PREDICTION OF ONCOMELANIA HUPENSIS DISTRIBUTION BASED ON REMOTE SENSING AND SPATIAL ANALYSIS TECHNOLOGY IN DONGTING LAKE REGION OF CHINA

Zhaoyan Liu[ab], Yuanyuan Jia[a], Hao Wu[a], Lingling Ma[a], Yonggang Qian[a], Lingli Tang[a],
[a]Earth observation Technology Application Department, Academy of Opto-electronics, CAS, Beijing, China
[b]University of Chinese Academy of Sciences, Beijing, China
(Email:zyliu@aoe.ac.cn)

KEY WORDS: Schistosomiasis japonica; Oncomelania hupensis; Remote sensing; Spatial analysis

ABSTRACT: The amphibious snail Oncomelania hupensis(O. hupensis) is the only intermediate host of Schistosomiasis japonica, which is a parasitic disease of considerable public health and economic significance. The survival of O. hupensis is governed by climatic and environmental factors, including vegetation, temperature, soil type and water level, etc. In this study, the environmental factors were derived from Landsat TM remote sensing image, and the relationship between environmental factors and the density of O. hupensis was analyzed by a multiple linear regression model. Although stepwise regression demonstrated that O. hupensis densities of live snails related significantly to the modified soil adjusted vegetation index (MSAVI), wetness index and land surface temperature (LST), the low correlation coefficient (0.3619) indicated that some important factors related to the abundance of snail had not yet been taken into account in the multiple linear regression model. Then, the great spatial analysis of the Geographic Information System (GIS) was used as a new tool to study the relationship between the distribution of O. hupensis and the surveillance of O. hupensis habitas. Spatial analysis of the regression residual was investigated by the semi-variogram method, and the spatial variation of O. hupensis density attributed to the spatial autocorrelation was estimated by ordinary kriging. Therefore, remote sensing and spatial analysis were both employed to predict the distribution of O. hupensis. Following this approach, O. hupensis in Dongting Lake region, China was predicted and the prediction results are validated with field data. The prediction results indeed improved considerably.

1. INTRODUCTION

Schistosomiasis japonica is one of many zoonotic parasitic diseases in the south of China. The central Government has, however, noted the serious situation and the national disease control programme has recently instituted a high-priority approach with regard to the major, communicable diseases in the area, in the particular schistosomiasis, HIV/AIDS, tuberculosis(Jiang, Wang et al. 2002; Chen, Wang et al. 2005; Li, Zhao et al. 2005; Utzinger, Zhou et al. 2005). In spite of great efforts and the remarkable progress made over the past 50 years since the inception of the national programme on schistosomiasis control, hyper-endemic areas still remain in lake and marshland regions, as well as in some of the mountainous regions in seven provinces of southern China, especially, in the lake and marshland regions, for example, Dongting Lake(Yuan, Jiang et al. 2002; Zhou, Wang et al. 2005). Therefore, it is necessary to predict the distribution of S. japonicum for sustained control of schistosomiasis, under the current situation.

The amphibious snail Oncomelania hupensis is the only intermediate host of Schistosoma japonicum, and its spatial distribution corresponds strongly with that of S. japonica in China(Zhang, Ong et al. 2008; Zhao 1994). This is obviously so because the survival of O. hupensis is governed by climatic and environmental factors, including vegetation, temperature, soil type and water level. The slightest variation of one factor or another can alter the distribution of the intermediate host snail, and hence the transmission dynamics of S. japonicum. Therefore, the relationship between environmental factors and the abundance of O. hupensis can be used not only for prediction of snail distribution, but also for mapping its endemic areas.

The use of remote sensing environmental data derived from satellite images to determine vector-borne diseases is widely documented. Several successful applications have been reported in the literature with an emphasis on schistosomiasis in different ecological and epidemiological settings (Hay, Packer et al. 1997; Kristensen, Malone et al. 2001). Remote sensing is an excellent tool for the collection of data, which facilitate the quantification of environmental factors that are a key way to understand the distribution of the intermediate host snail of schistosomiasis.

This study presented a method to predict the underlying geographic distribution and density of the intermediate host snail of S. japonicum in Dongting Lake based on remotely sensed environmental data, and using spatial analysis as a tool to improve the prediction accuracy.

