Bayesian geostatistical modelling for mapping schistosomiasis transmission

P. VOUNATSOU[1]*, G. RASO[2,3], M. TANNER[1], E. K. N'GORAN[4,5] *and* J. UTZINGER[1]

[1] *Department of Public Health and Epidemiology, Swiss Tropical Institute, P.O. Box, CH-4002 Basel, Switzerland*
[2] *Division of Epidemiology and Social Medicine, School of Population Health, The University of Queensland, Public Health Building, Herston Road, Brisbane, Queensland 4006, Australia*
[3] *Molecular Parasitology Laboratory, Queensland Institute of Medical Research, 300 Herston Road, Brisbane, Queensland 4006, Australia*
[4] *UFR Biosciences, Université de Cocody-Abidjan, 22 BP 582, Abidjan 22, Côte d'Ivoire*
[5] *Centre Suisse de Recherches Scientifiques, 01 BP 1303, Abidjan 01, Côte d'Ivoire*

(*Received 24 October 2008; revised 24 February 2009; accepted 27 February 2009; first published online 2 June 2009*)

SUMMARY

Progress has been made in mapping and predicting the risk of schistosomiasis using Bayesian geostatistical inference. Applications primarily focused on risk profiling of prevalence rather than infection intensity, although the latter is particularly important for morbidity control. In this review, the underlying assumptions used in a study mapping *Schistosoma mansoni* infection intensity in East Africa are examined. We argue that the assumption of stationarity needs to be relaxed, and that the negative binomial assumption might result in misleading inference because of a high number of excess zeros (individuals without an infection). We developed a Bayesian geostatistical zero-inflated (ZI) regression model that assumes a non-stationary spatial process. Our model is validated with a high-quality georeferenced database from western Côte d'Ivoire, consisting of demographic, environmental, parasitological and socio-economic data. Nearly 40% of the 3818 participating schoolchildren were infected with *S. mansoni*, and the mean egg count among infected children was 162 eggs per gram of stool (EPG), ranging between 24 and 6768 EPG. Compared to a negative binomial and ZI Poisson and negative binomial models, the Bayesian non-stationary ZI negative binomial model showed a better fit to the data. We conclude that geostatistical ZI models produce more accurate maps of helminth infection intensity than the spatial negative binomial ones.

Key words: Schistosomiasis, *Schistosoma mansoni*, Bayesian geostatistics, non-stationarity, overdispersion, zero-inflated model, infection intensity, Côte d'Ivoire.

INTRODUCTION

Empirical maps of schistosomiasis transmission are important tools in guiding control interventions. Usually, these maps are based on hospital records or data arising from cross-sectional epidemiological surveys carried out over a number of locations within a designated study area (Doumenge *et al*. 1987; Brooker *et al*. 2000; Brooker, 2007). The data are spatially correlated because common exposures influence transmission similarly at neighbouring locations. Among other factors, these common exposures include climatic and environmental features governing the survival and longevity of the intermediate host snails (Stensgaard *et al*. 2006) and proximity of human habitations to transmission sites (Booth *et al*. 2004; Kitron *et al*. 2006). Risk maps of schistosomiasis are produced by predicting the transmission outcome at non-sampled locations. These predictions are more accurate when they are based on models relating transmission to known environmental predictors of schistosomiasis, and when they make use of the spatial correlation present in the data, which filters the noise and highlights the existing patterns. The standard regression models assume independence of the data, leading to inaccurate estimation of the precision of the parameter estimates and of the predictions when they are applied to spatially-correlated data (Cressie, 1991).

In this paper, we first summarize how our ability to map and predict the distribution of schistosomiasis transmission has improved as a result of advances made with Bayesian geostatistical approaches. However, research has mainly focused on mapping and prediction of prevalence data. In view of morbidity control being the declared goal of national schistosomiasis control programmes, new research is needed for modeling infection intensity data (WHO, 2002; Bergquist, Johansen and Utzinger, 2009). Thus far, only one attempt has been made to predict infection

* Corresponding author: Penelope Vounatsou, Department of Public Health and Epidemiology, Swiss Tropical Institute, P.O. Box, CH-4002 Basel, Switzerland. Tel: +41 61 284-8109. Fax: +41 61 284-8105. E-mail: penelope.vounatsou@unibas.ch

Parasitology (2009), **136**, 1695–1705. © Cambridge University Press 2009
doi:10.1017/S003118200900599X Printed in the United Kingdom

https://doi.org/10.1017/S003118200900599X Published online by Cambridge University Press

intensity of *Schistosoma mansoni* using Bayesian inference (Clements, Moyeed and Brooker, 2006*b*). We examine the underlying assumptions of that model, i.e. stationarity and negative binomial distribution of egg counts. We then propose an approach that might predict the infection intensity of *S. mansoni* more accurately; namely a Bayesian non-stationary zero-inflated (ZI) negative binomial model. We compare the performance of this model with a negative binomial and a ZI Poisson model on the basis of credible intervals of the predictive ability of the models, using an existing high-quality georeferenced database from western Côte d'Ivoire. Finally, we discuss future research directions to further improve upon the mapping of infection intensity of schistosomiasis and other tropical diseases that are often neglected.

BAYESIAN APPROACHES FOR MAPPING AND PREDICTING SCHISTOSOMIASIS

Spatial geostatistical models introduce at each observed location an additional parameter, the so-called random effect, and build spatial correlation on the distribution of the random effects. The latter is done by assuming that the distribution arises from a multivariate normal distribution with correlation matrix often to be a parametric decreasing function of distance between any pair of locations. Fitting these models, however, is not straightforward, which is illuminated on the following grounds. First, the models include a large number of parameters, proportional to the number of observed locations. Second, computational challenges arise in relation to the large correlation matrices. Bayesian formulation of these models (Diggle, Moyeed and Tawn, 1998), facilitates parameter estimation via Markov chain Monte Carlo (MCMC) simulation methods. The availability of software for fitting these models, although for a relatively small number of locations (Lunn *et al*. 2000), together with the earlier established geographical information system (GIS) software and remote sensing (RS) tools enabled spatial analyses of cross-sectional prevalence data and the generation of model-based schistosomiasis risk maps (Raso *et al*. 2005, 2006*a*; Yang *et al*. 2005*a*; Clements *et al*. 2006*a*, 2008; Beck-Wörner *et al*. 2007; Brooker and Clements, 2009). These risk maps, emphasizing areas where the prevalence of schistosome infections is higher relative to other locations, are useful tools for spatial targeting of control interventions and to enhance cost-effectiveness (Brooker *et al*. 2008, 2009).

Interestingly, only a single effort has been made to date for mapping and predicting schistosomiasis transmission intensity levels (Clements *et al*. 2006*b*). In that study, the authors made the following assumptions. Firstly, the spatial correlation of *S. mansoni* infection intensity is stationary. Secondly, the distribution of excreted eggs, which is a proxy for quantifying transmission intensity, shows a negative binomial distribution. The assumption of negative binomial distribution was tested against ZI Poisson and ZI negative binomial models in a non-spatial context. Both assumptions warrant scrutiny, before presenting a promising approach for mapping *S. mansoni* infection intensity, i.e. a Bayesian non-stationary ZI geostatistical model.

