Using Geographic Information System-based Ecologic Niche Models to Forecast the Risk of Hantavirus Infection in Shandong Province, China

Lan Wei,[†] Quan Qian,[†] Zhi-Qiang Wang, Gregory E. Glass, Shao-Xia Song, Wen-Yi Zhang, Xiu-Jun Li, Hong Yang, Xian-Jun Wang, Li-Qun Fang,* and Wu-Chun Cao*

State Key Laboratory of Pathogen and Biosecurity, Beijing Institute of Microbiology and Epidemiology, Beijing, People's Republic of China; Shandong Center for Disease Control and Prevention, Jinan, People's Republic of China; Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland; Institute of Disease Control and Prevention of Chinese People's Liberation Army

Abstract. Hemorrhagic fever with renal syndrome (HFRS) is an important public health problem in Shandong Province, China. In this study, we combined ecologic niche modeling with geographic information systems (GIS) and remote sensing techniques to identify the risk factors and affected areas of hantavirus infections in rodent hosts. Land cover and elevation were found to be closely associated with the presence of hantavirus-infected rodent hosts. The averaged area under the receiver operating characteristic curve was 0.864, implying good performance. The predicted risk maps based on the model were validated both by the hantavirus-infected rodents' distribution and HFRS human case localities with a good fit. These findings have the applications for targeting control and prevention efforts.

INTRODUCTION

Hantaviruses are rodent-borne viruses belonging to the family *Bunyaviridae*, and can cause two kinds of human diseases, i.e., hemorrhagic fever with renal syndrome (HFRS) in Europe and Asia and hantavirus pulmonary syndrome (HPS) in the Western Hemisphere.¹ Each hantavirus species is predominately associated with a distinct or a few related rodents as its primary natural reservoirs. Humans usually acquire hantavirus infection by contact or inhalation of aerosols and secretions from infected rodent hosts. In China, Hantaan virus (HTNV) and Seoul virus (SEOV), the recognized causative agents of HFRS, are respectively associated with *Apodemus agrarius* and *Rattus norvegicus*.²

Hemorrhagic fever with renal syndrome is a significant public health problem in mainland China. The number of incident cases accounts for 90% of the total reported cases in the world.² Shandong Province is one of the most severe endemic areas in mainland China. Although integrated control measures including vaccination have been carried out for years, the disease incidence has not substantively decreased and new foci continuously emerged.

The incidence and geographic distribution of HFRS are known to be mainly determined by the population density, infection rate, and distribution of rodent hosts, which are usually influenced by the natural habitat structure including the availability of cover and burrow space, etc.^{3–8} Environmental changes in habitat could lead to an increase in virus transmission risk from infected rodents to humans.^{4–9} Studies on HPS indicated that the emergence of hantavirus was associated with weather and climatic events like El Niño Southern Oscillation.^{3,10,11} The prevalence of HFRS in China was determined by various environmental factors such as elevation, precipitation, temperature, vegetation, and soil types.^{12–14} Our previous study revealed that precipitation, humidity, and temperature was associated with the seasonal variation of HFRS incidence in Shandong Province.¹⁵ In this study, we used the surveillance data of the province to identify the risk factors

and affected areas of hantavirus infections in rodent hosts based on the ecologic niche modeling (ENM) approach.

MATERIALS AND METHODS

Collection of surveillance data. Shandong Province is located on the eastern edge of the North China Plain (N34°22'52"–N38°15'02", E114°19'53"–E122°43'), the downstream of the Yellow River. Surveillance on hantavirus infections in rodent hosts were carried out in four administrative regions including Jiaonan City, Zhangqiu City, Qingzhou City, and Jynan County (Figure 1), according to the protocol made by Chinese Centers for Disease Control and Prevention. Briefly, several sites were selected for the surveys in spring and late autumn to early winter, when there were epidemic peaks of HFRS in the province. A total of 100–150 traps per patch with peanuts as bait were placed for two to three consecutive nights.¹⁶ Universal precautions were strictly followed including wearing protective clothing, footwear, surgical masks, and gloves. After identification of species and sex, blood samples were collected and lung tissues were removed from the captured rodents and stored in liquid nitrogen until tested. The rodent carcasses were disinfected and buried deeply in the field. For unidentified rodent species in the field, the craniums were brought to the laboratory for further identification.

