An integrated approach to identify distribution of *Oncomelania hupensis*, the intermediate host of *Schistosoma japonicum*, in a mountainous region in China

Kun Yang [a,b], Xian-Hong Wang [a], Guo-Jing Yang [b], Xiao-Hua Wu [a], Yun-Liang Qi [c], Hong-Jun Li [c], Xiao-Nong Zhou [a,*]

[a] *National Institute of Parasitic Diseases, Chinese Center for Disease Control and Prevention, Shanghai 200025, People's Republic of China*
[b] *Jiangsu Institute of Parasitic Diseases, Yangxiang 117, Meiyuan, Wuxi, Jiangsu 214064, People's Republic of China*
[c] *Eryuan Institute of Schistosomiasis Control, Eryuan, Yunnan 671200, People's Republic of China*

Abstract

The aim of this study is to better understand ecological variability related to the distribution of *Oncomelania hupensis*, the snail intermediate host of *Schistosoma japonicum*, and predict the spatial distribution of *O. hupensis* at the local scale in order to develop a more effective control strategy for schistosomiasis in the hilly and mountainous regions of China. A two-pronged approach was applied in this study consisting of a landscape pattern analysis complemented with Bayesian spatial modelling. The parasitological data were collected by cross-sectional surveys carried out in 11 villages in 2006 and mapped based on global positioning system (GPS) coordinates. Environmental surrogates and landscape metrics were derived from remotely-sensed images and land-cover/land-use classification data. Bayesian non-spatial and spatial models were applied to investigate the variation of snail density in relation to environmental surrogates and landscape metrics at the local scale. A Bayesian spatial model, validated by the deviance information criterion (DIC), was found to be the best-fitting model. The mean shape index (MSI) and Shannon's evenness indexes (SEI) were significantly associated with snail density. These findings suggest that decreasing the heterogeneity of the landscape can reduce snail density. A prediction maps were generated by the Bayesian model together with environmental surrogates and landscape metrics. In conclusion, the risk areas of snail distribution at the local scale can be identified using an integrated approach with landscape pattern analysis supported by remote sensing and GIS technologies, as well as Bayesian modelling.
© 2008 Australian Society for Parasitology Inc. Published by Elsevier Ltd. All rights reserved.

Keywords: Schistosomiasis; Bayesian model; Geographic information system (GIS); Landscape pattern analysis; Mountainous regions; *Oncomelania hupensis*; *Schistosoma japonicum*

1. Introduction

Schistosomiasis japonica is a serious zoonotic parasitic disease. The causative organism, *Schistosoma japonicum*, and its snail intermediate host, *Oncomelania hupensis*, are mainly distributed along the Yangtze River valley and in southern China. Its endemic areas have been classified into marshland and lake regions, hilly and mountainous regions, and plain and water network regions (Zhou et al., 2007). Despite remarkable progress made in the national programme of schistosomiasis control over more than 50 years, schistosomiasis had re-emerged in some of the mountainous regions of Sichuan Province by the end of 2004 (Liang et al., 2006, 2007). Most probably, environmental and socio-political changes are responsible, particularly for the emergence of new snail habitats and hence the occurrence of acute cases (Yuan et al., 2002; Utzinger et al., 2005; Clennon et al., 2007).

It is well-known that the transmission of *S. japonicum* is closely related to the distribution of the snail intermediate host, which depends on geographical and environmental characteristics such as vegetation coverage, land-use patterns, quality and humidity of the soil (Engels et al., 2002; Ross et al., 2002). In the hilly and mountainous regions, snails are distributed along the valleys, on grassy hill slopes, the vertical sides of terraced fields and in irrigation ditches. These snail habitats are normally isolated from one another, which results in a patchy transmission situation with scattered risk areas intersected by safe land (Zhou et al., 2005; Liang et al., 2006). Therefore, re-emerging foci of schistosomiasis are difficult to predict, even with geospatial technologies, e.g. remote sensing (RS) and geographic information systems (GIS), believed to be useful for mapping and prediction of intermediate host distribution and disease transmission (Zhou et al., 2001; Guo et al., 2005; Yang et al., 2005b; Gong et al., 2006; Steinmann et al., 2007). In addition, large-scale analyses cannot easily capture the small scale information. This is a typical epidemiological feature of schistosomiasis transmission in the hilly and mountainous regions (Brooker et al., 2006; Koukounari et al., 2006; Brooker, 2007). The spatial distribution of snail intermediate hosts, reservoir hosts and the parasite itself may be limited by the heterogeneity of the landscape as well as climatic and anthropogenic factors (Calvete et al., 2004; Deter et al., 2006; Kiss et al., 2006). Due to the variability of snail distribution, mapping at the local scale can appear confusing or misleading (Zhang et al., 2003). Furthermore, some studies showed that snail populations, like other organisms in the natural environment, tend to be clustered in the field. Therefore, the contribution of spatial autocorrelation should be taken into account when predicting snail abundance (Spear et al., 2004; Xu et al., 2004; Wintle and Bardos, 2006). However, the spatial distribution of snails is still not well understood (Kariuki et al., 2004).

Two approaches have been widely used for prediction of disease transmission in different aspects at the local scale. One is landscape pattern analysis which can provide a conceptual approach to study environmental risk factors and related disease transmission (Kitron, 1998; Graham et al., 2005; Jackson et al., 2006; Russell et al., 2006). The other is Bayesian spatial modelling employing Markov chain Monte Carlo (MCMC) simulation, which has been recognised as a powerful means to provide more robust estimates as it takes into account possible correlations and covariate effects (Gangnon and Clayton, 2000; Elliott and Wartenberg, 2004; Boyd et al., 2005; Feltbower et al., 2005; Ashby, 2006).

The purpose of our study was to explore the possibility of using a two-pronged approach, consisting of landscape pattern analysis and Bayesian spatial modelling approaches, to predict the spatial distribution of *O. hupensis* snails at the local scale. This was done in order to better understand the ecological variability of snail distribution, and to contribute to a more effective control strategy to prevent the transmission of *S. japonicum* in the hilly and mountainous regions of China.

