Mapping risk of leptospirosis in China using environmental and socioeconomic data

Jian Zhao†, Jishan Liao‡†, Xu Huang†, Jing Zhao†, Yeping Wang†, Jinghuan Ren†, Xiaoye Wang† and Fan Ding†*

Abstract

Background: Leptospirosis is a water-borne and widespread spirochetal zoonosis caused by pathogenic bacteria called leptospira. Human leptospirosis is an important zoonotic infectious disease with frequent outbreaks in recent years in China. Leptospirosis’s emergence has been linked to many environmental and ecological drivers of disease transmission. In this paper, we identified the environmental and socioeconomic factors associated with leptospirosis in China, and predict potential risk area of leptospirosis using predictive models.

Methods: Leptospirosis incidence data were derived from the database of China’s web-based infectious disease reporting system, a national surveillance network maintained by Chinese Center for Disease Control and Prevention. We built statistical relationship between occurrence of leptospirosis and nine environmental and socioeconomic risk factors using logistic regression model and Maxent model.

Results: Both logistic regression model and Maxent model have high performance in predicting the occurrence of leptospirosis, with AUC value of 0.95 and 0.96, respectively. Annual mean temperature (Bio1) and annual total precipitation (Bio12) are two most important variables governing the geographic distribution of leptospirosis in China. The geographic distributions of areas at risk of leptospirosis predicted from both models show high agreement. The risk areas are located mainly in seven provinces of China: Sichuan Province, Chongqing Municipality, Hunan Province, Jiangxi Province, Guangdong Province, Guangxi Province, and Hainan Province, where surveillance and control programs are urgently needed. Logistic regression model and Maxent model predicted that 403 and 464 counties are at very high risk of leptospirosis, respectively.

Conclusions: Our results highlight the importance of socioeconomic and environmental variables and predictive models in identifying risk areas for leptospirosis in China. The values of Geographic Information System and predictive models were demonstrated for investigating the geographic distribution, estimating socioeconomic and environmental risk factors, and enhancing our understanding of leptospirosis in China.

Keywords: Leptospirosis, Ecological niche modeling, Logistic regression, Maximum Entropy model, Geographic Information System

Background

Human leptospirosis is a waterborne infectious disease caused by pathogenic bacteria called leptospira. Currently, it is one of the most common widespread spirochetal zoonosis and a growing worldwide public health concern [1]. Leptospirosis can be transmitted directly or indirectly from animals to humans, human-to-human transmission occurs very rarely [2]. In recent years, leptospirosis has gained increased attention because of many outbreaks of leptospirosis around the world [3]. Those outbreaks pose a great burden on health systems and cause significant economic and social disruption [4]. The WHO has identified leptospirosis as a neglected tropical disease of global importance, requiring further research to understand its epidemiology, ecology and the disease burden that it causes around the world [2].

In China, a total of 176,450 cases were confirmed in 29 out of the 34 provinces and municipalities from 1991 to 2010, resulting in an average annual incidence rate of 0.70 cases per 100,000 people [5]. Human leptospirosis,
Leptospirosis is a multi-factorial infectious disease that is highly associated with both environmental and socioeconomic factors [13].

**Climatic variables**

Although occurs worldwide, leptospirosis is most prevalent in tropical and subtropical areas with high rainfall [2]. Heavy rainfall events and floods increase the risk of leptospirosis by bringing leptospira and their animal hosts including rat and pig into closer contact with humans [14]. A high correlation between seroprevalence of infection and heavy rainfall and flood events in China was also reported [15]. Leptospires are able to survive for prolonged periods of time in higher temperatures and humid environments [1, 16]. In addition, we found that 37% cases occurred in summer and 50% cases in autumn (Fig. 2) in China. Therefore, four variables (annual mean temperature, temperature seasonality, annual precipitation and precipitation seasonality) were obtained from Worldclim database [17], which was interpolated from weather station records from 1950–2000.

