District prediction of cholera risk in China based on environmental factors

XU Min[1], CAO ChunXiang[1*], WANG DuoChun[2], KAN Biao[2], JIA HuiCong[1], XU YunFei[1] & LI XiaoWen[1]

[1] *State Key Laboratory of Remote Sensing Science, Jointly Sponsored by the Institute of Remote Sensing and Digital Earth of Chinese Academy of Sciences and Beijing Normal University, Beijing 100101, China;*
[2] *State Key Laboratory for Infectious Disease Prevention and Control, Institute for Infectious Disease Control and Prevention, Chinese Center for Disease Control and Prevention, Beijing 102206, China*

The epidemics of cholera are impacted by many climatic and environmental factors such as precipitation, temperature, elevation and so on. The paper analyzed the suitable degree of *V. cholerae* in China using MaxEnt based on some geographic and climatic factors, and predicted the cholera risk in each district of China according to the suitable degree. The result shows that the areas in coastal southeast, central China and western Sichuan Basin are relatively suitable for *V. cholerae* and the suitable degree is higher in the Xinjiang Basin than in surrounding areas. The variables of precipitation, temperature and DEM are three main environmental risky factors that affecting the distribution of cholera in China. The variables of relative humidity, the distance to the sea and air pressure also have impacts on cholera, but sunshine duration and drainage density have little impact. The AUC value of MaxEnt based model is above 0.9 which indicates a high accuracy.

MaxEnt, cholera, environmental factors, district prediction

Cholera was first introduced to China in 1820 during its first pandemic. It caused tremendous losses to China in the followed pandemic. The horizontal incidence of cholera in China has steadily been at a low level in recent years. However, there would be an outbreak every 3–4 years due to the impact of extreme weather conditions or natural disasters [1]. In 1998, China was hit by a once-in-a-century flood, which led to the Yangtze River, the Yellow River and some other rivers burst their banks. The flood carrying with *vibrio cholerae* (shortly *V. cholerae*) was washed into the residential areas especially the towns and villages, which caused the food and water to be contaminated. As a result, the incidence of cholera in China increased significantly in 1998 and 1999 [2]. The overall cholera incidence was relatively low during the subsequent decades, but there was a small peak appeared in 2001 and 2005 respectively. The cholera cases outbreak in 2001 were mainly distributed in Guangdong, Sichuan, Zhejiang, but in 2005, they were mainly distributed in Fujian, Zhejiang and Guangdong. What's more, the dominant serotype O139 was transformed into serotype Inaba which had been caused the cholera pandemics in history. Hainan Province suffered a once-in-30-year heavy rainfall, which caused floods in many areas. Because of the seawater intrusion and the hydrops being difficult to discharge, the natural *V. cholerae* contaminated the living environment of the residents. Also, there were cholera epidemics outbreak in Mengcheng, Anhui and Huai'an, Jiangsu in August and September, 2010 respectively [1].

Cholera epidemic is affected by climate, geography and some other natural environmental factors. It also has connections with lifestyle of the inhabitants, sanitation, population and some other socio-environmental factors [3]. The

*Corresponding author (email: cao413@irsa.ac.cn)

relationship between socio-environmental factors and cholera epidemic is very complex. The socio-environmental factors are difficult to be quantitatively described and the relevant data is difficult to acquire. However, most of the natural environmental factors can be acquired with the help of RS and GIS techniques, and they are easy to be quantitatively expressed [4,5]. Therefore, the paper focused on the analysis of the influences of natural environmental factors such as geography and climate on the cholera epidemic, and predicted the potential spatial distribution of cholera in China based on these natural environmental factors.