2. Materials and methods

2.1. Study site

The study area consisted of 11 villages surrounding Cibihu Lake in the valley of a mountainous area of southwestern China, with an elevation range of 1500–2700 m above sea level, at the eastern edge of Eryuan county, Yunnan Province (Fig. 1). The site is subtropical with an annual average temperature of 17 °C and an annual rainfall of about 1000 mm, >90% of which falls between the beginning of June and the end of October. The main agricultural products are rice, corn, wheat, beans, garlic, eggplants and tomatoes. People become infected with *S. japonicum* whilst planting crops and vegetables in the lowland/upland fields or in the terraces. They are also at risk when washing their feet and working utensils along the ditches which are part of the irrigation system, or when children play in the riparian zones along riverbeds.

Eryuan county was one of the highly endemic counties for *S. japonicum* in Yunnan Province (Dong et al., 2006). The recent re-emergence of schistosomiasis was influenced by a significant increase in snail density and snail infection rate in the county which, in turn, was affected by environmental and social factors. Although the snail infection rate was reduced by 17% (from 0.41% to 0.34%) between 2000 and 2004, the density of live snails increased by 588% (from 0.134 to 0.924 per 0.11 m^2). As a consequence, the density of infected snails increased by 524% (from 0.0005 to 0.0031 per 0.11 m^2) (Li, 2006).

2.2. Data

2.2.1. Parasitological data

The snail data (infected and total numbers) were taken from a field survey, performed in 2006 from March to April
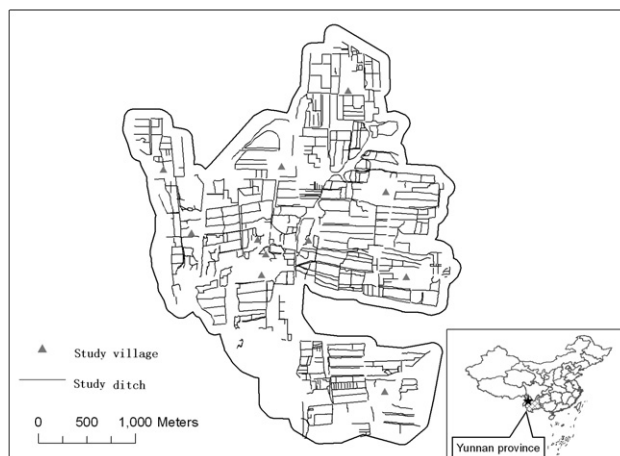

Fig. 1. Study site and distribution of study ditches in Eryuan county, Yunnan province, China in 2006.

using a stratified sampling method. A square frame of $0.11 \, \text{m}^2$ was set every 10 m along the ditches. All snails within the frame were collected and counted. The collected snails were crushed and examined by microscopy to detect schistosome infections. The density of live snails and infected snails were calculated using the following formulas:

$$\text{Live snail density(snails per } 0.11 \, \text{m}^2)$$
$$= \frac{\text{Total number of live snails}}{\text{Total number of detection frames}}$$
$$\text{Infected snail density(infected snails per } 0.11 \, \text{m}^2)$$
$$= \frac{\text{Total number of infected snails}}{\text{Total number of detection frames}}$$

All field data were double-checked and entered into a computer equipped with FoxPro 6.0 software (Microsoft Corporation, USA) to establish a parasitological database.

2.2.2. Spatial environmental data

Coordinated records of study site. The ditch network in the study site was traced on foot with a GPS unit (Trimble GeoExplorer, Trimble Navigation Limited, Sunnyvale, CA, USA). The GPS data were differentially corrected with data from a stationary-based station (Trimble 4000 SSI, Trimble Navigation Limited). An accurate map of the ditch structure was created after the use of an error-correcting process including various characteristics such as ditch type, construction type, width and depth. In addition, the data were identified as they were recorded (collection identification, date and time) and the central points of the ditches were recorded with their coordinates. The coordinates were determined on a Universal Transverse Mercator (UTM) projection system with zone North 49. The attributes were entered and organised in an electronic spreadsheet. All recorded GPS data coupled with filed data were imported into the GIS database.

Satellite image. A remotely-sensed image of SPOT 5, acquired by SPOT Earth observation satellites on March 16, 2006 with path 255/298, was purchased from China Remote Sensing Satellite Ground Station (China RSGS). This image offers a resolution of 5 m in panchromatic mode and 10 m in coloured mode. A sub-scene of the satellite image containing the study site was re-sampled using the subset image module of ERDAS 8.6 software (Leica Geosystems, Atlanta, USA), geometrically corrected to the UTM projection. The coefficients of the linear transformation equations were obtained by linear regression with several well-distributed control points from 1:10,000 topographical digital maps and ground control points from a GPS survey. A root mean square error (RMSE) evaluation was subsequently performed to assess the image-to-map accuracy. The RMSE for the transformed image and standardised maps was <1 pixel. The resolution of the coloured image was enhanced using the 5 m panchromatic band and then coloured high-resolution images were obtained.

2.3. Variables derived from the remotely-sensed image and GIS database

2.3.1. Land-cover/land-use classification

A standard 'supervised' classification method was used to produce a map of land-cover type from the composite map of the coloured SPOT 5 image and GPS map, where a maximum likelihood classification was performed to allocate each image pixel to one of a small number of known categories. The main classes of the land-cover types identified were water, dry field, paddy field, agrarian road, hill forest, ditch with concrete bank, grassland, residential area and hill. Separate training and test samples were chosen for each class based on field visits to the study site. The overall accuracy of the land-cover categories was measured by the Kappa coefficient.

2.3.2. Normalized difference vegetation index (NDVI)

NDVI was widely used in vegetation analyses due to its power in highlighting differences in plant cover not perceivable in the original data. The following formula was used to calculate the NDVI:

$$\text{NDVI} = \frac{(\text{Near-infrared} - \text{red})}{(\text{Near-infrared} + \text{red})}$$

2.3.3. Slope of each ditch

The slope of each ditch within the study site was obtained from a digital elevation model (DEM), downloaded from the US Geological Survey (USGS) website (www.usgs.gov) using the ArcGIS 9.0 (ESRI Corp., USA) spatial module. Because all ditches were constructed to utilise the slope for improved water distribution, we calculated the minimum, maximum and range of slope for each ditch.

2.3.4. Distance from ditches to village

The distance from each ditch to the centre of a village was estimated in order to understand human activities in relation to water contact. We drew a boundary for each village based on high-resolution satellite imagery using ArcGIS 9.0 and then calculated the distance between ditches and a village.

2.4. Data analysis

The data analysis was performed using the following two-pronged approach in order to understand the risk areas and predict the snail or infected snail density related to the transmission of *S. japonicum*.