Lung tissues were examined for hantavirus antigens at the Center for Disease Control and Prevention of Shandong Province using indirect immunofluorescent assay (IFA) previously described by Lee and others.¹⁷ In addition, data on reported HFRS cases were retrieved from the Shandong Notifiable Disease Surveillance System.

Collection and management of environmental data. Data on eco-geographical variables (EGV) possibly contributing to transmission of hantavirus in rodent hosts were collected (Table 1).^{13,18–20} The data of each variable were converted to the study's geographic projection (Krasovsky_1940_Albers) and clipped to the study area. Elevation value was derived from a shuttle radar topography mission database with a spatial resolution of 1 km (<http://srtm.csi.cgiar.org>). Slope and aspect values were calculated from elevation using tools in ARCGIS 9.1 (Environmental Sciences Research Institute, Redlands, CA). Data on precipitation, relative humidity (RH), and temperature

*Address correspondence to Wu-Chun Cao and Li-Qun Fang, 20 Dong-Da-Jie Street, Feng-Tai District, Beijing 100071, P. R. China. E-mails: caowc@nic.bmi.ac.cn and fanglq@nic.bmi.ac.cn

[†]The first two authors contributed equally to this study.

points for training. The data for evaluation were merged with 10,000 randomly selected background points and were entered into a receiver operating characteristic (ROC) plot analysis to derive AUC.^{24,25,35} Ten random partitions rather than a single one were made to allow for assessing the average results of the algorithms and statistical testing of differences in performance.²⁵ The 10 replicates multivariate Maximum Entropy models were run with linear, quadratic, and hinge features, and tested with ROC plots. Model accuracy using AUC was characterized as 0.50–0.60 = insufficient; 0.60–0.70 = poor; 0.70–0.80 = average; 0.80–0.90 = good; 0.90–1 = excellent.^{36,37}

RESULTS

A total of 3,729 rodents of 8 species (except for 2 unidentified) were captured out of 415,890 traps at 57 trapping sites (Supplemental Appendix: Table A1). Fifty-three rodents from 33 sites were positive for hantavirus by IFA including 37 *R. norvegicus*, 11 *Mus musculus*, 4 *Rattus rattus*, and 1 *A. agrarius* (Supplemental Appendix: Tables A2 and A3).

Preliminary modeling and analyses indicated that stepwise elimination of precipitation, aspect, slope, and LSTD improved the model either because their inclusion generated a negative test gain or the AUC value was higher when the variables were removed.³⁸ The remaining 6 EGVs were associated with the presence of infected rodents, and their percent contribution was listed in Table 1. The relative importance of each

variable to the presence of infected rodents was evaluated by jackknife plots of training gain, test gain, and AUC (Figure 2). Elevation and land cover provided reasonably good fits to the training data, indicating they contained the most useful information that was not already contained in the other variables. Although RH provided only a little gain (Figure 2A), omitting it decreased the training gain considerably (see the lighter blue bars), implying it was necessary for estimating the distribution of the rodents. The response curves indicated nonlinear associations between the probability of infected rodents and variation of each EGV (Figure 3). Two types of lands, i.e., “Rainfed croplands” (code 14) and “Mosaic Vegetation/Croplands” (code 30) were highly associated with the presence of hantavirus-infected rodent hosts (Figure 3A). The probability of infected rodents increased with elevation, peaked at about 180 m, and decreased subsequently (Figure 3B). The response curves of NDVI, LSTN, and temperature show similar patterns (Figure 3C, D, and F), in which the risk for presence of infected rodent hosts was initially increased, peaked at a certain value, and declined thereafter. The risk for presence of infected rodents was highest when RH was around 55%, dropped to the lowest at 66%, increased to the second peak at 71%, and declined again (Figure 3E).

