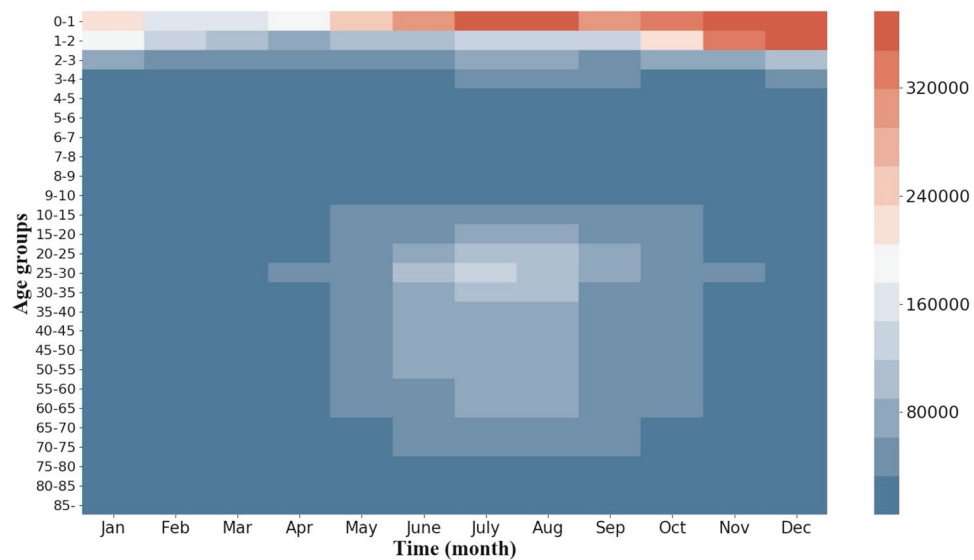


Figure 2. Seasonal and age-specific numbers of infectious diarrhea cases in mainland of China, 2005–2017. (See Supplementary Table 1 in the Supplementary material to obtain the detailed data). The heat map is created by package seaborn in Python⁶².



Figure 3. The yearly averaged numbers of infectious diarrhea cases in each province of mainland of China during 2005–2017. The map is generated by an open source package called pyechart in Python⁶³.

The seasonal, age-specific numbers of infectious diarrhea cases are demonstrated in Fig. 2. From this figure, we observe that the population clustered into three age groups: under 5, 5–10, and over 10 years age. The incidence of infectious diarrhea in the group with age under 5 years old increased mainly in the early autumn and winter, while the incidence in the group with ages over 10 years old increased mainly in summer. In comparison, the incidence of infectious diarrhea in the 5–10 age group stayed quite stable and was less sensitive to season change.

The provincial situations about yearly infectious diarrhea cases are demonstrated in Figs. 3, 4, 5. The yearly averaged number of infectious diarrhea cases in mainland of China during 2005–2017 (shown in Fig. 3) illustrates significant differences between provinces. The top 4 provinces with the most infected cases are Guangdong, Zhejiang, Anhui and Shandong provinces. Figure 4 provides provincial numbers of cases in four concrete years (2005, 2009, 2013 and 2017), indicating that the number of infectious diarrhea cases increased in most of the provinces during 2005–2017. The primary cases lied in southeast coastal provinces when considering the increase

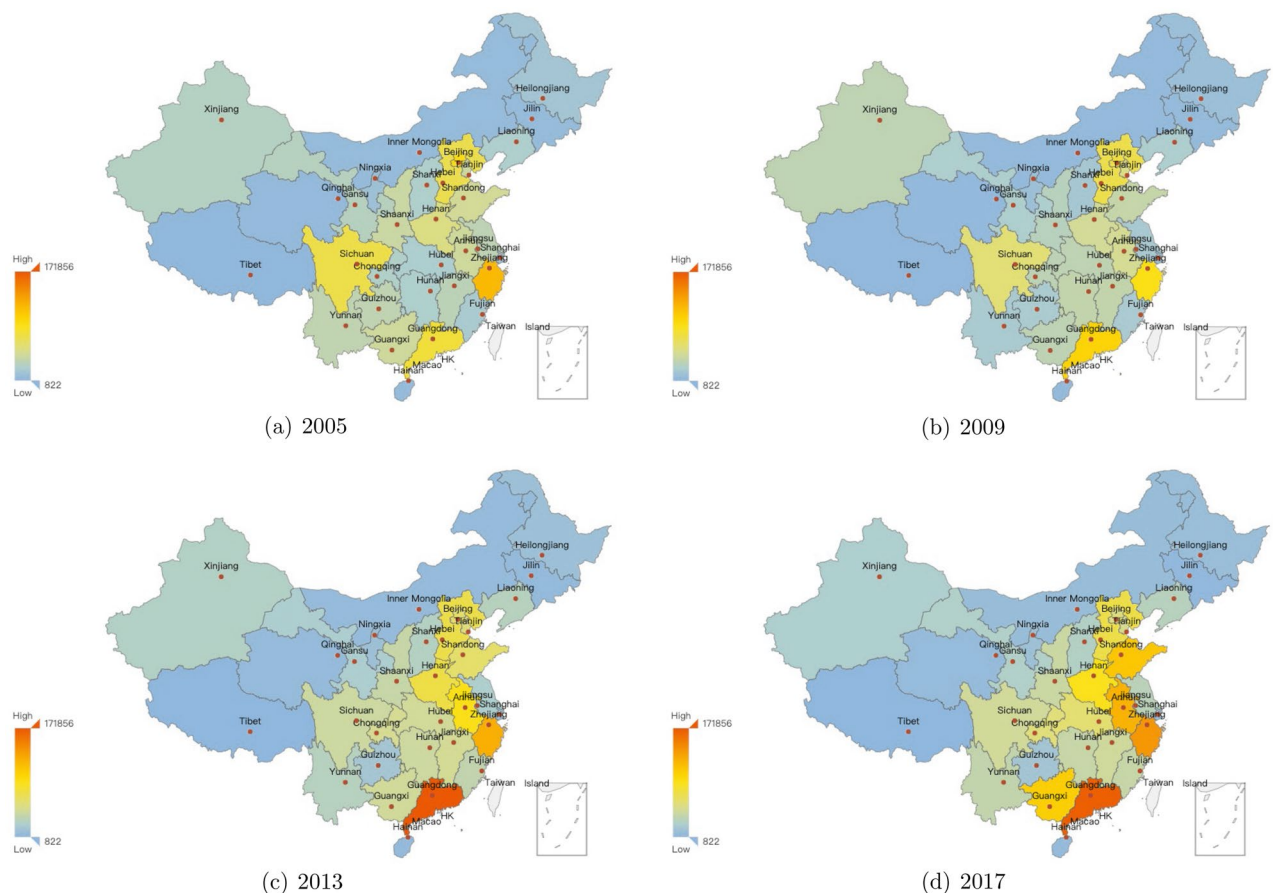


Figure 4. Provincial numbers of infectious diarrhea cases in mainland of China, in four different years during 2005–2017. The map is generated by an open source package called `pyechart` in Python⁶³.

of all age groups. Figure 5 focuses on two age groups: under and older than 5 years old, and demonstrate the provincial number of infectious cases in 2005 and 2017. The increased cases of under 5 years old are majorly distributed in Guangdong and Zhejiang provinces. For the age group over 5 years old, the increased cases are mainly distributed in Anhui province. Only among the over-5 age group we observe the reduced number of infectious diarrhea cases, which happened in Beijing and Zhejiang province.