1 The analysis on the natural environmental factors of cholera

Temperature is one of the most important climatic factors which affect the living and reproducing of *V. cholerae* in the natural environment [6]. The temperature of 37°C is the most suitable for the growth and reproduce of *V. cholerae*. In addition, the temperature between 16°C and 42°C is also very suitable for the growth and reproduce of *V. cholerae*. Although the *V. cholerae* can also survive a certain amount of time out of this interval of temperature, the environment temperature above or below this temperature will inhibit its reproduction [7]. Because of the global industrialization process, a large number of greenhouse gases were increased. The trend of global warming is increasingly evident. The average temperature of the tropical regions becomes closer to the most suitable breeding temperature of the bacterium. Therefore, the epidemic of cholera in these regions especially in Africa countries which have poor basic health conditions will be more frequent. Fernández et al. [8] analyzed the relationship among cholera cases in Lusaka, Zambia in weeks during 2003–2006 and the climate factors such as rainfall, temperature and so on. The results showed that when the temperature of six weeks ago increased 1 degree, the cases of cholera increased 5.2 percent. Gil et al. [9] found there was a significant positive correlation between the sea surface temperature and the local cholera incidence in Peru by continuous collected the water sample in four locations along the coast of Peru. Precipitation is also an important climate factor which affect the cholera epidemic [10,11]. The high-incidence period of cholera is always after the rainy season. That's because the large number of rain lead to the raise of local rivers and lakes, even caused the heavy flood. A large number of residential areas were submerged. That would increase the probability of the residents contacting to the contaminated water, which led to the increase of the cholera occurrence. The researches in Hue, Bangladesh and KwaZulu-Nuatal, South Africa showed that the local incidence of cholera had significantly related to the precipitation two months ago [10,12]. Pascual et al. [13] found that the cholera mortality in nine provinces of India is significantly related to the local precipitation.

As an important geographic environmental element, elevation may significantly impact the distribution of variety of diseases. In general, low attitude areas are prone to the outbreak of cholera. The reason is that *V. cholerae* mainly spread through the water, the areas with a low elevation are vulnerable to the floods. Thus the local residents have higher probability to contact the water or food with bacterium. Drainage density is an important indicator of the basin structural features. It reflects the precipitation and the underlying surface condition to a certain extent, which exerts great influence on the risk of cholera outbreaks. The regions of high rainfall and weak permeability have high risk with the flooding which associated with the outbreak of cholera. Therefore, drainage density may be an indicator that indirectly reflects the risk of cholera outbreaks. The ocean is an important aquatic environment for the survival and reproduction for *V. cholerae*. The cholera cases in China and many other countries are mainly distributed in coastal and offshore areas. The rising sea levels caused by global warming makes the coastal areas in the low-elevation more vulnerable to the invasion of seawater, resulting in a large number of marine saline water being injected into fresh water such as groundwater. This will raise the water salinity of coastal areas to be close to the suitable salinity for *V. cholerae*, thus resulting in the rise of the cholera incidence [14]. Borroto et al. [15] researched the spatial model of cholera incidence of Mexican states during 1991 and 1996 using the method of spatial autocorrelation. The result showed the cholera incidence in coastal areas is 2.47 times that in inland areas.

2 Data processing

2.1 Cholera and demographic data

The cholera data is acquired from the Institute for Infectious Disease Control and Prevention, Chinese Center for Disease Control and Prevention. It is a compilation of data which was submitted and summarized through administrative levels. The data set includes the record of cholera incidence and deaths of cholera in each county of China during 2001–2008. Each record contains the number of cholera incidence and deaths per month and the corresponding county name and district code. The demographic data is from the Speediness Collected Data of National Population of Fifth Census, which can be downloaded from the website of National Bureau of Statistics (http://www.stats.gov.cn/). The paper collected the demographic data of each county in China in the fifth national census. Each record of demographic data includes the natural population, county name and district code.

2.2 Geographic data

The basic geographic data is collected from the 1 : 1000000 national basic scale electronic maps, which are provided by

the National Geomatics Center. This data set includes the administrative margin vector of each county, and the National Rivers distribution vector. We established a connection between the administrative boundary vector and the cholera and demographic data by using the field of "district code", and generated a new vector which contained the number of cholera cases and population in each county, thus, realized the spatialization of the cholera and demographic data. We unified the coordinate system through the projection conversion for each layer, and then established the temporal and spatial database of cholera in China.

