INTERNATIONAL CONFERENCE ON RESEARCH IN SCIENCE AND TECHNOLOGY
14 December 2015, Kualalumpur –Malaysia

Assessment of environmental factors associated with rural endemics of Leptospirosis in Guilan Province, Iran

Ali Mohammadinia
Student at Faculty of Geodesy and Geomatics Eng., K. N. Toosi University of Technology, Valiasr Street, Mirdamad Cross, Tehran 19967-15433, Iran
ali.mohamadinia@email.kntu.ac.ir

Abbas Alimohammadi
Teacher at Faculty of Geodesy and Geomatics Eng., K. N. Toosi University of Technology, Valiasr Street, Mirdamad Cross, Tehran 19967-15433, Iran
Alimoh_abb@kntu.ac.ir

Roya Habibi
Student at Faculty of Geodesy and Geomatics Eng., K. N. Toosi University of Technology, Valiasr Street, Mirdamad Cross, Tehran 19967-15433, Iran
rhabibi@mail.kntu.ac.ir

Abstract

Leptospirosis is a globally emerging zoonotic disease depending on environmental conditions with social and economic burdens on rural areas. Regarding to worldwide prevalence, animals and humans who are in contact with contaminated water or soil especially at temperate and tropical regions with high rainfall and floods are the most significant targets of this disease. Major hosts of infection are wild and domestic animals that hold pathogenic leptospires. Exposure of humans to the urine of infected animals directly or indirectly is the main way of transmission occurrence. Due to suitable conditions, leptospirosis is endemic in north of Iran (Caspian Region) include Mazandaran, Golestan and especially Guilan provinces. Guilan included a lot of paddy fields and farmers and had the most reported cases among all provinces during 2009-2011. In this study, we are going to assess the influence of 9 environmental parameters on leptospirosis occurrence throughout 109 rural districts of Guilan during three years. For this purpose, spatial statistics approaches such as Geographically Weighted Regression (GWR) and Ordinary Least Squares (OLS) were utilized to investigate correlation between number of disease cases and environmental factors. However, assessment results showed better success of GWR versus OLS as regressions were performed. Finally, due to the economic and health burdens of leptospirosis on rural areas, understanding of effective parameters and their influences are so helpful for decision makers and privacy policies in prevention planning.

Keywords: Infectious disease, GIS, Spatial Statistics, Leptospirosis

Introduction

Leptospirosis is an important global public health problem with significant mortality and morbidity especially in developing countries. Over a hundred years since its discovery at coal mine worker in Japan, it has been little understood for some reasons (Ko and Picardeau, 2009). Regarding to World Health Organization (WHO), the number of human cases is not known precisely and disease is underreported. But, available reports indicate annual incidence rates vary from 0.1-1 per 100,000 in temperate climates to 10-100 per 100,000 in the humid tropics and more than 100 per 100,000 during outbreaks and in high exposure risk groups (WHO, 2015). Leptospirosis is highly dependent on environmental factors and would be occurred in rural or urban contexts with temperate and tropical climates (Wiwanitkit, 2006). Most of wild and domestic mammalians are major hosts of leptospirosis especially dogs and rats. Major way of transmission is contact with existent pathogenic bacteria in the urines of infected animals either by direct contact or through contaminated soil or water (Faine, 1994). Due to similar behavior with other diseases such as Dengue fever, lack of early diagnosis or on time treatment would lead to death in some cases. Moreover, it is considered as an occupational disease around the world especially in south east of Asia (Victoriano et al, 2009).

Currently, leptospirosis has been posed significant public health problem in some provinces of Iran endemically in Caspian Region with temperate and humid climate. According to acquisition data which are provided by Iran’s Ministry of Public Health surveillance system, Guilan as the second province in rice production with many rural areas and rice fields, had the most reported cases during 2009-2011 among provinces. Clinical aspects of human and especially animal leptospirosis have been studied in various provinces of Iran including the Guilan (Mansour-Ghanaei et al., 2005), Bushehr (Esmaeli et al., 2014), Golestan (Dashliboron et al., 2013).