The influence of social factors with regard to different age groups. The social factors we take into consideration are: the quantity of centers for disease control and prevention (CDCP), the quantity of health supervision institute (HSI), the quantity of health technicians (HT), the amount of wastewater discharge (WD), chemical oxygen demand emissions (CODE), nitrogen output (NO), the weight of feces that are harmfully treated (FHT), and the forest coverage rate (FCR). The yearly data of social factors as well as the number of infectious cases in different age groups are illustrated later in Table 1. The influence of social factors on the number of infectious diarrhea cases in different age groups is shown in Fig. 6 and we leave the detailed data in Supplementary Table 2 of the Supplementary material.

Generally speaking, social factors have different influence on infectious diarrhea cases of different age groups in mainland of China. Some age groups have principal influential factors while other age groups do not. For adolescents with age between 10 and 25, HT is the dominant factor. For middle aged people with years between 45 and 60, NO seems to affect those group more. For elderly aged over 75, WD matters the most.

From the perspective of social factor, CDCP, HSI, HT and WD affect the number of infectious diarrhea cases in almost the whole age groups. FCR has more significant impact on ages between 5 and 40 than other age groups. FHT primarily strikes the old and young. CODE has a main contribution in age groups of 45–50 and 55–60. NO has the most clear influence to people between 70 and 75 years old.

Predicting future numbers of infectious diarrhea cases by meteorological and social factors. In this article, we adopt meteorological factors and social factors to predict future number of infectious diarrhea cases and get reasonable results. The monthly data of infectious diarrhea cases in mainland of China during 2005–2015, which were divided by age groups and provinces, are adopted as training data. The rest data of 2016–2017 are adopted as prediction test data.

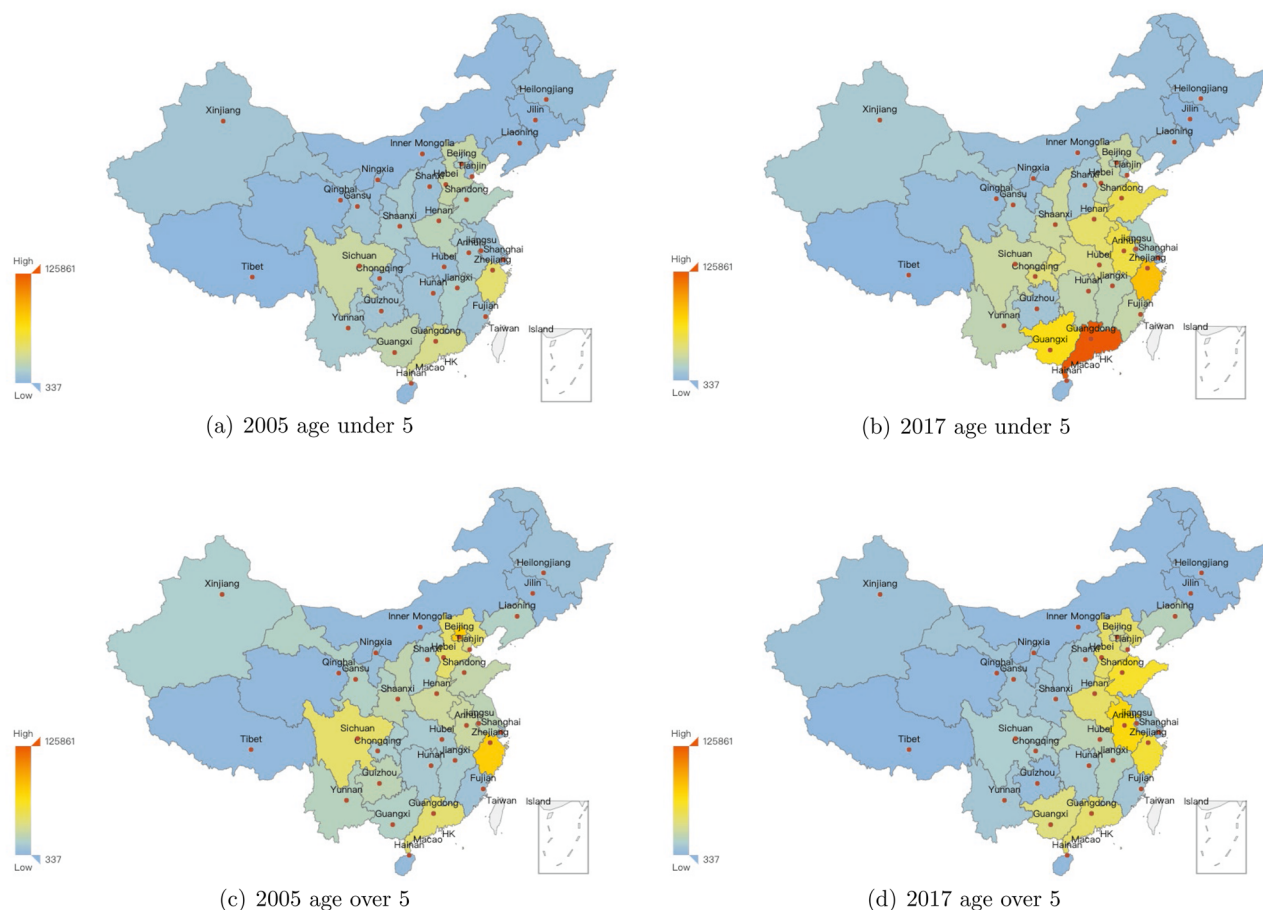


Figure 5. Provincial numbers of infectious diarrhea cases with regard to different age groups in mainland of China, in 2005 and 2017. The age groups are divided into two: under and over 5 years. The map is generated by an open source package called pycchart in Python⁶³.

Year	# of cases		CDCP ^a	HSI ^b	HT ^c	POFC ^d	FHT ^e	TWD ^f	CODE ^g	NO ^h
	age < 5	age >= 5								
2005	366,976	656,625	3585	1702	456.41	18.2	1090.8	5,245,089	1414.2	149.78
2006	417,873	759,072	3548	2097	472.84	18.2	1275.3	5,144,802	1414.2	149.78
2007	449,620	676,076	3585	2553	491.32	18.2	1598	5568494.16	1381.8	132.34
2008	473,458	580,909	3534	2675	517.45	20.4	1426.8	5,716,801	1320.7	126.97
2009	415,755	523,964	3536	2809	553.51	21.6	1294.9	5890877.25	1277.5	122.61
2010	479,839	528,847	3513	2992	587.62	21.6	1259.7	6,172,562	1238.1	120.29
2011	507,217	576,101	3484	3022	620.29	21.6	1310.1	6591922.44	2499.86	260.44
2012	521,157	580,842	3490	3088	667.55	21.6	1010.4	6847612.14	2424	253.59
2013	640,027	572,002	3516	2967	721.06	21.6	1004.6	6954432.7	2352.7	245.66
2014	517,576	514,027	3490	2975	758.98	21.6	860	7161750.53	2294.6	238.53
2015	555,034	530,361	3478	2986	800.75	21.6	763.1	7353226.83	2223.5	229.91
2016	587,744	562,750	3481	2986	845.44	21.6	652.1	7110953.88	1046.53	141.78
2017	801,257	601,451	3456	2992	898.82	21.6	540	6996609.97	1021.97	139.51

Table 1. Yearly data of number of infectious diarrhea cases in two age groups, as well as the social factors. ^aThe quantity of centers for disease control and prevention. ^bThe quantity of health supervision institute. ^cThe quantity of health technicians (10,000). ^dThe forest coverage rate (%). ^eThe weight of fecal harmful treatment (10,000 tons). ^fThe amounts of wastewater discharge (10,000 tons). ^gChemical oxygen demand emissions (10,000 tons). ^hNitrogen output (10000 tons).