**Animal host density variable**

Leptospirosis is found mainly in area where humans come into contact with the urine of infected animals or urine-contaminated environments [2]. Rodents and some domestic animals such as pigs and dogs are considered to be the most dangerous for transmitting leptospires to humans [18]. Leptospirosis is considered the only epidemic-prone infection that can be transmitted to humans directly from water, soil and food contaminated urine from infected animal hosts [1]. We considered only the effect of pig density on transmission of leptospirosis. The pig density data was obtained from the Food and Agriculture Organization (FAO) (available at http://www.fao.org/ag/againfo/resources/en/glw/GLW dens.html).

**Water reservoir variable**

Pathogenic leptospires was reported to be associated with the presence of water bodies including streams, lakes and springs. Therefore, areas with high density of water bodies might be at higher risk of leptospirosis occurrence [3]. Average river density was calculated by dividing the total river length in each community by the community area.
Fig. 1 a Number of reported cases in each county. b Proportion of risk area in each county predicted from Maxent model. c Proportion of risk area in each county predicted from logistic regression model.
Land cover type
Land cover and land cover change (LUCC) have been related to the incidence of leptospirosis [3, 8, 14]. Land cover type data was derived from GlobCover Land Cover version V2.3 released by European Space Agency (available at http://www.esa.int/esaEO/SEMXB7TTGOF_index_0.html).

Economic variable
Leptospirosis cases have been reported in a variety of settings ranging from large urban centers after heavy rainfall events to remote rural areas with limited access to clean drinking water. Leptospirosis is often considered a disease of poverty in middle and low income countries because it affects mostly vulnerable populations [19]. Gross Domestic Product (GDP) is a measurement of the economic performance and living standard. GDP data at county level was obtained from National Bureau of Statistics. Human population density could serve as proxy for the exposure to infected livestock. Gridded human population density data was obtained from Data Sharing Infrastructure of Earth System Science (available at http://www.geodata.cn/Portal/metadata/viewMetadata.jsp?id=100101-38).

Population density
National 1 km*1 km gridded population density data of 2003 in China was obtained from Data Sharing Infrastructure of Earth System Science (available at http://www.geodata.cn/Portal/metadata/viewMetadata.jsp?id=100101-38).

Data processing
All data layers were projected in Albers coordinates and resampled to spatial resolution of ~1 × 1 km in ArcGIS Desktop 9.3 (Environmental Systems Research Institute, Redlands, CA, USA) environment. We log-transformed three variables including pig density, GDP and human population density. The final set of predictors is listed in Table 1. We calculated correlation matrix for all variable. The highest correlation coefficient is between log-transformations of pig density and human population density (0.635) (Table 2). Therefore, there is low multicollinearity among predictors.

Ecological niche modeling
Ecological niche modeling (ENM) has been proved to be an effective tool to investigate the geography and ecology of disease transmission [6, 10]. A logistic regression model and Maxent model were used in this study to predict the potential risk of leptospirosis distribution in China.

Logistic regression model
Logistic regression is often used to determine the probability of a binary dependent variable based on one or several independent variables. The binary dependent variable represents presence or absence of leptospirosis points. The independent variables are a set of environmental and socioeconomic factors thought to determine distribution of leptospirosis. The advantage of logistic regression is that, through the addition of a suitable link function to the usual linear regression model, the variables may be either continuous or discrete, or any
combination of both types [20]. Logistic regression analysis is explained as a linear equation given below.

\[ Y = \logit(p) = \ln \left( \frac{p}{1-p} \right) \]  

\[ Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n \]  

Where \( p \) is the probability that the dependent variable (\( Y \)) is 1, \( p/(1 - p) \) is the so-called odd or frequency ratio, \( \beta_0 \) is the intercept, and \( \beta_1, \beta_2, \ldots, \beta_n \) are coefficients, which measure the contribution of the independent variables (\( X_1, X_2, \ldots, X_n \)) to the variations in \( Y \).

