Available online at www.sciencedirect.com
SCIENCE DIRECT®
Acta Tropica 96 (2005) 205–212
ACTA TROPICA
www.elsevier.com/locate/actatropica

Remote sensing and spatial statistical analysis to predict the distribution of *Oncomelania hupensis* in the marshlands of China

Zhang Zhi-Ying [a], Xu De-Zhong [a,*], Zhou Xiao-Nong [b], Zhou Yun [c], Liu Shi-Jun [d]

[a] *Department of Epidemiology, Fourth Military Medical University, Shaanxi Province 710032, China*
[b] *National Institute of Parasitic Diseases, Chinese Center for Disease Control and Prevention, Shanghai 200025, China*
[c] *Center for Disease Control and Prevention of Jiangning County, Jiangsu Province 211100, China*
[d] *Anti-epidemic Team of Nanjing Military Area, Jiangsu Province 210014, China*

Available online 16 September 2005

Abstract

Remote sensing and spatial statistical analysis were employed to predict the distribution of *Oncomelania hupensis*, the intermediate host snail of *Schistosoma japonicum*, in the marshlands of Jiangning county in China. Surrogate indices related to environmental factors in the marshlands were derived from a Landsat 7 ETM+ image, and the relationship between environmental covariates and the density of *O. hupensis* was analyzed by stepwise regression models and ordinary kriging. Although stepwise regression demonstrated that *O. hupensis* densities of live snails in the marshlands related significantly to the modified soil-adjusted vegetation index, wetness and land surface temperature, the correlation coefficient was low (0.282). Therefore, spatial patterns of the regression residual were investigated by the semi-variogram method, and the spatial variation of *O. hupensis* density attributed to the spatial autocorrelation was estimated by ordinary kriging. The regression model of the snail density and ordinary kriging of its spatial variation were then combined with the aim of improving the prediction of *O. hupensis*. Following this approach, the prediction indeed improved considerably (0.852). Our results show that it is possible to predict the distribution of *O. hupensis* in these marshlands by using remotely sensed environmental indices, and that spatial statistical analyses are capable of improving prediction accuracy. These findings are of relevance for mapping and prediction of schistosomiasis japonica in China, and hence the national control programme.

Keywords: Schistosomiasis japonica; *Oncomelania hupensis*; Remote sensing; Spatial statistics; Ordinary kriging; Snail control

1. Introduction

The amphibious snail *Oncomelania hupensis* is the only intermediate host of *Schistosoma japonicum*, and its spatial distribution corresponds strongly with that of schistosomiasis japonica in China (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*.

* Corresponding author. Tel.: +86 29 83374955; fax: +86 29 83251302.
E-mail address: xudezh@fmmu.edu.cn (D.Z. Xu).

0001-706X/$ – see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.actatropica.2005.07.027

Therefore, the relationship between environmental factors and abundance of *O. hupensis* can be used not only for prediction of snail distribution, but also for mapping its endemic areas.

The use of remotely sensed environmental data derived from satellite images to determine vector-borne diseases is widely documented (Hay et al., 1997; Kristensen et al., 2001). Several successful applications have been reported in the literature with an emphasis on schistosomiasis in different ecological and epidemiological settings (Malone et al., 1994; Bavia et al., 2001; Zhou et al., 2002a). Remote sensing is an excellent tool for the collection of data, which facilitate the quantification of environmental factors that are key to understand the distribution of the intermediate host snail of schistosomiasis (Brooker, 2002; Guo et al., 2005; Malone, 2005; Raso et al., 2005; Yang et al., 2005).

Despite significant achievements in the control of schistosomiasis japonica in China over the past 50 years, the disease remains a public health problem in the Yangtze River basin, including the Dongting and Poyang lakes (Zhou et al., 2005). The study presented here was designed to develop computer models to predict the underlying geographic distribution of the intermediate host snail of *S. japonicum* in these areas based on existing historical information and remotely sensed environmental data.

