SPATIO-STATISTICAL MODELING OF HUMAN BRUCELLOSIS USING ENVIRONMENTAL PARAMETERS: A CASE STUDY OF NORTHERN IRAN

N. Seyedalizadeh [1, *], A. A. Alesheikh [1], M. Ahmadkhani [2]

[1] Faculty of Geodesy and Geomatics Engineering, K. N. Toosi University of Technology, Tehran, Iran
nadia_alizadeh@email.kntu.ac.ir , alesheikh@kntu.ac.ir
[2] Dept. of Geography, Environmental and Society, University of Minnesota, USA– ahmad178@umn.edu

KEY WORDS: Spatial-statistical modeling, Health GIS, Maximum Entropy, Jackknife, Brucellosis, Environment

ABSTRACT:

Brucellosis is one of the most important zoonotic diseases which is endemic in Iran. This disease is considered a significant hazard to citizens' health and imposes heavy economic burdens, hence, requires a thorough control and management plan. The aims of this study are identifying the areas having the highest risk of brucellosis, as well as discovering the contributing environmental factors. The maximum entropy (MaxEnt) method was used to model the probability of brucellosis in Golestan, Mazandaran, and Guilan provinces. The possible contribution of 12 environmental parameters in this disease was also measured using the Jackknife method. The results showed that the highest risk of brucellosis is located in southern Golestan, East, and West of Mazandaran, and south of Guilan province, and moisture, slope, vegetation and elevation are the most effective environmental factors on the spatial distribution of the disease. In addition, the probability of the disease in northern Iran increases from west to east. These findings could assist the public health managers and decision-makers in organizing a more efficient public health system.

1. INTRODUCTION

Finding pathogens, knowing the pattern of spatial spread of the disease and areas at risk, are prerequisites in planning for the health system. Preventive measures can be taken by providing a system by which the outbreak can be predicted (Gordis, 2009).
The presence of natural factors such as mountains and seas on the one hand and the dependence of some pathogens on specific environmental and climatic conditions on the other, have limited some diseases to specific regions of the Earth, so geography and location in disease studies are of particular importance (Mesgari and Massoomy, 2006).
Brucellosis is a common disease between humans and animals. It is usually transmitted to humans through contaminated livestock or consumption of unpasteurized dairy products. Due to various clinical manifestations and relatively long physical complications, it is difficult to diagnose patients correctly.
Brucellosis affects several organs of the body. In acute cases, it can be a malignant disease. If left untreated, it can become a chronic illness with severe complications and in some cases, death (Cripps, 2000).
The disease is affected by economic, social and environmental conditions. Due to the physical and mental complications of patients, the costs of disease, abortion in livestock, reduced milk production and animal infertility have put a heavy cost and pressure on many countries (Dean et al., 2012; Kirk et al., 2015; Pappas et al., 2006).
It is one of the most common zoonotic diseases that has resulted in the death of more than 500 million people worldwide, with an annual death rate of approximately 370,000 per year (Dean et al., 2012; Hotez et al., 2012).
Figure 1 shows the global distribution of brucellosis and the critical situation in Iran globally. Although it is controlled in most developed countries, it is still an endemic disease in Iran (Pappas et al., 2006).
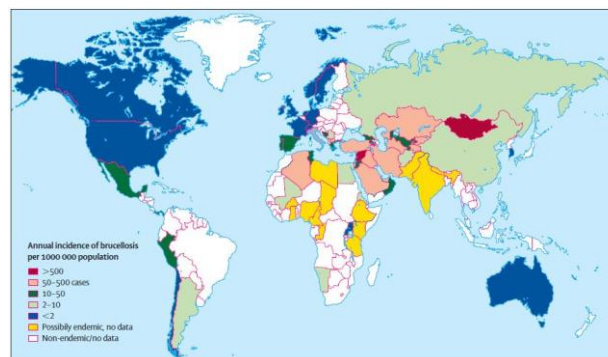

Figure 1. Global Distribution of Brucellosis (Pappas et al., 2006)

