

*Geospatial Health* 8(1), 2013, pp. 125-132

## Comparison of data-fitting models for schistosomiasis: a case study in Xingzi, China

Yi Hu<sup>1,2,3,\*</sup>, Cheng-Long Xiong<sup>1,2,\*</sup>, Zhi-Jie Zhang<sup>1,2,3,4</sup>, Robert Bergquist<sup>5</sup>, Zeng-Liang Wang<sup>1,2,3</sup>, Jie Gao<sup>1,2,3</sup>, Rui Li<sup>1,2,3</sup>, Bo Tao<sup>6</sup>, Qiu-Lin Jiang<sup>6</sup>, Qingwu Jiang<sup>1,2,3</sup>

<sup>1</sup>Department of Epidemiology, School of Public Health, Fudan University, Shanghai 200032, People's Republic of China; <sup>2</sup>Key Laboratory of Public Health Safety, Ministry of Education, Shanghai 200032, People's Republic of China; <sup>3</sup>Laboratory for Spatial Analysis and Modeling, School of Public Health, Fudan University, Shanghai 200032, People's Republic of China; <sup>4</sup>Biomedical Statistical Center, Fudan University, Shanghai 200032, People's Republic of China; <sup>5</sup>Ingerod, Brastad, Sweden; <sup>6</sup>Xingzi Station for Schistosomiasis Prevention and Control, Xingzi 332800, People's Republic of China

**Abstract.** When modelling prevalence data, epidemiological studies usually employ either Gaussian, binomial or Poisson models. However, reasons are seldom given in the literature why the chosen model was felt to be the most appropriate. In this study, we compared all three models for fitting schistosomiasis risk in Xingzi county, Jiangxi province, People's Republic of China. Parasitological data from conventional surveys were available for 36,208 individuals aged between 6 and 65 years from 42 sampled villages and used in combination with environmental data to map the spatial patterns of schistosomiasis risk. The results show that the Poisson model fitted the data best and this model identified the role of environmental risk factors in explaining the geographical variation of schistosomiasis risk. These factors were further used to develop a predictive map, which has important implications for the control and eventual elimination of schistosomiasis in the People's Republic of China.

**Keywords:** *Schistosoma japonicum*, environmental factors, geostatistics, spatial prediction, geographical information systems, People's Republic of China.

### Introduction

It is well recognised that reliable maps of the spatial pattern of parasitic diseases facilitate both planning and evaluation of disease control. During the past decades, there has been a growing research interest in modelling disease risk (Hay et al., 2006). Most of these studies have modelled prevalence data (Thomson et al., 1999; Brooker et al., 2006; Diggle et al., 2007; Zhang et al., 2009; Ekpo et al., 2013), since this indicator is readily available and thus easily collected (Clements et al., 2006). Different models, such as Gaussian (Bavia et al., 2001), binomial (Diggle et al., 2007) and Poisson (Zhang et al., 2012) have been used to fit prevalence data for the exploration of the association between prevalence of parasitic infections and risk factors, thus producing predictive maps of the relative risk in various geographical areas. However, these studies generally forego a discussion of which

type of modelling would fit the prevalence data best, thus there is no discussion supporting the choice of statistical approach, be it binomial, Gaussian or Poisson distribution.

A fully Bayesian geostatistical approach has been increasingly adopted by studies modelling the spatial distribution of parasitic diseases (Carabin et al., 2003; Diggle et al., 2007; Clements et al., 2009). This method explicitly incorporates the spatial correlation structure of the data, and it has the ability to include covariate effects as well as a comprehensive representation of uncertainty in the model outputs. Recently, Diggle and Jr.Ribeiro (2007) have presented a general framework of generalised, linear models for geostatistical data, which provides a reliable way to compare different models when fitting prevalence data. The classic (non-spatial) regression model, however, should be given priority as it fits the data better than the spatial one if model residuals show no spatial correlation (Fischer and Wang, 2011).