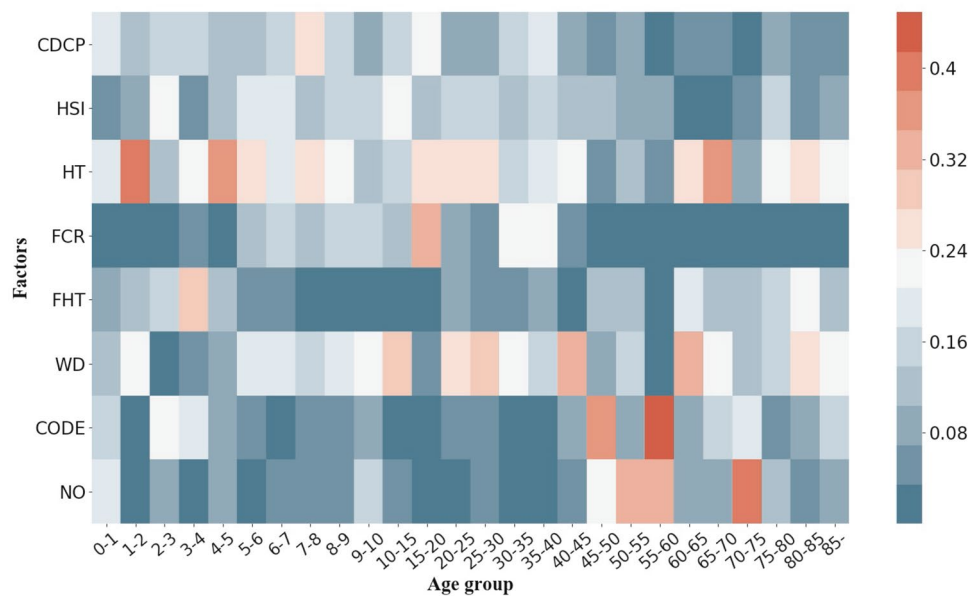


Figure 6. Heat maps of influence of social features on the number of infectious diarrhea cases in different age groups in mainland of China, from 2005 to 2017. (See Supplementary Table 2 in the Supplementary material to obtain the detailed data). The heat map is created by package seaborn in Python⁶².

When predicting future infectious diarrhea cases, we use an existing package called scikit-learn in Python to implement Random Forest regression. For each province in mainland of China, we divided the diarrheal patients into 26 age groups (0–1, 1–2, 2–3, 3–4, 4–5, 5–6, 6–7, 7–8, 8–9, 9–10, 10–15, 15–20, 20–25, 25–30, 30–35, 35–40, 40–45, 45–50, 50–55, 55–60, 60–65, 65–70, 70–75, 75–80, 80–85, 85–). For each age–province combination, we apply the `cross_val_score` function in scikit-learn to tune the parameters by 5-fold cross validation. The parameters being tuned are the number of trees in the forest (tuning scale: 50–300, tuning step:10) and the maximum depth (tuning scale: 5–30, tuning step:2), while other parameters are set to the default values (see Supplementary Table 3 in Supplementary material to obtain the tuned parameters for each age–province combination and Supplementary Table 4 to obtain the exhaustive list of other parameters). After the parameter tuning, we run 10 times of train–test implementation and obtain the mean error and an indication of the variance around that mean (see Supplementary Table 5 in Supplementary material to obtain the detailed data).

We apply a method called nMAE as predicting evaluation, which is a normalized metric, defined by⁶⁴:

$$\text{nMAE} = \frac{\text{MAE}}{\bar{x}}. \quad (1)$$

Here, \bar{x} is the average value of x , and MAE is the mean absolute error, which is defined by:

$$\text{MAE} = \frac{1}{N} \times \sum_{i=1}^N |\text{predicted}_i - \text{true}_i|. \quad (2)$$

Approximately, the average numbers of predicted infectious diarrhea cases coincide with those of the real situation in twelve months, which are April to October, December of 2016 and July to September, November of 2017 (shown in Fig. 7). In those twelve months, the difference between predicted data and real data are less than 10%. There are large gaps between predicted data and real data in seven months, which are January, February and November of 2016, and January to March, December of 2017. In those seven months, the difference between the two groups of data are more than 30% (see Supplementary Table 6 in Supplementary material to obtain the detailed data). The best predicting result (nMAE = 0.088) and the worst predicting result (nMAE = 2.65) are shown in Fig. 8). The detailed data about the worst and best predictions are given by Supplementary Tables 7 and 8 respectively in Supplementary material.

The top four provinces with the most infectious diarrhea cases are further observed (shown in Fig. 9). Compared with social factors, the meteorological factors contribute to diarrhea morbidity with greater impacts. The vapor pressure is the crucial feature and mainly affects children with ages between 2 and 9 in Guangdong province. The temperature has a primary influence on middle aged people with ages between 40–65 in Zhejiang province. Meanwhile, the junior are more vulnerable to vapor pressure while the senior are more sensitive to temperature in Anhui and Shandong provinces. Although the social factors have a weaker connection with diarrhea, some factor still has strong influence on some age groups. The quantity of health technicians has considerable effect on people with ages between 25–35 in Guangdong province. Generally speaking, the quantity of

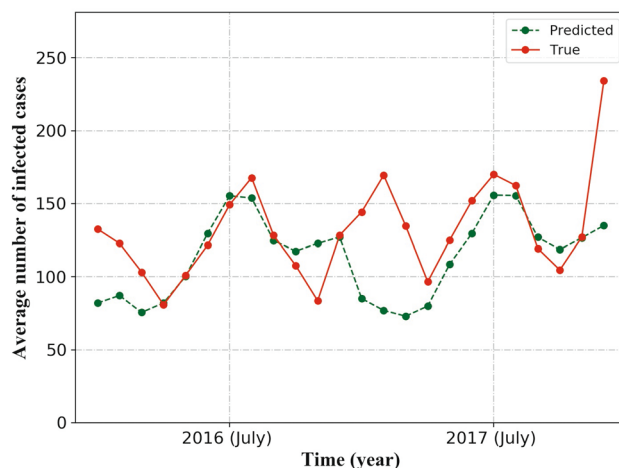
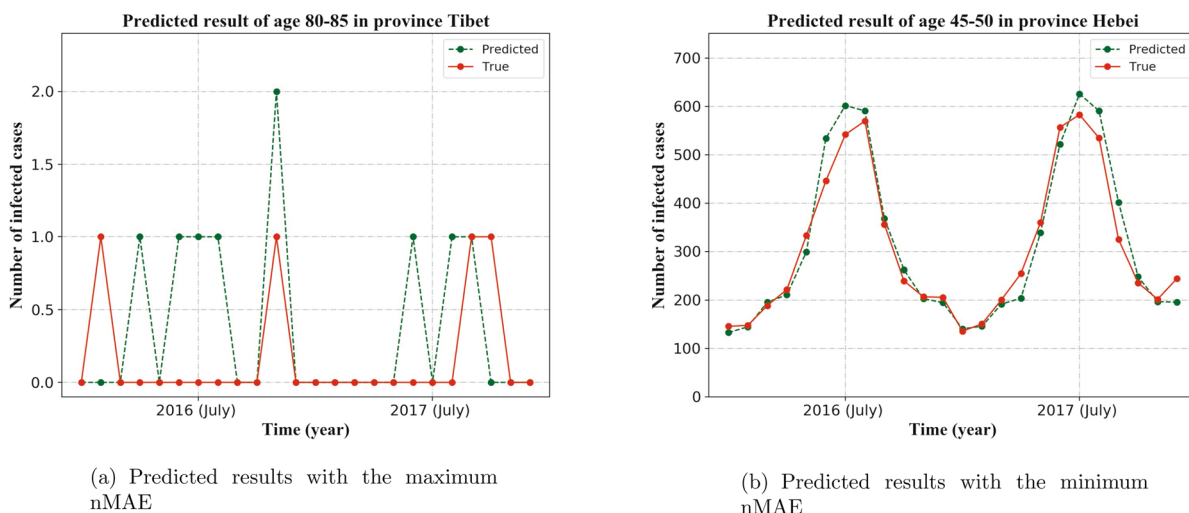


Figure 7. The comparison between average number of cases in the predicted data and the counterparts in the true data. The red line stands for true value while the green line stands for predicted value. (See Supplementary Table 6 in Supplementary material to obtain the detailed data).