The probability of leptospirosis occurrence can be calculated from 9 environmental and socioeconomic factors (Table 1) using the logistic regression equation and coefficients obtained in R.

Maxent model

Maximum entropy (Maxent) model [21] is a very popular ecological niche model, which outputs the maximum entropy distribution that satisfies the constraints determined by a set of environmental variables. Maxent uses presence and background points (pseudo-absence points) randomly drew in study area to estimate probability of distribution of leptospirosis. 75 % occurrence points (1597) were randomly selected for training model, and the remaining 25 % points (532) were used for assessing model performance. A total of 21290 pseudo-absence points were randomly sampled from provinces where no case were reported in the period of 2010 to 2014. Nine environmental and socio-economic variables (Table 1) were used to build models to predict the geographic distribution of leptospirosis in China. To measure the relative importance of each environmental and socio-economic variable in Maxent model, a jackknife manipulation was performed.

Model assessment

Receiver operating characteristic (ROC) analysis was used to evaluate the discrimination ability of the two models, and to determine the optimal threshold to convert probability of distribution of leptospirosis to binary outputs (presence and absence of leptospirosis). The threshold-independent indicator, AUC (area under the Receiver Operating Characteristic curve), was used to evaluate model performance. AUC greater than 0.7 indicates an adequate predictive ability of the model [22]. The optimal threshold that equals test sensitivity and specificity, which gives equal weight to sensitivity and specificity.
specificity, was selected to identify risk areas of leptospirosis in China. The probability of leptospirosis presence could be converted into presence and absence using this threshold. A set of threshold-dependent assessment indices including sensitivity, specificity, Kappa, true skill statistic, were calculated.

Results
Model assessment
Both logistic regression model and Maxent model have high performance in predicting the occurrence of leptospirosis in China, with AUC value of 0.95 and 0.96, respectively. The optimal threshold that equals sensitivity and specificity is 0.05 for logistic regression model and 0.34 for Maxent model, respectively. Five threshold-dependent model assessment indices are shown in Table 3. Our results show that Maxent model performs better than logistic regression model regarding all model assessment indices.

Main risk factors of leptospirosis
Annual mean temperature (Bio1) and annual total precipitation (Bio12) are two most important variables governing the geographic distribution of leptospirosis in China, with permutation importance of 43.5 % and 37.8 % from Maxent model, respectively (Table 1). They produced the best predictions when used alone from Jackknife analysis. The response curves of the two most important variables reveal that the transmissions of leptospirosis mainly happen in hot and relatively wet regions in China. The probability of leptospirosis presence increases as annual mean temperature increases from 12 °C to 30 °C (Fig. 3). The risk of leptospirosis outbreak increases significantly as annual precipitation increases from 1000 mm to 2000 mm. Compared with the linear relationship between response variable and predictors in logistic regression model, response curve from Maxent model showed that Maxent can build nonlinear and complex relationships between probability of occurrence of leptospirosis and predictors.

Leptospirosis risk mapping
We converted the predicted maps of probability of leptospirosis presence from the two models to presence and absence using the corresponding optimal thresholds. The areas predicted to be present with leptospirosis are considered as risk area of leptospirosis. Our results revealed that there is high spatial variability of leptospirosis seroprevalence in China. The risk areas are located mainly in seven provinces of China: Sichuan Province, Chongqing Municipality, Hunan Province, Jiangxi Province, Guangdong Province, Guangxi Province, and Hainan Province, where surveillance and control programs are urgently needed (Fig. 4). There is high agreement between the prediction maps from logistic regression model and Maxent model. But the logistic regression model predicted larger risk areas than Maxent model, especially in Yunnan Province. We also calculated the proportion of risk area in each county of all provinces which have reported cases in the period of 2010–2014. Based on the proportion of risk area, we classified all counties into six risk categories: no risk (risk area = 0), very low (0 < risk area <= 0.2), low (0.2 < risk area <= 0.4), medium (0.4 < risk area <= 0.6), high (0.6 < risk area <= 0.8), very high (0.8 < risk area <= 1.0) (Fig. 5). The risk maps show very apparent spatial clustering of counties at very high risk of leptospirosis in China. There are 403 and 464 counties in very high risk category from logistic regression model and Maxent, respectively (Fig. 5). In those counties, disease survey and monitoring program are urgently needed, especially from August to October when most leptospirosis infections happen.