2. Materials and methods

2.1. Data collection and preparation

The study areas were selected in the marshlands of Jiangning county, Jiangsu province, which is located in the lower reaches of the Yangtze River. This selection was based on the fact that, historically, the marshlands were areas highly endemic for schistosomiasis japonica. The data on snail abundance used for the present study had been obtained previously by an institution in Jiangning county belonging to the new Center for Disease Control and Prevention system. Digitized maps at a scale of 1:50,000 with topographical sheets of Jiangning county were used for mapping of this information.

2.2. Satellite image

A Landsat 7 ETM+ scene over Jiangning county, taken on 11 November 2000, was used for the current analysis. The satellite image was geo-referenced with the digitized topographical map of Jiangning county. The study area was bounded by latitudes 31.78–31.96°N and longitudes 118.49–118.61° E.

2.3. Environmental data extraction from satellite image

Environmental data were extracted from the Landsat 7 ETM+ image and, using the ERDAS Imagine v. 8.5 package (Leica Geosystems, Atlanta, GA, USA), three indices were calculated. First, the second modified soil-adjusted vegetation index (MSAVI2). The general expression of MSAVI2 is given below:

$$\mathrm{MSAVI2} = \frac{2\mathrm{NIR}+1-\sqrt{(2\mathrm{NIR}+1)^2-8(\mathrm{NIR}-\mathrm{RED})}}{2},$$

where NIR (near infrared) and RED refer to bands 4 and 3 of the ETM+ image, respectively (Zhou et al., 1999).

Second, the land surface temperature (LST). Remote sensing of the earth's surface temperature from space can be carried out using a specific portion of the electromagnetic spectrum such as band 6 of the ETM+ scene. We used a formula put forth by NASA (1998) to calculate the temperature from ETM+ band 6 image, as follows:

$$T = \frac{K_2}{\ln(K_1/L_\lambda + 1)} - 273,$$

where T is the at-satellite temperature measured in degrees Celsius, K_1 and K_2 are the calibration constants for ETM+ sensor, which are approximately 666.093 W/(m^2 s μm) and 1282.7108 K, respectively, and L_λ the spectral radiance in W/(m^2 s μm) of the ETM+ band 6.

Third, surrogate indices of environmental factors from the tasseled-cap transformed ETM+ scene. 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, and the derived data respond to particular physical scene class characteristics and capture 95% or more of the total data variability in the raw spectral bands (Qiu et al., 1998). The transformation formula for ETM+ scene is defined as (Leica Geosystems, 2003):

Brightness = $0.3561CH_1 + 0.397CH_2 + 0.390CH_3 + 0.697CH_4 + 0.229CH_5 + 0.160CH_7$; greenness = $-0.334CH_1 - 0.354CH_2 - 0.456CH_3 + 0.697CH_4 - 0.024CH_5 - 0.0263CH_7$; and wetness = $0.2626CH_1 + 0.214CH_2 + 0.093CH_3 + 0.066CH_4 - 0.763CH_5 - 0.539CH_7$, where CH_1 to CH_7 refer to the original ETM+ image.

2.4. Development of the model and statistical analysis

According to the theory of mathematical statistics (Johnston and Ver Hoef, 1999; Wang, 1999), the random variable (Z) can be decomposed into a constant mean (μ) for the data, the random errors (ε) and the error (ε'') for the spatial dependence, which can be presented as $Z = \mu + \varepsilon + \varepsilon''$. Both linear regression and spatial analysis methods were used for model development to predict the cluster distribution of snails in space.

In the first stage, an ordinary linear regression analysis was performed to determine the relationship between snail abundance and the environmental indices extracted from the Landsat 7 ETM+ image. In the regression analysis, the square-root transformed snail density in the habitats was used as the dependent variable and the environmental indices from the satellite image as independent variables. In the second stage, the spatial pattern of the regression residuals was analysed by the semi-variogram technique using the spatial analysis extension of the ArcView v. 8.1 package (Esri Inc., Redlands, CA, USA), and the residual spatial dependence in the scale of its range α was computerised by ordinary kriging. After the model had been established, cross-validation was performed to determine the fitness of the kriged model for the regression residual. The goal for "good-to-fit" is to have a standardised mean prediction error near 0, to obtain small root-mean square prediction errors, to make average standard error near root-mean square prediction errors and standardised root-mean square prediction errors near 1.