The issue of stationarity

Stationarity is a common assumption in geostatistical modelling. It asserts that spatial correlation is only a function of distance between locations and independent of the location itself. The covariance is the same between any two points that are at the same distance apart no matter which two points are chosen. However, local effects such as man-made ecological transformations (e.g. water resources development and management projects; for recent reviews see Steinmann *et al*. 2006; Li *et al*. 2007), climate change (Suthers, 2004; Yang *et al*. 2005*b*; Zhou *et al*. 2008) or disease control interventions (Yang *et al*. 2005*a*; Brooker, 2007) may alter correlation, and hence resulting in non-stationarity.

Non-stationary models in disease mapping have been developed and find increasing application in malaria (Gemperli *et al*. 2004; Gosoniu *et al*. 2006; Silué *et al*. 2008). Recently, these models have been extended from malaria to helminth infections, including schistosomiasis (Raso *et al*. 2005, 2006*b*; Beck-Wörner *et al*. 2007).

The issue of overdispersion

The transmission intensity of schistosomiasis is a function of the parasitic worm load within a group of individuals, which can indirectly be quantified by the number of eggs that are excreted. Host heterogeneities in exposure and susceptibility to infection lead to an aggregated distribution of worm burden across individuals (Bradley, 1972; Polderman, 1979; Anderson and May, 1985). Hence, a few individuals harbour large numbers of worms, whilst the majority of individuals are uninfected or only carry a low worm burden. In addition, widely used diagnostic approaches for schistosomiasis (e.g. the Kato-Katz technique for *S. mansoni* diagnosis) fail to detect some infected individuals, particularly when only a single stool sample is examined and infection intensities are light (de Vlas and Gryseels, 1992; Engels, Sinzinkayo and Gryseels, 1996; Utzinger *et al*. 2001; Booth *et al*. 2003). Due to these two issues, often a large proportion of individuals are considered as 'zero egg excretors'.

The standard Poisson distribution, which assumes equal variance and mean and is commonly employed to model count data, is not appropriate to fit the

https://doi.org/10.1017/S003118200900599X Published online by Cambridge University Press

observed egg counts since the variance of the counts is much larger than their mean. Three decades ago, the use of the negative binomial distribution was proposed to model the excessive variation, known as overdispersion, in helminth egg counts (Cohen, 1977). In the meantime, negative binomial regression models have been widely used to analyse helminth infection intensity data (Utzinger *et al*. 2002; Scott *et al*. 2003; Yapi *et al*. 2005; Brooker *et al*. 2006).

ZI MODELS

Rationale and previous applications

Negative binomial models are not the only approach to tackle overdispersed count data. Recently, there has been considerable interest in modelling count data with an excessive number of zeros, and the use of ZI models is particularly noteworthy. These models assume that a proportion of individuals have no chance to be infected, as they are not exposed. In other words, there is a process which determines whether an individual is likely to be infected at all and a second process determining the number of excreted eggs among those who are at risk of infection. ZI Poisson (ZIP) models assume that the number of excreted eggs follows a Poisson distribution. ZI negative binomial (ZINB) models assume that the number of worms among those who are at risk of infection has a negative binomial distribution. ZI models have been initially developed by Mullahy (1986) and further extended by Lambert (1992). A Bayesian analysis of ZIP models is given in Rodrigues (2003) and of ZINB models in Denwood *et al*. (2008). To our knowledge, Agarwal, Gelfand and Citron-Pousty (2002) were the first to employ ZIP models for stationary count data in a Bayesian framework.

Motivating example

The data which motivated the current Bayesian geostatistical application stem from a study for mapping and predicting the spatial distribution of *S. mansoni* and hookworm monoinfection and single infection (Raso *et al*. 2005, 2006*b*, 2007), *S. mansoni*-hookworm co-infection (Raso *et al*. 2006*a*) and *Plasmodium falciparum* infection in the Man region, western Côte d'Ivoire (Silué *et al*. 2008). Details of the study area, population surveyed, geostatistical analyses used and implications for schistosomiasis, soil-transmitted helminthiasis and malaria control have been described in previous publications. In brief, the field work was carried out between May and August 2002 in the Man region, which is the major focus of intestinal schistosomiasis in Côte d'Ivoire (Doumenge *et al*. 1987; Utzinger *et al*. 2000; Raso *et al*. 2005). All children attending grades 3–5 from 55 rural schools were enrolled. Demographic data (name, age and sex) were obtained from existing registries for the respective school year. Children's socio-economic status was indirectly assessed by means of a questionnaire that collected information on household assets ownership (e.g. possession of a radio) and household characteristics (e.g. walls constructed with bricks). Parasitological data were obtained following the microscopic examination of a single Kato-Katz thick smear per child, using a standardised, quality-controlled method (Katz, Chaves and Pellegrino, 1972; Raso *et al*. 2005). The number of *S. mansoni* eggs was counted and the number of eggs per gram of stool (EPG) recorded. *S. mansoni*-infected children were treated with a single 40 mg/kg oral dose of praziquantel (WHO, 2002). The *S. mansoni* risk map was discussed with local and national health authorities and governed the spatial targeting of prevention and control interventions, facilitated by the establishment of village health committees and improved access to anthelminthic drugs.

Geographical coordinates of each school were collected using a hand-held global positioning system (GPS; Thales Navigation, Santa Clara, CA, USA). Streets, village boundaries, rivers, elevation lines and soil types were digitized from existing ground maps. A GIS database was built linking the parasitological data with RS environmental and climatic factors associated with transmission. In particular, normalized difference vegetation index (NDVI) and land surface temperature (LST) were extracted at 1×1 km spatial resolution from Moderate Resolution Imaging Spectroradiometer (MODIS) from USGS EROS Data Centre. Rainfall estimate (RFE) data with an 8×8 km spatial resolution from Meteosat 7 satellite were obtained from the Africa Data Dissemination Service (ADDS). A digital elevation model (DEM) was employed originating from the Shuttle Radar Topography Mission (SRTM) to delineate watersheds and rivers (Beck-Wörner *et al*. 2007). Rivers were ordered after a system proposed by Strahler some 50 years ago (Beck-Wörner *et al*. 2007).

Model specification

Negative binomial model. Let Y_{ij} be the *S. mansoni* egg count of child j in village i. We assumed that Y_{ij} arises from a negative binomial distribution, $Y_{ij} \sim Nb(\mu_{ij}, r)$ with mean μ_{ij}, dispersion parameter r and probability density function

$$f(Y_{ij}=y_{ij}|r,\mu_{ij}) = \frac{(y_{ij}+r-1)!}{y_{ij}!(r-1)!}\left(\frac{r}{r+\mu_{ij}}\right)^{r} \times \left(\frac{\mu_{ij}}{r+\mu_{ij}}\right)^{y_{ij}}, r>0 \quad (1).$$

The negative binomial model assumes that the variance of the counts, $\mathrm{var}(Y_{ij})$ is equal to

$$\mathrm{var}(Y_{ij}) = \mu_{ij} + k * \mu_{ij}^{2} \quad (2)$$

https://doi.org/10.1017/S003118200900599X Published online by Cambridge University Press

with $k=1/r$, known as aggregation parameter. The Poisson distribution arises as $r\to\infty$ (or equivalently $k\to 0$) and thus $\text{var}(Y_{ij})=\mu_{ij}$.