In the present study, we employed all three models (Gaussian, binomial and Poisson) to investigate the spatial distribution of the prevalence of *Schistosoma japonicum* infection in Xingzi county, Jiangxi province, People's Republic of China (P.R. China). The predictive map of schistosomiasis risk across the

Corresponding author:

Zhi-Jie Zhang

Department of Epidemiology, School of Public Health,  
Fudan University, Shanghai 200032, People's Republic of China  
Tel./Fax +86 21 5423 7410

E-mail: epistat@gmail.com

\*contributed equally to this work

region was then produced, including the corresponding uncertainties. Finally, we compared results of the three models to give some insights for model selection in further studies. This analysis provides estimates of the effects of factors that dominate the spatial patterns of schistosomiasis infection, e.g. climatic conditions and social variables.

## Materials and methods

### Study area

Xingzi county, located in the northern part of Jiangxi province, has a total population of 25 million with 16 million living in areas with ongoing schistosomiasis transmission. Historically, Xingzi county is one of the most schistosomiasis-affected regions in P.R. China (Zhou et al., 2005). Snail habitats there have been characterised as “land in winter, water in summer”, i.e. water level rises and submerges the whole bottomland in the wet season (April - October), while the water recedes and the grasslands emerge in the dry season (November-March), thus providing an ideal environment for growth and reproduction of the snail intermediate host.

### Parasitological data

A cross-sectional study of *Schistosoma japonicum* infection was conducted in 42 randomly selected villages in Xingzi county by health professional from the

local anti-schistosomiasis control station. The study took place in 2008 and included a total of 36,208 residents, aged 6–65 years. The indirect haemagglutination assay test (Yu, 1981) was first applied to screen all individuals for anti-schistosomiasis antibodies and this was followed by stool microscopy with the Kato-Katz technique (Feldmeier and Poggensee, 1993) using three thick smears from one stool specimen from all sero-positive individuals. Only egg-positive individuals were considered to be infected with *S. japonicum*.

### Environmental data

Elevation data covering the study area were obtained from the Global Land Information System (GLIS) of the United States Geological Survey (<http://edcwww.cr.usgs.gov/landdaac/gtopo30/>). Geomorphology data were provided by the State Key Laboratory of Resources and Environmental Information Systems (LREIS) from the Institute of Geographic Sciences and Natural Resources Research (IGSNRR), Chinese Academy of Science (CAS). Monthly data of rainfall and hours of daylight in 2008 were obtained from China Meteorological Data Sharing Service System (<http://cdc.cma.gov.cn/home.do>). Inverse distance weighting (IDW) interpolation was used to derive data at the study area.

Considering the effects of overall climate condition and seasonal variation in local climate, four indexes (minimum, maximum, mean and standard deviation in 12 months) were used here. All 8-day, global 1 km

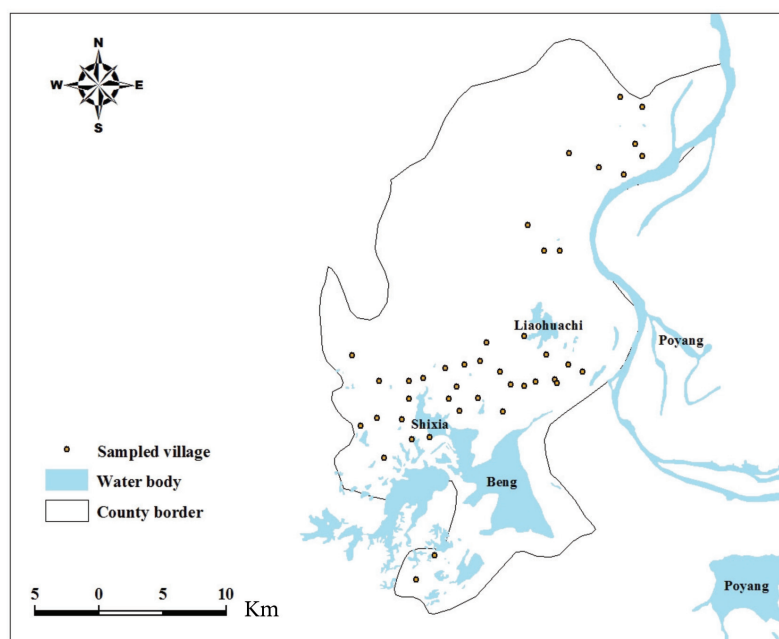


Fig. 1. Distribution of sampled villages in the study area (Xingzi county, Jiang Xi province, P.R. China).

products for land surface temperature (LST) in 2008 and the monthly global 1 km products for the normalized difference vegetation index (NDVI) covering Xingzi county were downloaded from the Level 1 and Atmosphere Archive and Distribution System (<http://ladsweb.nascom.nasa.gov/data/search.html>). Similarly, the four climate indexes mentioned above were obtained for the study area.