3. Results

3.1. Environmental indices derived from Landsat 7 ETM+ in marshlands

Based on the annul surveillance in Jiangning county, 25 suitable habitats of *O. hupensis* were found in the marshlands for the year 2000. On average, live snails per $0.1\,m^2$ were found in 7.75% of the surveyed locations, and the average snail density was 1.69 snails/$0.1\,m^2$. Habitat-specific snail data were used for subsequent analysis with an emphasis on the relationship between snail density and environmental features. All indices related to the environmental factors extracted from the satellite image for snail habitats varied in a certain range, the values of which were smaller than those of the marshlands (Table 1). For example, the mean of MSAVI2 in snail habitats was 0.316 with values for 80% of the habitats ranging from 0.204 to 0.447. These results translate to a more narrow scale than that observed for the whole marshland (i.e. range from −1.703 to 0.652). Our findings underscore that *O. hupensis* only breed and live in certain parts of the marshlands where they encounter suitable microenvironmental conditions.

Table 1
Environmental indices derived from a Landsat 7 ETM+ image in the marshlands of Jiangning county

Surrogate index	Whole marshland	Snail habitats	
		Mean ± S.D.	Range for 80% habitats
Tasseled-cap transformation			
Wetness	−0.523–0.137	−0.191 ± 0.039	−0.238, −0.134
Brightness	0.331–1.076	0.704 ± 0.049	0.635, 0.763
Greenness	−0.290–0.321	−0.010 ± 0.043	−0.056, 0.055
MSAVI2[a]	−1.703–0.652	0.316 ± 0.094	0.204, 0.447
LST[b]	9.95–28.28	13.845 ± 0.786	12.858, 14.787

[a] MSAVI2: second modified soil-adjusted vegetation index.
[b] LST: land surface temperature.

Table 2
Pearson correlation analysis of the relationship between the snail abundance and environmental indices from a Landsat 7 ETM+ image in the habitats

Surrogate index	Pearson coefficient	*P*-value
Tasseled-cap transformation		
Wetness	−0.333	<0.001
Brightness	0.400	<0.001
Greenness	0.429	<0.001
MSAVI2[a]	0.442	<0.001
LST[b]	−0.209	0.021

[a] MSAVI2: second modified soil-adjusted vegetation index.
[b] LST: land surface temperature.

3.2. Pearson bivariate correlation analysis

Table 2 summarized the Pearson correlation analysis between snail abundance and environmental features derived from the satellite image. Negative correlations were found between the square root transformed density of snails and both LST and wetness of habitats. On the other hand, positive correlations were found between snail density and both brightness and greenness of the habitats.

3.3. Regression analysis

Table 3 shows the results of the prediction model for the snail abundance using stepwise regression analysis. Three of the five environmental factors were selected by the stepwise regression procedure to be included in the model; namely (i) MSAVI2 (x_1), (ii) wetness (x_2), and (iii) LST (x_3). The model can be presented as follows: $Y_1 = 2.481 + 3.219x_1 - 9.143x_2 - 0.261x_3$, where Y_1 refers to the square root of snail density, and the correlation coefficient of the model is 0.282 ($F = 15.30$, $p < 0.0001$).

3.4. Spatial pattern investigation of the regression residual and development of the prediction model for distribution of *O. hupensis* in the marshlands

The semi-variogram for the regression residual was a spherical model with a sill value of 0.989, a range of 0.116, an arch of 0.591 and 0.398 nuggets. The formula of the model is as follows:

$$\gamma(h) = \begin{cases} 0 & (h = 0) \\ 0.398 + 0.591\left(\dfrac{3}{2}\dfrac{h}{0.116^2} - \dfrac{1}{2}\dfrac{h^3}{0.116^3}\right) & (0 < h < 0.116) \\ 0.989 & (h > 0.116) \end{cases}$$

This spherical model indicates that the variation of the regression residual attributed to spatial dependence can be estimated by the semi-variogram when the spatial scales are less than 0.116. The ratio of arch and the sill was 0.598. In other words, 59.8% of the variation of the regression residuals was attributed to the spatial autocorrelation, and hence could be estimated using the variogram model.