ZI models. The ZI count models have mixed specifications that add extra weight to the probability of observing a zero (Lambert, 1992). In particular they are mixture models having two components and mixing probability, π related to the proportion of non-infected individuals who have no chance to be infected. $(1-\pi)$ corresponds to the probability of observing a positive egg count arising from an adopted count distribution $f(y_{ij})$ such as Poisson or negative binomial. In the general form, the model can be written as:

$$P(Y_{ij}=y_{ij})=\pi I_{\{0\}}(y_{ij})+(1-\pi)f(y_{ij}) \quad (3)$$

where $I_{\{0\}}(y_{ij})$ is the one-point distribution, putting all its mass at zero, that is $I_{\{0\}}(y_{ij})=1$ if $y_{ij}=0$ and zero otherwise. Equivalently the model can be specified by:

$$P(Y_{ij}=y_{ij})=\begin{cases}\pi+(1-\pi)f(y_{ij}), & y_{ij}=0\\ (1-\pi)f(y_{ij}), & y_{ij}>0\end{cases} \quad (4).$$

The ZIP model has the Poisson density $f(y_{ij})=(1-\pi)\exp(-\mu_{ij})\mu_{ij}^{y_{ij}}/y_{ij}!$, $y_{ij}>0$ and the ZINB model has the negative binomial density function which is given in equation (1). The mean of the ZI model is equal to $\pi I_{\{0\}}(y_{ij})+(1-\pi)\mu_{ij}(1-I_{\{0\}}(y_{ij}))$ and the variance is $Var(Y_{ij})=(1-\pi)^2 Var_f(Y_{ij})$, where $Var_f(Y_{ij})=\mu_{ij}$ for the ZIP model and it is given in equation (2) for the ZINB model.

Bayesian non-stationary overdispersed count model. In the above negative binomial and ZI models we introduce covariates $\underline{X}_{ij}$ and village-specific spatial random effect ϕ_i on the $\log(\mu_{ij})$, that is $\log(\mu_{ij})=\underline{X}_{ij}^T\underline{\beta}+\phi_i$, where $\underline{\beta}$ is the vector of regression coefficients. We assume that the random effects model a continuous spatial process that is $\underline{\phi}=(\phi_1, \phi_2, \ldots, \phi_N)^T\sim MVN(\underline{0},\Sigma)$, has a multivariate normal distribution with variance-covariance matrix $\Sigma_{il}=\sigma^2\exp(-\rho d_{il})$, where d_{il} is the shortest straight-line distance between villages i and l, σ^2 is the geographic variability known as the sill, and ρ is a smoothing parameter that controls the rate of correlation decay with increasing distance. To take into account non-stationarity, we partitioned the study area in K ecologic sub-regions, i.e. watersheds of the local hydrology (Beck-Wörner *et al*. 2007), and assumed a local stationary spatial process $\underline{\omega}_k$ in each sub-region $k=1, 2, \ldots, K$. We then viewed spatial correlation in our area as a mixture of the different spatial processes and modelled the spatial random effect ϕ_i at location i as a weighted average of the sub-region-specific (independent) stationary processes as follows: $\phi_i=\sum_{k=1}^{K} a_{ik}\omega_{ki}$, with weights a_{ik}, which are decreasing functions of the distance between location i and the centroids of the sub-regions k (Banerjee *et al*. 2004). Assuming $\underline{\omega}_k\sim MVN(0,\Sigma_k)$, $(\Sigma_k)_{il}=\sigma_k^2 exp(-\rho_k d_{il})$, we have $\underline{\phi}=N(\underline{0}, \sum_{k=1}^{K} A_k^T\Sigma_k A_k)$, where $A_k=diag\{a_{1k}, a_{2k}, \ldots, a_{nk}\}$. The range is defined as the minimum distance at which spatial correlation between locations is below 5% and it can be calculated as $\xi_k={}^3/_{\rho_k}$.

Model fit and implementation. Model fit was carried out in WinBUGS version 1.4 (Imperial College & Medical Research Council, London, UK) and in specialized Bayesian geostatistical codes written in Fortram 95 by the authors. Following a Bayesian model specification, we adopted prior distributions for the model parameters. We choose vague Normal distributions for the $\underline{\beta}$ parameters with large variances (i.e. 10 000), gamma prior for r with large variance, inverse gamma priors for σ_k^2 and uniform priors for ρ_k, $k=1, 2, \ldots, K$. MCMC simulation was employed to estimate the model parameters (Gelfand and Smith, 1990). We ran a single chain sampler with a burn-in of 5000 iterations. Convergence was assessed by inspection of ergodic averages of selected model parameters. Covariates from the multivariate model were used to generate a smooth risk map for *S. mansoni* infection intensity using Bayesian kriging (Diggle *et al*. 1998).

Model validation. For the model validation a training sample from the current database was used. From the 55 schools, 43 schools were randomly selected and fitted into the models. The remaining 12 schools were used for validation purposes. Validation was done at the individual level to take into account age, sex and socio-economic status as these factors significantly influenced *S. mansoni* infection intensity ($n=731$ children). The predictive ability of the models was assessed by calculating for each model credible intervals (the equivalent of confidence interval in Bayesian statistics) with probability coverage varying between 1% and 100% of the posterior predictive distribution of test data. The model with the best predictive ability was the one with the highest percentage of locations within the interval of smallest coverage.

Results

Complete demographic, socio-economic and parasitological data were available for 3818 schoolchildren. The *S. mansoni* infection prevalence was 38·9% with a mean egg count among infected children of 162 EPG (range: 24 to 6768 EPG). Fig. 1 shows the average *S. mansoni* infection intensity at the unit of the school; the mean egg excretion ranged from 0 to 875 EPG.

The non-stationary ZINB model showed a better fit to the data than the ZIP model and the negative binomial model. Table 1 summarizes the results

https://doi.org/10.1017/S003118200900599X Published online by Cambridge University Press