(a) Predicted results with the maximum nMAE

(b) Predicted results with the minimum nMAE

Figure 8. Demonstration of worst and best predictions. (See Supplementary Tables 7, 8 in Supplementary material to obtain the detailed data). The predicted result is measured by nMAE.

health technicians and the amounts of wastewater discharge affect people of 5–45 years old in Zhejiang province. The number of elderly people infected by diarrhea in Anhui province is connected with the quantities of health technicians, health supervision institutes and the amount of wastewater discharge. The social factors seem to contribute much less in Shandong province.

Discussion

Our observations illustrate the vulnerability of children under 5 years old to infectious diarrhea. This group not only accounts for almost half of infectious diarrhea cases, but also is much more sensitive to pernicious exposure. In addition, our findings confirm the distinctive seasonal pattern of infectious diarrhea cases in the mainland of China while demonstrate the association between climate variation and epidemiological morbidity. In specific, we take different age groups and provinces into consideration to illustrate the change of diarrheal morbidity over time. Meteorological and social factors have been applied for prediction of number of diarrheal cases. The effects of respective factors are analyzed in different provinces, respectively. Furthermore, the influence of those factors on different age groups are evaluated, which underscore the potential use of those factors on future diversified public health management.

Infectious diarrhea prevention for children under 5 years old. Infectious diarrhea is quite common in infancy and early childhood⁶⁵. Children's diarrhea is usually caused by a viral infection⁶⁶ and is likely to

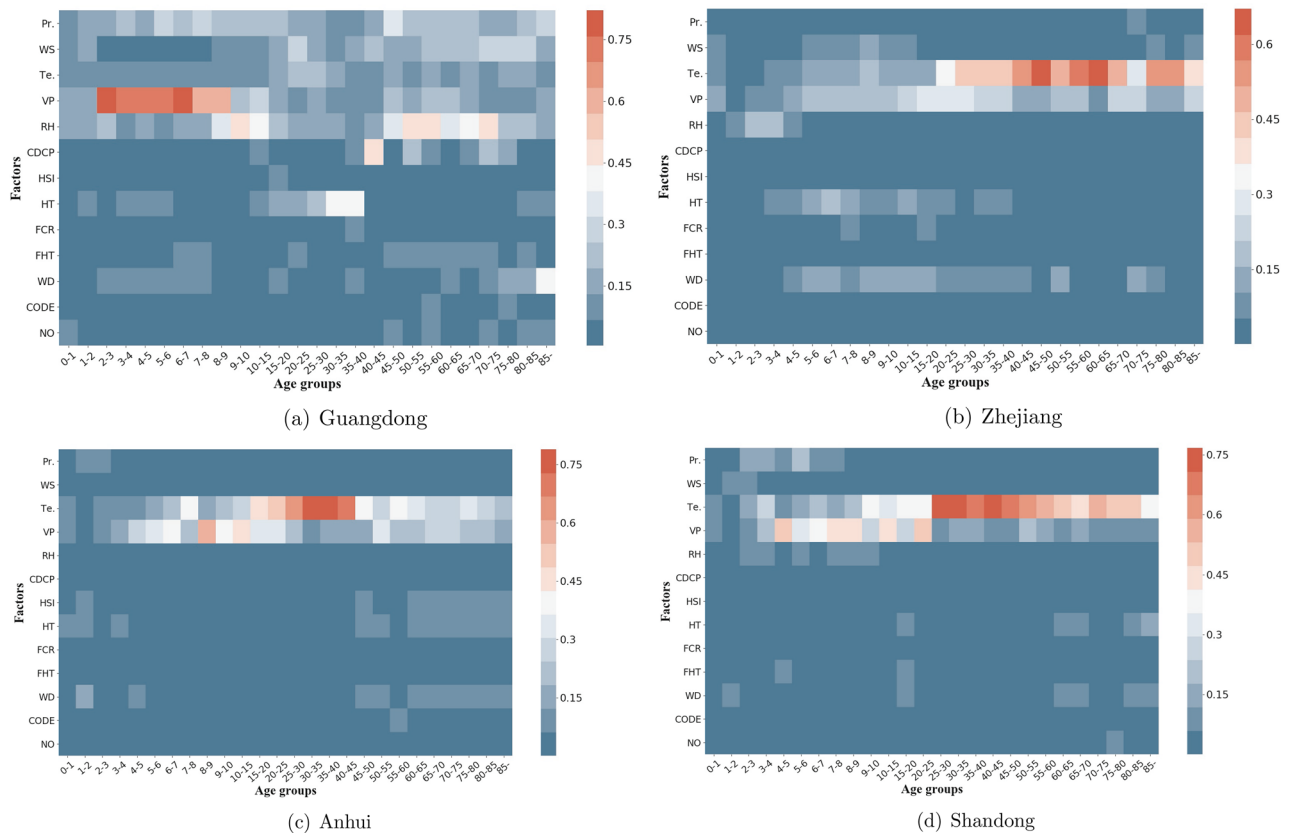


Figure 9. Heat maps of influence of potential factors on number of infected diarrhea cases of different age groups in the top four provinces with most infectious cases in mainland of China, from 2005 to 2017. (See Supplementary Tables 9, 10, 11, 12 in Supplementary material to obtain the detailed data). The heat map is created by package seaborn in Python⁶². Meteorological factors Pr.: precipitation, WS: wind speed, Te.: temperature, VP: vapor pressure, RH: relative humidity Social factors CDCP: the quantity of centers for disease control and prevention, HSI: the quantity of health supervision institute, HT: the quantity of health technicians, WD: the amounts of wastewater discharge, CODE: chemical oxygen demand emissions, NO: nitrogen output, FHT: the weight of feces that are harmfully treated, FCR: forest coverage rate.

damage their growth, electrolyte, and nutrient balance⁶⁷. More seriously, episodes of diarrhea may predispose to pneumonia in undernourished children⁶⁸, which triggers further damage to children's health.

Numerous countermeasures have been proven to effectively reduce the high burden of infectious diarrhea in children. This includes rotavirus vaccine⁶⁹, improvement of zinc nutrition⁷⁰, lactobacillus therapy⁷¹, and dietary counseling⁷².

Our study suggests that raising the number of health technicians and improving the way of treating feces could be effective for reducing the number of children infected with diarrhea, which also agrees with some existing researches. For instance, Stephen et al.⁷³ found out that participant of field workers at promoting hand washing among neighborhoods reduces the incidence of infectious diarrhea. Jai et al.⁷⁴ showed that insecticide spraying may reduce diarrhea in children in a region where there were clear connection with fly numbers and associated diarrhea. Although the authors did not further investigate the origin of flies, it seems reasonable to recognize the relationship between the appearance of flies and the way how the feces are treated.

Meteorological factors and their impacts on infectious diarrhea. Our observation shows that the total case of infectious diarrhea mainly increase from May to September, when the temperature and the relative humidity of the whole country increase significantly.

When considering climate-related diarrhea among different sub-population, some researches suggest that children and older adults be the most susceptible to diarrhea caused by climatic variation⁷⁵. Both the intensity and frequency of climate changes have significant effects on childhood diarrhea⁷⁶. Furthermore, the climate effect on infectious diarrhea for men, teenagers, and young adults (10–29 years) were higher than those for other populations⁷⁷. Our study replenish those researches. We found that for children under 5 years old, the infectious diarrhea morbidity is apparently distinct among different seasons. And the number of young adults (20–35 years) with infectious diarrhea raises evidently in summer and early autumn, which may be caused by the lifestyle of youths but not confirmed.