2.4.1. Landscape pattern analysis

The effect of landscape on the spatial structure of snails was examined through a landscape ecology approach, or landscape metrics. The landscape metrics were divided into three levels, namely (i) individual patches (patch-level indices), (ii) land-cover types (class-

level indices) and (iii) the entire landscape (landscape-level indices). For each ditch, a buffer was created to generate a polygon describing the area within a 50 m distance from a ditch. The polygonal buffers were overlaid on the landscape map. More than 20 relevant landscape metrics were developed using Patch Analyst 2.1 (http://flash.lakeheadu.ca/~rrempel/patch/) based on land-cover/land-use image. Patch Analyst is an extension to the ArcView GIS system that facilitates the spatial analysis of landscape patches and modelling of attributes associated with patches. Due to possible autocorrelation between these indices (Graham et al., 2004; Kumar et al., 2006), we combined the results of a Spearman non-parametric correlation analysis and selected the five variables that best represented the whole landscape status, namely, mean shape index (MSI), Shannon's evenness index (SEI), slope, proportion of paddy fields, and proportion of agrarian roads. The details of these landscape metrics are shown in the following.

The MSI is a metric measurement of shape complexity and is defined by the following formula:

$$\text{MSI} = \frac{\left[\sum P(x) / \left(4\sqrt{\sum A(x)}\right)\right]}{\sum x}$$

where P is the perimeter, A is the area and x is the landscape patch size. When $\text{MSI} = 1$, the patches within the landscape are circular or square; the complexity increases with an increasing MSI value.

SEI is a measure of patch distribution and abundance.

$$\text{SEI} = \frac{-\sum_{i=1}^{m}(P_i \ln P_i)}{\ln m}$$

where P_i is the proportion of the landscape occupied by patch type (class) i and m is the number of patch types (classes) present in the landscape, excluding the landscape border if present.

As SEI increases, the patch types become more diverse. When SEI approaches one the patch type will be the most diverse.

2.4.2. Bayesian modelling

Conventional statistic analysis. Several variables, such as environmental data, landscape metrics and snail data (density and infectivity) were employed for the multivariate statistical analysis (Table 1). Since the variables were not normally distributed, a Spearman non-parametric correlation analysis was applied to explore the relationships between the density of live and infected snails and the environmental and landscape metrics. The variables which had statistical significance in the Spearman non-parametric correlation analyses were used for the Bayesian models. All the statistical analyses were performed using SPSS 11 software (SPSS Inc, USA).

Bayesian model formulation. Bayesian non-spatial and spatial Poisson models were fitted in WinBUGS 1.1.4 (MRC Biostatistics Unit, Cambridge, UK). The deviance

information criterion (DIC) was used to compare the fit of the models (Spiegelhalter et al., 2002). Models with a smallest DIC are considered to reveal the best fit. The parameters of the models were estimated using MCMC simulation.

We assume that the observed counts of snails in ditch i follow a poisson distribution, that is, $O_i \sim \text{Poisson}(\mu_i)$, where O_i is observed snail count in ditch i. We introduce the covariates and spatial effect on the $\log(\mu_i)$ and develop two models for $\log(\mu_i)$.

(1) $\log(\mu_i) = \log(E_i) + \alpha_0 + \alpha_1 \cdot \text{MSI}_i + \alpha_2 \cdot \text{SEI}_i + \alpha_3 \cdot P_{1i} + \alpha_4 \cdot P_{2i} + \alpha_5 \cdot \text{slope}_i$ (non-spatial model).

(2) $\log(\mu_i) = \log(E_i) + \alpha_0 + \alpha_1 \cdot \text{MSI}_i + \alpha_2 \cdot \text{SEI}_i + \alpha_3 \cdot P_{1i} + \alpha_4 \cdot P_{2i} + \alpha_5 \cdot \text{slope}_i + \phi_i$ (Spatial model)

where E_i represents the expected snail count in ditch i; α_0 represents the snail density over all ditches when all covariates have zero value, α_{1-4} are the coefficients of landscape metrics and environmental factors MSI, SEI, P_1, P_2 and slope, respectively; MSI and SEI are the landscape metrics of buffer zone of ditch i; P_1 and P_2 are the proportion of paddy fields and agrarian road of buffer zone of ditch i; slope is the range of slope of ditch i; ϕ_i represents the spatial random effect for ditch i, namely, spatial effect of neighbouring locations of ditch i.

The spatial correlation of ϕ is assumed as a multivariate normal distribution, $\phi \sim \text{MVN}(0, \Sigma)$, with variance–covariance matrix Σ; the spatial process is assumed as an isotropic stationary process, $\Sigma_{ij} = \sigma^2 \exp((-ud_{ij})^h)$; d_{ij} is the Euclidean distance between i and j. σ^2 is the geographical variability known as the sill, u is a smoothing parameter that controls the rate of correlation decay with increasing distance and measures the range of geographical dependency, h is the scalar parameter controlling the amount of spatial smoothing. This is constrained to lie in the interval $(0, 2)$. When $h = 1$ the type of spatial correlation is an exponential function and when $h = 2$ the type of spatial correlation is a Gaussian function. For the prior distributions for the model parameters, we assume vague normal distributions for the α_{1-5} parameters with large variances, inverse gamma prior distributions for σ^2 and a uniform prior distribution for u. MCMC simulation was applied to fit the models. We ran a two chains sampler with a burn-in of 30,000 iterations.