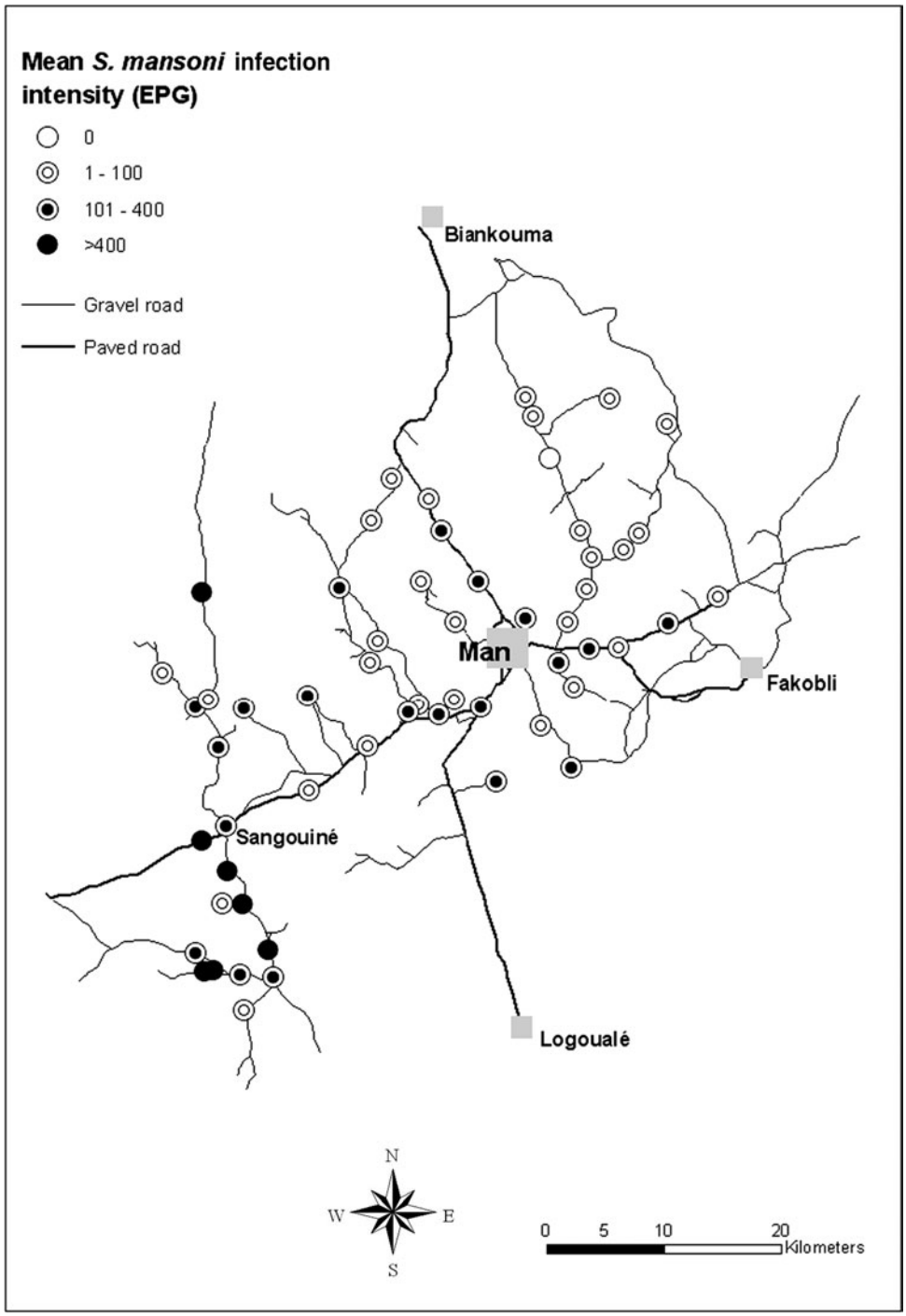


Fig. 1. Observed mean *S. mansoni* infection intensity at the unit of the school in the region of Man, western Côte d'Ivoire.

from the model validation, according to the percentage of test data with *S. mansoni* infection intensity falling within credible intervals of probability coverage ranging between 10% and 40%. The ZINB model included 100% of the test data in the narrowest 15% credible interval.

Table 2 displays the results from the best fitting non-stationary ZINB regression model. The covariates that explained significantly the geographical variation of *S. mansoni* infection intensity were age group and watershed, since the mean coefficient estimates were within the credible intervals. The ZI mixing proportion parameter (π) was 0·723 (95% credible interval = 0·667–0·780) suggesting that the proportion of zero counts which was higher than that assumed by the negative binomial distribution was 72·3%. A possible interpretation could be that over half of children who tested negative had no chance to be infected with *S. mansoni* because they were not exposed. The aggregation parameter k was estimated equal to 1·386 (95% credible interval = 1·282–1·500), indicating that there is still over-dispersion even

https://doi.org/10.1017/S003118200900599X Published online by Cambridge University Press

Table 1. Absolute and cumulative frequency of *S. mansoni* infection intensity among 731 schoolchildren within the 12 test locations falling within credible intervals of probability coverage ranging between 10% and 40% of the posterior predictive distribution

Probability coverage of credible interval	Zero-inflated negative binomial (ZINB) model	Zero-inflated Poisson (ZIP) model	Negative binomial model
10%	659 (90·2%)		507 (69·4%)
11%	690 (94·4%)		522 (71·4%)
12%	704 (96·3%)		532 (72·8%)
13%	720 (98·5%)		553 (75·6%)
14%	729 (99·7%)	0 (0%)	570 (78·0%)
15%	731 (100%)	655 (89·6%)	591 (80·8%)
16%		722 (98·8%)	605 (82·8%)
17%		731 (100%)	617 (84·4%)
18%			636 (87·0%)
19%			649 (88·8%)
20%			657 (89·9%)
30%			722 (98·8%)
40%			731 (100%)

though some of the excessive zeros have been modelled separately. The range parameters ξ_k, $k = 1, 2, \ldots, K$ were similar for the three fixed tiles, indicating that spatial correlation was not significant at distances of approximately 2·6 km and above. In contrast, the geographical variability varied between the tiles from 0·364 to 4·496 indicating non-stationarity.

Fig. 2 displays the *S. mansoni* intensity risk obtained from the non-stationary ZINB regression model. The model predicted low intensities in the north-eastern part of the study area, and high infection intensities in the southern and north-western region. The highest intensities were predicted along the rivers.

The lowest prediction error of the infection intensity was estimated in the north-eastern part of the study area, where accordingly the infection intensity was predicted to be low (Fig. 3).

IMPLICATIONS OF OUR FINDINGS AND FUTURE RESEARCH DIRECTIONS

Areas of intensive schistosomiasis transmission are usually associated with a high morbidity, and hence burden of the disease. Although schistosomiasis prevalence maps are useful tools in guiding control interventions (Brooker *et al*. 2000; Raso *et al*. 2005; Clements *et al*. 2006*a*; Brooker, 2007), maps of transmission intensity provide additional information on the severity of the infection. Such maps allow identification of hot spots with potentially high morbidity and the largest attributable disease fraction, and hence provide insight about where to target 'preventive chemotherapy' for morbidity control. Geostatistical models which are based on realistic assumptions about the distributions which generated the intensity data, including the underlying spatial processes, will provide more accurate estimation of the environment-transmission relations, and hence improve the current prediction accuracies (Brooker, 2007).

A general feature of helminth egg output data is that approximately one-fifth of the population are responsible for an estimated 80% of the environmental contamination (Anderson and May, 1985). It follows that the egg count data which quantify transmission intensity usually include a large number of zeros or very light infections, which lead to overdispersion. Hence, the variance of such a typical helminth egg distribution is considerably larger than the one assumed by the Poisson distribution. Conventionally the negative binomial distribution has been adopted to take into account overdispersion by an alternative variance-mean specification (Cameron and Trivedi, 1986). However, this approach ignores that a proportion of individuals have no exposure to the disease and therefore neglects the process which generated the additional zeros. ZI models explicitly incorporate this process in the modelling framework. The models we have employed in the current application assume that exposed individuals have a probability to be negative according to the Poisson or negative binomial model which is applied to fit the counts. This may be true due to diagnostic error or possibly genetic factors. We could assess this assumption by applying an alternative class of models, the so-called hurdle, two-part or conditional models (Mullahy, 1986; Heilbron, 1994), which assume a zero-truncated standard distribution (Poisson or negative binomial) in the exposed group. Such a model would require further modification to take into account the possibility of false-negatives, which otherwise will be wrongly allocated to the unexposed group.