Not only the divergent sub-population, but also the divergent sub-locations show a distinctive relevance with the climate-related diarrhea. For instance, a study in Wuhan (the capital of Hubei province) suggests that

the central districts have high risk compared to other areas¹³. In our study, the various locations are categorized by province. Our finding reveals that coastal provinces in central and eastern China have higher numbers of infectious diarrhea cases, among which Guangdong province has the highest number of infectious diarrhea. The high morbidity of infectious diarrhea in those coastal provinces may be explained by relatively frequent contact of seafoods, but this needs further exploration.

The conjunction between infectious diarrhea and meteorological conditions may be explained by pathogens activity. Infectious diarrhea can be easily caused by microorganisms, which are closely dependent on the environmental situation⁷⁸. Among those microorganisms, rotavirus, calicivirus, enteropathogenic and enterotoxigenic *E. coli* cause more than half of all diarrheal deaths in children under 5 years in the world^{79–81}. The host, which those microorganisms parasitize, is directly or indirectly influenced by meteorological conditions. Microorganisms vary in the sensitivity to alterations in temperature, humidity, oxygen, light and nutrients^{82–84}, which leads to different survival and reproduction rates outside of the host. Subsequently, the appropriate temperature triggers expression of the related genes inside a host^{85–87}, which results in a successful infection of a host. Variations in rainfall and temperature also affect the possibility of fecal contamination^{10,87}, which is an immediate cause of infectious diarrhea^{48,88}. Heavy rainfalls and floods alter human exposure patterns and lead to population displacement, with a variety of resultant health impacts^{87,89,90}.

Since there exists a distinct connection between meteorological conditions and infectious diarrhea, it seems natural that meteorological factors can be applied for infectious diarrhea prediction. As a matter of fact, lots of statistic models have been utilized for infectious diarrhea prediction, such as linear regression model^{57,91}, multiple regression model⁹², poisson regression model^{21,75,93} and spatial panel regression model⁹⁴. However, traditional statistic models encounter inherent flaws when approximating complicated correlation. In order to fit the inter-relation between meteorological conditions and infectious diarrhea in a better way, we exploit the Random Forest Algorithm in our study. The Random Forest Algorithm is an “ensemble learning” algorithm, which was first proposed in 2001 by Breiman. It was then widely applied in multiple domains and was considered one of the most powerful machine learning algorithms^{95–98}. As a relatively new algorithm, it has been seldom employed in the field of public health. Our study shows the great potential of the Random Forest Algorithm in predicting the case number of infectious diarrhea.

Social factors and their impacts on infectious diarrhea. The social factors that we investigated are classified into 4 categories. They are resource of public health, safety of water, fecal contamination and local forest coverage. As shown in Fig. 6, the resource of public health and safety of water play more important roles in affecting diarrhea, compared to fecal contamination and local forest coverage.

The resource of public health, which are respectively the quantity of centers for disease control and prevention (CDCP), health supervision institute (HSI) and health technicians (HT), has obvious influence on the whole age groups. The positive influence of those factors may owe to their improvement of sanitation and hygiene, as infectious diarrhea could be easily caused by contaminated food^{43,99–101}, bad sanitation environment⁵⁴ and poor health situation^{45,46}. Those risk factors of infectious diarrhea can be mitigated by health intervene, like a good habit of hand washing^{54,102}, appropriate food interventions^{47,102}, a relatively rapid diagnose/solution¹⁰³ and more medical technicians⁴¹. Increasing the quantity of CDCP, HSI and HT provide more health education for public and more professional treatments of patients^{103,104}. Thus more resource of public health results in more effective diarrhea prevention.

As shown in Fig. 6, compared to other kinds of social factors, the safety of water has considerable impact on incidence of diarrhea, especially to the elderly. The major concern about the water safety is water pollution⁶¹. Existing researches demonstrate that not only the polluted water for drinking, but also the polluted water for using can bring infectious diarrhea^{42,52,53}. The water pollution may trigger diarrhea through pathogens contamination and cause physical function degeneration. A review study illustrates that the most common pollutants of water are pathogens (bacteria, viruses and protozoas)⁶¹. Those pathogens could lead to various kinds of infectious diarrhea. On the other hand, water pollution injury is likely to cause diseases like rheumatic, vascular diseases, myocardial infarct and nervous system damage, especially to the elderly¹⁰⁵. Those diseases could make the elderly more vulnerable to diarrhea¹⁰⁶.

The fecal contamination is represented by the weight of feces that are harmfully treated. The fecal contamination may result in diarrhea through pathogens contamination directly or indirectly. For instance, there are studies showing that the fecal contamination is associated with viral and bacterial infection^{51,54}. Some researches demonstrate that the fecal contamination lead to water pollution^{22,47}, which may further bring about diarrhea infection.

In contrast to other social factors, forest coverage has less impact on infectious diarrhea. The relationship between forest coverage and diarrhea may result from the the influence on environment from forest. For instance, there is research illustrating that forest contributes to diarrhea prevention by improving water quality⁵⁵. Another study shows that heatwaves might promote the occurrence of cholera, but this connection could be modified by forest coverage²⁴.

Our study shows that different social factors have different effects on specific age groups. Those findings shed new lights to infectious diarrhea prevention and may promote more flexible and efficient disease control measurement. For instance, our research indicates that the number of health technicians plays a more important role to children under 5 years old, youth with age between 15 and 30, and the elderly over 60 years old. A connection has been observed between the weight of feces that are harmfully treated and monthly numbers of infectious diarrhea cases of the old as well as children under 5 years old. Based on those findings, health technicians should put much more attention to those two age groups when considering infectious diarrhea prevention, which may bring about more accurate disease control as well as rational and effective management of human resource. Meanwhile, the weight of feces that are harmfully treated should be brought to the forefront.

Figure 9 further demonstrates that the impact of social factors differ in different provinces. For instance, the social factors have a strong connection with diarrhea in Guangdong province than in Shandong province. The regional difference may result from socio-demographic variousness¹⁰⁷. Although the mechanism remains unclear, it seems significant that geographical locations should be put on more attention when considering infectious diarrhea¹⁰⁸.

Figure 9 also shows that social factors contribute to diarrhea morbidity, but not in a dominant way. This conclusion coincides with some existing researches^{13,37}. Furthermore, some study reveals that the impact social factors have on diarrhea remains unstable¹⁰⁷. Further investigations are needed to explore major mechanisms underlying the association between infectious diarrhea and certain social factors.

Methods

Diarrhea data. In China, infectious diarrhea (excluding cholera, dysentery, typhoid, and paratyphoid) is an intestinal infectious disease with diarrhea and/or vomiting as the main symptom. Monthly data on infectious diarrhea during 2005–2017 in mainland of China were collected from the Nationwide Notifiable Infectious Diseases Reporting Information System (NIDRIS)¹⁰⁹. The information included time, age, and date of onset. The data are publicly available.

Meteorological data. Monthly climate data were obtained from the National Meteorological Information Center of China (<http://data.cma.cn/>), including precipitation, wind speed, temperature, vapor pressure, and relative humidity. The data are publicly available.

Social factors data. Monthly social factors data were obtained from the Chinese National Bureau of Statistics (<http://www.stats.gov.cn>), containing the quantity of centers for disease control and prevention, the quantity of health supervision institute, the quantity of health technicians, the amounts of wastewater discharge, chemical oxygen demand emissions, nitrogen output, the weight of Feces that are harmfully treated and the forest coverage rate. The data are publicly available.