2. MATERIALS AND METHODOLOGY

2.1Study areas and field survey data

Dongting Lake is located at 28°30′~30°20′ N and 111°40′~113°40′ E in the northeastern part of Hunan Province and covers a water surface area of 2,681 km2. Archaeological studies have shown that schistosomiasis japonica has been endemic in the Dongting Lake region for thousands of years. Nearly 20% buffaloes were infected, and about 200,000 human cases were reported in 2003(Zhao, Zhao et al. 2005).
Field survey data on snail abundance of 2009 had been obtained previously from National Institute of Parasitic Diseases, Chinese Center for Disease Control and Prevention.

2.2Environmental parameters extraction method from satellite image

Landsat 5 TM scene over Dongtiang Lake, taken on 15 April 2009, was used for the current analysis. Environmental parameters were extracted from the image by using the ENVI Version 4.3 (ITT Visual Information Solutions, USA), three indices were calculated. First, the modified soil-adjusted vegetation index (MSAVI). The general expression of MSAVI is given below:

$$\mathrm{MSAVI} = \frac{\left(2\mathrm{NIR}+1-\sqrt{(2\mathrm{NIR}+1)^2-8(\mathrm{NIR}-\mathrm{Red})}\right)}{2}. \quad (1)$$

where NIR(near infrared) and Red refer to bands 4 and 3 of the TM image, respectively(Qi, Chehbouni et al. 1994).
Second, the land surface temperature (LST). Remote sensing of the land surface temperature from space can be carried out using a specific portion of the electromagnetic spectrum such as band 6 of the TM scene. We used a formula to calculate the temperature from TM band 6, as follows:

$$T = 1260.53/\ln[1+60.776/(0.1238+0.00563256\,DN_{TM6})] \quad (2)$$

where T is at-satellite temperature measured in degrees Celsius, DN_{TM6} is the Digital Number value of the band 6, then by using a mono-window algorithm to retrieve land surface temperature(Qin Z 2001).
Third, wetness index, an index of environmental factor from the tasseled-cap transformed TM scene. There are three main features from the tasseled-cap transformation, i.e. brightness, greenness and wetness index. Brightness was designated to capture the main trend of variation in soil reflectance of barren land, the greenness was used as a proxy for the presence and density of vegetation, and the wetness index provided a measure of canopy and soil moisture content. Here we only selected the wetness index that can be explained: the content reported by the other two indices, i.e. brightness and greenness, have been captured by LST and MSAVI that have both been extracted from the satellite image and have also been used in the model (Guo, Vounatsou et al. 2005). Then tasseled-cap transformations were used to extract relevant variables related to environmental factors, since it is a linear combination of the original sensor bands to interpret the multi-spectral satellite image, the transformation formula of wetness index for TM scene is defined as (Price, Guo et al. 2002):

$$\mathrm{WI} = 0.1446L_{TM1} + 0.1761L_{TM2} + 0.3322L_{TM3} + 0.3396L_{TM4} - 0.6210L_{TM5} - 0.4186L_{TM7} \quad (3)$$

where TM1 to TM7 refer to the radiance of Landsat 5 TM image.

2.3Spatial analyses

According to the theory of Geostatistics(Wang 1999), the random variable (Z) can be decomposed into a constant mean (μ) for the data, the random errors (ε) and the error for the spatial dependence(ε'), which can be presented as $Z = \mu + \varepsilon + \varepsilon'$. Both multiple linear regression model and spatial analysis methods were used for model development to predict the cluster distribution of snails.
In the first stage, an ordinary multiple linear regression model analysis was performed to determine the relationship between snail abundance and the environmental indices extracted from the Landsat 5 TM image. In the regression analysis, the square-root transformed snail density in the habitats was used as the dependent variable and environmental indices from the satellite (MSAVI, LST and wetness index) as independent variables.
In the second stage, spatial correlation analysis was performed by the semivariogram model, which provides a measure of the variance as a function of distance between data points. Based on the semi-variogram model of regression residuals, the spatial autocorrelation was estimated and mapping using ording Kriging. The semivariance is calculated image as independent variables. As half of the mean-squared difference between two values separated by the distance h:

$$\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(x_i) - Z(x_{i+h})]^2. \quad (4)$$