Another extension of the ZI models discussed here arises when allowing covariates not only on the average count of the exposed group, but also on the mixing proportion, that is the proportion of the non-exposed individuals. An issue related with modelling covariates is that of linearity. Environmental covariates influencing the transmission intensity is less likely to have a linear effect on intensity. The standard approach of categorising the covariates relies on arbitrary cut-off values of those covariates. An alternative modelling approach that could be assessed in mapping infection intensity and other transmission-related outcomes could include the use of spline functions (Eubank, 1988; Dimatteo, Genovese and Kass, 2001). Our own experience with spline regression models in malaria mapping (Gosoniu *et al*. 2009) suggests that this is a promising approach, when taking into account non-linearity in disease mapping.

https://doi.org/10.1017/S003118200900599X Published online by Cambridge University Press

Table 2. Posterior summaries (medians and 95% credible intervals) of the parameters of the non-stationary ZINB multiple regression. The estimate of the regression coefficients parameters represent density ratio (DR) of excreted egg counts. The range parameters ξ_k, $k=1,2,3$ are given in km

	Non-stationary ZINB multiple regression	
Indicator	Estimate	95% credible interval
Age group (years)		
6–10	1	
11–16	1·172	1·025, 1·337
Sex		
Male	1	
Female	0·902	0·773, 1·029
Socio-economic status		
Most poor	1	
Very poor	1·052	0·852, 1·284
Poor	0·978	0·787, 1·197
Less poor	1·162	0·925, 1·437
Least poor	1·149	0·911, 1·425
Household within village boundary	0·895	0·711, 1·105
Elevation	0·770	0·486, 1·187
Stream order		
1	1	
2	2·202	0·878, 4·781
3	3·184	0·700, 10·315
Watershed[a]		
1	1	
2	6·512	1·464, 17·28
3	7·072	1·729, 24·49
4	7·926	0·878, 33·17
Normalized difference vegetation index	0·715	0·457, 1·074
Land cover		
Woody savannah	1	
Tropical forest	0·522	0·192, 1·174
Deforested savannah and crops	1·412	0·393, 3·809
Tropical rainforest	1·336	0·339, 3·679
Distance to permanent water bodies	1·095	0·685, 1·603
Distance to temporary water bodies (m)		
$\leqslant 200$	1	
201–500	1·184	0·527, 2·414
>500	2·119	0·545, 6·284
Mean land surface temperature (°C)		
$<25{\cdot}0$	1	
25·0–26·4	0·593	0·217, 1·366
$\geqslant 26{\cdot}5$	0·787	0·107, 2·645
Distance to dispensaries (km)		
$\leqslant 1$	1	
1·1–5	1·154	0·338, 2·645
>5	2·024	0·682, 4·885
Aggregation parameter (k)	1·386	1·282, 1·500
Zero-inflated proportion (π)	0·723	0·667, 0·780
Spatial correlation parameters		
ξ_1	2·250	1·364, 26·118
ξ_2	2·606	1·371, 21·076
ξ_3	2·404	1·364, 19·623
σ_1^2	0·364	0·007, 2·153
σ_2^2	4·496	0·815, 9·055
σ_3^2	0·822	0·014, 4·827

[a] Arbitrary measure (for further details; see Beck-Wörner *et al.* 2007).

https://doi.org/10.1017/S003118200900599X Published online by Cambridge University Press

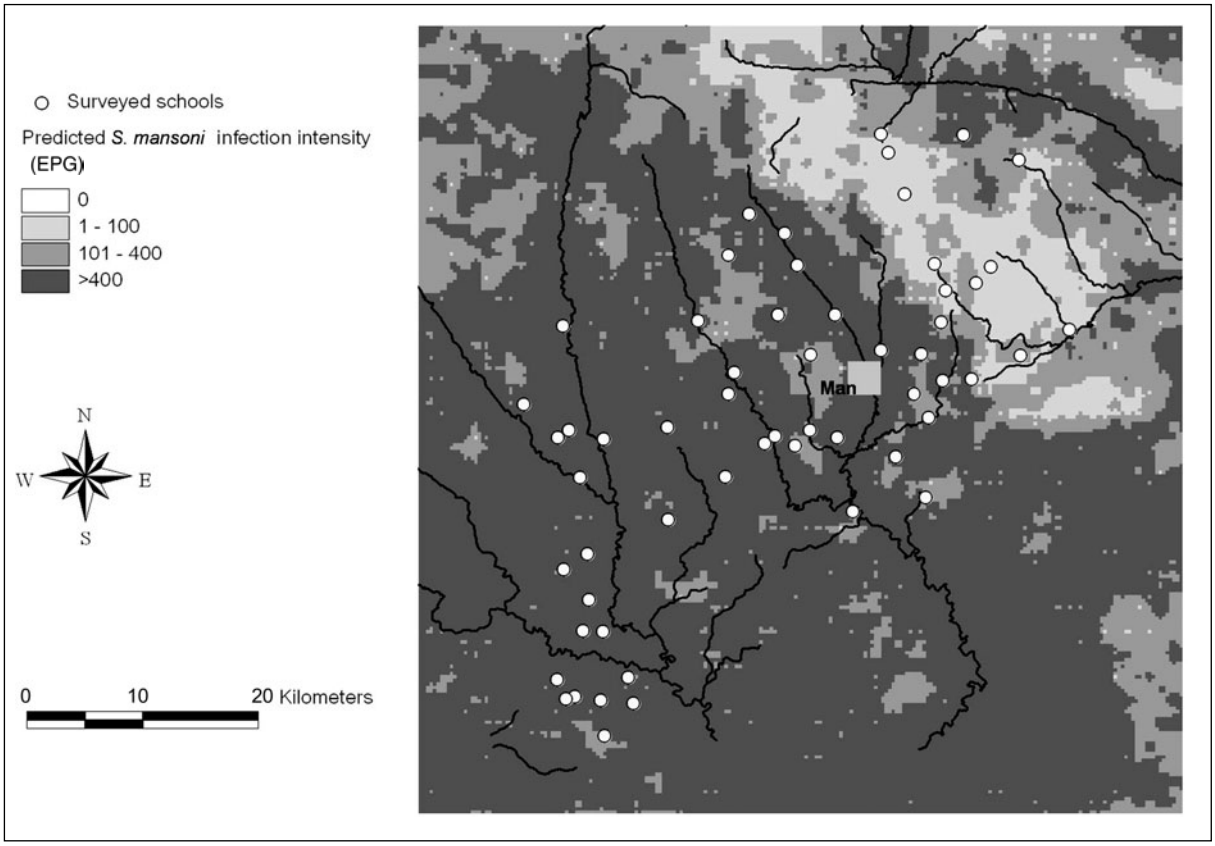


Fig. 2. Smoothed map of *S. mansoni* infection intensity based on the Bayesian geostatistical ZINB model.