Nevertheless, studies mostly overlooked the disease clinically and concentrated on characteristics of the individual patients and reservoirs. Regarding to our researches, no efforts have been made relating to spatial statistics about leptospirosis in this region or any parts of Iran. So, the main objective of this study is to analyze the correlation between 9 environmental parameters and counts of leptospirosis throughout the province by two popular regression approaches, GWR and OLS, in the context of comparison and analysis the results of both regressions.

Materials and methods

Study area

Guilan province have 1,231,933 male and 1,248,941 female populations according to 2011 enumeration, covers the area of 14,041 square kilometers and includes 109 rural districts (Figure 1). It is surrounded by Caspian Sea from north and Alborz Mountains from south with minimum height of -90 m and maximum height of 3687 m according to raster image of Modis satellite. Regarding to agricultural and littoral conditions, mean areas of the province included farm lands and have temperate and humid climates.

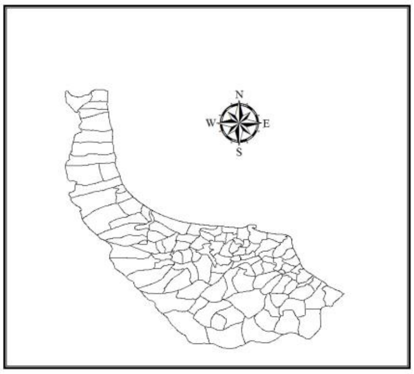

Figure 1. Rural districts of Guilan province

Data collection

Seven environmental parameters including minimum of daily temperature, maximum of daily temperature, average temperature, precipitation, average humidity, average wind speed and average vapor pressure were gained from National Weather Organization of Iran (NWOI). They were obtained from 12 synoptic weather stations which are distributed throughout Guilan province. Statistical Center of Iran (SCI) provided estimated counts of population census during 2009–2011. Leptospirosis reported cases were collected from Centers for Disease Control and prevention (CDCs) of 109 rural districts which are in cooperate with Ministry of Public Health of Iran (MPHI).

Data preparation

Census counts were imported to Microsoft Excel 2013 software to be prepared. Reported cases without address and deterministic laboratory confirmation were omitted for further analysis. As a normalization, number of reported cases per 100,000 were computed for each of 109 rural districts. Annual average of environmental factors for each station were computed. Also, Digital Elevation Model (DEM) of province was obtained utilizing raster image of Modis satellite. Eventually, environmental factors were interpolated and extracted to the centroid of any rural districts using ArcGIS 10.1 software.

Methods

Regression is the most common statistical modelling method that estimates relationship between dependent variable and some independent variables. Investigating spatial non-stationary in relationships was one of the reasons led to innovation of global and local regressions (Charlton et al., 2009). OLS and GWR are both popular and good examples of global and local regressions, respectively, that have been applied in many sciences and different researches. In the following, we will continue to explain about these two methods in more details.

Ordinary Least Squares (OLS)

OLS can be utilized for both single and multiple explanatory variables. Simplest relationship is a line of best fit as follows ‘Y=a+bX’, where ‘a’ denotes the intercept and ‘b’ is the regression coefficient. Major hypothesis in ordinary least squares regression is that explanatory variables are measured without error. As regards, when we have randomness in parameters, this assumption is often found incorrect (Rogerson, 2001). Fitness of model can be concluded from comparison and the differences (residuals) of observations and expected values of dependent variable. Squaring these residuals will omit negative ones and sum of all squared residuals indicate how well model fits the data (Craven and Islam, 2011). It is presented in equation (1)

$$Y_i = a_0 + \sum_k a_k x_k + \varepsilon_i$$

$$a = (x^t x)^{-1} x^t y \qquad (1)$$

Where ‘a’ represents the vector of global parameters to be estimated, ‘x’ is a matrix of independent variables with the elements of the first column set to 1, and ‘y’ represents a vector of observations on the dependent variable.