Bayesian model validation. Validation was undertaken using receiver operating characteristics (ROC) analysis, a method widely applied to diagnostic test evaluation and recently applied to validation of regression models (Brooker et al., 2001, 2002). Predicted density at validation locations was compared with observed density using ROC. The statistic used for the comparison was the area under the curve (AUC), which relates to the ROC curve, a plot of sensitivity versus one minus specificity. We validated the non-spatial and spatial models for live snails using an observed snail density threshold >0.5 per 0.11 m², and val-

Table 1
Spearman's correlation analysis: landscape metrics and environmental features correlated with live snail density and infected snail density

Variable	Source	Spearman's rho coefficient for live snails	Spearman's rho coefficient for infected snails
Total landscape area	Land cover	−0.100[a]	−0.031
No of patches	Land cover	−0.077	−0.044
Mean patch size	Land cover	−0.02	0.038
Total edge	Land cover	−0.058	−0.034
Edge density	Land cover	0.090[a]	−0.03
Mean shape index	Land cover	0.092[a]	0.033
Mean patch fractal dimension	Land cover	0.073	0.017
Shannon's diversity index	Land cover	0.054	−0.055
Shannon's evenness index	Land cover	0.100[a]	−0.025
Largest patch index	Land cover	−0.063	0.037
Landscape shape index	Land cover	−0.023	−0.037
Modified Simpson's diversity index	Land cover	0.072	−0.047
Patch richness	Land cover	−0.038	−0.049
Patch richness density	Land cover	0.106[a]	0.001
Shannon's evenness index	Land cover	0.100[a]	−0.025
Simpson's evenness index	Land cover	0.095[a]	−0.024
Modified Simpson's evenness index	Land cover	0.097[a]	−0.013
Proportion of glebe fields	Land cover	−0.054	0.062
Proportion of paddy fields	Land cover	0.081[a]	−0.043
Proportion of agrarian road	Land cover	0.170[b]	−0.007
Normalised difference vegetation index	SPOT 5	−0.015	0.045
The minimum distance to village	GIS database	0.055	0.022
The maximum distance to village	GIS database	0.029	0.026
Slope	DEM	−0.132[b]	−0.004

GIS, geographic information system; DEM, digital elevation model.
[a] Correlation is significant at the 0.05 level (2-tailed).
[b] Correlation is significant at the 0.01 level (2-tailed).

idated the spatial model for infected snails using the categorical data with or without infected snails.

3. Results

3.1. Correlation of landscape metrics and parasitological data

A ditch network map was created with a total of 554 ditches traced by GPS (Fig. 1). The overall accuracy of the land-cover categories was 0.7935 and the overall Kappa statistic was 0.7492. The accuracies for water, dry fields, paddy fields, agrarian roads, hill forests, ditches with concrete banks, residential areas and hills were 77.78%, 100.00%, 75.00%, 86.67%, 28.57%, 76.92%, 60.00% and 100.00%, respectively.

The parasitological survey data showed that the mean densities of live and infected snails were 0.394 and 0.002 per 0.11 m^2, respectively. The maximum densities of live snails and infected snails were 22.25 and 0.13 per 0.11 m^2, respectively (Figs. 2a and b).

Table 1 shows the correlation of landscape metrics and environmental features with the densities of live and infected snails. The outcome of the correlation analysis suggests that many landscape and environmental variables are significantly correlated with the density of live snails, i.e. the total number of snails recorded, but less so with regard to the number of infected snails present.

3.2. Bayesian modelling and validation

Table 2 shows parameter estimates from the non-spatial model for live snails as well as the spatial models for live snails and two spatial models for infected snails; one spatial model includes the variables which do not show significant Spearman correlations (Spatial model 1 for infected snails) whilst the other spatial model does not include these variables (Spatial model 2 for infected snails). The results obtained from the non-spatial analysis indicate that both the landscape metrics and the environmental covariates investigated were significantly related to snail distribution. In particular, the density of live snails increased with decreasing values of SEI, the proportion of paddy fields, slope, increasing values of MSI and the proportion of agrarian roads. However, in the spatial model, only MSI and SEI were significant, and both variables had significant positive effects. Furthermore, the effect of MSI in the spatial model was different from that in the non-spatial model. However, no landscape metrics or environmental variables were significantly associated with the density of infected snails in the spatial model 1 for infected snails. The model comparison shows that the spatial model has a small DIC value, therefore it has the best fit. In addition, this model has large spatially correlated random effects. For the infected snails, the spatial model 2 with the lower DIC value was selected to develop the prediction map of infected snails.
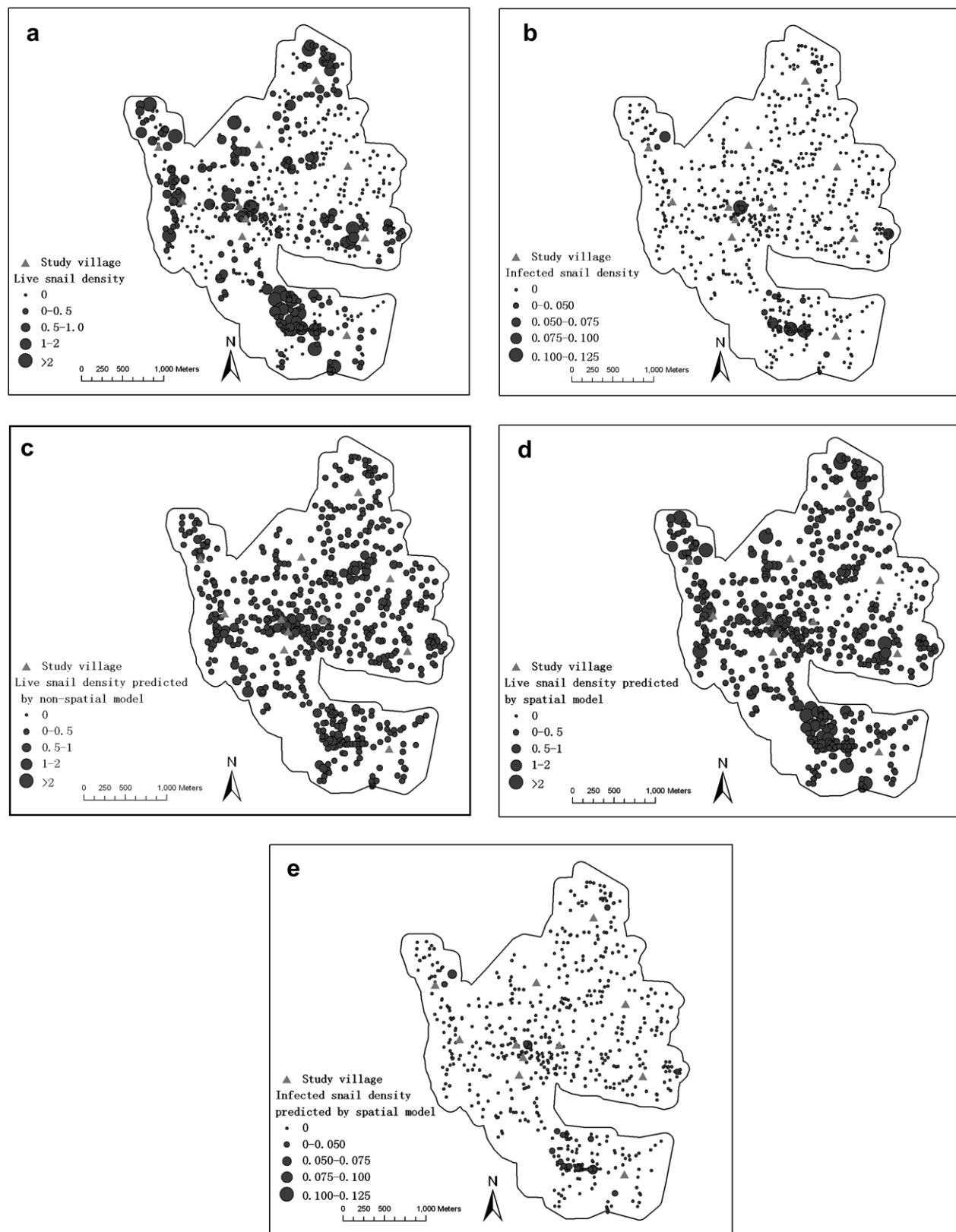
Fig. 2. Data from study ditches in Eryuan county, Yunnan province, China in 2006. Maps of live snails (a) and infected snails (b) distributed in the study ditches. Prediction maps of live snail distribution using the non-spatial model (c) and the spatial model (d). Prediction maps of infected snail distribution using the spatial model (e).