The average ROC for 10 replicate runs of the model was 0.864 with a standard deviation of 0.074, indicating the model performance was good. On the basis of the model, a map predicting the risk of hantavirus-infected rodent hosts was

FIGURE 2. Jackknife analysis results of training gain, test gain, and area under the curve (AUC). The blue, light blue, and red bar represents results of the model created with each individual variable, all the remaining variables and all variables.

FIGURE 3. Response curves for the variables related to presence of hantavirus-infected rodents. Red lines are mean values for the 10 Maxent runs and blue bars represent ± 1 SD.

constructed. To validate the model, 19 records of hantavirus-infected rodents at the field surveillance sites in 2009 were overlaid on the risk map. All points fell on the moderate and high risk areas and 13 points on the high risk area (Figure 4). The model was further validated for veracity by overlaying the locality of each human case in a different year onto the risk maps (Figure 5). The percentages of HFRS cases in predicted low, moderate, and high risk areas are listed in Table 2. For example, 41.9% cases in 2004 occurred in the predicted high risk areas, the area proportion of which was 24.3% of the province. In contrast, 29.4% cases were reported in low risk areas that accounted for 48.2% terrain. The results were

similar in other years (Table 2). Thresholds were determined using “maximum training sensitivity plus specificity logistic threshold” and “balance training omission, predicted area and threshold value logistic threshold” for classification.^{39–41} With the 10 partitions, the result gave average values for the cut points of 0.380 and 0.149, respectively.

DISCUSSION

Hemorrhagic fever with renal syndrome has been recognized as a serious public health problem in Shandong Province since the first human case was recognized in 1962. In this study,

FIGURE 4. The predicted risk map of hantavirus-infected rodents for 2009 overlaid with real infected hosts' localities from surveillance data.

we combined an eco-epidemiologic approach with GIS and remote sensing techniques to increase our understanding of the habitat of hantavirus-infected rodent hosts, and anticipate the natural foci of the disease in the province. The ENM offers a valuable approach to predict risk of an infectious disease, whenever the “absence” data are unavailable.

To predict the potential distribution of hantavirus-infected hosts within Shandong Province, we used field surveillance data from 2005 to 2008, and validated the model with the data of 2009. In addition, the predicted risk maps were consistent with the locations where human cases occurred from 2004 to 2007. On the basis of the prediction map of 2009 (Figure 4), the high risk area for hantavirus-infected hosts is clustered predominantly in the central and southern regions of the province. On the basis of the EGV's examined, land cover, elevation, NDVI, and LSTN were most strongly associated with distribution of

TABLE 2

Percentages of the human cases in low, moderate, and high predicted risk areas in 2004–2007

Year	Cases % (area %)		
	Low (< 0.149)	Moderate (0.149–0.380)	High (> 0.380)
2004	29.4% (48.2%)	28.7% (27.6%)	41.9% (24.3%)
2005	33.7% (57.0%)	33.8% (27.2%)	32.4% (15.8%)
2006	28.7% (40.8%)	24.7% (31.1%)	46.6% (28.0%)
2007	31.8% (46.1%)	28.6% (28.6%)	39.6% (25.3%)

Values in parentheses following the “Low, Moderate, High” are the risk probability and values in parentheses following the case percentage are the geographic extent of the respective areas.

hantavirus-infected rodents. Land cover usually influences the virus occurrence in hosts by affecting the contact rate between infected and susceptible rodents. Here, the suitable habitat for rodents is rain-fed croplands and mosaics of vegetation/croplands. Although a certain land cover type is often associated with the presence of hantavirus-infected rodents, it was not always the largest contributor.⁴² The NDVI correlated with the amount and productivity of vegetation and crops, which are the food source for rodents, and also significantly affected the presence of infected rodents. The LSTN rather than LSTD was also associated with the presence of infected rodents. Because *R. norvegicus* and *M. musculus* are commonly nocturnal, the nighttime land surface temperature may directly influence their activities. The nonlinearities in the effects of nighttime temperatures (Figure 3) imply that it's unlikely for hantavirus to be transmitted among rodent hosts at a too high or too low temperature.

