

Bayesian geostatistical prediction of the intensity of infection with *Schistosoma mansoni* in East Africa

A. C. A. CLEMENTS^{1,2}, R. MOYEED³ and S. BROOKER^{1*}

¹ Department of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, UK

² Schistosomiasis Control Initiative, Department of Infectious Disease Epidemiology, Imperial College London, London, UK

³ School of Mathematics and Statistics, University of Plymouth, Plymouth, UK

(Received 7 April 2006; revised 14 June 2006; accepted 16 June 2006; first published online 6 September 2006)

SUMMARY

A Bayesian geostatistical model was developed to predict the intensity of infection with *Schistosoma mansoni* in East Africa. Epidemiological data from purpose-designed and standardized surveys were available for 31 458 schoolchildren (90% aged between 6 and 16 years) from 459 locations across the region and used in combination with remote sensing environmental data to identify factors associated with spatial variation in infection patterns. The geostatistical model explicitly takes into account the highly aggregated distribution of parasite distributions by fitting a negative binomial distribution to the data and accounts for spatial correlation. Results identify the role of environmental risk factors in explaining geographical heterogeneity in infection intensity and show how these factors can be used to develop a predictive map. Such a map has important implications for schistosomiasis control programmes in the region.

Key words: Bayesian models, geostatistical prediction, negative binomial distribution, *Schistosoma mansoni*, schistosomiasis, East Africa.

INTRODUCTION

It is well recognized that reliable maps of the spatial distribution of parasitic diseases are required to help plan and evaluate disease control, and during the last decade there has been growing research interest in modelling disease risk (Hay *et al.* 2000, 2006). Most studies have modelled point prevalence data (Thomson *et al.* 1999; Kleinschmidt *et al.* 2001; Brooker *et al.* 2001, 2006; Clements *et al.* 2006) since this is an easily collected and readily available indicator. However, although prevalence data are often used for making recommendations for the need for control, data on the intensity of infection has greater relevance to the transmission dynamics of a given host-parasite system and for understanding the occurrence of morbidity (Anderson and May, 1991).

The modelling of infection intensity is nonetheless complicated by the fact that parasite distributions are highly aggregated within communities (Woolhouse *et al.* 1998; Shaw *et al.* 1998; Basáñez and Boussinesq, 1999) and by the presence of spatial correlation. In addition to parasite distributions, aggregated – or so-called overdispersed – distributions are a characteristic of several epidemiological phenomena including social networks (Zheng *et al.*

2006), infection exposure (Woolhouse *et al.* 1998) and wildlife distributions (Thogmartin *et al.* 2004). Aggregated distributions can often be fitted by the negative binomial distribution (Anderson, 1993) and previous analysis has employed negative binomial regression modelling to investigate the relationship between intensity of parasitic infection and a range of individual- and group-level risk factors (Nodtvedt *et al.* 2002; Yapi *et al.* 2005).

Studies modelling the spatial distribution of parasitic diseases have increasingly adopted a fully Bayesian geostatistical approach (Diggle *et al.* 2002; Carabin *et al.* 2003; Gemperli *et al.* 2004, 2006; Raso *et al.* 2005, 2006; Yang *et al.* 2005; Clements *et al.* 2006). This approach explicitly incorporates the spatial correlation structure of the data, the ability to include covariate effects and the comprehensive representation of uncertainty in the model outputs. Recently, Alexander and others (Alexander *et al.* 2000, 2003) have presented a negative binomial spatial model, which incorporates spatial correlation and adjusts for individual-level covariates such as age and sex. Their analysis, however, was explicative rather than predictive (limiting its utility for disease control decision-making), conducted on a small geographical scale and, unlike the Bayesian geostatistical modelling studies described above, did not include environmental factors which may have helped explain some of the observed heterogeneity.