#### *Distance to snail habitats and water bodies*

The location of snail habitats was recorded using a hand-held global positioning system (GPS) instrument (MobileMapper™, Thales Navigation Inc.). Water body data were extracted from the World Wildlife Fund's Conservation Science Data Sets (<http://world-wildlife.org>). For each sample village, the Euclidian distance to the nearest snail habitat and the nearest water body were calculated.

#### *Population density*

Population density (measured as the number of persons per km<sup>2</sup>) was downloaded from the Center for International Earth Science Information Network (CIESIN) at Columbia University (<http://sedac.ciesin.columbia.edu/data/collection/gpw-v3>).

#### *Statistical analysis*

All statistical analyses were implemented using R2.14.1 software (R Development Core Team, Vienna, Austria). Generalised linear models (Gaussian, binomial and Poisson) were fitted to the available schistosomiasis data to identify the significant environmental covariates. Note that for the Gaussian model, a Box-Cox transformation (Box and Cox, 1964) of the crude prevalence was performed to assume Gaussian model to hold. Initially, univariate analyses were conducted and variables with  $P > 0.2$  were excluded. Colinearity was investigated among all possible pairs of potential predictor variables and if any pair had a correlation coefficient  $> 0.2$ , the member of the pair that was thought less likely to be biologically important was excluded. With the remaining variables, backwards-stepwise regression was conducted using  $P > 0.1$  as the exit criterion and  $P \leq 0.05$  as the entry criterion. Finally, the residuals from this model were tested for spatial autocorrelation using variogram (Cressie, 1993), which was introduced in the model-based geostatistical model. If the residuals showed no spatial autocorrelation, this classic (non-

spatial) model was considered as the fitting model; otherwise, a Bayesian geostatistical approach was used instead.

#### *Model-based geostatistical model*

Let  $Y_i$  denote the number of positive samples out of  $n_i$  individuals tested in village  $i$ . The Gaussian model assumes that the transformed crude infection rates ( $Y_i/n_i$ ) are conditionally independent Gaussian variates, while the binomial and Poisson models assume that  $Y_i$  are conditionally independent binomial and Poisson variates, respectively, given an unobserved spatial stochastic process (denoted by  $S$ ), and that the mean response at  $i$  depends on explanatory variables observed at village  $i$  and on  $S(i)$ . The spatial random-effect  $S(i)$  was modelled as a stationary Gaussian process (signal) with mean 0, variance  $\sigma^2$  and correlation function  $\rho(u)$ , where  $u$  is the distance between two villages. The exponential correlation function used here is defined as follows:

$$\rho(u) = \exp(-u/\phi) \quad (1)$$

in which  $\phi > 0$  is a scale parameter with the dimensions of distance, measuring the rate at which the spatial correlation decays over distance. Under this function, the practical range (defined as the minimum distance at which spatial correlation between locations is below 5%) is approximately  $3\phi$ .

Hence, for the Gaussian model:

$$g(Y_i/n_i) = \mu(i) + S(i) \quad (2)$$

where  $g(Y_i/n_i)$  is a transformation function and  $\mu(i)$  refers to a trend surface model or covariate information.

For the binomial model:

$$\log\{p(i)/[1 - p(i)]\} = \mu(i) + S(i) \quad (3)$$

where  $p(i)$  is the probability that a randomly selected person at a village  $i$  will test positive for schistosomiasis.

For the Poisson model:

$$\log Y_i - \log n_i = \mu(i) + S(i) \quad (4)$$

where  $\log n_i$  is used as an offset. The role of  $S(i)$  in the model is to capture residual, spatial variation after adjusting for covariates, without which equations (2), (3) and (4) become generalised linear models.