The HTNV and SEOV, the two major causative agents of HFRS in mainland China, are respectively associated with distinct rodent species. In future studies, HTNV- and SEOV-related HFRS cases should be differentiated to explore the difference in the association between presence of hantavirus-infected rodents and related environmental factors; because

FIGURE 5. Comparison of the four predicted hantavirus risk maps from 2004 to 2007 by Maximum Entropy model and validation of the predicted risk maps by overlaying and comparing them with the human hemorrhagic fever with renal syndrome (HFRS) cases localities.

each rodent species has its own ecological tolerances.¹³ Further research based on the two hantavirus types may generate different prediction maps that provide clues to changes in the epidemic area.

Because using niche models to predict disease distributions in temporal and/or spatial domains can lead to variance of the output from each run of the model, we attempted to dampen this variability by averaging the model outputs based on 10 random replicates data sets. In fact, several studies have showed that different modeling approaches may yield substantially different predictions.^{22,42} Thus, future work could be using a suite of different modeling approaches for prediction.

Received June 2, 2010. Accepted for publication December 9, 2010.

Note: Supplemental appendix appears at www.ajtmh.org.

Acknowledgments: We thank Feng-min Hui at College of Global Change and Earth System Science, Beijing Normal University for his assistance in processing the remote sensing data. We also thank Sake J. de Vlas at Department of Public Health of Erasmus Medical Center (MC), University Medical Center Rotterdam for his comments on the manuscript.

Financial support: The study was supported by the Chinese National Science Fund for Distinguished Young Scholars (no. 30725032), Special Program for Prevention and Control of Infectious Diseases in China (no. 2008ZX10004-012, no. 2009ZX10004-720), Natural Science Foundation of China (no. 30590374, no. 30972521).

Authors' addresses: Lan Wei, Quan Qian, Xiu-Jun Li, Hong Yang, Li-Qun Fang, and Wu-Chun Cao, Beijing Institute of Microbiology and Epidemiology, State Key Laboratory of Pathogen and Biosecurity, Feng-Tai District, Beijing, Peoples' Republic of China, E-mails: weilanisme@gmail.com, qianquanmail@yahoo.com.cn, xjli@sdu.edu.cn, anni_hong@163.com, fanglq@nic.bmi.ac.cn, and caowc@nic.bmi.ac.cn. Zhi-Qiang Wang, Shao-Xia Song, and Xian-Jun Wang, Shandong Center for Disease Control and Prevention, Jinan, People's Republic of China, E-mails: wzq3678@126.com, songsong7921@163.com, and xjwang62@163.com. Gregory E. Glass, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, E-mail: ggurriel@jhsph.edu. Wen-Yi Zhang, Institute of Disease Control and Prevention of Chinese People's Liberation Army, Feng-Tai District, Beijing, Peoples' Republic of China, E-mail: zwy0419@126.com.