Geographically Weighted Regression (GWR)

GWR has been used in many science fields. It is a statistical method that models processes which are not constant and vary over space (Wheeler, 2014). Despite global regressions, estimation of regression coefficients in this method are locally. One of the main deficiencies of global regressions such as OLS is that the processes assumed to be constant over space which led to innovation of GWR technique (Fotheringham et al., 2003). The model is presented in equation (2) (Fotheringham, 2003):

$$y_i = a_0(u_i, v_i) + \sum_k a_k(u_i, v_i)\, x_{ik} + \varepsilon_i \quad (2)$$

where (u_i,v_i) denotes the coordinates of the i^{th} point in space and $a_k(u_i,v_i)$ is a realization of the continuous function $a_k(u,v)$ at point i. It can be seen that global model is a special case of the GWR model.

IDW

Different interpolation methods have been produced and used in researches with various advantages and disadvantages. Inverse Distance Weighting (IDW) is probably the simplest technique which computes unknown values based on utilizing known values. According to the assumptions of this method, neighbor features which are nearer have more effects in applying interpolation to unknowns (Zimmerman et al., 1999). The equation (3) presents (Shepard, 1968):

$$Z(x) = \sum_{i=1}^{n} w_i z_i \Big/ \sum_{i=1}^{n} w_i \quad (3)$$

$$w_i = d_i$$

Where 'Z(x)' is predicted value at an interpolated point, 'Z_i' is predicted value at known point i, 'n' is total number of known points, 'd_i' is distance between point i and prediction point, 'w_i' is weight assigned to point i.

Results and discussion

According to a lot of rice fields and workers in Guilan, leptospirosis is considered as an occupational disease in this province. It is an endemic disease in Guilan, and occurring annually in definite months when rice field farmers begin and end their work. Gathering and analyzing reported cases from CDCs of this province shows men are almost two times magnitude of women at risk on leptospirosis (Figure 2).

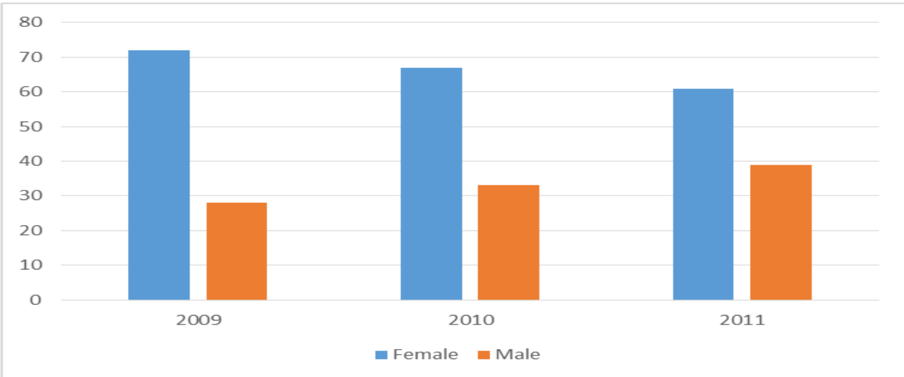

Figure 2. Percent of leptospirosis occurrence in both male (blue) and female (red).

Appling Microsoft Excel 2013, number of leptospirosis monthly cases for all of rural districts were computed and were divided to the population of rural districts multiply 100 000 for normalization. So, leptospirosis incidence rates were concluded for each month during 2009-2011 (Figure 3). As you can see, leptospirosis prevalence were happened almost from April to November and in 2010 the most cases have been reported especially at Jun. Our results were exactly coinciding the reality and showed that leptospirosis is a periodic disease and at the other months we have no report in the study area.

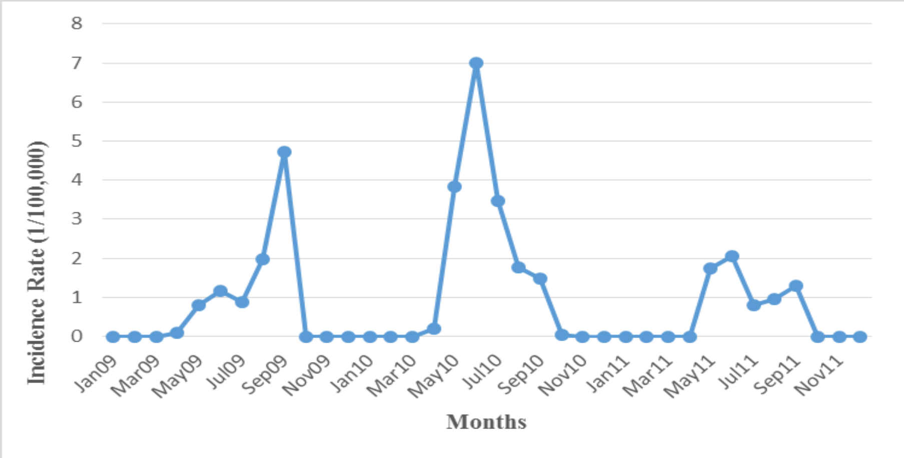

Figure 3. Leptospirosis monthly incident rate for all of rural districts of Guilan