Some researchers in Iran and around the world have been interested in studying brucellosis. For example, a recent study using multiple linear regression and pearson's correlation examined the association of brucellosis with four climatic factors between 2009 and 2012. Results showed that the incidence of disease was positively correlated with temperature, wind speed and vegetation and negatively correlated with precipitation (Ahmadkhani and Alesheikh, 2017). In a study using moran's I and spatial and space-time scan, brucellosis clusters were identified in Iran from 2009 to 2012. Results showed that the disease clusters are located in west, northwest and northeast of Iran (Mollalo et al., 2014). Using GIS, spatio-temporal analysis of brucellosis was performed in Iran from 2011 to 2014. Results indicated that the highest incidence was observed in Koohrang city of Chaharmahal Bakhtiari province. The maximum and minimum incidence occurs in mid-summer and mid-winter. The pattern of disease distribution in Iran is clustered so that it is more prevalent in the mountains, especially in the cities near the Zagros Mountains. The main limitation of this study was the lack of consideration of

---

* Corresponding author

ecological parameters (Pakzad et al., 2018). The spatial distribution of brucellosis and the influence of geographical parameters with pearson correlation in Chaharmahal va Bakhtiari province during 2008-2011 was investigated. There was a significant relationship between temperature, precipitation and humidity and the incidence of brucellosis (Entezari et al., 2016).

In this study, we investigated the epidemiology of brucellosis using GIS and the MaxEnt method. To this end, the main research questions are: Where did the brucellosis clusters occur? What is the probability of brucellosis in the region? What is the relationship between brucellosis and ecological parameters in northern Iran?

2. MATERIALS AND METHODS

2.1 Research Methodology

The trend of the research method is shown in Figure 2. In this study, cases were collected with environmental parameters. Then, in the pre-processing section of probable errors, the descriptive information link to the spatial information, the fitting of the raster surface to the meteorological station points, and the correction required on the data were also performed. In the next step, spatial-statistical modeling of the disease was carried out to estimate the probability of brucellosis occurring throughout the region. Finally, the model was evaluated and the relationship of each environmental factor to the spatial distribution of the disease was determined.
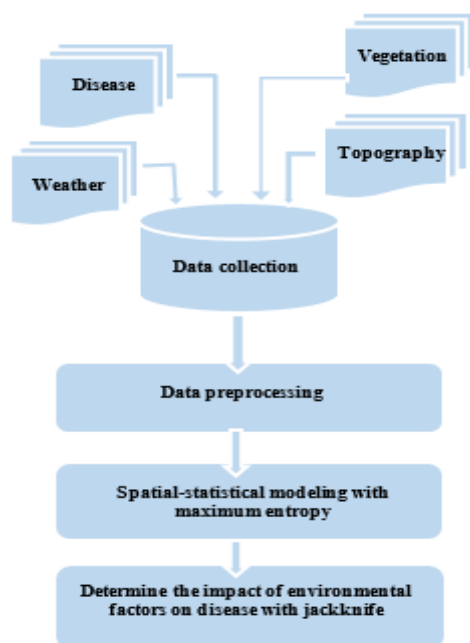
Figure 2. Flowchart of research methodology

2.2 Study area

The Caspian coastal provinces including Golestan, Mazandaran and Guilan are considered as the study area. These provinces lying within the 35°46' to 38°27' N and 48°34' to 56°19' E, and is located in the northern parts of Iran (Figure 3). According to the 2016 census, the total population of these three provinces reaches more than 768 million people and covers an area of more than 58,000 square kilometres.

Figure 3. The study area in northern Iran

2.3 Data collection and preparation

2.3.1 Disease data: The disease data used in this study were obtained from the Iranian Ministry of Health. The data included 6895 cases of human brucellosis that were received monthly for each county during April 2009 to March 2017. Figure 4 bar graph shows the cumulative number of patients per year in the study area.
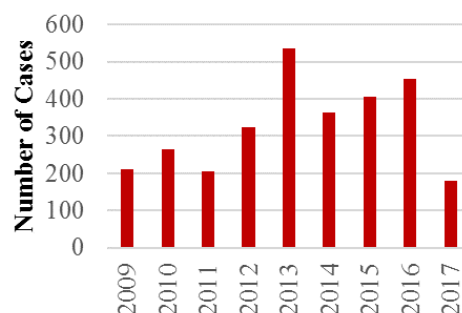
Figure 4. Cumulative number of patients per year