where Z is a value at a particular location, N(h) the number of paired data at the distance h, and γ(h) is the semivariance. Denoting the spatial correlation parts of Z, it equals the expected squared difference value of observed points based on the fixed distance h. Semivariogram, a graph of semivariance plotted against separation distance h, conveys information about the continuity and spatial variability of the process. If observations close together are more alike than those farther apart, the semivariance increases as the separation distance increases, reflecting the decline of spatial autocorrelation with distance. Often, the semivariance will level off to nearly a constant value (called the sill) at a large separation distance (called the range). Beyond this distance, observations are spatially uncorrelated, reflected by a (near) constant variance in paired differences. We used the spatial analyst module of ArcGIS 9.0 and selected the exponential model instead of the spherical model to fit the spatial correlation of O. hupensis. Since the semivariance in our study did not really level off to a constant value, but increased very slowly beyond the range of distance, the formula is as follows:

$$\gamma(h) = \begin{cases} C_0 + C_e[1 - \exp(-h/a_e)] & h > 0 \\ 0 & h = 0 \end{cases} \quad (5)$$

where C_0 is the nugget effect and C_e is the partial sill (so C_0+ Ce is the sill). Although a_e is called the range, the "effective range" is $3a_e$ since the semivariance γ(h) approaches the sill (C_0+ C_e) asymptotically, i.e. the minimum distance at which spatial autocorrelation becomes less than 0.05 is $3a_e$.
Finally, the regression model of the snail abundance and the kriged prediction of its spatial variation were used to develop the prediction model of O. hupensis distribution in the Dongting Lake region. The final prediction model is of the follows: Y = a + bx1 +bx2 +bx3 + kriged residual, where Y is the square root transformed snail density, and x1, x2 and x3 stand for LST, MSAVI and wetness index, respectively.

3. RESULTS

3.1Environment parameters derived from Landsat 5 TM image

By using above materials and methodology, the environment indices (MSAVI, LST, and Wetness index) were derived from Landsat 5 TM image as Figure 1.
The multiple linear regression model analysis was performed to determine the relationship between square-root transformed snail density (dependent variable) and parameters including LST, MSAVI and wetness index (independent variables).The model can be presented as follows:

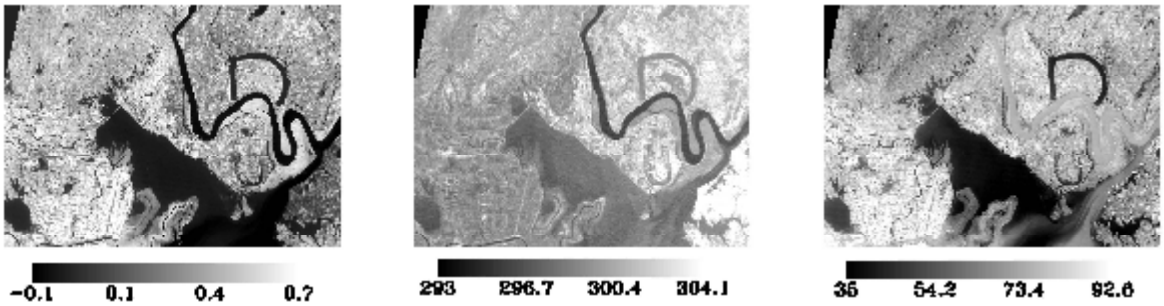
$$Y = 6.2794 - 0.01403x1 + 0.8985x2 - 0.03486x3 + \gamma(h) \quad (6)$$

Where Y refers to the square root transformed snail density, γ(h) is regression residual, and the correlation coefficient of the model is 0.3619, and hence the model only explained 36.19% of the total variation of snail abundance in our study area.

3.2Spatial analysis of the regression residual and prediction model for distribution of O. hupensis

The semi-variogram for the regression residual was an exponential model with a sill value of 0.09966, the nugget value 0.04282, and the range was 46.9235 m .The formulation of the residual is as equation (7).
Based on the semi-variogram model of regression residuals, the spatial variation of O. hupensis abundance attributed to the spatial autocorrelation was estimated and mapped using ordinary kriging (Figure 2).

$$\gamma(h) = \begin{cases} 0.04282 + 0.09966[1 - \exp(-h/a_e)] & h > 0 \\ 0 & h = 0 \end{cases} \quad (7)$$



(a) MSAVI (b) LST (k) (c) Wetness index

Figure 1. Environment indices (MSAVI, LST, and Wetness index) derived from Landsat 5 TM image