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 (Fig. 1).

Table 3
An analysis of the relationship between environmental indices derived from a Landsat 7 ETM+ image and snail density in the habitats using stepwise regression

	Non-standardised coefficients (*B*)	Standardised coefficients (β)	*T*	*P*-value
Constant	2.481		1.360	0.177
MSAVI2[a]	3.219	0.298	3.273	0.001
Tasseled-cap transformation wetness	−9.143	−0.351	−3.756	<0.001
LST[b]	−0.261	−0.207	−2.085	0.039

[a] MSAVI2: second modified soil-adjusted vegetation index.
[b] LST: land surface temperature.

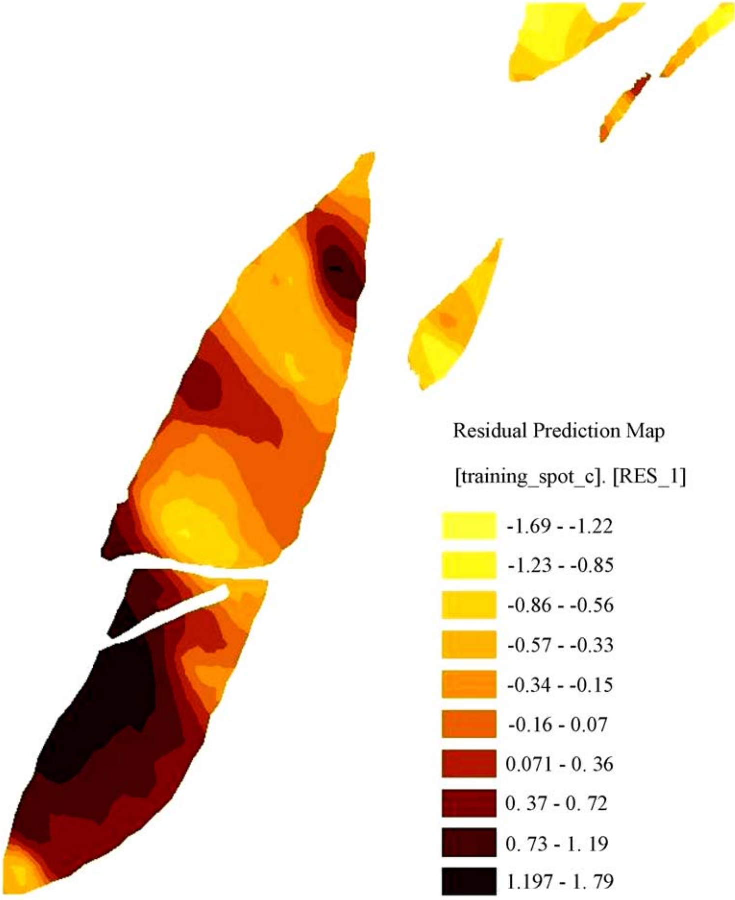

Fig. 1. Prediction map of regression residuals using ordinary kriging.

The cross-validation showed that the prediction map was a fitness estimate of the regression residual with a standardised mean prediction error of −0.00902, and a standardised root-mean square prediction error of 0.962.

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 marshlands of Jiangning county. The final prediction model is of the following kind: $Y = 2.481 + 3.219x_1 - 9.143x_2 - 0.261x_3 +$ kriged-residual, where Y is the square root transformed snail density, and x_1, x_2 and x_3 stand for MSAVI2, wetness and LST, respectively. The correlation coefficient of the final model was 0.852, which was greatly improved over that of the linear regression model. Fig. 2 shows the final prediction map of *O. hupensis* distribution in the marshlands of Jiangning county.