For further analysis, we need to calculate the incidence of the disease. For this purpose, information on population censuses was obtained from the Iranian Statistical Center. Then the incidence of disease, which is equal to the ratio of the number of patients to the total population at a given time, was calculated, usually expressed in every 100,000 people. Figure 5 shows the average incidence of disease in each city.
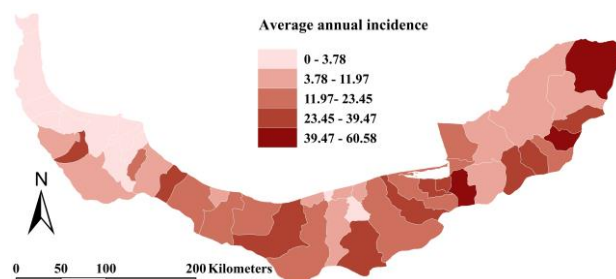
Figure 5. Average incidence of Brucellosis

MaxEnt modeling requires the presence of disease sites. For this purpose, the average annual incidence of each city in ArcGIS software was classified into 5 classes. Then for the second to fifth grades, the presence of brucellosis in the center of the city was considered from 1 to 4 points, respectively.

2.3.2   Environmental data: Environmental data were used to perform spatial-statistical modeling. These data included elevation, slope, evaporation, temperature, humidity, precipitation, wind speed and vegetation.

Satellite images were used to investigate the effect of vegetation on the prevalence of brucellosis. Due to the 8-year period, the area of study area, high number of images and high volume, normalized vegetation index of Modis satellite images were used. These images, with a spatial resolution of 250 m, were sampled at 16-day intervals from http://modis.gsfc.nasa.gov during the period of 2009 to 2017 in the central month of each season. NDVI values for each city were averaged over all seasons by performing multiple analyzes on images obtained using ArcGIS software (Fig. 6).
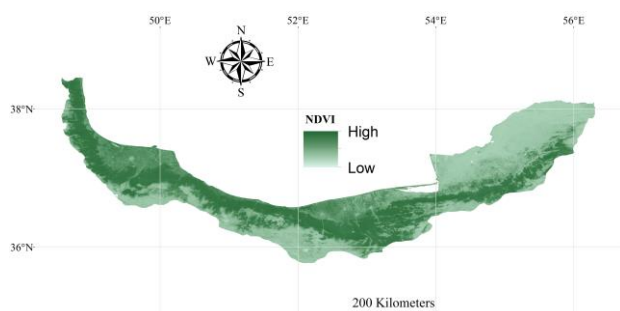

Figure 6. Average NDVI for the years 2009-2017

In order to investigate the effect of climate factors on the prevalence of brucellosis, data from synoptic stations of Golestan, Mazandaran and Guilan provinces during 2009 to 2017 were used. This information was collected monthly from the Iranian Meteorological Organization. Because the meteorological data were point-based, interpolation was performed for a continuous surface. For this purpose, Inverse Distance Weighting (IDW) interpolation was used for each meteorological factor (Fig. 7).
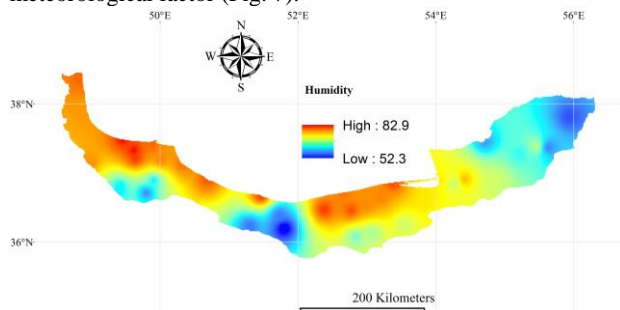