The elevation data is acquired from SRTM (Shuttle Radar Topography Mission) with a spatial resolution of 90 m. The DEM data of SRTM divided the global scale into raster maps with 1 degree in the length and width according to the latitude and longitude. The paper collected the DEM raster maps of SRTM which covered the whole China. We spliced these images using the software of ENVI 4.5, and then cut the mosaic data based on the administrative boundary vector layer of China. By this way, we got the elevation raster layer of China.

The drainage density was calculated based on the National Rivers distribution vector in the 1:1000000 national basic scale geographic datasets. It was expressed as the ratio of average river length l and the average adjacent area a. If there are n rivers in the area A, and the whole rivers length is L. Then $l=L/n$, $a=A/n$. The format of drainage density can be expressed as

$$D = \frac{L}{A} = \frac{l}{a},\qquad(1)$$

namely the drainage density can be expressed as the total length of all the rivers per unit area.

The distance from each county to the ocean is calculated using the coast map of China as the input layer. By computing the distance of the central point of each county to the ocean based on spatial analysis, we generated the corresponding raster layer of the distance to the ocean.

2.3 Climate data

The climate data in the paper was downloaded from the website of China Meteorological Data Sharing Service (http://cdc.cma.gov.cn/). Five climate elements were collected in this paper, including temperature, precipitation, relative humidity, air pressure, sunshine duration from the annual standard data in the surface climate dataset during 2001 and 2008. The dataset originated from the surface meteorological records which was submitted and summarized through administrative levels by the meteorology departments of each province, city and county. The type of original climate data is ASCII file. It includes the summary data from the 722 basic surface meteorological observation stations.

The climate data were processed as follows. Firstly, we

did something about the missing data processing to the climate data, and deleted the station records which contained blank value or invalid value in the original file. Secondly, we converted the ASCII file into the file of floating data with a standard unit to each element. Then, we calculated the average annual values of temperature, precipitation, relative humidity, barometric pressure, sunshine duration. Finally, we spatialized all the climate data.

The process of spatialization establishes the connection between the climate data of each station and its location by latitude and longitude file based on the site ID code, and then generates the vector layer for each meteorological station. In order to acquire the continuous planar layer of the climate data, the paper spatial interpolated the climate elements layers using the method of IDW(Inverse Distance Weighted). Based on the above processing, the paper acquired the raster layers of average annual temperature, precipitation, relative humidity, barometric pressure, sunshine duration in mainland China between 2001 and 2008.

3 The MaxEnt

The Ecological Niche Model (ENM) is based on the theory of niche, which is the general term of the environment for special species and their living habits [16,17]. ENM was first used to research on the potential geographical distribution of the species. The basic principle of its prediction is based on the living environment of biological or microbiological special. Firstly, using a suitable mathematical model to summarize or simulated its niche needs from its known distribution, namely the living environment. Then, predicting the suitable distribution area of the target area according to various environmental elements of the target area.

The MaxEnt is a niche prediction model based on the principle of maximum entropy. It extrapolates or predicts from deficient known information. It explores the non- random relationship between the environmental characteristics of the known species distribution and the study area according to the known species distribution data and environmental layers. And find the maximum entropy probability distribution as optimal solution for the prediction of species suitable area under a certain constraints [18]. The mathematical principle is as follows [19]. Assuming that the set of X present the points which are transformed from the grid cells of whole study area, $x_1, x_2, \cdots x_n$ are the finite positions of the known biodistribution. n is the number of the known distributed points. Namely, the series of $x_1, x_2, \cdots x_n$ are samples of X which present the set of all points in the study area. These samples are dominated by the probability distribution of a position which we denote π. Therefore, we should build an approximate value of $\hat{\pi}$ to estimate π. We assume that f_j is the function of environment variable of all the target point-of-presence. Then the expectation of f_j is