Discussion
In this study, we compiled surveillance data of leptospirosis in China from 2010 to 2014, and collected a number of environmental and socioeconomic variables thought to determine the geographic distribution of leptospirosis. We used ecological niche modeling approach to better understand the relationship between occurrence of leptospirosis and environmental and socioeconomic variables, and predicted the geographic distribution of leptospirosis in China. Because the reliable absence data are hardly available, random selected pseudo-absence points in provinces having no reported case was used in logistic regression model. According to the model performance assessment, both model can predict the geographic distribution of leptospirosis in China very successfully. Maxent model has better performance compared with logistic regression model, because it can handle non-linear relationships between response variable and predictors.

Pervious studies of epidemiology of leptospirosis have highlighted the importance of socioeconomic and
Environmental drivers [13]. Environmental variables play a significant role in natural-focal diseases by affecting pathogens directly, or influencing the distribution and abundance of disease hosts and vectors [23]. Four biological meaningful climatic variables including annual mean temperature (Bio1), annual precipitation (Bio12), temperature seasonality and precipitation seasonality, which represent the annual trend and seasonality of climatic condition of leptospirosis transmission, were selected to build predictive models for leptospirosis occurrence. Annual mean temperature (Bio1) and annual precipitation (Bio12) are the two most important factors determining the geographic distribution of leptospirosis in China. Because leptospirosis is often considered as a water-borne infectious disease, we used average river density to quantify the impacts of water condition on leptospirosis transmission indirectly. However, river density variable has no significant effect on predicting the geographic distribution of leptospirosis, possibly due to the coarse spatial resolution variables and the river density is high across the south of China.

Leptospirosis is highly correlated with socioeconomic factors. Leptospirosis occurrence in developing countries is related with intense and rapid urbanization without adequate infrastructure, resulting in sanitation problems, especially in poor vulnerable areas (i.e., slums) located close to rivers or channels, prone to periodical flooding. GDP was a meaningful variable explaining the relationship between leptospirosis and local economic condition. But the GDP data in China was calculated at county level, not at fine spatial resolution of 1 km * 1 km like other variables used in this study. The large number of pigs and backyard piggeries have previously been implicated in leptospirosis transmission [6]. However, we found that no socioeconomic variable was selected as the most important variable in determining the geographic distribution of leptospirosis in China. Other sanitation related variables and population socioeconomic status variables might preform better than the socioeconomic variables used in this study, but those kind of data are not available at fine spatial resolution, and especially at national scale.

Along with the unprecedented global climate change [24], the geographic distribution of leptospirosis would be changed, which need further study. A study revealed that climate change likely contributed to an increase in rodent populations, and therefore the number of outbreaks of leptospirosis [25]. Climate change may increase the incidence rate of leptospirosis infections in the current transmission area and cause the leptospirosis spread to new area, especially in developing countries. Surveillance program and control program is urgent needed in the counties identified as high risk county or very high risk county.

A comprehensive approach involving humans, animals and the environment is key to developing targeted control program [26]. The China CDC and its branches are officially required to report monthly and yearly information on leptospirosis, including the incidence, mortality and major animal hosts, to the Ministry of Health of China. We successfully identified the most influential risk factors of transmission of leptospirosis, and the risk level of counties in China. Our results could help to improve the surveillance program and control program of leptospirosis in China. The current available prevention and control program include improving in sanitation and living conditions, reducing the leptospirosis infection rate in animal hosts, and targeting vaccinations to high risk populations. Our risk maps of leptospirosis could help to optimize the allocation of public health resources, particularly in areas with limited financial and public health resources.
Fig. 4 Risk map of leptospirosis in China. a Area at risk of leptospirosis in China predicted from Maxent model. b Area at risk of leptospirosis in China predicted from logistic regression model.
Conclusion
Ecological niche modeling approach is a great tool to enhance our understanding of the drivers of leptospirosis transmission at large scale, and predict the geographic distribution of leptospirosis in China. Our risk mapping results highlighted the importance of socioeconomic and environmental variables in identify risk areas for leptospirosis. Future study should be aimed at including more meaningful risk factors and using spatial-temporal modeling to investigate the seasonal and interannual variability of leptospirosis risk in China.