Figure 7. Average Humidity for the years 2009-2017

Topographic data were used to produce elevation and slope maps. For this purpose, the digital elevation model extracted by SRTM satellite images with a resolution of 30 m was used to obtain altitude data in the study area. The height and slope raster were cut according to the national divisions at the level of the study area (Fig. 8). Then the environment layers in the ArcGIS software were converted to ASCII format to enter the MaxEnt software.
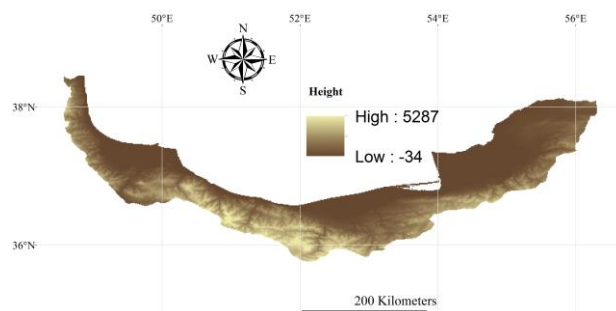

Figure 8. Digital elevation model of northern Iran

2.4 Proposed method

2.4.1   MaxEnt: MaxEnt is a machine learning approach that originates from statistical mechanics and theoretical information and has been successfully applied in a wide range of applications, including modeling the distribution of vector species that cause disease. The principles of this method go back to maximum entropy or proximity to reality. Shannon's predictable distribution describes entropy as the amount of chance that an event occurs (Shannon, 1951).

The maximum entropy model is a statistical model and in order to obtain species distribution it must establish a link between the data model and the ecological model. In the maximum entropy model, the presence points of the disease from X_1 to X_M are used to obtain an unknown distribution. This modeling program was written by Steven Phillips in 2006. This method does not assume that observations are independent because they can integrate interactions between variables and use continuous and classified data and integrate a number of features to constrain the disease area. Properties such as linear (one continuous variable), square (power of 2 continuous variables), multiplicative (multiplied by two continuous variables), threshold and binary. Modelers need to choose the features they think apply the most constraint. Benefits of maximum entropy can be successfully observed in samples with low numbers (less than 100 observations) and even observations as high as 10 records (Phillips et al., 2006).

2.4.2   Jackknife: Jackknife is one of the major re-sampling methods. The purpose of the jackknife method is to estimate the error rate and standard error of an estimator. In this method there are no parametric assumptions about the distribution of society. Suppose the parameter estimation target θ and $\hat{\theta} = T(x_1, \ldots, x_n)$ is the estimator for this parameter. In the Jackknife method one of the data is extracted from the original sample, then the estimator value is calculated. Repeat this until all the sample data is extracted once, followed by a sequence of n replicates. The Jackknife estimator for the community parameter is then calculated from the equation (1). Also, the statistical error estimates are derived from the equation (2) and the standard error estimated from the equation (3) (Efron, 1982).

$$\theta_{jack} = \frac{1}{n}\sum \theta_{-i} \tag{1}$$

$$bias_{jack}(\theta) = (n-1)(\theta_{jack} - \theta) \tag{2}$$

$$SE_{jack}(\theta) = \frac{n-1}{n}\sum (\theta_{-i} - \theta_{jack})^2 \tag{3}$$

3. RESULTS

Maximum entropy method was used to perform spatio-statistical modeling and to determine the probability of Brucellosis in Caspian region. Figure 9 shows the map of this modeling. Maximum entropy model showed that the highest probability of occurrence of brucellosis is located in the south of Golestan province, so that from west to east of the province in all southern regions the probability of disease is high (probability> 0.8). East and west of Mazandaran and south of Guilan province are also more likely to have brucellosis.