4. Discussion

Traditionally, the schistosome-endemic areas in China were classified according to epidemiological patterns of the disease and ecological requirements of its intermediate host snail, i.e. *O. hupensis*. Following this approach, the *S. japonicum*-endemic settings in China can be stratified into three types, namely (i) plain regions with waterway networks, (ii) marshland and lake regions, and (iii) hilly and mountainous regions. It is well documented that specific micro-environmental conditions in the marshlands are essential for survival

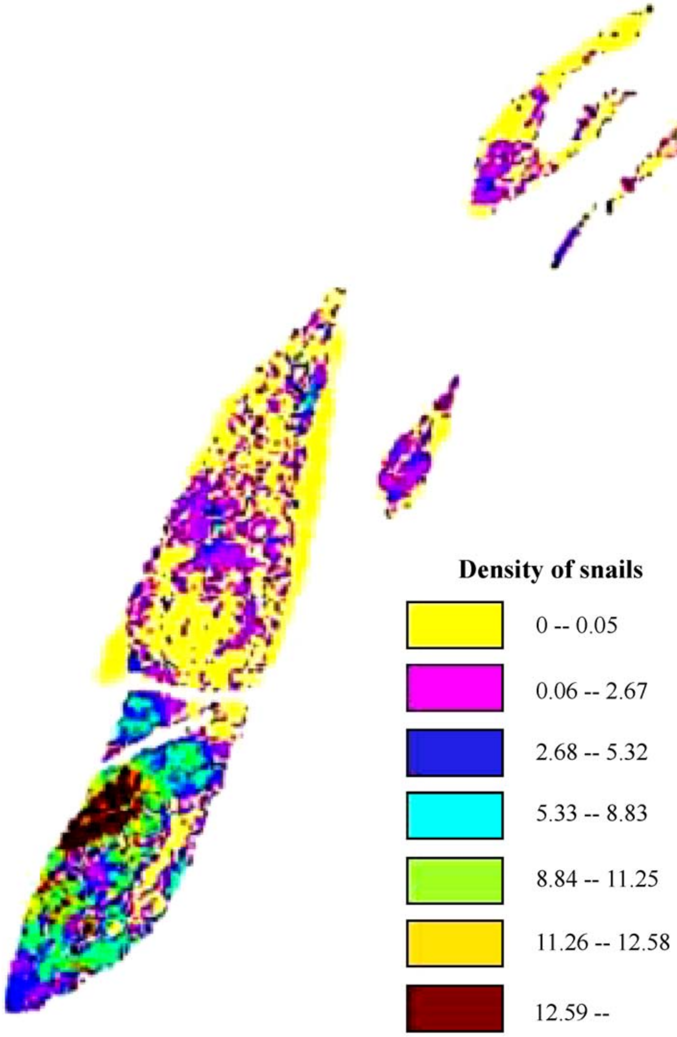

Fig. 2. Prediction map of *Oncomelania hupensis* distribution in marshland of Jiangning county, using a Landsat 7 ETM+ image.

of *O. hupensis*, and over- or undershooting of just one of these factors would preclude snail breeding (Zhao, 1994; Wu et al., 1999; Zhang et al., 1999).

This study extracted several environmental features from an available multi-temporal Landsat 7 ETM+ satellite image for identification and prediction of *O. hupensis* habitats. Employing stepwise regression analysis, we found that MSAVI2, LST and wetness were the variables with the highest predictive power for mapping snail abundance. Our study therefore confirms previous research with an emphasis on predicting habitats of *Biomphalaria pfeifferi*, the intermediate host snail of *Schistosoma mansoni*, carried out in Ethiopia (Kristensen et al., 2001). In the current study, the MSAVI2 was used to estimate ground vegetation for its normally positive correlation with the dynamic range of the vegetation sensitivity to the soil background brightness and to minimize soil background influences on the vegetation signal. All three main features from the tasseled-cap transformation, i.e. brightness, greenness and wetness, were investigated. 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 provided a measure of canopy and soil moisture content. In the linear regression analysis, however, only the wetness was selected. This can be explained: the content reported by the other two indices, i.e. brightness and greenness, have been captured by LST and MSAVI2 that have both been extracted from the satellite image and have also been used in the model. Guo et al. (2005) were able to predict *O. hupensis* habitats in the Poyang Lake area by means of the normalized difference vegetation index and the tasseled-cap transformation wetness feature, both derived from Landsat TM images. Furthermore, a group of researchers reported a successful application of a tasseled-cap transformation for the detection of coastal saline land uses in Shangyu city, China (Zhou et al., 2002b).