defined as $\sum_{x \in X} \pi(x) f_j(x)$ which denoted by $\pi[f_j]$. The feature expectations $\pi[f_j]$ can be approximated using the known sample points. The empirical average of f_j is $\frac{1}{n}\sum_{j=1}^{n} f_j(x_j)$, which we can write as $\overline{\pi}[f_j]$. Therefore, the constraint condition can be expressed as $\hat{\pi}[f_j] = \overline{\pi}[f_j]$. In fact, the empirical means has a deviation with the true means of f_j. So a neighborhood constraint condition is required to make the empirical values to be close to the true values.

$$\left| \hat{\pi}[f_j] - \overline{\pi}[f_j] \right| \leqslant \beta_j. \tag{2}$$

According to the convex duality theory, $\hat{\pi}$ can be described by a special Gibbs distribution.

$$q_\lambda(x) = \frac{e^{\lambda * f(x)}}{Z_\lambda} = \frac{e^{\sum_{j=1}^{N} \lambda_j * f_j(x)}}{Z_\lambda}, \tag{3}$$

where λ_j is the weight coefficient of index j; Z_λ is a constant which are used to insure that $\sum q_\lambda = 1$. This distribution is namely Gibbs distribution. The probability distribution of the maximum entropy equals to the special Gibbs distribution. And the special Gibbs distribution should ensure that the value of $\overline{\pi}[-\ln(q_\lambda)]$ is the minimum.

The prediction of MaxEnt is divided into the following four steps. Firstly, collect the investigation records of the species distribution or refer to the relevant historical literature records, and obtain the data of the known distribution area. Secondly, select the appropriate environment variables which have a direct relationship with the accuracy of the prediction results from the environmental data set of species. Then, input the environment variables set of the known distribution data into the model and get the prediction result of the target area, namely the potential geographic distribution of the species. Finally, analyze and evaluate the prediction model.

4 Result and discussion

4.1 The suitable degree of *V. cholerae* in China

The district prediction of cholera risk based on environmental factors is mainly to estimate the probability of the epidemic outbreak in the unknown areas according to the geographic distribution and the risky factors of known samples. Zo et al. [20] proved that the outbreak of cholera epidemic was consistent with the increase of the amount of the *V. cholerae* in the natural environment. So there must be a number of *V. cholerae* in the area where cholera cases outbreak. Therefore, we can use the niche model to predict the suitable degree of *V. cholerae* in different areas based on

the geographic and environmental factors, and realize the district partition of cholera risk in China according to the suitable degree distribution.

We regarded the distribution layer of cholera cases for all the counties in China between 2001 and 2008 as the known distribution data in the ENM. Then we randomly extracted 25% samples of the known cholera cases as the test data, and the other 75% as the training data. The average values of annual precipitation, temperature, sunshine duration, air pressure, relative humidity, DEM, drainage density and the distance from the ocean were input as environmental variables. The output suitable degree distribution of *V. cholerae* in China has a value in the range of 0–1. As shown in Figure 1, the southeast coastal China, central China and the western Sichuan Basin have a relatively high suitable degree of *V. cholerae*. However, there is a relatively low suitable degree of *V. cholerae* in northeast and northwest China. In northwest areas, the Xinjiang Basin has a relatively higher suitable degree of *V. cholerae* than surrounding areas.