The predicted map of prevalence of *S. japonicum* infection, including prediction uncertainties, was then produced. The predictive approach was chosen as follows: if the residuals of the generalised linear model showed no spatial autocorrelation over the study area, this non-spatial model was chosen as the predictive model; otherwise the Bayesian geostatistical approach was used. For prediction, the study region was divided at the level of 1-km resolution, and all the corresponding, environmental variables mentioned above were obtained.

Finally, we used an indicator of the mean squared error (MSE), which was defined as follows, to assess the goodness-of-fit for the three models:

$$MSE = \frac{1}{m} \sum_{i=1}^m (\hat{p}_i - Y_i/n_i)^2 \quad (5)$$

where  $m$  is the number of sampled villages and  $\hat{p}_i$  denotes the predicted prevalence at village  $i$  using different models. The smaller the indicator value, the better the fit of the model.

## Results

Tables 1, 2 and 3 report the results of the Gaussian, binomial, and Poisson models, respectively. For each model, both  $\sigma^2$  and  $\phi$  are less than 0.01, indicating that no spatial trends are present in the residuals after adjusting for the environmental factors. The coefficients of the Gaussian model showed that the minimum rainfall, the minimum hours of daylight, the maximum LST at daytime, and the maximum and minimum LST at night were all significantly associated with schistosomiasis, while those of both the binomial and the Poisson models showed that distance to snail habitats, the maximum and minimum rainfall,

mean hours of daylight, the maximum LST at daytime, the maximum and minimum LST at night and the mean NDVI were significantly associated with schistosomiasis risk and that the estimates of the coefficients changed only little.

Fig. 2 illustrates the predicted prevalence maps of *S. japonicum* infection. Map (a), produced with the Gaussian model, shows that two regions with a relatively high schistosomiasis prevalence (from 0.01 to 0.10) were found (one near Poyang Lake and the other around Beng Lake) and that the schistosomiasis risk was low in the rest of the county (<0.01). Map (b), produced with the binomial model, also showed two regions with relatively high risk of schistosomiasis near the Poyang and Beng lakes, but the spatial pattern was quite different: the areas were smaller compared to map (a), but the risks were higher (from 0.08 to 0.22). Map (c), produced with the Poisson model, showed a pattern similar to that of map (b), but with a slightly higher risk at Poyang Lake (0.27).

Fig. 3 represents the prediction uncertainties, quantified by standard deviation. As seen in map (a), a low level of uncertainty was apparent in locations close to the sampled villages and a high level of uncertainty was present in locations distant from these villages. In maps (b) and (c), one region of higher level of uncertainty nearby Poyang Lake stood out. However, while the uncertainty in map (c) was high, the uncertainties in the rest of the other areas were very low. In summary, uncertainties were high in map (a), while those in map (b) and map (c) were low across the study region. Finally, model comparison revealed in Tables 1, 2 and 3 that Poisson model had the smallest MSE value. However, all the MSEs were in the same order of magnitude ( $10^{-3}$ ).

Table 1. Environmental determinants of schistosomiasis japonica: the Gaussian regression model.

Parameters	Estimate	Standard error	$t$	P
Intercept	-8.32	3.36	-2.47	0.02
Population density	1.58e-04	9.98e-05	1.59	0.12
Rainfall (min)	-2.72	0.73	-3.72	<0.01
Hours of daylight (min)	0.34	0.13	2.51	0.02
LST at daytime (max)	0.02	0.01	2.00	0.04
LST at night (min)	1.21e-03	4.21e-04	2.88	<0.01
LST at night (max)	1.34e-02	5.68e-03	2.37	0.02
$\sigma^2$ (residuals)	<0.01	-	-	-
$\phi$ (residuals)	<0.01	-	-	-
MSE	5.73e-03	-	-	-

Table 2. Environmental determinants of schistosomiasis japonica: the binomial regression model.