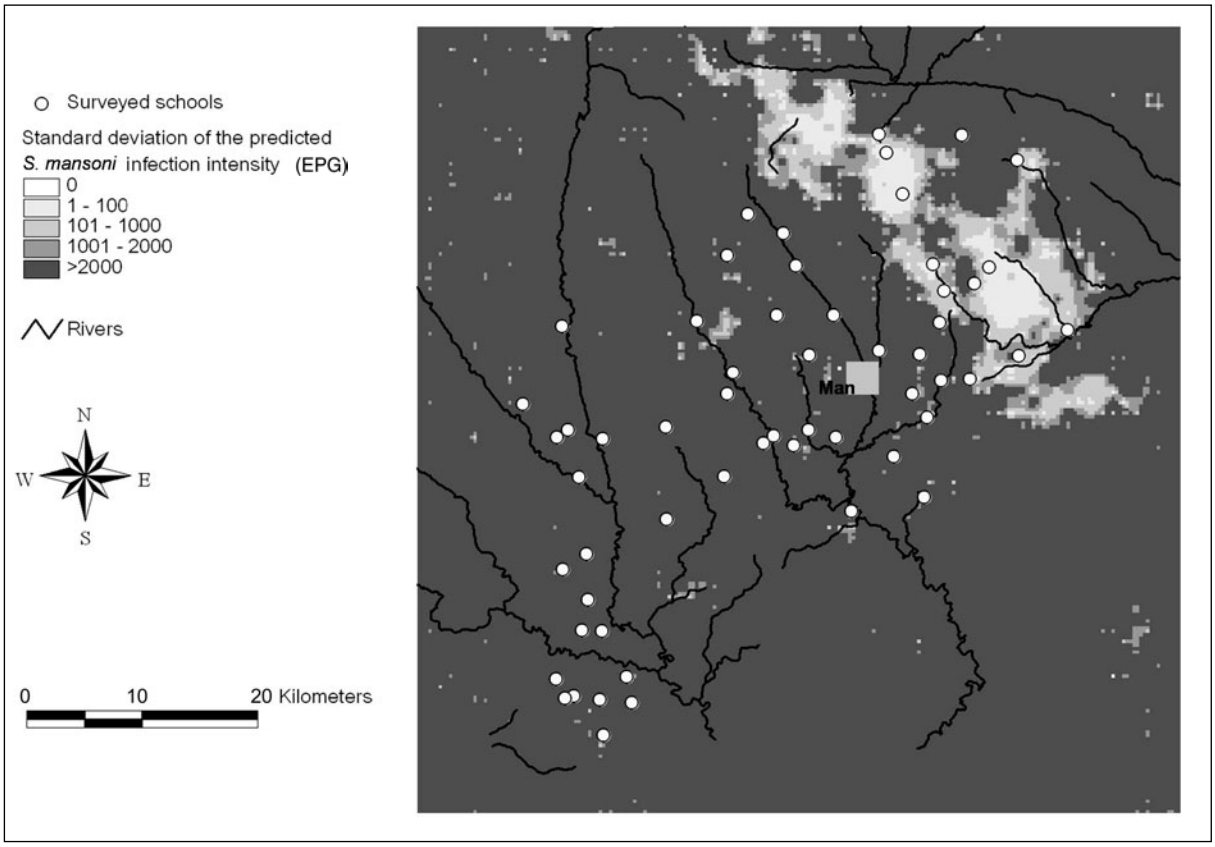


Fig. 3. Prediction error of *S. mansoni* infection intensity based on the Bayesian geostatistical ZINB model.

https://doi.org/10.1017/S003118200900599X Published online by Cambridge University Press

Non-stationarity is an important aspect of geographically-explicit databases, which should be considered, especially when fitting spatial models over large areas. The non-stationary modelling approach we have adopted here relies on a partition of the study region into meaningful ecological sub-regions, such as those governed by water catchment areas (Beck-Wörner *et al*. 2007). This approach is more appropriate when modelling schistosomiasis data over large areas covering different ecological zones which define the fixed partitions. An extension of the model will allow different covariate effects in each zone since it is likely to have an interaction effect between the zone and environmental effects on transmission. The model, in addition, smoothes the estimates at the border of the regions, and hence avoids discontinuities, which would otherwise arise (Sogoba *et al*. 2007).

Another form of non-stationarity is anisotropy which implies that spatial correlation depends not only on distance between any pair of locations but also on their relative orientation. To our knowledge, anisotropy has not been considered in any of the schistosomiasis transmission mapping exercises carried out thus far. We speculate that the spatial correlation is stronger on directions towards transmission sites rather than in the opposite direction, governed by hydrological factors upon which the intermediate host snails depend (Kitron *et al*. 2006; Stensgaard *et al*. 2006; Clennon *et al*. 2007).

Modelling multiple species parasitic infections is receiving increasing interest recently (Raso *et al*. 2006*a*; Brooker and Clements, 2009) as it has been recognised that in tropical and sub-tropical environments multiparasitism is the norm rather than the exception (Raso *et al*. 2004; Utzinger and de Savigny, 2006; Brooker and Utzinger, 2007; Steinmann *et al*. 2008). Individual exposures or genetic factors are likely to introduce positive or negative correlation in the infection intensities of the different parasites. It is conceivable that control measures can be employed in a more cost-effective manner, once areas of high infection intensities of multiple parasites have been identified and the underlying risk factor(s) determined, so that an integrated control approach can be envisaged. These analyses will require extending the ZIP and the ZINB models to their multivariate analogues, as well as considering multivariate spatial processes (Majumdar and Gelfand, 2007) to model spatial correlation of co-infection intensity data.

ACKNOWLEDGEMENTS

We thank Dr. J. Russell Stothard for inviting us to prepare this article for a special issue of *Parasitology*. With regard to the motivating example, we thank the education officers, the school directors and teachers, the participating schoolchildren, and the field and laboratory team for their commitment. PV (project no. 3252B0-102136), GR (project no. PBBSB-109011) and JU (project no. PPOOB-102883, PPOOB-119129) are financially supported by the Swiss National Science Foundation. GR also received funds from Novartis Foundation, the Roche Research Foundation through a fellowship and the University of Queensland for a postdoctoral research fellowship.

REFERENCES

Agarwal, D. K., Gelfand, A. E. and Citron-Pousty, S. (2002). Zero-inflated models with application to spatial count data. *Environmental and Ecological Statistics* **9**, 341–355.

Anderson, R. M. and May, R. M. (1985). Helminth infections of humans: mathematical models, population dynamics and control. *Advances in Parasitology* **24**, 1–101.

Banerjee, S., Gelfand, A. E., Knight, J. R. and Sirmans, C. F. (2004). Spatial modeling of house prices using normalized distance-weighted sums of stationary processes. *Journal of Business and Economic Statistics* **22**, 206–213.

Beck-Wörner, C., Raso, G., Vounatsou, P., N'Goran, E. K., Rigo, G., Parlow, E. and Utzinger, J. (2007). Bayesian spatial risk prediction of *Schistosoma mansoni* infection in western Côte d'Ivoire using a remotely-sensed digital elevation model. *American Journal of Tropical Medicine and Hygiene* **76**, 956–963.

Bergquist, R., Johansen, M. V. and Utzinger, J. (2009). Diagnostic dilemmas in helminthology: what tools to use and when? *Trends in Parasitology* **25**, 151–156.