REFERENCES

- Schmaljohn C, Hjelle B, 1997. Hantaviruses: a global disease problem. *Emerg Infect Dis* 3: 95–104.
- Luo CW, Chen HX, 2003. Study on the factors influenced epidemic of hemorrhagic fever with renal syndrome. *Chin J Vector Biol Control* 14: 451–454.
- Engelthaler DM, Mosley DG, Cheek JE, Levy CE, Komatsu KK, Ettestad P, Davis T, Tanda DT, Miller L, Frampton JW, Porter R, Bryan RT, 1999. Climatic and environmental patterns associated with hantavirus pulmonary syndrome, Four Corners region, United States. *Emerg Infect Dis* 5: 87–94.
- Jonsson CB, Figueiredo LT, Vapalahti O, 2010. A global perspective on hantavirus ecology, epidemiology, and disease. *Clin Microbiol Rev* 23: 412–441.
- Langlois JP, Fahrig L, Merriam G, Artsob H, 2001. Landscape structure influences continental distribution of hantavirus in deer mice. *Landscape Ecol* 16: 255–266.
- Suzan G, Marce E, Giermakowski JT, Armien B, Pascale J, Mills J, Ceballos G, Gomez A, Aguirre AA, Salazar-Bravo J, Armien A, Parmenter R, Yates T, 2008. The effect of habitat fragmentation and species diversity loss on hantavirus prevalence in Panama. *Ann N Y Acad Sci* 1149: 80–83.
- Koch DE, Mohler RL, Goodin DG, 2007. Stratifying land use/land cover for spatial analysis of disease ecology and risk: an example using object-based classification techniques. *Geospat Health* 2: 15–28.
- Goodin DG, Paige R, Owen RD, Ghimire K, Koch DE, Chu YK, Jonsson CB, 2009. Microhabitat characteristics of *Akodon montensis*, a reservoir for hantavirus, and hantaviral seroprevalence in an Atlantic forest site in eastern Paraguay. *J Vector Ecol* 34: 104–113.
- Mills JN, 2005. Regulation of rodent-borne viruses in the natural host: implications for human disease. *Arch Virol Suppl* 19: 45–57.
- Gubler DJ, Reiter P, Ebi KL, Yap W, Nasci R, Patz JA, 2001. Climate variability and change in the United States: potential impacts on vector- and rodent-borne diseases. *Environ Health Perspect* 109 (Suppl 2): 223–233.
- Hjelle B, Glass GE, 2000. Outbreak of hantavirus infection in the Four Corners region of the United States in the wake of the 1997–1998 El Niño-southern oscillation. *J Infect Dis* 181: 1569–1573.
- Bi P, Tong S, Donald K, Parton K, Ni J, 2002. Climatic, reservoir and occupational variables and the transmission of haemorrhagic fever with renal syndrome in China. *Int J Epidemiol* 31: 189–193.
- Yan L, Fang LQ, Huang HG, Zhang LQ, Feng D, Zhao WJ, Zhang WY, Li XW, Cao WC, 2007. Landscape elements and hantaan virus-related hemorrhagic fever with renal syndrome, People's Republic of China. *Emerg Infect Dis* 13: 1301–1306.
- Zhang WY, Guo WD, Fang LQ, Li CP, Bi P, Glass GE, Jiang JF, Sun SH, Qian Q, Liu W, Yan L, Yang H, Tong SL, Cao WC, 2010. Climate variability and hemorrhagic fever with renal syndrome transmission in northeastern China. *Environ Health Perspect* 118: 915–920.
- Fang LQ, Wang XJ, Liang S, Li YL, Song SX, Zhang WY, Qian Q, Li YP, Wei L, Wang ZQ, Yang H, Cao WC, 2010. Spatiotemporal trends and climatic factors of hemorrhagic fever with renal syndrome epidemic in Shandong Province, China. *PLoS Negl Trop Dis* 4: e789.
- Jiang JF, Zuo SQ, Zhang WY, Wu XM, Tang F, De Vlas SJ, Zhao WJ, Zhang PH, Dun Z, Wang RM, Cao WC, 2008. Prevalence and genetic diversities of hantaviruses in rodents in Beijing, China. *Am J Trop Med Hyg* 78: 98–105.
- Lee HW, Lee PW, Johnson KM, 1978. Isolation of the etiologic agent of Korean hemorrhagic fever. *J Infect Dis* 137: 298–308.
- Glass GE, Cheek JE, Patz JA, Shields TM, Doyle TJ, Thoroughman DA, Hunt DK, Ensore RE, Gage KL, Irland C, Peters CJ, Bryan R, 2000. Using remotely sensed data to identify areas at risk for hantavirus pulmonary syndrome. *Emerg Infect Dis* 6: 238–247.
- Martínez-Freiría F, Sillero N, Lizana M, Brito JC, 2008. GIS-based niche models identify environmental correlates sustaining a contact zone between three species of European vipers. *Divers Distrib* 14: 452–461.
- Zhang WY, Fang LQ, Jiang JF, Hui FM, Glass GE, Yan L, Xu YF, Zhao WJ, Yang H, Liu W, Cao WC, 2009. Predicting the risk of hantavirus infection in Beijing, People's Republic of China. *Am J Trop Med Hyg* 80: 678–683.
- Peterson AT, 2006. Ecologic niche modeling and spatial patterns of disease transmission. *Emerg Infect Dis* 12: 1822–1826.
- Elith J, Graham CH, Anderson RP, Dudik M, Ferrier S, Guisan A, Hijmans RJ, Huettmann F, Leathwick JR, Lehmann A, Li J, Lohmann LG, Loiselle BA, Manion G, Moritz C, Nakamura M, Nakazawa Y, Overton JM, Peterson AT, Phillips SJ, Richardson K, Scachetti-Pereira R, Schapire RE, Soberon J, Williams S, Wisz MS, Zimmermann NE, 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29: 129–151.
- Gibson L, Barrett B, Burbidge A, 2007. Dealing with uncertain absences in habitat modelling: a case study of a rare ground-dwelling parrot. *Divers Distrib* 13: 704–713.
- Hernandez PA, Franke I, Herzog SK, Pacheco V, Paniagua L, Quintana HL, Soto A, Swenson JJ, Tovar C, Valqui TH, Vargas J, Young BE, 2008. Predicting species distributions in poorly-studied landscapes. *Biodivers Conserv* 17: 1353–1366.
- Phillips SJ, Anderson RP, Schapire RE, 2006. Maximum entropy modeling of species geographic distributions. *Ecol Modell* 190: 231–259.
- Phillips SJ, Dudik M, 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography* 31: 161–175.
- Thorn JS, Nijman V, Smith D, Nekaris KA, 2009. Ecological niche modelling as a technique for assessing threats and setting conservation priorities for Asian slow lorises (Primates: Nycticebus). *Divers Distrib* 15: 289–298.
- Wisz MS, Hijmans RJ, Li J, Peterson AT, Graham CH, Guisan A, Durrant NP, 2008. Effects of sample size on the performance of species distribution models. *Divers Distrib* 14: 763–773.