ROC curve was used to evaluate this model. The area under the curve (AUC) was 0.88 for the training data and 0.79 for the test data.
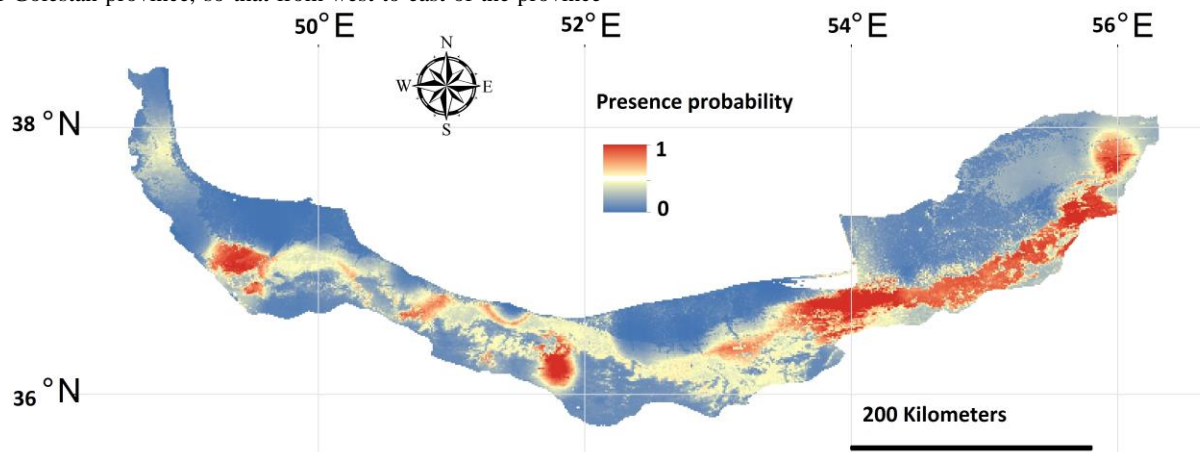

Figure 9. Predicted distribution of Brucellosis in northern Iran using MaxEnt model

The Jackknife result is shown in Figure 10. In this figure, the vertical axis represents the independent variable used in the modeling, and the horizontal axis is the extent to which these variables affect the model, with and without any other variables. The Jackknife test showed that humidity, land slope, vegetation and elevation are the most effective environmental factors on the spatial distribution of Brucellosis. It also states that the removal of the vegetation parameter from the modeling process reduces the accuracy of the model more than the other parameters.
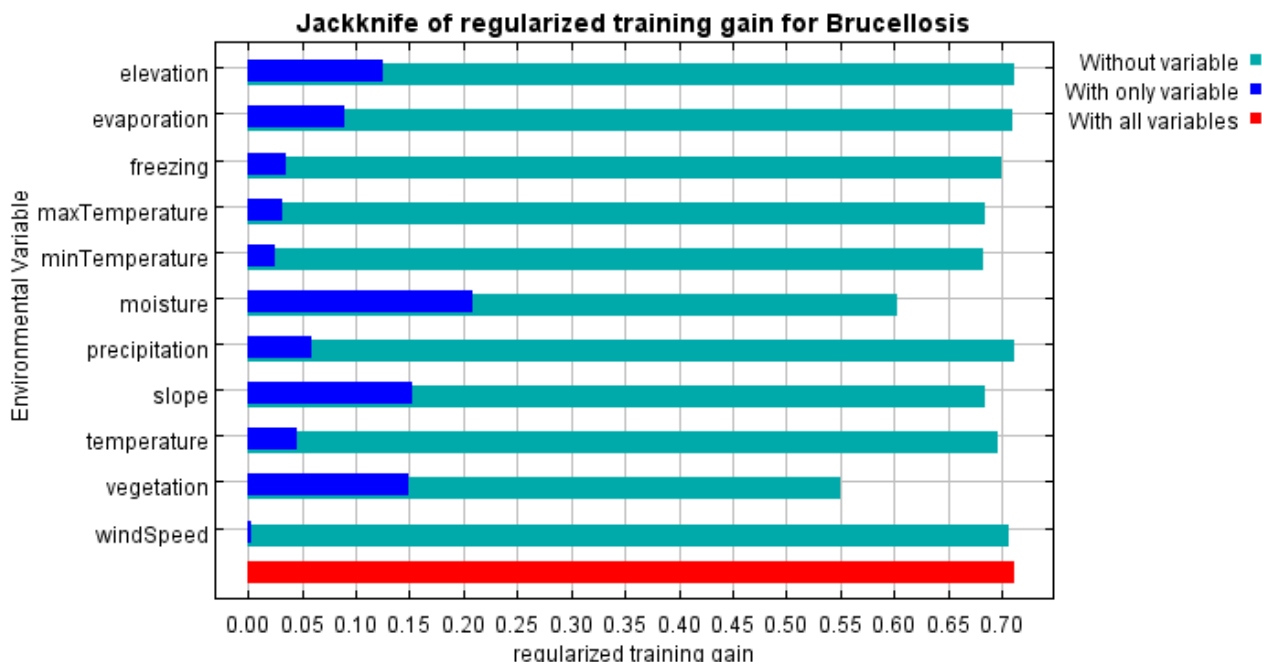