Parameters	Estimate	Standard error	$z$	P
Intercept	21.75	65.01	0.34	0.74
Distance to snail habitat	-2.04e-04	4.64e-05	-4.39	<0.01
Rainfall (min)	-61.47	17.79	-3.45	<0.01
Rainfall (max)	-0.24	0.09	-2.60	<0.01
Hours of daylight (mean)	-3.74	1.31	-2.86	<0.01
LST at daytime (max)	0.26	0.10	2.61	<0.01
LST at night (min)	0.02	3.48e-03	4.33	<0.01
LST at night (max)	0.15	0.04	3.32	<0.01
NDVI (mean)	2.73e-04	1.21e-04	2.25	0.02
$\sigma^2$ (residuals)	<0.01	-	-	-
$\phi$ (residuals)	<0.01	-	-	-
MSE	4.49e-03	-	-	-

## Discussion

This study compared three models in assessing the distribution of risk factors and predicting schistosomiasis risks in an area in P.R. China, where re-emergence and transmission of *S. japonicum* are evident. Non-spatial regression models were found to fit the data well, indicating that the environmental factors used in the study perfectly explain the geographical variation of schistosomiasis risks and can therefore be used to develop predictive maps. The Poisson model was shown to have the smallest MSE and also to fit the data best. Hence, the predictive map produced by this model provides a more reliable basis the identification of priority areas for schistosomiasis control and appraisal of potential impact of control strategies applied.

In terms of inference, the Gaussian model is the sim-

plest among those tried, since prevalence is a continuous variable. It can be directly described by normal distribution through transformation functions, while the binomial and the Poisson models are better used to describe discrete variables rather than continuous variables such as the number of people infected within a population. Based on large-sample theories, one of the three distributions can be approximated by another. The binomial distribution, for example, is approximately normal for large populations and for success probability ( $p$ ) not too close to zero or one (Box et al., 1978). Field data in practice, however, usually cannot satisfy large-sample theories. In practice, it is therefore necessary to investigate which model fits the field data best. Our study shows that the binomial and the Poisson models fit the data better than the Gaussian model and further investigation indicated that the

Table 3. Environmental determinants of schistosomiasis japonica: the Poisson regression model.

Parameters	Estimate	Standard error	$t$	P
Intercept	21.64	64.52	0.34	0.74
Distance to snail habitat	-2.00e-04	4.60e-05	-4.35	<0.01
Rainfall (min)	-60.86	17.69	-3.44	<0.01
Rainfall (max)	-0.24	0.09	-2.60	<0.01
Hours of daylight (mean)	-3.71	1.30	-2.85	<0.01
LST at daytime (max)	0.26	0.10	2.61	<0.01
LST at night (min)	0.02	3.46e-03	4.32	<0.01
LST at night (max)	0.14	0.04	3.30	<0.01
NDVI (mean)	2.70e-04	1.21e-04	2.23	0.03
$\sigma^2$ (residuals)	<0.01	-	-	-
$\phi$ (residuals)	<0.01	-	-	-
MSE	4.46e-03	-	-	-