Table 2
Posterior estimates for model parameters

Variable	Non-spatial model for live snails		Spatial mode for live snails		Spatial mode 1 for infected snails[e]		Spatial mode 2 for infected snails[f]	
	Median	95% CI[a]	Median	95% CI[a]	Median	95% CI[a]	Median	95% CI[a]
Intercept	−2.071	(−2.288, −1.831)	−5.884	(−7.874, −4.250)	−1.85	(−4.934, 1.411)	−4.56869	(−8.661, −2.569)
Mean shape index (MSI)	1.902	(1.746, 2.052)	1.195	(0.146, 2.512)	1.983	(−0.598, 3.807)		
Shannon's evenness index (SEI)	−0.3577	(−0.5532, −0.1666)	1.082	(0.106, 2.350)	−0.713	(−3.181, 2.063)		
Proportion of paddy fields	−3.103	(−3.307, −2.906)	1.5	(−0.226, 3.673)	−0.83	(−3.355, 1.533)		
Proportion of agrarian road	3.435	(3.145, 3.728)	0.775	(−1.601, 3.844)	−0.491	(−4.538, 3.226)		
Slope	−0.1195	(−0.154, −0.088)	−0.014	(−0.164, 0.133)	−0.355	(−1.208, 0.054)		
Spatial variation (σ^2)			36.67	(11.827, 256.739)	11.24	(6.451, 21.545)	10.43	(5.390, 22.773)
u[b]			0.027	(0.004, 0.085)	0.0782	(0.009, 0.106)	0.0672	(0.010, 0.091)
h[c]			0.657	(0.258, 1.381)	1.215	(0.503, 1.876)	1.179	(0.482, 1.914)
DIC[d]	21,220		1794.91		609.34		565.34	

[a] Confidence intervals (or posterior intervals).

[b] u is a scalar parameter representing the rate of decline of correlation with distance between points.

[c] h is the scalar parameter controlling the amount of spatial smoothing. This is constrained to lie in the interval (0, 2), when $h = 1$ the type of spatial correlation is an exponential function and when $h = 2$ the type of spatial correlation is a Gaussian function.

[d] The deviance information criterion (DIC), a smaller DIC value indicates a better fitting model.

[e] The model includes these variables which do not show significant Spearman correlations of Table 1.

[f] The model does not include these variables which do not show significant Spearman correlations of Table 1.

The smoothed maps of live snail density at the study site, which were obtained from the non-spatial and the spatial models, are shown in Fig. 2c and d. The prediction map of the spatial model shows the clustered snail distribution, whilst the non-spatial one indicates that the snails are randomly scattered with an average density less than 0.5 per 0.11 m². A high consistency between the observed snail density and predicted snail density was found in the spatial model. Fig. 2e shows the smoothed map of the density of infected snails in the study site obtained from the spatial model. The predicted clusters of infected snails were found to be distributed in the regions where the infected snails were found.

Fig. 3a shows the ROC analysis results for the non-spatial and the spatial models for live snails. The solid black reference lines represent equal trade-offs between the sensitivity and specificity of the prediction modelling. The area under the ROC curve (AUC) is 0.500, and the other lines are ROC lines for the non-spatial and the spatial models.
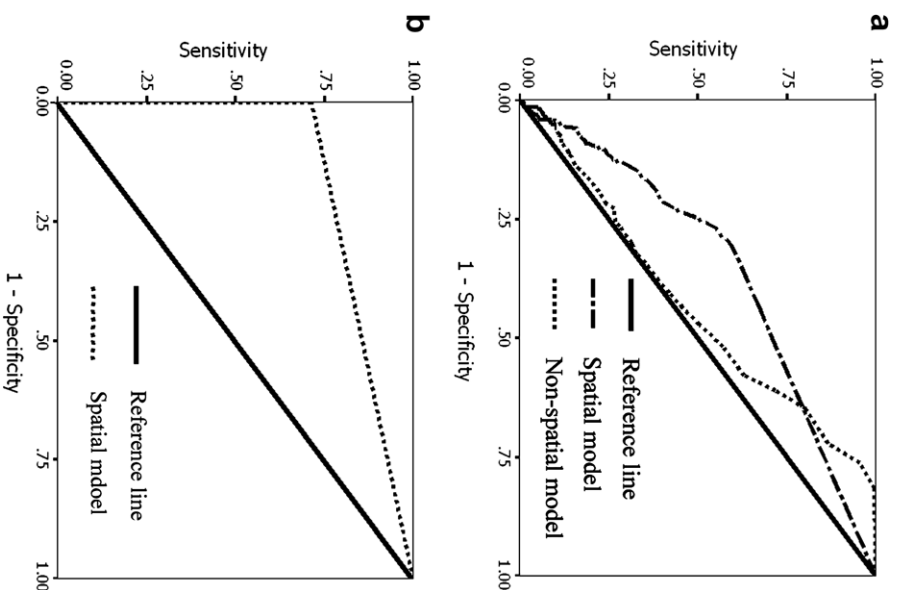
Fig. 3. Receiver operating characteristics (ROC) analysis for the non-spatial and spatial models of live snail density (a) and the spatial model of infected snail density (b). The solid black line is the reference line which represents equal trade-offs between the sensitivity and specificity of prediction modelling; the area under the ROC Curve (AUC) is 0.500; other lines are the ROC lines for non-spatial and spatial models. The AUC of the spatial model and the non-spatial model for live snails was 0.790 and 0.540, respectively. The AUC of the spatial model for infected snail was 0.857.