Annual average of seven environmental parameters including minimum of daily temperature, maximum of daily temperature, average temperature, precipitation, average humidity, average wind speed and average vapor pressure from 12 synoptic weather stations which are distributed throughout Guilan province, were used for interpolation. Also, a free raster image of Modis satellite were used for producing DEM and slope through the whole province. All these results were extracted to centroid of each rural district for further analysis. However, rural districts with maximum and minimum values of environmental parameters were listed in table 4.

Table 4. Rural districts include minimum and maximum values of environmental parameters.

Rural district/ Parameters	Year minT	Year maxT	Year avgT	Year precip	Year avgHum	Year Wspeed	Year avgVpr	Height	Slope
Max	Chahar Farizeh	Kaleshtar	Ali Abad Ziba Kenar	Kaleshtar	Homeh	Jirandeh	Ali Abad Ziba Kenar	Licharaki Hasan Rood	Mollasara
Min	Daylaman	Daylaman	Daylaman	Chahar Farizeh	Jirandeh	Layalestan	Jirandeh	Shoeil	Oshian

For understanding the correlation of environmental parameters and detecting which parameters have more effects on the number of disease cases, data analysis was applied by Microsoft Excel 2013. It can be considered from figure 5 that during three years, height, slope and annual wind speed had always negative correlation with the number of disease cases and annual minimum temperature, annual average temperature, annual average precipitation, annual average humidity and annual average vapor pressure had always positive correlation. As the only parameter that had a different behavior during three years, annual maximum temperature was negatively correlated in 2009 and 2010 and was positively correlated in 2011. Regardless of negative or positive signs, height, slope, annual average humidity and annual average precipitation were more correlated parameters all over the period and annual average temperature was the least correlated.

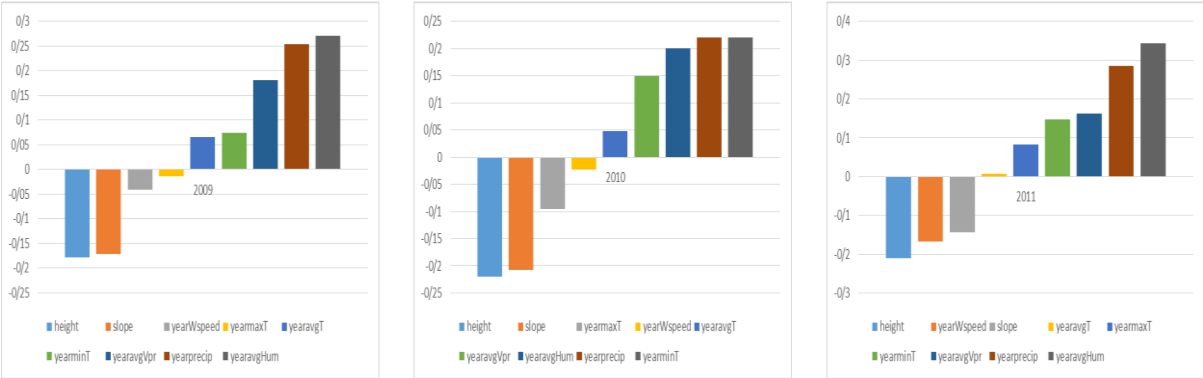

Figure 5. Correlation of parameters with number of disease cases during 2009-2011