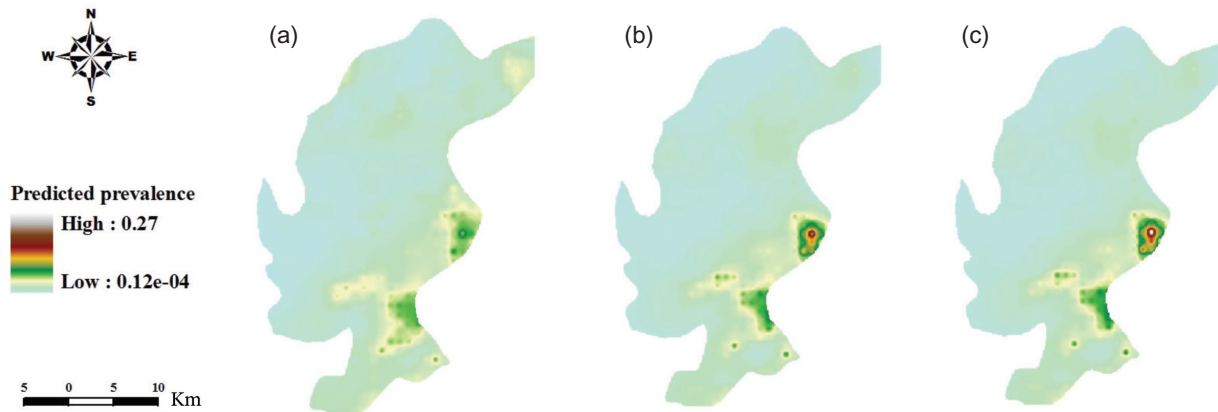


Fig. 2. Predicted prevalence of schistosomiasis japonica using (a) a Gaussian model, (b) a binomial model, and (c) a Poisson model.

Poisson model is optimal. When assessing association between schistosomiasis risk and environmental factors, the three different models gave similar, but different results. The Gaussian model indicated that the minimum rainfall, the minimum hours of daylight, the maximum LST at daytime and the maximum and minimum LST at night were significant risk factors with respect to schistosomiasis, while both the binomial and the Poisson models reported that distance to snail habitats, the maximum and minimum rainfall, mean hours of daylight, the maximum LST at daytime, the maximum and minimum LST at night as well as the mean NDVI were significantly associated with risk. These identified factors are already not only well-known but also consistent and interpretable with the epidemiology of schistosomiasis and snail biology. Evidently, the binomial and Poisson models identified more environmental factors than the Gaussian model associated with schistosomiasis. Even so, examination of the spatial correlation of residuals indicated that the covariates in the Gaussian model can

still explain the geographical variation of schistosomiasis risk equally well as those in the binomial and Poisson models. In Fig. 3, however, large areas with relatively high levels of uncertainty in map (a), compared to the same areas in maps (b) and (c), indicate that the binomial and Poisson models perform better than the Gaussian model with regard to prediction.

When fitting prevalence data of parasitic diseases, previous studies are usually based on the assumption that the response variable follows a certain distribution. However, the reason for this assumption was seldom provided. We believe that a comparison of different models is needed to find the optimal model. Importantly, such a model facilitates the detection of accurate associations between health outcomes and risk factors. In our study, the Poisson model detected more significant, environmental risk factors than the Gaussian model. Based on the results of the Gaussian model only, some important risk factors (e.g. NDVI and snail habitats) would be missed. Second, this model is

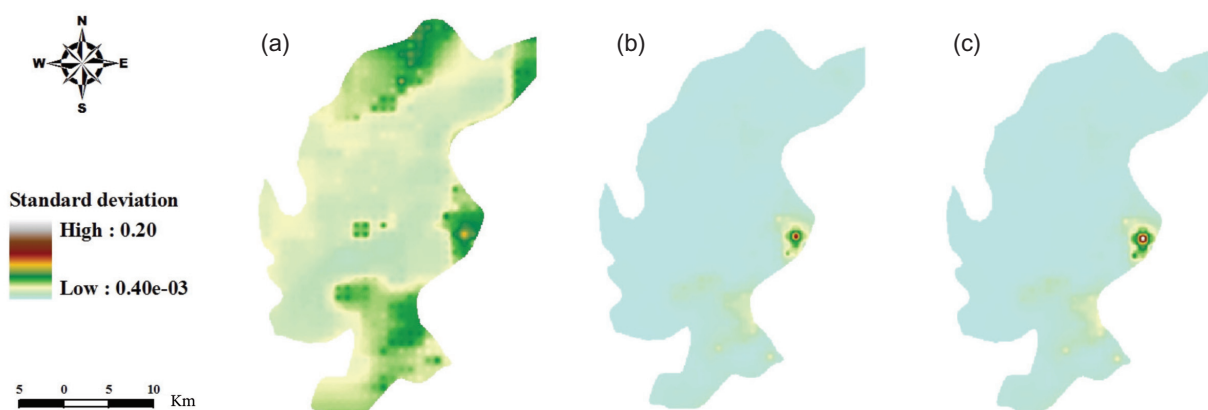


Fig. 3. Uncertainties of the model predictions of the prevalence of schistosomiasis japonica in Xingzi county using (a) a Gaussian model, (b) a binomial model and (c) a Poisson model.

useful for the assembly of an accurate predictive map of health outcomes. Such a risk map is vital for decision-makers to identify priority areas for schistosomiasis control. Hence, we believe map (c) in Fig. 3 to be the most reliable map. It seems that the differences between maps (b) and (c) are slight, but the MSE revealed that the Poisson model fitted the data better.