Linear regression analysis alone was not sufficient to accurately predict snail abundance. In fact, the coefficient of the regression model was only 0.282, and hence the model only explained 28.2% of the total variation of snail abundance in our study area. The large regression residual indicated that some important factors related to the abundance of snails had not yet been taken into account in the model. The relatively low value of the regression coefficient may be attributed to a non-random distribution of *O. hupensis*, as reported before (Zhao, 1994; Zhang et al., 2002). Another possible explanation is that the regression coefficient and corresponding statistical test are invalid since a bias may be present due to the ordinary linear regression analysis only providing the covariate adjustment and prediction of mean risk in this area, and hence ignoring the spatial dimension of the data (Walter, 1993; Kitron, 1998; Thomson and Connor, 2000). In order to overcome these potential biases, we used the semi-variogram technique to investigate the spatial dimension of the regression residuals and estimated the variation of snail distribution for spatial dependence by ordinary kriging. A substantial improvement was shown in the final prediction model, which combined the regression model and the kriged-residual model. The correlation coefficient was as high as 0.852, more than 3-fold higher than in the regression model. The same approach has been used by Kleinschmidt et al. (2000) to develop a prediction model for malaria in Mali, West Africa.

We conclude that our model presented here and that of others (Guo et al., 2005) hold promise for prediction of *O. hupensis* habitats in schistosome-endemic marshlands of China. Further validation is warranted as these models could become important tools for the national schistosomiasis control programme, particularly in areas where regular surveillance is a difficult task, such as during the rainy season when areas are often flooded.

Acknowledgements

The work was supported in part by the obligatory budget of the People's Liberate Army (PLA) of China, grant No. 01L078. The authors thank Dr. Zhou Yun and colleagues from the Center of Disease Control and Prevention of Jiangning county for providing data on the prevalence of *S. japonicum* and density of *O. hupensis* and for help in the field investigations. The authors also thank Drs. R. Bergquist and J. Utzinger for revising this manuscript.

References

Bavia, M.E., Malone, J.B., Hale, L., Dantas, A., Marroni, L., Reis, R., 2001. Use of thermal and vegetation index data from earth observing satellites to evaluate the risk of schistosomiasis in Bahia, Brazil. Acta Trop. 79, 79–85.

Brooker, S., 2002. Schistosomes, snails and satellites. Acta Trop. 82, 207–214.

Guo, J.G., Vounatsou, P., Cao, C.L., Utzinger, J., Zhu, H.Q., Anderegg, D., Zhu, R., He, Z.Y., Li, D., Hu, F., Chen, M.G., Tanner, M., 2005. A geographic information and remote sensing based model for prediction of *Oncomelania hupensis* habitats in the Poyang Lake area, China. Acta Trop. 96, 213–222.

Hay, S.I., Packer, M.J., Roger, D.J., 1997. The impact of remote sensing on the study and control of invertebrate intermediate hosts and vectors for disease. Int. J. Remote Sens. 18, 899–930.

Johnston, K., Ver Hoef, J.M., 1999. Using ArcGIS: Geostatistical Analysis. Esri Inc., Redlands, USA, pp. 49–79.

Kitron, U., 1998. Landscape ecology and epidemiology of vector-borne diseases: tools for spatial analysis. J. Med. Entomol. 35, 435–445.

Kleinschmidt, I., Bagayoko, M., Clarke, G.P.Y., Craig, M., Le Sueur, D., 2000. A spatial statistical approach to malaria mapping. Int. J. Epidemiol. 29, 355–361.