a

b

The AUC of the spatial and the non-spatial models are 0.79 and 0.54, respectively. The overall predictive accuracy of the spatial model is higher than that of the non-spatial model. Up to 79% of locations with a live snail density >0.5 per 0.11 m^2 were correctly predicted. Fig. 3b shows the ROC analysis results for the spatial models for infected snails. The AUC of 0.857 indicates that the correct prediction rate of infected snails is as high as 85.7%.

4. Discussion

Numerous investigations have been carried out by using GIS, RS coupled with statistical methodology to enhance the understanding of the distribution of schistosomiasis, to provide relevant data for policy modifications, and to guide interventions at different spatial scales (Zhou et al., 2001; Brooker et al., 2002; Yang et al., 2005b; Gong et al., 2006; Chen et al., 2007). However, the utility of such studies are often limited due to low spatial resolution of remotely-sensed environmental data at the macro-scale (national, provincial or county level) rather than at the micro-scale (the village level and lower) (Clennon et al., 2004). The scale upon which a decision is based is crucial, particularly with regard to designing intervention programmes (Gurarie and King, 2005; Yang et al., 2005b; Brooker, 2007). With the recent launch of new and powerful satellites, access to higher resolution RS data has become possible at relatively low cost. Xu et al. (2004) extracted land-cover and land-use information from a 4 m resolution IKONOS image to predict snail density in the mountainous regions in southwest Sichuan Province, China. However, their predictive capacity could have been further improved if the heterogeneous landscape effect on component structures, and the spatial effect on various geographical representations, had been included in their study.

The former effect can be estimated by landscape pattern analysis which is capable of explaining the outcome of continuous interactions amongst their biotic and abiotic components. This has been applied to various vector-borne diseases (Kitron, 1998; Brooker et al., 2002), and also to a spatially heterogeneous of disease distribution pattern on a range of scales (Françoise and Jacques, 2003). The latter effect can be tackled by Bayesian hierarchical modelling (Boyd et al., 2005; Feltbower et al., 2005; Wang et al., 2006a). Snails tend to be clustered in the natural environment. Bayesian spatial hierarchical modelling takes into account spatial correlations between each pair of locations and can integrate prior information into observed data. Recently developed computational tools such as the Gibbs sampler (Gelfand et al., 1992), as well as MCMC simulation allow the possibility of posterior analyses (Yang et al., 2005a; Ashby, 2006; Raso et al., 2006). Therefore, we integrated both approaches in our study by considering both effects on the prediction of snail distribution.

From landscape pattern analysis, we found that decreasing the heterogeneity of the landscape can decrease the snail density. Both the non-spatial and the best-fitting spatial models reveal a significant positive association between MSI and live snail density. This positive association can be understood as most snails cluster along irregular patch shapes (such as irrigation ditches) and are less attracted to flat regions such as crop fields (Graham et al., 2004). The irregular shapes of many ditches increase the chances of snail survival in two ways. One is that ditches are often divided into many parts and create more snail habitats at a smaller scale (Li et al., 1999a; Xu et al., 2004; Liang et al., 2007). The other is that irregular shapes increase the survival chances of snails by decreasing the efficiency of molluscicides (Huang et al., 2003; Wang et al., 2006b). In addition, results from the best fit spatial model show that live snail density increases with an increasing value of SEI. The higher the SEI value, the more diversity or the more heterogeneous the distribution of types of patches becomes (Françoise and Jacques, 2003), which in turn increases snail abundance by enlarging the suitable snail habitats.

The results of the Bayesian spatial model suggest that the spatial pattern of the density of live snails at the local scale is mainly governed by heterogeneity of the landscape (represented by the indicator *u*). Thus, *u* represents the rate of decline of the correlation with distance between the centre points of the ditches. The small value of *u* (0.027) in the present study indicates that the minimum distance in metres at which the spatial correlation between locations drops below 5% is 110 m, known as the range of geographical dependency and expressed by 3/*u* (Raso et al., 2006). The distance is lower than the nearest distance between study villages. This suggests that the variability of the distribution pattern of live snails mainly occurs within the village boundary, and that this variability is closely related to the agriculture structure and therefore is important in relation to snail control. Our work suggests that means used for snail control, when applied in this type of environment, should consider the heterogeneity of the landscape. The range of geographical dependency of the density of infected snails is about 44.6 m, which is smaller than that of the density of live snails; it represents the clustering effect of infected snails and it appears to be stronger than that of uninfected snails. Other authors have expressed the notion that many factors influence the distribution of infected snails and that the situation is much more complicated, e.g. the distribution of infected cattle may contribute greatly to the distribution of infected snails, in addition to the environmental factors (Li et al., 1999a; Li et al., 1999b). However, the work presented here indicates that the environmental factors may have the strongest impact, at least in mountainous areas.

The smoothed maps of the predicted snail distribution with live and infected snail data in 2006, based on the spatial correlation structures, provided far more information than the limited survey data. The perceived risk at locations without data can be presented in a prediction map. Hence, a dramatic saving of scarce labour resources could be achieved in future snail surveys if they were guided by

stratified sampling based on prediction maps with land-cover/land-use and landscape features. In addition, these maps can be used by decision-makers for the design and implementation of control interventions for *S. japonicum* in high-risk regions characterised by a complex geographical environment.

In contrast to other studies, there was no significant correlation between NDVI, distance to village and density of live snails in our study (Utzinger et al., 2005; Yang et al., 2005b). This could be due to study scales, i.e. when a study is undertaken at the macro-scale, and the size of study target, e.g. the snail population, is very small (or the field survey is un-informative), there is always a large variability in the estimated variables. Our study was carried out at a small, local scale, which is very different from the studies cited above which were undertaken on a larger scale.