The suitable degree probabilities of cholera that the MaxEnt obtained are continuous variables in the range of 0–1. We need select an appropriate threshold in order to get the risky levels of the cholera incidence in each region. China was divided into non-incidence area, low-risk area, medium-risk area and high-risk area according to the suitable degree of *V. cholerae* using empirical method though consulting with the epidemiological experts. The area which had a value of suitable degree less than 0.01 was regarded as no incidence area. The range of 0.01–0.1 was regarded as low-risk area and 0.1–0.3 as medium-risk area, while the value over 0.3 was high-risk area. According to the partition, we got the risky map of cholera in China as shown in Figure 2. Most regions of the Southeast coastal provinces such as Guangdong, Fujian, Zhejiang, Jiangsu, Hainan and Central China such as Anhui, Jiangxi, Hunan and Hubei are high-risk areas of cholera. Sichuan Province which is located in West China has very complex geographical and climatic
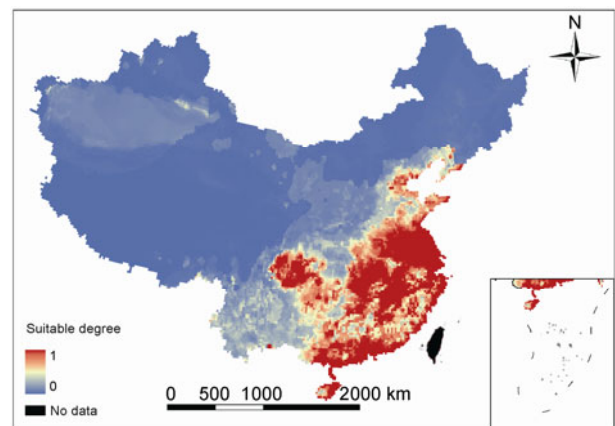

Figure 1 Prediction of the suitable of cholera in China. The warm colour represents the area with high suitable degree; the cold colour represets the area with low suitable degree; the black colour represents no data area.
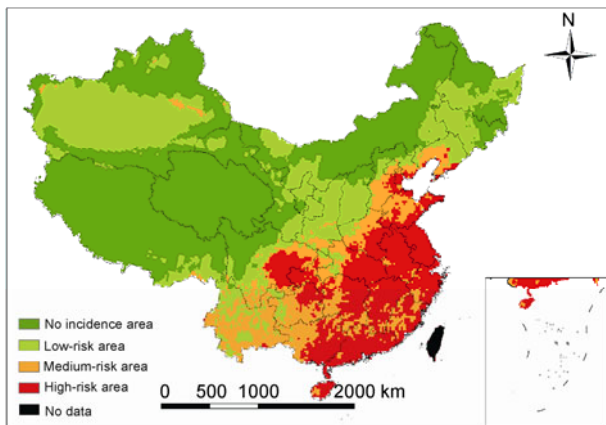
Figure 2 District prediction of cholera risk in China. The red colour represent high-risk area; the orange colour represent medium-risk area; the reseda colour represent low-risk area; the bottle green colour represent non-incidence area and the black colour represent no data area.

Figure 3 The ROC curve and AUC value of prediction, the red lines represent the ROC curve of the training samples, and the blue line represents the ROC curve of the test data.

environments, and it covers all the four levels of risk. Most of the Northwest regions except Xinjiang are no incidence areas. The Northeast regions except coastal Liaoning are no incidence or low-risk areas.

4.2   The evaluation of prediction model

The accuracy and error sources of the prediction model are analyzed through model evaluation process. Generally, the Receiver Operating Characteristic Curve (ROC Curve) is widely used for the prediction model evaluation [21,22]. It is a comprehensive indicator that reflects the continuous variables of sensitivity and specificity. The basic principle is as follows. To suppose a threshold value is the correct criteria for prediction. The predicted value below this threshold is regarded as wrong prediction, and the predicted value over the threshold is regarded as correct prediction. We separately obtained a series of correct data and wrong data by analyzing the predictive results, computed the true positive rate and the false positive rate by contrasting the true value.