Abbreviations
AUC, area under the receiver operating characteristic curve; Bio1, annual mean temperature; Bio12, annual precipitation; Bio15, precipitation seasonality (Coefficient of Variation); Bio4, temperature seasonality (standard deviation *100); China CDC, Chinese Center for Disease Control and Prevention; ENM, ecological niche modeling; FAO, Food and Agriculture Organization; GDP, gross domestic product; GIS, geographic information system; Landcover, land cover type globcover; Log_GDP, gross domestic product of county; Log_Pig, distribution of pig density from FAO; LUCC, land cover and land cover change; Maxent, maximum entropy model; River Density, average river density; ROC, receiver operating characteristic ; RS, remote sensing; WHO, World Health Organization

Acknowledgements
We thank Dr Chunxiang CAO, Dr Min XU, Dr Wei CHEN at State Key Laboratory of Remote Sensing Science, Jointly Sponsored by Institute of Remote Sensing Applications, Chinese Academy of Sciences and Beijing Normal University for valuable suggestions of data analysis and comments on draft of the paper.

Funding
This study was supported by the Special Fund for Health Sector, People’s Republic of China (No. 2013020004 and 201202021).

Availability of data and materials
The risk factors were described in methods, but the incidence data could not be shared for the policy reason.

Authors’ contributions
JZ carried out the modeling and drafted the manuscript. JL provided guidance for the whole project and polished the manuscript. XH, JZ, YW, JR and XW participated in data preparation and data analysis. FD designed this study and gave final approval of the version to be published. All authors read and approved the final manuscript.

Competing interests
All authors declare that they have no competing interests.

Consent for publication
Not applicable.

Ethics approval declarations
This study involved no human participants, human data or human tissue concern.

Author details
1Disease Control and Emergency Response Office, Chinese Center for Disease Control and Prevention, Beijing, China. 2Department of Biological Sciences, University of Notre Dame, Notre Dame, USA.