Data analysis. In our study, the Random Forest Algorithm was applied for data analysis, whose standard procedure is repeated as follows¹¹⁰:

1. Draw n_{tree} bootstrap samples from the original data.
2. For each of the bootstrap samples, grow an unpruned classification or regression tree, with the following modification: at each node, rather than choosing the best split among all predictors, randomly sample m_{try} of the predictors and choose the best split from among those variables. (Bagging can be thought of as the special case of random forests obtained when $m_{try} = p$, the number of predictors.)
3. Predict new data by aggregating the predictions of the n_{tree} trees (i.e., majority votes for classification, average for regression).

An estimate of the error rate can be obtained, based on the training data, by the following:

1. At each bootstrap iteration, predict the data not in the bootstrap sample (what Breiman calls “out-of-bag”, or OOB, data) using the tree grown with the bootstrap sample.
2. Aggregate the OOB predictions. (On the average, each data point would be out-of-bag around 36% of the times, so aggregate these predictions.) Calculate the error rate, and call it the OOB estimate of error rate.

Received: 30 March 2021; Accepted: 20 October 2021

Published online: 03 December 2021

References

1. Khalil, I. A. *et al.* Morbidity and mortality due to shigella and enterotoxigenic Escherichia coli diarrhoea: The Global Burden of Disease Study 1990–2016. *Lancet Infect. Dis.* **18**, 1229–1240 (2018).
2. Swerdlow, D. L. *et al.* Waterborne transmission of epidemic cholera in Trujillo, Peru: Lessons for a continent at risk. *Lancet* **340**, 28–32 (1992).
3. Troeger, C. *et al.* Estimates of global, regional, and national morbidity, mortality, and aetiologies of diarrhoeal diseases: A systematic analysis for the Global Burden of Disease Study 2015. *Lancet Infect. Dis.* **17**, 909–948 (2017).
4. Fischer Walker, C. L. *et al.* Global burden of childhood pneumonia and diarrhoea. *Lancet* **381**, 1405–1416 (2013).
5. Yang, S. *et al.* Epidemiological features of and changes in incidence of infectious diseases in China in the first decade after the SARS outbreak: an observational trend study. *Lancet Infect. Dis.* **17**, 716–725 (2017).
6. Lipp, E. K., Huq, A. & Colwell, R. R. Effects of global climate on infectious disease: The cholera model. *Clin. Microbiol. Rev.* **15**, 757–770 (2002).
7. Wu, X., Lu, Y., Zhou, S., Chen, L. & Xu, B. Impact of climate change on human infectious diseases: Empirical evidence and human adaptation. *Environ. Int.* **86**, 14–23 (2016).
8. Cook, S. M., Glass, R. I., Lebaron, C. W. & Ho, M. S. Global seasonality of rotavirus infections. *Bull. World Health Org.* **68**, 171–177 (1990).
9. Karen, L., Hubbard, A. E. & Eisenberg, J. N. Seasonality of rotavirus disease in the tropics: A systematic review and meta-analysis. *Int. J. Epidemiol.* **6**, 1487–1496 (2008).

10. Alexander, K. A., Carzolio, M., Goodin, D. & Vance, E. Climate change is likely to worsen the public health threat of diarrheal disease in Botswana. *Int. J. Environ. Res. Public Health* **10**, 1202–1230 (2013).
11. Pascual, M., Rodo, X., Ellner, S. P., Colwell, R. & Bouma, M. J. Cholera dynamics and El Niño-Southern Oscillation. *Science* **289**, 1766–1769 (2000).
12. Onozuka, D. Effect of non-stationary climate on infectious gastroenteritis transmission in Japan. *Sci. Rep.* **4**, 1–6 (2014).
13. Li, Z. *et al.* Identifying high-risk areas of bacillary dysentery and associated meteorological factors in Wuhan, China. *Sci. Rep.* **3**, 1–6 (2013).
14. Ni, W., Ding, G., Li, Y., Li, H. & Jiang, B. Impacts of floods on dysentery in Xinxiang city, China, during 2004–2010: A time-series Poisson analysis. *Global Health Act.* **7**, 23904 (2014).
15. Zhou, X. *et al.* High temperature as a risk factor for infectious diarrhea in Shanghai, China. *J. Epidemiol.* **23**, 418–423 (2013).
16. Liu, Z. *et al.* Daily temperature and bacillary dysentery: Estimated effects, attributable risks, and future disease burden in 316 Chinese cities. *Environ. Health Perspect.* **128**, 057008 (2020).
17. Zhang, Y., Bi, P., Hiller, J. E., Sun, Y. & Ryan, P. Climate variations and bacillary dysentery in northern and southern cities of China. *J. Infect.* **55**, 194–200 (2007).
18. Liu, Z. *et al.* Effect of ambient temperature and its effect modifiers on bacillary dysentery in Jinan, China. *Sci. Total Environ.* **650**, 2980–2986 (2019).
19. Hao, Y. *et al.* Effects of ambient temperature on bacillary dysentery: A multi-city analysis in Anhui Province, China. *Sci. Total Environ.* **671**, 1206–1213 (2019).
20. Yi, L. *et al.* The impact of climate variability on infectious disease transmission in China: Current knowledge and further directions. *Environ. Res.* **173**, 255–261 (2019).
21. Singh, R. B. *et al.* The influence of climate variation and change on diarrheal disease in the Pacific Islands. *Environ. Health Perspect.* **109**, 155–159 (2001).
22. Camacho, A. *et al.* Cholera epidemic in Yemen, 2016–18: An analysis of surveillance data. *Lancet Global Health* **6**, e680–e690 (2018).
23. Lobitz, B. *et al.* Climate and infectious disease: Use of remote sensing for detection of *Vibrio cholerae* by indirect measurement. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 1438–1443 (2000).
24. Wu, J., Yunus, M., Ali, M., Escamilla, V. & Emch, M. Influences of heatwave, rainfall, and tree cover on cholera in Bangladesh. *Environ. Int.* **120**, 304–311 (2018).
25. Jagai, J. S. *et al.* Seasonality of rotavirus in south asia: A meta-analysis approach assessing associations with temperature, precipitation, and vegetation index. *PLoS ONE* **7**, e38168 (2012).
26. Onozuka, D., Gasparrini, A., Sera, F., Hashizume, M. & Honda, Y. Modeling future projections of temperature-related excess morbidity due to infectious gastroenteritis under climate change conditions in Japan. *Environ. Health Perspect.* **127**, 077006 (2019).
27. Nkoko, D. B. *et al.* Dynamics of cholera outbreaks in great Lakes region of Africa, 1978–2008. *Emerg. Infect. Dis.* **17**, 2026–2034 (2011).
28. Liu, Z. D. *et al.* Distributed lag effects and vulnerable groups of floods on bacillary dysentery in Huaihua, China. *Sci. Rep.* **6**, 1–8 (2016).
29. Liu, Z. *et al.* Identifying different types of flood-sensitive diarrheal diseases from 2006 to 2010 in Guangxi, China. *Environ. Res.* **170**, 359–365 (2019).
30. Zhang, N. *et al.* The impact of the flood event in Anhui Province, China on infectious diarrhea disease: An interrupted time-series study. *Environ. Int.* **127**, 801–809 (2016).
31. Gong, L. *et al.* Short-term effects of moderate and severe floods on infectious diarrheal diseases in Anhui Province, China. *Sci. Total Environ.* **675**, 420–428 (2019).
32. Teshima, A., Yamada, M., Hayashi, T., Wagatsuma, Y., Terao, T. Climate impact on seasonal patterns of diarrhea diseases in Tropical area, Technical Report.
33. Kelly-Hope, L. A. *et al.* Temporal trends and climatic factors associated with bacterial enteric diseases in Vietnam, 1991–2001. *Environ. Health Perspect.* **116**, 7–12 (2008).
34. Kulinkina, A. V. *et al.* Seasonality of water quality and diarrheal disease counts in urban and rural settings in south India. *Sci. Rep.* **6**, 1–12 (2016).
35. Fang, X. *et al.* Epidemiology of infectious diarrhoea and the relationship with etiological and meteorological factors in Jiangsu Province, China. *Sci. Rep.* **9**, 1–9 (2019).
36. Heaney, A. K., Shaman, J. & Alexander, K. A. El Niño-Southern oscillation and under-5 diarrhea in Botswana. *Nat. Commun.* **10**, 1–9 (2019).
37. Zhang, S. X. *et al.* Impact of co-infections with enteric pathogens on children suffering from acute diarrhea in southwest China. *Infect. Dis. Poverty* **5**, 64 (2016).
38. Patel, M. M. *et al.* Global seasonality of rotavirus disease. *Pediatric Infect. Dis. J.* **32**, e134–47 (2013).
39. Mellor, J. E. *et al.* Planning for climate change: The need for mechanistic systems-based approaches to study climate change impacts on diarrheal diseases. *Sci. Total Environ.* **548**, 82–90 (2016).
40. Zhang, H., Si, Y., Wang, X. & Gong, P. Patterns of bacillary dysentery in China, 2005–2010. *Int. J. Environ. Res. Public Health* **13**, 164 (2016).
41. Ma, Y., Zhang, T., Liu, L., Lv, Q. & Yin, F. Spatio-temporal pattern and socio-economic factors of bacillary dysentery at county level in Sichuan Province, China. *Sci. Rep.* **5**, 1–9 (2015).
42. Shahinian, M. L. *et al.* Of cabbages and chlorine: Cholera in Peru. *Lancet* **340**, 20–21 (1992).
43. Vollaard, A. M. *et al.* Risk factors for typhoid and paratyphoid fever in Jakarta, Indonesia. *J. Am. Med. Assoc.* **291**, 2607–2615 (2004).
44. Dagnew, A. B. *et al.* Prevalence of diarrhea and associated factors among under-five children in Bahir Dar city, Northwest Ethiopia, 2016: A cross-sectional study. *BMC Infect. Dis.* **19**, 1–7 (2019).
45. Gunn, R. A. *et al.* Bottle feeding as a risk factor for cholera in infants. *Lancet* **314**, 730–732 (1979).
46. Shahinian, M. L. *et al.* *Helicobacter pylori* and epidemic *Vibrio cholerae* 01 infection in Peru. *Lancet* **355**, 377–378 (2000).
47. Kur, L. *et al.* Cholera outbreak-southern Sudan, 2007. *MMWR Morb. Mortal Wkly. Rep.* **58**, 337–341 (2009).
48. The, H. C. *et al.* Assessing gut microbiota perturbations during the early phase of infectious diarrhea in Vietnamese children. *Gut Microbes* **9**, 38–54 (2018).
49. Grimwood, K. & Forbes, D. A. Acute and persistent diarrhea. *Pediatric Clin. N. Am.* **56**, 1343–1361 (2009).
50. Ts, S., Nm, T. & Rl, G. Protozoal agents: What are the dangers for the public water supply?. *Ann. Rev. Med.* **48**, 329–340 (1997).
51. Borchardt, M. A., Chyou, P. H., DeVries, E. O. & Belongia, E. A. Septic system density and infectious diarrhea in a defined population of children. *Environ. Health Perspect.* **111**, 742–748 (2003).
52. Birmingham, M. E. *et al.* Epidemic cholera in Burundi: Patterns of transmission in the Great Rift Valley Lake region. *Lancet* **349**, 981–985 (1997).
53. Gauld, J. S. *et al.* Domestic river water use and risk of typhoid fever: Results from a case-control study in Blantyre, Malawi. *Clin. Infect. Dis.* **70**, 1278–1284 (2020).