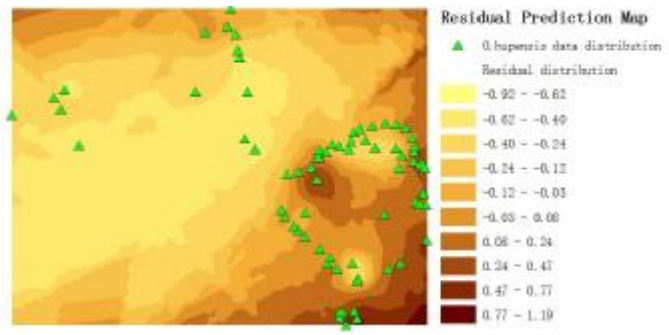
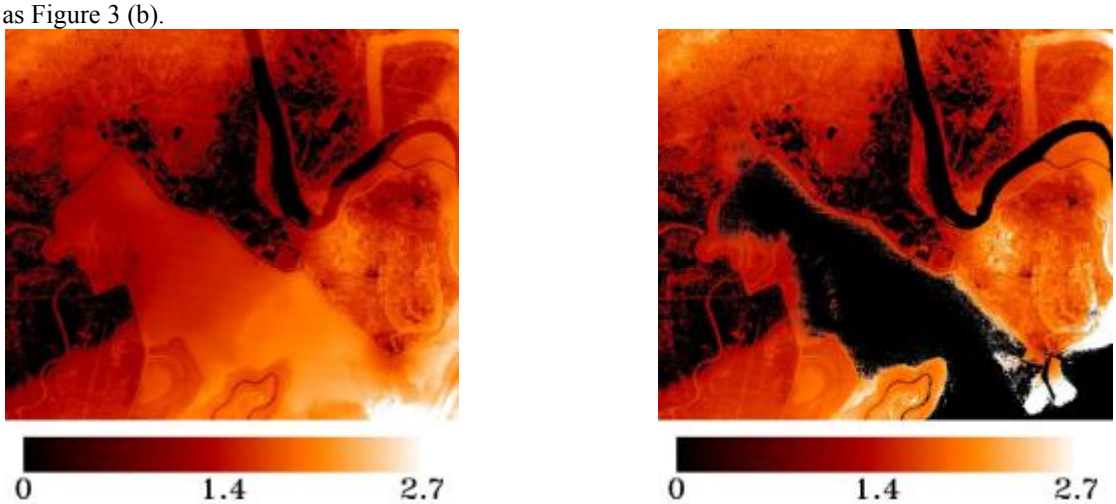


Figure 2. Prediction map of regression residuals using ordinary kriging

Finally, through combine a multiple linear regression with kriged residual model, the prediction results are validated with field data, and the model of living snail density can get a high correlation coefficient as 0.835. Figure 3 (a) shows the prediction map of O. hupensis distribution density in Dongting Lake, China. Although the correlation coefficient is higher than just only using linear regression, we still can find that there is some mistake in the prediction map. For example, water areas of lake should not have snails. Possible explanation is that this method is invalid since some special situation due to it just for considerate the relationship with snails living environment, and hence ignoring some restrictive conditions. It can get a good result at the snail living region, but unlikely include all factors. Because snail is hard to survive at no vegetation areas, we use the MASVI to extract the no vegetation areas, then the final prediction map is as Figure 3 (b).



(a) Prediction map of not extract no vegetation areas (b) Prediction map of extract no vegetation areas

Figure 3. Prediction map of O. hupensis distribution densities in Dongting Lake

4. DISCUSSION AND CONCLUSION

This study extracted several environmental features from an available Landsat 5 TM satellite image for identification and prediction of O. hupensis habitats. We used MSAVI, LST and wetness index to carry out multiple linear regression analysis as these variables with the highest predictive power for mapping snail abundance. In the current

study, the MSAVI was used to estimate ground vegetation with its normally positive correlation with the soil background brightness and minimize soil background influences on the vegetation signal. Wetness index provided a measure of canopy and soil moisture content. Linear regression by environmental factors alone was not sufficient to accurately predict snail abundance as the correlation coefficient of the linear regression model is only 0.3619. Large regression residual indicated that some important factors related to the abundance of snails had not yet been taken into account in the model(Zhang, Xu et al. 2005). In order to overcome this potential biases, we used the semi-variogram technique to investigate the spatial dimension of the regression residuals and estimated the variation of snail spatial distribution in ordinary kriging method. A substantial improvement was shown in the prediction model, which combined the regression model and the kriged-residual model. The prediction results are validated with field data,and the correlation coefficient was as high as 0.835. From the results analysis, we also found that this method can capture the relationship between O. hupensis distribution density and environmental factor, such as vegetation, temperature, spatial distribution etc, but at some special, for example water area, it performed not well. This method ignored some restriction conditions of O. hupensis survivability, so we should study the limited condition of relative environment factors, and try to use the limited factors to modify the results, like MSAVI in this study.