Received: 17 December 2015 Accepted: 7 June 2016
Published online: 22 July 2016

References
1. Levett P. Leptospirosis. Clin Microbiol Rev. 2001;14:296–326.
2. World Health Organization: Human Leptospirosis: Guidance for Diagnosis, Surveillance and Control. 2003. http://apps.who.int/iris/bitstream/10665/42667/1/WHO_CDS_CSR_EPH_2003.23.pdf.
3. Vega-Corredor M, Opadeyi J. Hydrology and public health: linking human leptospirosis and local hydrological dynamics in Trinidad, West Indies. Earth Perspectives. 2014;1:3.
4. Sejvar J, Bancroft E, Winthrop K, Bettiger J, Bajani M, Bragg S, Shutt K, Kaiser R, Marano N, Popovic T. Leptospirosis in “Eco-Challenge” athletes, Malaysian Borneo. Emerg Infect Dis. 2003;9:702–7.
5. Zhang C, Wang H, Yan J. Leptospirosis prevalence in Chinese populations in the last two decades. Microbes Infect. 2012;14:317–23.
6. Lau C, Clements A, Skelly C, Dobson A, Smythe L. Leptospirosis in American Samoa – Estimating and Mapping Risk Using Environmental Data. PLoS Negl Trop Dis. 2012;6(5), e1669.
7. Lau C, Lee D, Scott B, Philip W. Climate change, flooding, urbanisation and leptospirosis: fuelling the fire? Trans R Soc Trop Med Hyg. 2010;104:631–8.
8. Bharti A, Nally J, Ricaldi J, Matthias M, Diaz M, Lovett M. Leptospirosis: a zoonotic disease of global importance. Lancet Infect Dis. 2003;3:757–71.
9. Stevens K, Pfeiffer D. Spatial modelling of disease using data- and knowledge-driven approaches. Spat Spatio-Tempor Epidemiol. 2011;2(3):125–33.
10. Qian Q, Zhao J, Fang L, Zhou H, Zhang W. Mapping risk of plague in Qinghai-Tibetan Plateau, China. BMC Infect Dis. 2014;14:382.
11. Xu M, Cao C, Wang D, Kan B. Identifying environmental risk factors of cholera in a coastal area with geospatial technologies. Int J Environ Res Public Health. 2015;12:354–70.
12. Sasaki S, Suzuki H, Igarashi K. Spatial analysis of risk factor of cholera outbreak for 2003–2004 in a peri-urban area of Lusaka, Zambia. Am J Trop Med Hyg. 2008;79:414–35.
13. Gracie R, Barcellos C, Magalhães M, Souza-Santos R, Barrocas P. Geographical scale effects on the analysis of Leptospirosis determinants. Int J Environ Res Public Health. 2014;11:10366–83.
14. Lau C, Smythe L, Weinstein P. Leptospirosis- an emerging disease in travellers. Travel Med Infect Dis. 2010;8:33–9.
15. Ren J, Gu L, Liu H, Wang J, Wang J, Wu J. Study on a monitoring program regarding leptospirosis in some fore-and-after flood-affected along large rivers in Anhui province [in Chinese]. Chinese J Epidemiology. 2005;26:960–3.
16. Victoriano A, Smythe L, Giorian-Barraga N, Cavinta L, Kasai T, Limpakarnjanarat K. Leptospirosis in the Asia Pacific region. BMC Infect Dis. 2005;12:147.
17. Hijnmans R, Cameron S, Parra J, Jones P, Janvs A. Very high resolution interpolated climate surfaces for global land areas. Int J Climatology. 2005;25:1965–78.
18. Yan J, Dai B, Yu E. Leptospira and Leptospirosis, second ed [in Chinese].
   Beijing: The People’s Healthy Publishing House; 2006.
19. Schneider M, Jancloes M, Buss D, Aldighieri S, Bertherat E, Najera P, Galan D,
   Durski K, Espinal M. Leptospirosis: A silent epidemic disease. Int J Environ
   Res Public Health. 2013;10:7229–34.
20. Lee S. Application of logistic regression model and its validation for
   landslide susceptibility mapping using GIS and remote sensing data. Int J
   Remote Sens. 2005;26(7):1477–91.
21. Steven J, Miroslav D, Robert E. A maximum entropy approach to species
   distribution modeling. Machine Learning. In: Proceedings of the Twenty-
   First International Conference on. 2004. p. 655–62.
22. Hastie T, Tibshirani R, Friedman J. The elements of statistical learning: data
   mining, inference, and prediction. New York: Springer; 2001.
23. Walther G, Post E, Convey P, Menzel A, Parmesan C, Bebee T, Fromentin J,
   Guldberg O, Bairlein F. Ecological responses to recent climate change.
   Nature. 2002;416:389–95.
24. Watts N, Adger W, Agnolucci P, Blackstock J, Byass P. Health and climate
   change: policy responses to protect public health. Lancet. 2015;386:1861–914.
25. Desai S, Treck U, Lierz M. Resurgence of field fever in a temperate country:
   an epidemic of leptospirosis among seasonal strawberry harvesters in
   Germany in 2007. Clin Infect Dis. 2009;48:691–7.
26. Ricaldi J, Swancutt M, Matthias M. Current trends in translational research in
   leptospirosis. Current Opinion Infect Dis. 2013;26(5):399–403.