The area under ROC curve is the AUC (Area Under Curve) value. AUC is one of the most commonly evaluation indicators for its impregnability to the threshold. The AUC value close to 1 indicates the predictive result of the model is good. Generally speaking, when AUC value is between 0.5 and 0.7, the accuracy of predictive model is believed relatively low. The AUC value between 0.7 and 0.9 indicates a moderate accuracy of the predictive model. However, while the AUC value is greater than 0.9, the accuracy is relatively high [23]. We evaluated the predictive result using ROC curve and computed the AUC value. The result is shown in Figure 3. The red lines represent the ROC curve of the training samples, and the blue line represents the ROC curve of the test data. Both of them are greater than 0.9 which are respectively 0.937 and 0.944. Therefore, it is believed that the predictive model has a high accuracy.
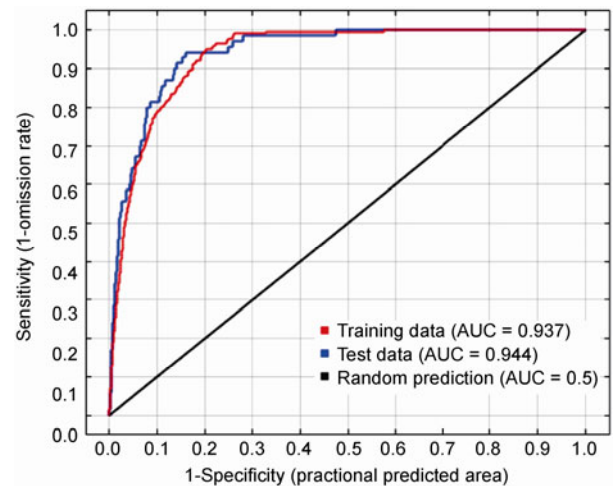
4.3   The contribution analysis of environmental risk factors

The MaxEnt evaluates the contribution of each environmental risk factor to the predictive model using the indicator of relative contribution, namely the importance of each environmental factor to the predicted results. The relative contribution of the environmental variable increases as the absolute value of the weight coefficient increases and it decreases as the weight coefficient decreases in each training iteration process.

This paper calculated the contribution of each environmental factor to the cholera predictive model. The results showed that the precipitation was the most important environmental variable that affected cholera incidence which had a relative contribution of 24.3%. The temperature (23.8%) and elevation (17.9%) has a little weaker importance than precipitation. These three environmental factors accounted for more than 65% of the relative contribution. The relative humidity (12.2%), the distance to the ocean (9.4%), the air pressure (7.8%) and the sunshine duration (2.8%) are also important environmental risk factors. The drainage density has the smallest contribution to the model which is only 1.8%, namely the drainage density has nearly no effect on the cholera incidence in China.

The precipitation has the greatest impact on the cholera incidence in China. The reason is that the floods induced by the continuing large amount of precipitation rushed the river water with *V. cholerae* into the neighborhoods, and contaminated the food and water which made the residents be prone to infect cholera. The temperature is also a major environmental factor of cholera, mainly due to the survival, and reproduction of the *V. cholerae* has great sensitivity to the environmental temperature. Only under suitable condition of temperature can *V. cholerae* survive and reproduce. The elevation also effect cholera a lot. Because the neigh-

borhoods in low altitude areas are more prone to be invade by the floods and the coastal waters which will increase the probability of the residents to contact with the *V. cholerae*.

4.4 The statistical analysis of the cholera risk

The paper analyzed the area ratio of different cholera risk levels for all the regions in mainland China. The statistical results are shown in Table 1. The high-risk areas of cholera are generally in the coastal or offshore areas such as Jiangsu, Shanghai, Anhui, Guangdong, Jiangxi, Zhejiang, Hainan, Tianjin, Fujian, etc. The top-3 high-risk areas we predicted are Anhui, Jiangsu and Shanghai. Mengcheng, Anhui and Huai'an, Jiangsu respectively had an outbreak of cholera in August and September 2010. The cholera outbreak in a Middle School of Huai'an, Jiangsu had infected total 19 students, while the cholera outbreak in Mengcheng, Anhui had reached 33 cases. Shanghai may reduce the risk of

cholera outbreak to some extent due to its good basic sanitation. In the following five regions, Jiangxi has similar geographical location and climate conditions with Anhui, and has a relatively backward basic sanitation. The probability of cholera outbreak in the future is very high for Jiangxi. While the cholera incidence of Guangdong, Zhejiang, Hainan, Tianjin and Fujian has been at the forefront in the country recently. They have relatively high-risk of cholera over a long period of time. Tibet and Qinghai are the lowest risk regions of cholera in China. Moreover, Tibet is the only region in China which has no case reported since the 7th global epidemic of cholera.