By using GWR in the ArcToolbox of ArcGIS 10.1 software, number of disease cases was selected as the dependent variable and each of environmental parameters as the explanatory variable separately. In this model, the kernel type (which specifies if the kernel is constructed as a fixed distance (Fixed), or if it is allowed to vary in extent as a function of feature density (adaptive)) was selected adaptive. Also, 'Band Width' specifies how the extent of the kernel should be determined. Three kinds of band width exist. At 'AICc', the extent of the kernel is determined using the Akaike Information Criterion (AICc). 'CV' indicates that the extent of the kernel is determined using Cross Validation. The extent of the kernel is determined by a fixed distance or a fixed number of neighbors at 'Bandwidth Parameter'. As the suitable band width is not known, AICc was selected as the band width all over the study. For 2009, 2010 and 2011 results are shown in tables 6, 7 and 8 respectively. Annual maximum temperature, annual average temperature and annual average humidity are three parameters that had no effects on construction of GWR and could not contribute to model leptospirosis. So, these parameters were not included the tables. According to the factor of R2 in GWR model, height parameter was the best coefficient and annual average vapor pressure was the most unreliable parameter among others during 2009-2011. Residual squares are low in 2009, 2011, and 2010 consecutively. Also, different values of neighbors are for selection of adaptive kernel type. Moreover, in 2010, the residual squares are so high which is not favorable.

Table 6. Results of GWR model for 2009

Cases/2009	Year minT	Year precip	Year Wspeed	Year avgVpr	Height	Slope
Neighbors	48	51	35	108	22	35
Residual Squares	99.86	97.60	95.92	133.06	79.30	98.99
AICc	321.05	318.10	324.67	339.14	324.07	331.38
R^2	0.32	0.34	0.35	0.10	0.46	0.33

Table 7. Results of GWR model for 2010

Cases/2010	Year minT	Year precip	Year Wspeed	Year avgVpr	Height	Slope
Neighbors	41	89	39	106	30	35
Residual Squares	3587.03	4486.07	3690	4845	3101.82	3362.77
AICc	715.57	724.97	719.09	731	709	715.65
R^2	0.31	0.14	0.29	0.07	0.41	0.36

Table 8. Results of GWR model for 2011

Cases/2011	Year minT	Year precip	Year Wspeed	Year avgVpr	Height	Slope
Neighbors	84	91	72	108	41	85
Residual Squares	248.93	243.19	241.10	262.01	219.39	246.28
AICc	411.56	406.93	409.78	413.02	410.82	411.47
R^2	0.11	0.13	0.14	0.07	0.22	0.12

OLS has been selected in comparison to GWR. Results of OLS model by utilizing ArcGIS 10.1 software were listed in table 9, 10 and 11 for 2009-2011 separately. Annual maximum temperature and annual average temperature were used in the OLS model but because the R2 factor were equal to zero, both of them have been omitted from results the same as GWR model. In contrast to GWR, annual average humidity showed good correlation to the number of disease cases on OLS model. By looking the values of R2 in the tables for both OLS and GWR, it is obvious that GWR results indicate a very better improvement in modelling leptospirosis disease.

Table 9. Results of OLS model for 2009

Cases/2009	Year minT	Year precip	Year avgHum	Year Wspeed	Year avgVpr	Height	Slope
AICc	344.86	342.36	341.28	349.42	345.95	346.09	346.35
R^2	0.04	0.06	0.07	0.00	0.03	0.03	0.02

Table 10. Results of OLS model for 2010

Cases/2010	Year minT	Year precip	Year avgHum	Year Wspeed	Year avgVpr	Height	Slope
AICc	732.69	732.74	733.71	738.10	735.67	732.72	733.37
R^2	0.04	0.04	0.03	0.00	0.02	0.04	0.04

Table 11. Results of OLS model for 2011

Cases/2011	Year minT	Year precip	Year avgHum	Year Wspeed	Year avgVpr	Height	Slope
AICc	416.78	409.88	405.43	416.03	416.23	414.20	416.85
R^2	0.02	0.08	0.11	0.02	0.02	0.04	0.02

Numerous assumptions and limitations have been considered in this study. According to collected data from CDCs, only the home addresses of patients were available and the location of disease occurrence were not definite. So, we considered the addresses as point features and the analysis were performed on them. It means that we ignored the distance of disease occurrence location and patient addresses just because the homes of farmers and workers are almost near rice fields. Also, we omitted the urban patients in GIS ready process for two reasons. First, because leptospirosis occurs in rural areas especially rice fields. Second, urban reported patients were so few to contribute the results. Moreover, we considered the centroid of each rural district for extracting the results of interpolation and other analyses which is untrue. On the other word, we suppose that for example one temperature value throughout each rural district which is not real because parameters vary along an area. Finally, number of disease cases were little and the approaches would have more suitable results on massive data.