54. Baker, K. K. Sanitation and hygiene-specific risk factors for moderate-to-severe diarrhea in young children in the global enteric multicenter study, 2007–2011: Case-Control Study. *PLoS Med.* **13**, e1002010 (2016).
55. Rasolofson, R. A., Ricketts, T. H., Johnson, K. B., Jacob, A. & Fisher, B. Forests moderate the effectiveness of water treatment at reducing childhood diarrhea. *Environ. Res. Lett.* **16**, 064035 (2021).
56. Xu, Z. *et al.* Exploration of diarrhoea seasonality and its drivers in China. *Sci. Rep.* **5**, 1–9 (2015).
57. Kolstad, E. W. & Johansson, K. A. Uncertainties Associated with quantifying climate change impacts on human health: A case study for Diarrhea. *Environ. Health Perspect.* **119**, 299–305 (2011).
58. Li, Z. J. *et al.* Nonlinear and threshold of the association between meteorological factors and bacillary dysentery in Beijing, China. *Epidemiol. Infect.* **143**, 3510–3519 (2015).
59. Wang, Y., Li, J., Gu, J., Zhou, Z. & Wang, Z. Artificial neural networks for infectious diarrhea prediction using meteorological factors in Shanghai (China). *Appl. Soft Comput. J.* **35**, 280–290 (2015).
60. Moore, S. M. *et al.* El Niño and the shifting geography of cholera in Africa. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 4436–4441 (2017).
61. Azizullah, A., Khattak, M. N. K., Richter, P. & Häder, D. P. Water pollution in Pakistan and its impact on public health: A review. *Environ. Int.* **37**, 479–497 (2011).
62. Waskom, M. L. seaborn: Statistical data visualization. *J. Open Source Softw.* **6**, 3021 (2021).
63. chenjiandongx. <https://github.com/pyecharts/pyecharts>.
64. Benali, L., Notton, G., Fouilloy, A., Voyant, C. & Dizene, R. Solar radiation forecasting using artificial neural network and random forest methods: Application to normal beam, horizontal diffuse and global components. *Renew. Energy* **132**, 871–884 (2019).
65. Navaneethan, U. & Giannella, R. A. Mechanisms of infectious diarrhea. *Nat. Clin. Pract. Gastroenterol. Hepatol.* **5**, 637–647 (2008).
66. Pérez-Gaxiola, G., Cuello-García, C. A., Florez, I. D., Pérez-Pico, V. M., Smectite for acute infectious diarrhoea in children. (2018).
67. Thiagarajah, J. R. *et al.* Advances in evaluation of chronic diarrhea in infants. *Gastroenterology* **154**, 2045–2059 (2018).
68. Schlaudecker, E. P., Steinhoff, M. C. & Moore, S. R. Interactions of diarrhea, pneumonia, and malnutrition in childhood: Recent evidence from developing countries. *Curr. Opin. Infect. Dis.* **24**, 496 (2011).
69. Hoshino, Y. & Kapikian, A. Z. Rotavirus vaccine development for the prevention of severe diarrhea in infants and young children. *Trends Microbiol.* **2**, 242–249 (1994).
70. Black, R. E. Therapeutic and preventive effects of zinc on serious childhood infectious diseases in developing countries. *Am. J. Clin. Nutr.* **68**, 476S–479S (1998).
71. Van Niel, C. W., Feudtner, C., Garrison, M. M. & Christakis, D. A. Lactobacillus therapy for acute infectious diarrhea in children: A meta-analysis. *Pediatrics* **109**, 678–684 (2002).
72. Stauffer, W. M., Konop, R. J. & Kamat, D. *Traveling with Infants and Young Children.* (Travelers' diarrhea, 2002).
73. Luby, S. P. *et al.* Effect of intensive handwashing promotion on childhood diarrhea in high-risk communities in Pakistan: A randomized controlled trial. *J. Am. Med. Assoc.* **291**, 2547–2554 (2004).
74. Das, J. K., Hadi, Y. B., Salam, R. A., Hoda, M., Lassi, Z. S., Bhutta, Z. A. Fly control to prevent diarrhoea in children (2018).
75. Chou, W. C. *et al.* Modeling the impact of climate variability on diarrhea-associated diseases in Taiwan (1996–2007). *Sci. Total Environ.* **409**, 43–51 (2010).
76. Wu, J., Yunus, M., Streatfield, P. K. & Emch, M. Association of climate variability and childhood diarrhoeal disease in rural Bangladesh, 2000–2006. *Epidemiol. Infect.* **142**, 1859–1868 (2014).
77. Wang, H. *et al.* Association of meteorological factors with infectious diarrhea incidence in Guangzhou, southern China: A time-series study (2006–2017). *Sci. Total Environ.* **672**, 7–15 (2019).
78. Shope, R. Global climate change and infectious diseases. *Environ. Health Perspect.* **96**, 171–174 (1991).
79. Duan, Z. *et al.* Hospital-Based Surveillance of Rotavirus Diarrhea in the People's Republic of China, August 2003–July 2007. *J. Infect. Dis.* **200**, S167–S173 (2009).
80. Lanata, C. F. *et al.* Global causes of diarrheal disease mortality in children <5 years of age: A systematic review. *PLoS ONE* **8**, e72788 (2013).
81. ...Yu, J. *et al.* Etiology of diarrhea among children under the age five in China: Results from a five-year surveillance. *J. Infect.* **71**, 19–27 (2015).
82. Murphy, C., Carroll, C. & Jordan, K. N. Environmental survival mechanisms of the foodborne pathogen *Campylobacter jejuni*. *J. Appl. Microbiol.* **100**, 623–632 (2006).
83. Rzezutka, A. & Cook, N. Survival of human enteric viruses in the environment and food. *FEMS Microbiol. Rev.* **28**, 441–453 (2004).
84. Polgreen, P. M. & Polgreen, E. L. Infectious diseases, weather, and climate. *Clin. Infect. Dis.* **66**, 815–817 (2017).
85. Harvell, C. D. *et al.* Climate warming and disease risks for terrestrial and marine biota. *Science* **296**, 2158–2162 (2002).
86. Shapiro, R. S. & Cowen, L. E. Thermal control of microbial development and virulence: Molecular mechanisms of microbial temperature sensing. *MBio* **3**, e00238-12 (2012).
87. Levy, K., Woster, A. P., Goldstein, R. S. & Carlton, E. J. Untangling the impacts of climate change on waterborne diseases: A systematic review of relationships between diarrheal diseases and temperature flooding, and drought, rainfall. *Environ. Sci. Technol.* **50**, 4905–4922 (2016).
88. Lauerma, J. Weathering diarrheal illness effects of El Niño in the South Pacific. *Environ. Health Perspect.* **109**, A84–A85 (2001).
89. Ahern, M., Kovats, R. S., Wilkinson, P., Few, R. & Matthies, F. Global health impacts of floods: Epidemiologic evidence. *Epidemiol. Rev.* **27**, 36–46 (2005).
90. Alderman, K., Turner, L. R. & Tong, S. Floods and human health: A systematic review. *Environ. Int.* **47**, 37–47 (2012).
91. Lloyd, S., Kovats, R. & Armstrong, B. Global diarrhoea morbidity, weather and climate. *Clim. Res.* **34**, 119–127 (2007).
92. Zhao, N., Ma, X., Gan, L., Wei, Y., Sun, P., Zhang, D. Research on the application of medical-meteorological forecast model of infectious diarrhea disease in Beijing. In: *Proceedings 2010 IEEE 5th International Conference on Bio-Inspired Computing: Theories and Applications, BIC-TA*. 100–103 (2010).
93. Hashizume, M. *et al.* Association between climate variability and hospital visits for non-cholera diarrhoea in Bangladesh: Effects and vulnerable groups. *Int. J. Epidemiol.* **36**, 1030–1037 (2007).
94. McCormick, B. J., Alonso, W. J. & Miller, M. A. An exploration of spatial patterns of seasonal diarrhoeal morbidity in Thailand. *Epidemiol. Infect.* **140**, 1236–1243 (2012).
95. Diaz-Uriarte, R. & Alvarez de Andrés, S. Gene selection and classification of microarray data using random forest. *BMC Bioinform.* **7**, 1–13 (2006).
96. Jiang, P. *et al.* MiPred: classification of real and pseudo microRNA precursors using random forest prediction model with combined features. *Nucleic Acids Res.* **35**, W339–W344 (2007).
97. Mutanga, O., Adam, E. & Cho, M. A. High density biomass estimation for wetland vegetation using worldview-2 imagery and random forest regression algorithm. *Int. J. Appl. Earth Observ. Geoinform.* **18**, 399–406 (2012).
98. Belgiu, M. & Drăgu, L. Random forest in remote sensing: A review of applications and future directions. *ISPRS J. Photogr. Remote Sens.* **114**, 24–31 (2016).
99. Kirk, M. D. *et al.* World Health Organization estimates of the global and regional disease burden of 22 foodborne bacterial, protozoal, and viral diseases, 2010: A data synthesis. *PLoS Med.* **12**, e1001921 (2015).
100. Kanungo, S., Dutta, S. & Sur, D. Epidemiology of typhoid and paratyphoid fever in India. *J. Infect. Dev. Ctries.* **2**, 454–460 (2008).