Figure 10. Jackknife test in MaxEnt model for Brucellosis.

4. CONCLUSION

From this research, it can be concluded that in the northern region of Iran, the risk of brucellosis in the east is more than the west, and the mountainous areas are more at risk than coastal areas. This information can be used by public health managers and decision makers to prevent and control this dangerous disease. Increasing the awareness of people at risk is also a very effective way to prevent the disease.

ACKNOWLEDGEMENTS

REFERENCES

Ahmadkhani, M., & Alesheikh, A. (2017). Space-time analysis of human brucellosis considering environmental factors in Iran. *Asian Pacific Journal of Tropical Disease*, *7*(5), 257–265.

Cripps, P. J. (2000). Veterinary education, zoonoses and public health: a personal perspective. *Acta Tropica*, *76*(1), 77–80.

Dean, A. S., Crump, L., Greter, H., Schelling, E., & Zinsstag, J. (2012). Global Burden of Human Brucellosis: A Systematic Review of Disease Frequency. *PLoS Neglected Tropical Diseases*, *6*(10), e1865.

Efron, B. (1982). *The Jackknife, the Bootstrap, and Other Resampling Plans - Bradley Efron - Google Books*. *Siam*.

Entezari, M., Moradpour, S., & Amiri, M. (2016). Spatial distribution and the impact of geographical factors on brucellosis in Chaharmahal and Bakhtiari Province, Iran. *International Journal of Epidemiologic Research*, *3*(2), 98–105.

Gordis, L. (2009). Epidemiology (4th edn). *Philadelphia, PA: Saunders Elsevier.*

Hotez, P. J., Savioli, L., & Fenwick, A. (2012). Neglected Tropical Diseases of the Middle East and North Africa: Review of Their Prevalence, Distribution, and Opportunities for Control. *PLoS Neglected Tropical Diseases*, *6*(2), e1475.

Kirk, M. D., Pires, S. M., Black, R. E., Caipo, M., Crump, J. A., Devleesschauwer, B., … Angulo, F. J. (2015). World Health Organization Estimates of the Global and Regional Disease Burden of 22 Foodborne Bacterial, Protozoal, and Viral Diseases, 2010: A Data Synthesis. *PLOS Medicine*, *12*(12), e1001921.

Mesgari, M. S., & Massoomy, Z. (2006). Detection of lung cancer clusters in Iran using GIS and statistical analyses. *International Journal of Geoinformatics*, *2*(4).

Mollalo, A., Alimohammadi, A., & Khoshabi, M. (2014). Spatial and spatio-temporal analysis of human brucellosis in Iran. *Transactions of The Royal Society of Tropical Medicine and Hygiene*, *108*(11), 721–728.

Pakzad, R., Pakzad, I., Safiri, S., Shirzadi, M. R., Mohammadpour, M., Behroozi, A., … Janati, A. (2018). Spatiotemporal analysis of brucellosis incidence in Iran from 2011 to 2014 using GIS. *International Journal of Infectious Diseases*, *67*, 129–136.

Pappas, G., Papadimitriou, P., Akritidis, N., Christou, L., &

Tsianos, E. V. (2006). The new global map of human brucellosis. *Lancet Infectious Diseases*.

Phillips, S. J., Anderson, R. P., & Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, *190*(3–4), 231–259.

Shannon, C. E. (1951). Prediction and Entropy of Printed English. *Bell System Technical Journal*, *30*(1), 50–64.