Conclusion

It was discovered that leptospirosis would be occurred at special months from April to November simultaneously with paddy season. So, this period is so important in leptospirosis prevalence and needs more preventive proceedings. Applying GWR and OLS regression indicated more success of local regressions against global. Also, results expressed that some parameters had more effect on leptospirosis prevalence. At some parameters, results had two or even three degrees of magnitude improvement in modelling leptospirosis by GWR as were mentioned before. It is true that early diagnosis would lead to treatment and few fatality cases reports every year, but, leptospirosis have socioeconomic consequences by making workers out of work and reducing force work especially at paddy season. On the other hand, the cost of health cares and privacy policies poses a burden on workers or hospitals and makes leptospirosis a significant health problem in Guilan province. At the end, we hope studies like this help decision makers and privacy policies to plan for controlling and prevention of leptospirosis occurrence and reduce annual reported cases. Future studies could be performed on investigating more parameters and their effects on leptospirosis especially socioeconomic factors. Also, more statistical methods would be applied for better understanding hidden aspects of this disease in more details.

References

Ko, A. I., Goarant, C., & Picardeau, M. (2009). Leptospira: the dawn of the molecular genetics era for an emerging zoonotic pathogen. Nature Reviews Microbiology, 7(10), 736-747.

World Health Organization (WHO), 2015. Leptospirosis Burden Epidemiology Reference Group (LERG). Available at: http://www.who.int/zoonoses/diseases/lerg/en/index2.html.

Wiwanitkit, V. (2006). A note from a survey of some knowledge aspects of leptospirosis among a sample of rural villagers in the highly endemic area, Thailand. Rural remote health, 6(1), 526.

Faine S., 1994. Leptospira and leptospirosis. CRC, London.

Victoriano, A. F., Smythe, L. D., Gloriani-Barzaga, N., Cavinta, L. L., Kasai, T., Limpakarnjanarat, K., ... & Adler, B. (2009). Leptospirosis in the Asia Pacific region. BMC Infectious Diseases, 9(1), 147.

Mansour-Ghanaei, F., Sarshad, A., Fallah, M. S., Pourhabibi, A., Pourhabibi, K., & Yousefi-Mashhoor, M. (2005). Leptospirosis in Guilan, a northern province of Iran: assessment of the clinical presentation of 74 cases. Medical science monitor, 11(5), CR219-CR223.

Esmaeli, H., Moradi Geravand, M., Hamedi, M., Sharifi, A., & Kalateh Rahmani, H. (2014). Outbreak of leptospirosis in cattle, sheep and goats of Hilla wetland in Bushehr in 2003. Veterinary Journal.

Dashliboron, O. J., Hassanpour, A., & Abdollahpour, G. R. (2013). Serological Study of Leptospirosis in Horses in Gonbad, Iran.

Charlton, M., Fotheringham, S., & Brunsdon, C. (2009). Geographically weighted regression. White paper. National Centre for Geocomputation. National University of Ireland Maynooth.

Rogerson, P. (2001). Statistical methods for geography. Sage.

Craven, B. D., & Islam, S. M. (2011). Ordinary least-squares regression (pp. 224-228). SAGE Publications.

Wheeler, D. C. (2014). Geographically weighted regression. In Handbook of Regional Science (pp. 1435-1459). Springer Berlin Heidelberg.

Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2003). Geographically weighted regression. John Wiley & Sons, Limited.

Zimmerman, D., Pavlik, C., Ruggles, A., & Armstrong, M. P. (1999). An experimental comparison of ordinary and universal kriging and inverse distance weighting. Mathematical Geology, 31(4), 375-390.

Shepard, D. (1968). A two-dimensional interpolation function for irregularly-spaced data. In Proceedings of the 1968 23rd ACM national conference (pp. 517-524). ACM.