Kristensen, T.K., Malone, J.B., McCarroll, J.C., 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 Trop. 79, 73–78.

Leica Geosystems, 2003. Tasseled-cap transformation for Landsat 7 ETM+ sensor. USA, ETM+ http://gis.leica-geosystems.com/Support/Downloads/dwnld_detail.aspxETM+?download_ID=9.

Malone, J.B., 2005. Biology-based mapping of vector-borne parasites by geographic information systems and remote sensing. Parassitologia 47, 27–50.

Malone, J.B., Huh, O.K., Fehler, D.P., Wilson, P.A., Wilensky, D.E., Holmes, R.A., Elmagdoub, A.I., 1994. Temperature data from satellite imagery and the distribution of schistosomiasis in Egypt. Am. J. Trop. Med. Hyg. 50, 714–722.

NASA, 1998. Landsat 7 Science data user's handbook, USA http://ltpwww.gsfc.nasa.gov/IAS/handbook/handbook_htmls/chapter11/chapter11.html.

Qiu, Z.Y., Li, J., Guo, H.J., 1998. Application of the Remote Sensing Technique. Wuhan University Press, Wuhan, China, pp. 97–98.

Raso, G., Matthys, B., N'Goran, E.K., Tanner, M., Vounatsou, P., Utzinger, J., 2005. Spatial risk prediction and mapping of *Schistosoma mansoni* infections among schoolchildren living in western Côte d'Ivoire. Parasitology 131, 97–108.

Thomson, M.C., Connor, S.J., 2000. Environmental information systems for the control of arthropod vectors of disease. Med. Vet. Entomol. 14, 227–244.

Walter, S.D., 1993. Assessing spatial patterns in disease rates. Stat. Med. 12, 1885–1894.

Wang, Z.Q., 1999. Geostatistics and its Application to the Ecology. Science Press, Beijing, pp. 153–154.

Wu, G., Su, R.P., Zhang, X.D., 1999. The relationship between the snail abundance and the vegetation type in marshland of lower reach of Yangtze River. Acta Ecol. Sin. 19, 118–122.

Yang, G.J., Vounatsou, P., Zhou, X.N., Tanner, M., Utzinger, J., 2005. Bayesian-based approach for spatio-temporal modeling of county level prevalence of *Schistosoma japonicum* infection in Jiangsu province, China. Int. J. Parasitol. 35, 155–162.

Zhang, X.D., Yang, X.C., Peng, Z.H., 1999. The relationship between the environmental factors and the distribution of snails. Acta Ecol. Sin. 19, 265–269.

Zhang, Z.Y., Zhou, Y., Xu, D.Z., Sun, Z.D., Zhou, X.N., Gong, Z.L., 2002. Spatial analysis of snail distribution in Jiangning county. J. Milit. Med. Coll. PLA 17, 88–91.

Zhao, W.X., 1994. Human Parasitology, second ed. Publishing House for People's Health, Beijing, pp. 391–404.

Zhou, C.H., Luo, J.C., Yang, X.M., 1999. Interpretation and Analysis of Remote Sensing Images in Geoscience. Science Press, Beijing press, Beijing, pp. 15–16.

Zhou, S., Wang, R.C., Huang, M.X., Landgraf, D., 2002b. Detection of coastal saline land uses with multi-temporal Landsat images in Shangyu city, China. Environ. Manage. 30, 142–150.

Zhou, X.N., Lin, D.D., Yang, H.M., Chen, H.G., Sun, L.P., Yang, G.J., Hong, Q.B., Brown, L., Malone, J.B., 2002a. Use of Landsat TM satellite surveillance data to measure the impact of the 1998 flood on snail intermediate host dispersal in the lower Yangtze River basin. Acta Trop. 82, 199–205.

Zhou, X.N., Wang, L.Y., Chen, M.G., Wu, X.H., Jiang, Q.W., Chen, X.Y., Zheng, J., Utzinger, J., 2005. The public health significance and control of schistosomiasis—then and now. Acta Trop. 96, 97–105.