Booth, M., Vennervald, B. J., Kenty, L., Butterworth, A. E., Kariuki, H. C., Kadzo, H., Ireri, E., Amaganga, C., Kimani, G., Mwatha, J. K., Otedo, A., Ouma, J. H., Muchiri, E. and Dunne, D. W. (2004). Micro-geographical variation in exposure to *Schistosoma mansoni* and malaria, and exacerbation of splenomegaly in Kenyan school-aged children. *BMC Infectious Diseases* **4**, 13.

Booth, M., Vounatsou, P., N'Goran, E. K., Tanner, M. and Utzinger, J. (2003). The influence of sampling effort and the performance of the Kato-Katz technique in diagnosing *Schistosoma mansoni* and hookworm co-infections in rural Côte d'Ivoire. *Parasitology* **127**, 525–531.

Bradley, D. J. (1972). Regulation of parasite populations. A general theory of the epidemiology and control of parasitic infections. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **66**, 697–708.

Brooker, S. (2007). Spatial epidemiology of human schistosomiasis in Africa: risk models, transmission dynamics and control. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **101**, 1–8.

Brooker, S., Alexander, N., Geiger, S., Moyeed, R. A., Stander, J., Fleming, F., Hotez, P. J., Correa-Oliveira, R. and Bethony, J. (2006). Contrasting patterns in the small-scale heterogeneity of human helminth infections in urban and rural environments in Brazil. *International Journal for Parasitology* **36**, 1143–1151.

Brooker, S. and Clements, A. C. A. (2009). Spatial heterogeneity of parasite co-infection: determinants and geostatistical prediction at regional scales. *International Journal for Parasitology* **39**, 591–597.

https://doi.org/10.1017/S003118200900599X Published online by Cambridge University Press

Brooker, S., Gyapong, J. O., Kabatereine, N. B., Stothard, J. R. and Utzinger, J. (2009). Rapid assessment of schistosomiasis and other neglected tropical diseases in the context of integrated control programmes in Africa. *Parasitology* **136**, 1707–1718.

Brooker, S., Kabatereine, N. B., Fleming, F. and Devlin, N. (2008). Cost and cost-effectiveness of nationwide school-based helminth control in Uganda: intra-country variation and effects of scaling-up. *Health Policy and Planning* **23**, 24–35.

Brooker, S., Rowlands, M., Haller, L., Savioli, L. and Bundy, D. A. P. (2000). Towards an atlas of human helminth infection in sub-Saharan Africa: the use of geographical information systems (GIS). *Parasitology Today* **16**, 303–307.

Brooker, S. and Utzinger, J. (2007). Integrated disease mapping in a polyparasitic world. *Geospatial Health* **1**, 141–146.

Cameron, A. C. and Trivedi, P. K. (1986). Econometric models based on count data: comparisons and applications of some estimators. *Journal of Applied Economics* **1**, 29–53.

Clements, A. C. A., Brooker, S., Nyandindi, U., Fenwick, A. and Blair, L. (2008). Bayesian spatial analysis of a national urinary schistosomiasis questionnaire to assist geographic targeting of schistosomiasis control in Tanzania, East Africa. *International Journal for Parasitology* **38**, 401–415.

Clements, A. C. A., Lwambo, N. J. S., Blair, L., Nyandindi, U., Kaatano, G., Kinung'hi, S., Webster, J. P., Fenwick, A. and Brooker, S. (2006*a*). Bayesian spatial analysis and disease mapping: tools to enhance planning and implementation of a schistosomiasis control programme in Tanzania. *Tropical Medicine and International Health* **11**, 490–503.

Clements, A. C. A., Moyeed, R. and Brooker, S. (2006*b*). Bayesian geostatistical prediction of the intensity of infection with *Schistosoma mansoni* in East Africa. *Parasitology* **133**, 711–719.

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

Cohen, J. E. (1977). Mathematical models of schistosomiasis. *Annual Review of Ecology and Systematics* **8**, 209–233.

Cressie, N. (1991). *Statistics for Spatial Data*. Wiley & Sons, New York-Chichester-Toronto-Brisbane-Singapore.

Denwood, M. J., Stear, M. J., Matthews, L., Reid, S. W. J., Toft, N. and Innocent, G. T. (2008). The distribution of the pathogenic nematode *Nematodirus battus* in lambs is zero-inflated. *Parasitology* **135**, 1225–1235.

de Vlas, S. J. and Gryseels, B. (1992). Underestimation of *Schistosoma mansoni* prevalences. *Parasitology Today* **8**, 274–277.

Diggle, P. J., Moyeed, R. A. and Tawn, J. A. (1998). Model-based geostatistics. *Journal of the Royal Statistical Society. Series C, Applied Statistics* **47**, 299–326.

Dimatteo, I., Genovese, C. R. and Kass, R. E. (2001). Bayesian curve-fitting with free-knot splines. *Biometrika* **88**, 1055–1071.

Doumenge, J. P., Mott, K. E., Cheung, C., Villenave, D., Chapuis, O., Perrin, M. F. and Reaud-Thomas, G. (1987). Atlas of the global distribution of schistosomiasis. World Health Organization and Presses Universitaires de Bordeaux (WHO-CEGET-CNRS), Geneva.

Engels, D., Sinzinkayo, E. and Gryseels, B. (1996). Day-to-day egg count fluctuation in *Schistosoma mansoni* infection and its operational implications. *American Journal of Tropical Medicine and Hygiene* **54**, 319–324.

Eubank, R. L. (1988). *Spline Smoothing and Nonparametric Regression*. Decker, New York.

Gelfand, A. E. and Smith, A. F. M. (1990). Sampling based approaches to calculating marginal densities. *Journal of the American Statistical Association* **85**, 398–409.

Gemperli, A., Vounatsou, P., Kleinschmidt, I., Bagayoko, M., Lengeler, C. and Smith, T. (2004). Spatial patterns of infant mortality in Mali; the effect of malaria endemicity. *American Journal of Epidemiology* **159**, 64–72.

Gosoniu, L., Vounatsou, P., Sogoba, N., Maire, N. and Smith, T. (2009). Mapping malaria risk in West Africa using a Bayesian nonparametric non-stationary models. *Communications in Statistics – Simulation and Computation,* doi: 10.1016/j.csda.2009.02.022

Gosoniu, L., Vounatsou, P., Sogoba, N. and Smith, T. (2006). Bayesian modelling of geostatistical malaria risk data. *Geospatial Health* **1**, 127–139.

Heilbron, D. (1994). Zero-altered and other regression models for count data with added zeros. *Biometrical Journal* **36**, 531–547.

Katz, N., Chaves, A. and Pellegrino, J. (1972). A simple device for quantitative stool thick-smear technique in schistosomiasis mansoni. *Revista do Instituto de Medicina Tropical de São Paulo* **14**, 397–400.

Kitron, U., Clennon, J. A., Cecere, M. C., Gürtler, R. E., King, C. H. and Vazquez-Prokopec, G. (2006). Upscale or downscale: applications of fine scale remotely sensed data to Chagas disease in Argentina and schistosomiasis in Kenya. *Geospatial Health* **1**, 49–58.

Lambert, D. (1992). Zero-inflated Poisson regression with an application to defects in manufacturing. *Technometrics* **34**, 1–14.