Two limitations of the study deserve to be mentioned. One is that the optimal model or the “best-fit” defined here is measured by the MSE of the infection rate, which may be somewhat subjective. Even if some other indicator could be used, we prefer the MSE of the infection rate because it is not only a simple, widely used, and excellent metric in the context of optimization (Zhou and Bovik, 2009), but it can also easily be implemented in practise. The other is that the prevalence of schistosomiasis risk calculated here may not be accurate since we could not rule out diagnostic error completely. In recent years, infection with *S. japonicum* is generally low in the lower reaches of the Yangtze River, which results in uncertainty both for sensitivity and specificity of infection, particularly with respect to the former (Bergquist et al., 2009; Zhou et al., 2011; Bergquist, 2013). A Bayesian method was used to deal with this uncertainty in a previous study (Wang et al., 2008), assuming that the diagnostic sensitivity and specificity for all villages are the same. This assumption, however, is to some extent uncertain and might introduce new errors due to subjective prior knowledge for model parameters, which heavily influences the corresponding posterior distributions (Gelman, 2008). Further studies are needed to resolve the importance of taking the diagnostic error into account when modelling schistosomiasis risk.

## Conclusion

In our analysis, non-spatial regression models rather than geostatistical models were used for the spatial prediction of schistosomiasis risk, as we had access to a sufficiently large number of environmental factors to explain the spatial variation of schistosomiasis risk. In this case, classical regression models are powerful enough with respect to prediction, while this is not the case if only few risk factors are available. In general, Bayesian geostatistical models are valuable and widely used in spatial prediction of prevalence of parasitic diseases, given the complexity of disease-transmission processes where human, parasite, intermediate host and environment interact, the paucity of field data, and inherent uncertainties in the model inputs exist as pointed out by Diggle et al. (2007).

## Acknowledgements

This work was financially supported by the National Natural Science Foundation of P.R. China (grant nos. are 81102167 81172609, and J1210041), Specialised Research Fund for the Doctoral Program of Higher Education, SRFDP (grant no. 20110071120040), a Foundation for the Author of National Excellent Doctoral Dissertation of P.R. China (FANEDD) (201186), the National S&T Major Program (2012ZX10004-220, 2008ZX10004-011), and Ecological Environment and Humanities/Social Sciences Interdisciplinary Research Project of Tyndall Center of Fudan University (FTC98503A09). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## References

- Bavia ME, Malone JB, 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.
- Bergquist R, 2013. Good things are worth waiting for. *Am J Trop Med Hyg* 88, 409-410.
- Bergquist R, Johansen MV, Utzinger J, 2009. Diagnostic dilemmas in helminthology: what tools to use and when? *Trends Parasitol* 25, 151-156.
- Box GEP, Cox DR, 1964. An analysis of transformations (with discussion). *J R Stat Soc, Ser B* 26, 211-252.
- Box GEP, Hunter WG, Hunter JS, 1978. Significance tests and confidence intervals for means, variances, proportions, and frequencies. In: *Statistics for experimenters: an introduction to design, data analysis, and model building*. New York: John Wiley & Sons, 107-151.
- Brooker S, Clements ACA, Bundy DAP, 2006. Global epidemiology, ecology and control of soil-transmitted helminth infections. *Adv Parasitol* 62, 221-261.
- Carabin H, Escalona M, Marshall C, Vivas-Martínez S, Botto C, Joseph L, Basáñez MG, 2003. Prediction of community prevalence of human onchocerciasis in the Amazonian onchocerciasis focus: Bayesian approach. *Bull World Health Organ* 81, 482-490.
- Clements ACA, Firth S, Dembelé R, Garba A, Touré S, Sacko M, Landouré A, Bosqué-Oliva E, Barnett AG, Brooker S, Fenwick A, 2009. Use of Bayesian geostatistical prediction to estimate local variations in *Schistosoma haematobium* infection in West Africa. *Bull World Health Organ* 87, 921-929.
- Clements ACA, Moyeed R, Brooker S, 2006. Bayesian geostatistical prediction of the intensity of infection with *Schistosoma mansoni* in East Africa. *Parasitology* 133, 711-719.
- Cressie N, 1993. *Statistics for spatial data*. New York: John Wiley & Sons, Inc.