However, limitations still exist. Firstly, in addition to landscape metrics and environmental variables, many other factors, such as socio-economic impact, interventions, etc., should be taken into account in the prediction of snail distributions. Second, since the distribution and density of snails depend on its status in preceding years, temporal effects should also be considered.

In conclusion, the present work represents the first time an integrated approach has been tried in which landscape pattern analysis has been combined with Bayesian modelling. In our opinion, this is a powerful and statistically robust approach to identify high-risk areas with a heterogeneous, complex and not easily accessed environment. It has provided us with an efficient tool to predict snail density at the local scale in a mountainous region. Furthermore, the methodology developed in our paper holds promise for other vector-borne diseases.

Acknowledgements

References

Ashby, D., 2006. Bayesian statistics in medicine: a 25 year review. Stat. Med. 25, 3589–3631.

Boyd, H.A., Flanders, W.D., Addiss, D.G., Waller, L.A., 2005. Residual spatial correlation between geographically referenced observations: a Bayesian hierarchical modeling approach. Epidemiology 16, 532–541.

Brooker, S., Hay, S.I., Issae, W., Hall, A., Kihamia, C.M., Lwambo, N.J., Wint, W., Rogers, D.J., Bundy, D.A., 2001. Predicting the distribution of urinary schistosomiasis in Tanzania using satellite sensor data. Trop. Med. Int. Health 6, 998–1007.

Brooker, S., Hay, S.I., Bundy, D.A., 2002. Tools from ecology: useful for evaluating infection risk models? Trends Parasitol. 18, 70–74.

Brooker, S., Clements, A.C., Bundy, D.A., 2006. Global epidemiology, ecology and control of soil-transmitted helminth infections. Adv. Parasitol. 62, 221–261.

Brooker, S., 2007. Spatial epidemiology of human schistosomiasis in Africa: risk models, transmission dynamics and control. Trans. R. Soc. Trop. Med. Hyg. 101, 1–8.

Calvete, C., Blanco-Aguiar, J.A., Virgos, E., Cabezas-Diaz, S., Villafuerte, R., 2004. Spatial variation in helminth community structure in the red-legged partridge (*Alectoris rufa* L.): effects of definitive host density. Parasitology 129, 101–113.

Chen, Z., Zhou, X.N., Yang, K., Wang, X.H., Yao, Z.Q., Wang, T.P., Yang, G.J., Yang, Y.J., Zhang, S.Q., Wang, J., Jia, T.W., Wu, X.H., 2007. Strategy formulation for schistosomiasis japonica control in different environmental settings supported by spatial analysis: a case study from China. Geospatial Health 2, 223–231.

Clennon, J.A., King, C.H., Muchiri, E.M., Kariuki, H.C., Ouma, J.H., Mungai, P., Kitron, U., 2004. Spatial patterns of urinary schistosomiasis infection in a highly endemic area of coastal Kenya. Am. J. Trop. Med. Hyg. 70, 443–448.

Clennon, J.A., King, C.H., Muchiri, E.M., Kitron, U., 2007. Hydrological modelling of snail dispersal patterns in Msambweni, Kenya and potential resurgence of *Schistosoma haematobium* transmission. Parasitology 134, 683–693.

Deter, J., Berthier, K., Chaval, Y., Cosson, J.F., Morand, S., Charbonnel, N., 2006. Influence of geographical scale on the detection of density dependence in the host–parasite system, *Arvicola terrestris* and *Taenia taeniaeformis*. Parasitology 132, 595–605.

Dong, Y., Feng, X.G., Yang, C.G., Dong, X.Q., 2006. Analysis of geographic information system on transmission of schistosomiasis in Yunnan Province. Chin. J. Schisto. Control 18, 109–112.

Elliott, P., Wartenberg, D., 2004. Spatial epidemiology: current approaches and future challenges. Environ. Health Perspect 112, 998–1006.

Engels, D., Chitsulo, L., Montresor, A., Savioli, L., 2002. The global epidemiological situation of schistosomiasis and new approaches to control and research. Acta Trop. 82, 139–146.

Feltbower, R.G., Manda, S.O., Gilthorpe, M.S., Greaves, M.F., Parslow, R.C., Kinsey, S.E., Bodansky, H.J., McKinney, P.A., 2005. Detecting small-area similarities in the epidemiology of childhood acute lymphoblastic leukemia and diabetes mellitus, type 1: a Bayesian approach. Am. J. Epidemiol. 161, 1168–1180.

Françoise, B., Jacques, B., 2003. Landscape Ecology: Concepts, Methods and Applications. Science Publishers, St. Albans, UK.

Gangnon, R.E., Clayton, M.K., 2000. Bayesian detection and modeling of spatial disease clustering. Biometrics 56, 922–935.

Gelfand, A.E., Smith, A.F.M., Lee, T.M., 1992. Bayesian analysis of constrained parameter and truncated data problems using gibbs sampling. J. Am. Stat. Assoc. 87, 523–532.

Gong, P., Xu, B., Liang, S., 2006. Remote sensing and geographic information systems in the spatial temporal dynamics modeling of infectious diseases. Sci. China C. Life Sci. 49, 573–582.

Graham, A.J., Danson, F.M., Giraudoux, P., Craig, P.S., 2004. Ecological epidemiology: landscape metrics and human alveolar echinococcosis. Acta Trop. 91, 267–278.

Graham, A.J., Danson, F.M., Giraudoux, P., 2005. Ecological epidemiology: the role of landscape structure in the transmission risk of the fox tapeworm *Echinococcus multilocularis* (Leukart 1863) (Cestoda:Cyclophyllidea:Taeniidae). Prog. Phys. Geog. 29, 77–79.

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.

Gurarie, D., King, C.H., 2005. Heterogeneous model of schistosomiasis transmission and long-term control: the combined influence of spatial variation and age-dependent factors on optimal allocation of drug therapy. Parasitology 130, 49–65.

Huang, Y.X., Sun, L.P., Hong, Q.B., 2003. Studies on molluscicidal effect of niclosamide ethanolamine salt soluble powder against *Oncomelania hupensis*. Chin. J. Schisto. Control 15, 255–258.

Jackson, L.E., Hilborn, E.D., Thomas, J.C., 2006. Towards landscape design guidelines for reducing Lyme disease risk. Int. J. Epidemiol. 35, 315–322.