In the present study, we extend the approach of Alexander and colleagues to model the spatial

* Corresponding author: Department of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, UK. Tel: +44 (0)20 7927 2614. Fax: +44 (0)20 7927 2918. E-mail: simon.brooker@lshtm.ac.uk

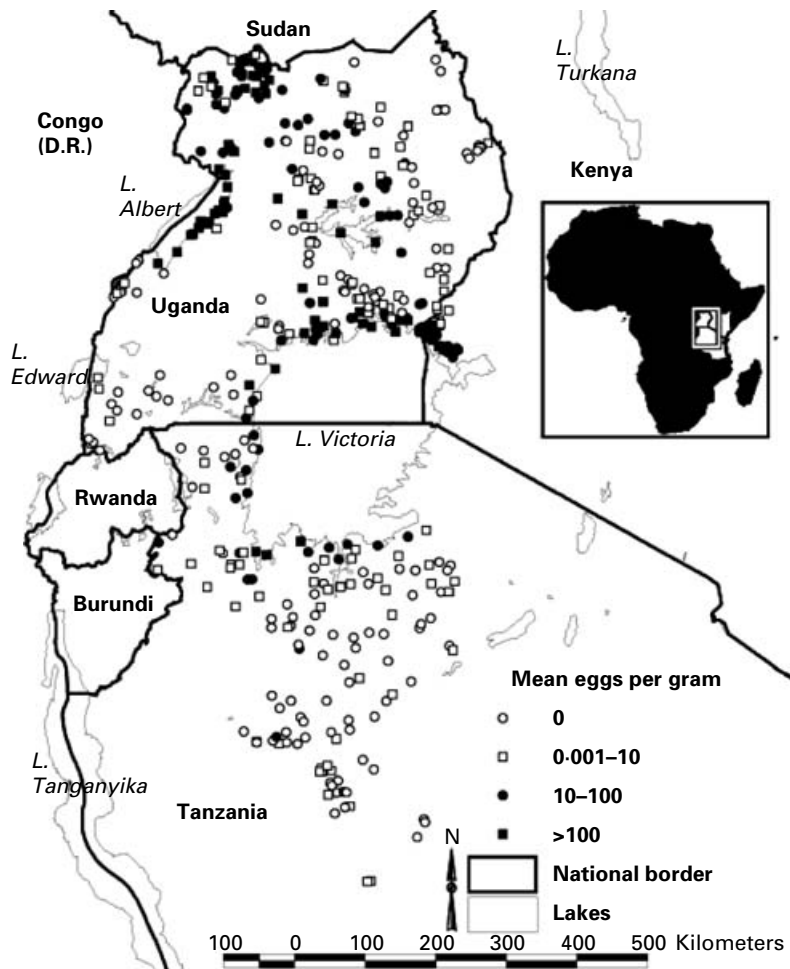


Fig. 1. Distribution of school surveys and observed intensity of *Schistosoma mansoni* (as measured by the mean number of eggs per gram faeces in each location) in 31 458 school children aged 5–22 years at 459 locations in East Africa conducted between 1999 and 2004. Locations of major lakes are also indicated.

distribution of the intensity of *Schistosoma mansoni* infection in East Africa. The modelling of intensity of infection, as measured by quantitative faecal egg counts, has particular relevance for the targeting of control activities since intensity is related to schistosome transmission dynamics and is a key determinant of schistosome-related morbidity with high levels likely to induce greater morbidity. Our adopted approach provides estimates of the effects of various factors influencing infection patterns including individual-level variables, such as age and sex, and location-level environmental variables. We then follow a Bayesian geostatistical approach to generate predictive maps of infection intensity across the region, and assess uncertainty associated with predictions.

PATIENTS AND METHODS

Data sources

Individual-level data on intensity of *Schistosoma mansoni* infection were obtained from cross-sectional random samples of schoolchildren from dedicated

school surveys conducted between 1999 and 2004 at 459 locations by national research teams under the auspices of the Schistosomiasis Control Initiative (SCI) in Tanzania (Clements *et al.* 2006) and in Uganda (Kabatereine *et al.* 2004) and by research projects in western Kenya (Brooker *et al.* 2001; Clarke *et al.* 2005). The locations of surveyed schools are shown in Fig. 1 and details of the studies are shown in Table 1. These data cover a large contiguous area of East Africa covering an area of approximately 1260 km (North/South) \times 590 km (East/West) and, to our knowledge, represent some of the most comprehensive geo-referenced data on the intensity of infection for any parasitic disease. No previous mass treatment had been undertaken in any of the schools. The details of the survey design, data collection and acquisition of ethical approval are reported in the original publications. Intensity of infection was based on the microscopic examination of 2 Kato-Katz smears derived from a single stool specimen, and expressed as eggs per gram of faeces (epg). As well as information on quantitative egg counts, individual-level factors including age and sex were also included in the final dataset.