5 Conclusion

The paper used niche model to predict the suitable degree of *V. cholerae* in different areas based on the geographic and environmental factors, and realized the district partition of cholera risk in China according to the suitable degree distribution. Most regions of the Southeast coastal regions such as Guangdong, Fujian, Zhejiang, Jiangsu, Hainan, and Central China such as Anhui, Jiangxi, Hunan, Hubei are high-risk areas of cholera. Sichuan Province located in West China has very complex geographical and climate environments, and it covers all the four levels of risk. Most of the Northwest regions except Xinjiang are no incidence areas. ROC curve was used to evaluate the predictive result. Both of the training samples and the test data are greater than 0.9. It shows the predictive model has a high accuracy. The precipitation is the most important environmental variable that affects cholera incidence. The temperature and elevation have weaker importance than precipitation. These three environmental factors account for more than 65% of the relative contribution. The relative humidity, the distance to the ocean, the air pressure and the sunshine duration are also important environmental risk factors. The drainage density has the smallest contribution to the model, only 1.8%. The analysis on the area ratio of different cholera risk levels for all the administrative regions in mainland China shows that Jiangsu, Shanghai, Anhui, Guangdong, Jiangxi, Zhejiang, Fujian, Hainan, Tianjin are the highest risk regions of cholera in China.

Table 1 The area ratio of different cholera risk levels in each region

Region	Non-incidence area (%)	Low-risk area (%)	Medium-risk area (%)	High-risk area (%)
Jiangsu	0.00	0.00	0.00	100.00
Shanghai	0.00	0.00	0.00	100.00
Anhui	0.00	0.00	6.44	93.56
Guangdong	0.00	0.00	12.72	87.28
Jiangxi	0.00	0.00	13.69	86.31
Zhejiang	0.00	0.00	15.20	84.40
Hainan	0.00	0.00	16.18	83.82
Tianjin	0.00	0.00	25.81	74.19
Fujian	0.00	0.00	37.46	62.54
Hunan	0.00	0.39	39.61	60.00
Hubei	0.00	10.78	30.87	58.35
Chongqing	0.00	9.13	35.10	55.77
Shandong	0.00	0.00	46.44	53.56
Guangxi	0.00	1.28	47.35	51.37
Henan	0.00	16.47	42.82	40.71
Guizhou	0.00	10.00	60.23	29.77
Sichuan	34.13	24.92	21.30	19.65
Hebei	26.40	22.54	39.31	11.75
Liaoning	0.98	50.74	38.73	9.56
Beijing	0.00	42.86	55.10	2.04
Yunnan	1.46	38.50	59.37	0.67
Shaanxi	3.65	83.21	13.14	0.00
Xinjiang	48.62	50.29	1.10	0.00
Tibet	92.24	7.36	0.40	0.00
Shanxi	26.87	72.90	0.23	0.00
Ningxia	23.36	76.64	0.00	0.00
Jilin	34.35	65.65	0.00	0.00
Heilongjiang	61.74	38.26	0.00	0.00
Gansu	66.94	33.06	0.00	0.00
Inner Mongolia	79.89	20.11	0.00	0.00
Qinghai	100.00	0.00	0.00	0.00

1   Xu M. Study on spatio-temporal distribution and prediction of cholera in China based on spatial information technology (in Chinese). Dissertation for the Doctoral Degree. Beijing: Graduate University of Chinese Academy of Sciences, 2011. 1–8