Kariuki, H.C., Clennon, J.A., Brady, M.S., Kitron, U., Sturrock, R.F., Ouma, J.H., Ndzovu, S.T., Mungai, P., Hoffman, O., Hamburger, J., Pellegrini, C., Muchiri, E.M., King, C.H., 2004. Distribution patterns and cercarial shedding of *Bulinus nasutus* and other snails in the Msambweni area, Coast Province, Kenya. Am. J. Trop. Med. Hyg. 70, 449–456.

Kiss, I.Z., Green, D.M., Kao, R.R., 2006. Infectious disease control using contact tracing in random and scale-free networks. J.R. Soc. Interface. 3, 55–62.

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

Koukounari, A., Sacko, M., Keita, A.D., Gabrielli, A.F., Landoure, A., Dembele, R., Clements, A.C., Whawell, S., Donnelly, C.A., Fenwick, A., Traore, M., Webster, J.P., 2006. Assessment of ultrasound morbidity indicators of schistosomiasis in the context of large-scale programs illustrated with experiences from Malian children. Am. J. Trop. Med. Hyg. 75, 1042–1052.

Kumar, S., Stohlgren, T.J., Chong, G.W., 2006. Spatial heterogeneity influences native and nonnative plant species richness. Ecology 87, 3186–3199.

Li, F., Xia, D.G., Ma, C.H., Jia, X.M., Zhang, X.Z., 1999a. Longitudinal observation on infected snail post as a source of inaction in a mountainous village of Dali city. Chin. J. Parasitol. Parasit. Dis. 17, 165–166.

Li, F., Xia, D.G., Ma, C.H., Jia, X.M., Zhang, X.Z., 1999b. Reappearance of infected snail in positive snail spot in mountainous region of Yunnan province. J. Appl. Parasit. Dis. 7, 61–64.

Li, X.B., 2006. Analysis for schistosomiasos status of Eryuan county from 2000 to2004. Parasit. Infect. Dis. 4, 148–149.

Liang, S., Yang, C., Zhong, B., Qiu, D., 2006. Re-emerging schistosomiasis in hilly and mountainous areas of Sichuan, China. Bull. World Health Organ. 84, 139–144.

Liang, S., Seto, E.Y., Remais, J.V., Zhong, B., Yang, C., Hubbard, A., Davis, G.M., Gu, X., Qiu, D., Spear, R.C., 2007. Environmental effects on parasitic disease transmission exemplified by schistosomiasis in western China. Proc. Natl. Acad. Sci. 104, 7110–7115.

Raso, G., Vounatsou, P., Gosoniu, L., Tanner, M., N'Goran, E.K., Utzinger, J., 2006. Risk factors and spatial patterns of hookworm infection among school children in a rural area of western Cote d'Ivoire. Int. J. Parasitol. 36, 201–210.

Ross, A.G., Bartley, P.B., Sleigh, A.C., Olds, G.R., Li, Y., Williams, G.M., McManus, D.P., 2002. Schistosomiasis. N. Engl. J. Med. 346, 1212–1220.

Russell, C.A., Real, L.A., Smith, D.L., 2006. Spatial control of rabies on heterogeneous landscapes. PLoS ONE 1, e27.

Spear, R.C., Seto, E., Liang, S., Birkner, M., Hubbard, A., Qiu, D., Yang, C., Zhong, B., Xu, F., Gu, X., Davis, G.M., 2004. Factors influencing the transmission of *Schistosoma japonicum* in the mountains of Sichuan Province of China. Am. J. Trop. Med. Hyg. 70, 48–56.

Spiegelhalter, D.J., Best, N.G., Carlin, B.P., van der Linde, A., 2002. Bayesian measures of model complexity and fit. J.R. Stat. Soc. B. 64, 583–639.

Steinmann, P., Zhou, X.N., Matthys, B., Li, Y.L., Li, H.J., Chen, S.R., Yang, Z., Fan, W., Jia, T.W., Li, L.H., Vounatsou, P., Utzinger, J., 2007. Spatial risk profiling of *Schistosoma japonicum* in Eryuan county, Yunnan province, China. Geospatial Health 2, 59–73.

Utzinger, J., Zhou, X.N., Chen, M.G., Bergquist, R., 2005. Conquering schistosomiasis in China: the long march. Acta Trop. 96, 69–96.

Wang, X.H., Wu, X.H., Zhou, X.N., 2006a. Bayesian estimation of community prevalences of *Schistosoma japonicum* infection in China. Int. J. Parasitol. 36, 895–902.

Wang, X.Y., Ang, L.Q., Zhang, L.H., Ge, Q.J., 2006b. Progress of research on molluscicides. Chin. J. Schisto. Control 18, 474–476.

Wintle, B.A., Bardos, D.C., 2006. Modeling species-habitat relationships with spatially autocorrelated observation data. Ecol. Appl. 16, 1945–1958.

Xu, B., Gong, P., Biging, G., Liang, S., Seto, E., Spear, R., 2004. Snail density prediction for schistosomiasis control using IKONOS and ASTER images. Photogramm. Eng. Rem. S. 70, 1285–1294.

Yang, G.J., Vounatsou, P., Zhou, X.N., Tanner, M., Utzinger, J., 2005a. A 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.

Yang, G.J., Vounatsou, P., Zhou, X.N., Utzinger, J., Tanner, M., 2005b. A review of geographic information system and remote sensing with applications to the epidemiology and control of schistosomiasis in China. Acta Trop. 96, 117–129.

Yuan, H., Jiang, Q., Zhao, G., He, N., 2002. Achievements of schistosomiasis control in China. Mem. Inst. Oswaldo. Cruz. 97, 187–189.

Zhang, B., Zhang, Z.Y., Xu, D.J., Zhou, X.N., 2003. Study on the relationship between Terra-MODIS image and the snail distribution in marshland of Jiangning county, Jiangsu province. Chin. J. Epidemiol. 24, 257–261.

Zhou, X.N., Malone, J.B., Kristensen, T.K., Bergquist, N.R., 2001. Application of geographic information systems and remote sensing to schistosomiasis control in China. Acta Trop. 79, 97–106.

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 in China-then and now. Acta Trop. 96, 97–105.

Zhou, X.N., Guo, J.G., Wu, X.H., Jiang, Q.W., Zheng, J., Dang, H., Wang, X.H., Xu, J., Zhu, H.Q., Wu, G.L., 2007. Epidemiology of Schistosomiasis in the People's Republic of China, 2004. Emerg. Infect. Dis. 10, 1470–1477.