Li, Y. S., Raso, G., Zhao, Z. Y., He, Y. K., Ellis, M. K. and McManus, D. P. (2007). Large water management projects and schistosomiasis control, Dongting Lake region, China. *Emerging Infectious Diseases* **13**, 973–979.

Lunn, D. J., Thomas, A., Best, N. and Spiegelhalter, D. (2000). WinBUGS – a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing* **10**, 325–337.

Majumdar, A. and Gelfand, A. E. (2007). Multivariate spatial modeling for geostatistical data using convolved covariance functions. *Mathematical Geology* **39**, 225–245.

Mullahy, J. (1986). Specification and testing of some modified count data models. *Journal of Econometrics* **33**, 341–365.

https://doi.org/10.1017/S003118200900599X Published online by Cambridge University Press

Polderman, A. M. (1979). Transmission dynamics of endemic schistosomiasis. *Tropical and Geographical Medicine* **31**, 465–475.

Raso, G., Luginbühl, A., Adjoua, C. A., Tian-Bi, N. T., Silué, K. D., Matthys, B., Vounatsou, P., Wang, Y., Dumas, M. E., Holmes, E., Singer, B. H., Tanner, M., N'Goran, E. K. and Utzinger, J. (2004). Multiple parasite infections and their relationship to self-reported morbidity in a community of rural Côte d'Ivoire. *International Journal of Epidemiology* **33**, 1092–1102.

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

Raso, G., Vounatsou, P., Gosoniou, L., Tanner, M., N'Goran, E. K. and Utzinger, J. (2006*b*). Risk factors and spatial patterns of hookworm infection among schoolchildren in a rural area of western Côte d'Ivoire. *International Journal for Parasitology* **36**, 201–210.

Raso, G., Vounatsou, P., McManus, D. P. and Utzinger, J. (2007). Bayesian risk maps for *Schistosoma mansoni* and hookworm mono-infections in a setting where both parasites co-exist. *Geospatial Health* **2**, 85–96.

Raso, G., Vounatsou, P., Singer, B. H., N'Goran, E. K., Tanner, M. and Utzinger, J. (2006*a*). An integrated approach for risk profiling and spatial prediction of *Schistosoma mansoni*-hookworm coinfection. *Proceedings of the National Academy of Sciences, USA* **103**, 6934–6939.

Rodrigues, J. (2003). Bayesian analysis of zero-inflated distributions. *Communication in Statistics-Theory and Methods* **32**, 281–289.

Scott, J. T., Diakhaté, M., Vereecken, K., Fall, A., Diop, M., Ly, A., De Clercq, D., de Vlas, S. J., Berkvens, D., Kestens, L. and Gryseels, B. (2003). Human water contacts patterns in *Schistosoma mansoni* epidemic foci in northern Senegal change according to age, sex and place of residence, but are not related to intensity of infection. *Tropical Medicine and International Health* **8**, 100–108.

Silué, K. D., Raso, G., Yapi, A., Vounatsou, P., Tanner, M., N'Goran, E. K. and Utzinger, J. (2008). Spatially-explicit risk profiling of *Plasmodium falciparum* infections at a small scale: a geostatistical modelling approach. *Malaria Journal* **7**, 111.

Sogoba, N., Vounatsou, P., Bagayoko, M. M., Doumbia, S., Dolo, G., Gosoniu, L., Traoré, S. F., Touré, Y. T. and Smith, T. (2007). The spatial distribution of *Anopheles gambiae sensu stricto* and *An. arabiensis* (Diptera: Culicidae) in Mali. *Geospatial Health* **1**, 213–222.

Steinmann, P., Du, Z. W., Wang, L. B., Wang, X. Z., Jiang, J. Y., Li, L. H., Marti, H., Zhou, X. N. and Utzinger, J. (2008). Extensive multiparasitism in a village of Yunnan province, People's Republic of China, revealed by a suite of diagnostic methods. *American Journal of Tropical Medicine and Hygiene* **78**, 760–769.

Steinmann, P., Keiser, J., Bos, R., Tanner, M. and Utzinger, J. (2006). Schistosomiasis and water resources development: systematic review, meta-analysis, and estimates of people at risk. *Lancet Infectious Diseases* **6**, 411–425.

Stensgaard, A. S., Jørgensen, A., Kabatereine, N. B., Rahbek, C. and Kristensen, T. K. (2006). Modeling freshwater snail habitat suitability and areas of potential snail-borne disease transmission in Uganda. *Geospatial Health* **1**, 93–104.

Sutherst, R. W. (2004). Global change and human vulnerability to vector-borne diseases. *Clinical Microbiology Reviews* **17**, 136–173.

Utzinger, J., Booth, M., N'Goran, E. K., Müller, I., Tanner, M. and Lengeler, C. (2001). Relative contribution of day-to-day and intra-specimen variation in faecal egg counts of *Schistosoma mansoni* before and after treatment with praziquantel. *Parasitology* **122**, 537–544.

Utzinger, J. and de Savigny, D. (2006). Control of neglected tropical diseases: integrated chemotherapy and beyond. *PLoS Medicine* **3**, e112.

Utzinger, J., N'Goran, E. K., Ossey, Y. A., Booth, M., Traoré, M., Lohourignon, K. L., Allangba, A., Ahiba, L. A., Tanner, M. and Lengeler, C. (2000). Rapid screening for *Schistosoma mansoni* in western Côte d'Ivoire using a simple school questionnaire. *Bulletin of the World Health Organization* **78**, 389–398.

Utzinger, J., Vounatsou, P., N'Goran, E. K., Tanner, M. and Booth, M. (2002). Reduction in the prevalence and intensity of hookworm infections after praziquantel treatment for schistosomiasis infection. *International Journal for Parasitology* **32**, 759–765.

WHO (2002). Prevention and control of schistosomiasis and soil-transmitted helminthiasis: report of a WHO expert committee. *WHO Technical Report Series* **912**, 1–57.

Yang, G. J., Vounatsou, P., Zhou, X. N., Tanner, M. and Utzinger, J. (2005*a*). A Bayesian-based approach for spatio-temporal modeling of county level *Schistosoma japonicum* prevalence data in Jiangsu province, China. *International Journal for Parasitology* **35**, 155–162.

Yang, G. J., Vounatsou, P., Zhou, X. N., Tanner, M. and Utzinger, J. (2005*b*). A potential impact of climate change and water resource development on the transmission of *Schistosoma japonicum* in China. *Parassitologia* **47**, 127–134.

Yapi, Y. G., Briët, O. J. T., Diabate, S., Vounatsou, P., Akodo, E., Tanner, M. and Teuscher, T. (2005). Rice irrigation and schistosomiasis in savannah and forest areas of Côte d'Ivoire. *Acta Tropica* **93**, 201–211.

Zhou, X. N., Yang, G. J., Yang, K., Wang, X. H., Hong, Q. B., Sun, L. P., Malone, J. B., Kristensen, T. K., Bergquist, N. R. and Utzinger, J. (2008). Potential impact of climate change on schistosomiasis transmission in China. *American Journal of Tropical Medicine and Hygiene* **78**, 188–194.

https://doi.org/10.1017/S003118200900599X Published online by Cambridge University Press