Acknowledgment

Particular thanks to Prof. Zhou Xiaonong (National Institute of Parasitic Diseases, Chinese Center for Disease Control and Prevention) for his generous provision of field survey data and helps in developing the analysis. Thanks to Dr. Yang Guojing (Jiangsu Institute of Parasitic Diseases, China), who provided a lot of materials.
This study is supported by the funding from the Kitty Hawk Plan Program of AOE,CAS.

References

Chen, X. Y., L. Y. Wang, et al. (2005). "Schistosomiasis control in China: the impact of a 10-year World Bank Loan Project (1992-2001)." Bulletin of the World Health Organization 83(1): 43-48.
Guo, J. G., P. Vounatsou, et al. (2005). "A geographic information and remote sensing based model for prediction of Oncomelania hupensis habitats in the Poyang Lake area, China." Acta Tropica 96(2-3): 213-222.
Hay, S. I., M. J. Packer, et al. (1997). "The impact of remote sensing on the study and control of invertebrate intermediate hosts and vectors for disease." International Journal of Remote Sensing 18(14): 2899-2930.
Jiang, Q. W., L. Y. Wang, et al. (2002). "Morbidity control of schistosomiasis in China." Acta Tropica 82(2): 115-125.
Kristensen, T. K., J. B. Malone, et al. (2001). "Use of satellite remote sensing and geographic information systems to model the distribution and abundance of snail intermediate hosts in Africa: a preliminary model for Biomphalaria pfeifferi in Ethiopia." Acta Tropica 79(1): 73-78.
Li, Y. S., Z. Y. Zhao, et al. (2005). "Applications and outcomes of periodic epidemiological surveys for schistosomiasis and related economic evaluation in the People's Republic of China." Acta Tropica 96(2-3): 266-275.
Price, K. P., X. L. Guo, et al. (2002). "Optimal Landsat TM band combinations and vegetation indices for discrimination of six grassland types in eastern Kansas." International Journal of Remote Sensing 23(23): 5031-5042.
Qi, J., A. Chehbouni, et al. (1994). "A Modified Soil Adjusted Vegetation Index." Remote Sensing of Environment 48(2): 119-126.
Qin Z , K. A., Berliner P (2001). "A mono - window algorithm for retrieving land surface temperature from Landsat TM data and its application to the Israel - Egypt border region " International Journal of Remote Sensing 22(18): 3719 - 3746.
Utzinger, J., X. N. Zhou, et al. (2005). "Conquering schistosomiasis in China: the long march." Acta Tropica 96(2-3): 69-96.
Wang, Z. Q. (1999). Geostatistics and its Application to the Ecology. Beijing, Science Press.
Yuan, H. C., Q. W. Jiang, et al. (2002). "Achievements of schistosomiasis control in China." Memorias Do Instituto Oswaldo Cruz 97: 187-189.
Zhang, Z., S. H. Ong, et al. (2008). "A model for the prediction of Oncomelania hupensis in the lake and marshland regions, China." Parasitology International 57(2): 121-131.
Zhang, Z. Y., D. Z. Xu, et al. (2005). "Remote sensing and spatial statistical analysis to predict the distribution of Oncomelania hupensis in the marshlands of China." Acta Tropica 96(2-3): 205-212.
Zhao, G. M., Q. Zhao, et al. (2005). "Surveillance for schistosomiasis japonica in China from 2000 to 2003." Acta Tropica 96(2-3): 288-295.
Zhao, W. X. (1994). "Human Parasitology, seconded." Publishing House for People's Health pp391-404.
Zhou, X. N., L. Y. Wang, et al. (2005). "The public health significance and control of schistosomiasis in China - then and now." Acta Tropica 96(2-3): 97-105.