29. Hernandez PA, Graham CH, Master LL, Albert DL, 2006. The effect of sample size and species characteristics on performance of different species distribution modeling methods. *Ecography* 29: 773–785.
30. Saatchi S, Buermann W, Ter Steege H, Mori S, Smith TB, 2008. Modeling distribution of Amazonian tree species and diversity using remote sensing measurements. *Remote Sens Environ* 112: 2000–2017.
31. Tinoco BA, Astudillo PX, Latta SC, Graham CH, 2009. Distribution, ecology and conservation of an endangered Andean hummingbird: the violet-throated metaltail (*Metallura baroni*). *Bird Conserv Int* 19: 63–76.
32. Phillips SJ, Dudik M, Schapire RE, 2004. A maximum entropy approach to species distribution modeling. *Proceedings of the 21st International Conference on Machine Learning*. Banff, Alberta, Canada: ACM.
33. Gilberto P, Graziano R, Alessandro F, 2008. Toward improved species niche modelling: *Arnica montana* in the alps as a case study. *J Appl Ecol* 45: 1410–1418.
34. Linard C, Tersago K, Leirs H, Lambin EF, 2007. Environmental conditions and Puumala virus transmission in Belgium. *Int J Health Geogr* 6: 55.
35. Fielding AH, Bell JF, 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environ Conserv* 24: 38–49.
36. Swets JA, 1988. Measuring the accuracy of diagnostic systems. *Science* 240: 1285–1293.
37. Araújo MB, Guisan A, 2006. Six (or so) research priorities for species distribution modelling. *J Biogeogr* 33: 1677–1688.
38. Phillips SJ, 2009. A brief tutorial on Maxent, version: 3.3.1. Available at: <http://www.cs.princeton.edu/~schapire/maxent/>. Accessed August 19, 2009.
39. Cantor SB, Sun CC, Tortolero-Luna G, Richards-Kortum R, Follen M, 1999. A comparison of C/B ratios from studies using receiver operating characteristic curve analysis. *J Clin Epidemiol* 52: 885–892.
40. Cramer JS, 2003. *Logit Models: From Economics and Other Fields*. Cambridge, MA: Cambridge Univ. Press, 66–67.
41. Liu CR, Berry PM, Dawson TP, Pearson RG, 2005. Selecting thresholds of occurrence in the prediction of species distributions. *Ecography* 28: 385–393.
42. Araújo MB, Whittaker RJ, Ladle RJ, Erhard M, 2005. Reducing uncertainty in projections of extinction risk from climate change. *Glob Ecol Biogeogr* 14: 529–538.