2   Kan B, Wang D C. Cholera (in Chinese). Xi'an: Shaanxi Science and Technology Press, 2005

3   Emch M, Feldacker C, Yunus M, et al. Local environmental predictors of cholera in Bangladesh and Vietnam. Am J Trop Med Hyg, 2008, 78: 823–832
4   Chang C Y, Cao C X, Wang Q, et al. The novel H1N1 Influenza A global airline transmission and early warning without travel containments. Chin Sci Bull, 2010, 55: 3030–3036
5   Cao C X, Xu M, Chang C Y, et al. Risk analysis for the highly pathogenic avian influenza in mainland china using meta-modeling. Chin Sci Bull, 2010, 55: 4168–4178
6   West P A. The human pathogenic vibrios—A public health update with environmental perspectives. Epidemiol Infect, 1989, 103: 1–34
7   Borroto R. Global warming, rising sea level, and growing risk of cholera incidence: A review of the literature and evidence. Geol J, 1998, 44: 111–120
8   Fernández L, Angel M, Bauernfeind A, et al. Influence of temperature and rainfall on the evolution of cholera epidemics in Lusaka, Zambia, 2003–2006: Analysis of a time series. T Roy Soc Trop Med H, 2009, 103: 137–143
9   Gil A I, Louis V R, Rivera I N G, et al. Occurrence and distribution of *V. cholerae* in the coastal environment of Peru. Environ Microbiol, 2004, 6: 699–706
10 Emch M, Feldacker C, Islam M, et al. Seasonality of cholera from 1974 to 2005: A review of global patterns. Int J Health Geogr, 2008, 7: 31
11 Koelle K, Rod X, Pascual M, et al. Refractory periods and climate forcing in cholera dynamics. Nature, 2005, 436: 696–700
12 Mendelsohn J, Dawson T. Climate and cholera in KwaZulu-Natal, South Africa: The role of environmental factors and implications for epidemic preparedness. Int J Hyg Environ Heal, 2008, 211: 156–162
13 Pascual M, Bouma M, Dobson A. Cholera and climate: Revisiting the

quantitative evidence. Microbes Infect, 2002, 4: 237–245
14 Constantin de Magny G, Murtugudde R, Sapiano M R, et al. Environmental signatures associated with cholera epidemics. Proc Natl Acad Sci USA, 2008, 105: 17676–17681
15 Borroto R J, Martinez-Piedra R. Geographical patterns of cholera in Mexico, 1991–1996. Int J Epidemiol, 2000, 29: 764–772
16 Hutchinson G E. Concluding remarks. Cold spring harbor symposium Quant Biol, 1957, 22: 425–427
17 Wang Y S. Application of Niche models in the risk assessment of invasive alien species (in Chinese). Dissertation for the Doctoral Degree. Changsha: Hunan Agricultural University, 2007. 20–22
18 Philips S J, Anderson R P, Schapire R E. Maxent entropy modeling of species geographic distribution. Ecol Model, 2006, 190: 231–259
19 Qi C P. GIS-based modeling of the potential distribution and the potential net primary productivity of deciduous broadleaved forest in the western part of the Chinese Loess Plateau (in Chinese). Dissertation for the Doctoral Degree. Lanzhou: Lanzhou University, 2009. 58–62
20 Zo Y G, Rivera I N, Russek-Cohen E, et al. Genomic profiles of clinical and environmental isolates of *V. cholerae* O1 in cholera-endemic areas of Bangladesh. In: Proceedings of the National Academy of Sciences of the United States of America, 2002, 99: 12409–12414
21 Goodenough D J, Rossmann K, Lusted L B. Radiographic applications of receiver operating characteristic (ROC) curves. Radiology, 1974, 110: 89–95
22 Metz C E. Basic principles of ROC analysis. Seminars in Nuclear Medicine, 1978, 8: 283–298
23 Li M Y, Ju Y W, Sunil K, et al. Modeling potential habitat for alien species of Dreissena polymorpha in the continental USA (in Chinese). Acta Ecol Sin, 2008, 28: 4253–4258