101. Brockett, S. *et al.* Associations among water, sanitation, and hygiene, and food exposures and typhoid fever in case-control studies: A systematic review and meta-analysis. *Am. J. Trop. Med. Hyg.* **103**, 1020–1031 (2020).
102. Cairncross, S. *et al.* Water, sanitation and hygiene for the prevention of diarrhoea. *Int. J. Epidemiol.* **39**, i193–i205 (2010).
103. Ma, Y. *et al.* Effects of community health volunteers on infectious diseases of children under five in Volta Region, Ghana: Study protocol for a cluster randomized controlled trial. *BMC Public Health* **17**, 1–10 (2017).
104. Olaiz-Fernandez, G. A. *et al.* Historical overview of acute infectious diarrhea in Mexico and future preventive strategies. *Salud publica de Mexico* **62**, 25–35 (2020).
105. Hamzah, Z., Handayani, A. T. W., Prihatiningrum, B., Tohirun, T., Indartin, D., Ernanda, H., Kartikasari, N. Water Pollution in Bedadung Watersheds area and diseases in elderly people. In *Proceedings of the 2nd International Conference in Health Sciences (ICHHS)*. 38–43 (2018).
106. Slotwiner-Nie, P. K. & Brandt, L. J. Infectious diarrhea in the elderly. *Gastroenterol. Clin. N. Am.* **30**, 625–635 (2001).
107. Alebel, A. *et al.* Prevalence and determinants of diarrhea among under-five children in Ethiopia: A systematic review and meta-analysis. *PLoS ONE* **13**, e0199684 (2018).
108. Kim, S. *et al.* Spatial and temporal patterns of typhoid and paratyphoid fever outbreaks: A worldwide review, 1990–2018. *Clin. Infect. Dis.* **69**, S499–S509 (2019).
109. Yang, W., Li, Z., Lan, Y., Wang, J., Ma, J., Jin, L., Sun, Q., Lv, W., Lai, S., Liao, Y., Hu, W. A nationwide web-based automated system for early outbreak detection and rapid response in China. *West. Pac. Surveill. Response J.* **2** (2011).
110. Liaw, A., Wiener, M. Classification and Regression by RandomForest. Technical Report (2001).

Author contributions

X.Y. is the submitting author. X.Y. and J.H. designed the study and collected the data, X.Y., W.F.X. and T.Y.H. analyzed the data and drafted the manuscript, X.Y., W.F.X., T.Y.H. and J.H. revised the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-00932-0>.

Correspondence and requests for materials should be addressed to J.H.

Reprints and permissions information is available at www.nature.com/reprints.

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



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

© The Author(s) 2021