- Diggle PJ, Jr. Ribeiro PJ, 2007. Model-based geostatistics. Berlin: Springer.
- Diggle PJ, Thomson MC, Christensen OF, Rowlingson B, Obsomer V, Gardon J, Wanji S, Takougang I, Enyong P, Kamgno J, Remme JH, Boussinesq M, Molyneux DH, 2007. Spatial modelling and the prediction of *Loa loa* risk: decision making under uncertainty. *Ann Trop Med Parasitol* 101, 499-509.
- Ekpo UF, Hürlimann E, Schur N, Oluwole AS, Abe EM, Mafe MA, Nebe OJ, Isiyaku S, Olamiju F, Kadiri M, Poopola TO, Braide EI, Saka Y, Mafiana CF, Kristensen TK, Utzinger J, Vounatsou P, 2013. Mapping and prediction of schistosomiasis in Nigeria using compiled survey data and Bayesian geospatial modelling. *Geospat Health* 7, 355-366.
- Feldmeier H, Poggensee G, 1993. Diagnostic techniques in schistosomiasis control. *Acta Trop* 52, 205-220.
- Fischer MM, Wang JF, 2011. Spatial data analysis: models, methods and techniques. Berlin: Springer.
- Gelman A, 2008. Objections to Bayesian statistics. *Bayesian Anal* 3, 445-450.
- Hay S, Graham A, Rogers DJ, 2006. Global mapping of infectious diseases: methods, examples and emerging applications. *Adv Parasitol* 62, ix-xi.
- Thomson MC, Connor SJ, D'Alessandro U, Rowlingson B, Diggle P, Cresswell M, Greenwood B, 1999. Predicting malaria infection in Gambian children from satellite data and bed net use surveys: the importance of spatial correlation in the interpretation of results. *Am J Trop Med Hyg* 61, 2-8.
- Wang XH, Zhou XN, Vounatsou P, Chen Z, Utzinger J, Yang K, Steinmann P, Wu XH, 2008. Bayesian spatio-temporal modeling of *Schistosoma japonicum* prevalence data in the absence of a diagnostic 'gold' standard. *PLoS Negl Trop Dis* 2, e250.
- Yu S, 1981. Studies on the methodology of egg collection with nylon bag for the diagnosis of *Schistosoma japonicum* infections [in Chinese]. *Chin J Med Lab Technol* 4, 210-213 (in Chinese).
- Zhang Z, Zhu R, Ward MP, Xu W, Zhang L, Guo J, Zhao F, Jiang Q, 2012. Long-term impact of the World Bank Loan Project for schistosomiasis control: a comparison of the spatial distribution of schistosomiasis risk in China. *PLoS Negl Trop Dis* 6, e1620.
- Zhang ZJ, Carpenter TE, Lynn HS, Chen Y, Bivand R, Clark AB, Hui FM, Peng WX, Zhou YB, Hao GM, Jiang QW, 2009. Location of active transmission sites of *Schistosoma japonicum* in lake and marshland regions in China. *Parasitology* 136, 737-746.
- Zhou W, Bovik AC, 2009. Mean squared error: love it or leave it? A new look at Signal Fidelity Measures. *Signal Processing Magazine, IEEE* 26, 98-117.
- Zhou XN, Wang LY, Chen MG, Wu XH, Jiang QW, Chen XY, Zheng J, Utzinger J, 2005. The public health significance and control of schistosomiasis in China-then and now. *Acta Trop* 96, 97-105.
- Zhou YB, Zheng HM, Jiang QW, 2011. A diagnostic challenge for schistosomiasis japonica in China: consequences on praziquantel-based morbidity control. *Parasit Vectors* 4, 194.