Table 1. Summary of data on *Schistosoma mansoni* in East Africa used in the analysis

Country	No. of schools	No. of children	Mean age, years (range)	% Male	Overall prevalence (range by school)	Mean intensity (range by school)	Reference
Kenya	21	1,091	12.9 (10–20)	52.8	14.4 (0–30)	51.4 (0–141)	Clarke <i>et al.</i> (2005)
Kenya	24	1,6775	13.3 (6–20)	53.9	32.3 (0.2–97)	220.4 (3–1774)	Brooker <i>et al.</i> (2001)
Tanzania	143	8,617	14.0 (6–22)	49.6	4.4 (0–95)	17.6 (0–1384)	Clements <i>et al.</i> (2006)
Uganda	271	20,073	10.4 (5–21)	50.1	16.4 (0–97)	154.2 (0–3234)	Kabatereine <i>et al.</i> (2004)
Total	459	31,458	11.5 (5–22)	50.6	18.6 (0–97)	116.7 (0–3234)	

Long-term climatic data (1982–2000) were obtained from the National Oceanographic and Atmospheric Administration's (NOAA) Advanced Very High Radiometer (AVHRR) which provides indirect estimates of land surface temperature (LST) and normalized difference vegetation index (NDVI). Estimates of elevation were obtained from an interpolated digital elevation model from the Global Land Information System (GLIS) of the United States Geological Survey (<http://edcwww.cr.usgs.gov/landdaac/gtopo30/>) and the location of inland waterbodies, obtained from the US National Imagery and Mapping Agency (NIMA) and transformed by FAO/GIS at the Food and Agriculture Organisation (FAO). For each school, Euclidian distance to inland water bodies was calculated. The degree of urbanization, categorized into urban, periurban, rural and extreme rural, was derived from Tatem and colleagues (Tatem *et al.* 2005). These data were imported into the GIS ArcMap 8.1 (ESRI, Redlands, CA) and linked by geographical location to the parasitological data.

Statistical analysis

Initial analysis indicated that the distribution of egg counts were not Poisson distributed, with a predominance of zero and low values, such that the median egg count was 0 epg and the variance was 3.6×10^5 , clearly indicative of overdispersion. This justified a multi-variable regression approach (to account for some of the additional sources of heterogeneity) using the negative binomial distribution (to model the overdispersion) in our statistical modelling and prediction. Negative binomial regression models were fitted in Stata/SE version 8.1 (Stata Corporation, College Station, Texas). Initially, univariate analyses were conducted and variables with Wald's $P > 0.2$ were excluded from further analyses. With the remaining variables, backwards-stepwise regression was conducted using Wald's $P > 0.1$ as the exit criterion and $P \leq 0.05$ as the entry criterion. In the final model, sex was included as an individual-level covariate and elevation and distance to perennial water bodies were included as school-level covariates. Modelling either the overdispersion parameter as a function of a range

of covariates or the zero values separately from the positive values (using a zero-inflated model) did not improve the overall fit, justifying the subsequent use of a standard negative binomial regression model with a fixed value for the overdispersion parameter.

To provide an explanation of the overdispersion in the data and in particular to assess whether the overdispersion is spatially structured, we followed the approach of Alexander *et al.* (2000). In this, the total egg count of each individual was modelled as a negative binomial variate with overdispersion parameter $k > 0$ which incorporates extra-Poisson variation. Larger values of k indicate less variability, with the limiting case $k = \infty$ corresponding to the Poisson distribution. We model the logarithm of the mean of the distribution as an additive function of the individual-level covariate sex, the two school-level covariates elevation and distance to nearest inland perennial water body and a spatially-structured school-level random-effect. The spatial random-effect was modelled as a stationary Gaussian process with mean 0, variance σ^2 and correlation function $\exp(-d_{ij}/\alpha)$, where d_{ij} is the distance between villages i and j and the parameter α measures the rate at which the spatial correlation decays over distance, with $\alpha \log 2$ being a characteristic length, which we call the 'half-distance', over which the correlation reduces by half, and 3α being the distance at which the correlation reduces to 0.05.

Bayesian inference was implemented via a Markov chain Monte Carlo algorithm using the model-based geostatistics framework of Diggle *et al.* (1998). For the particular application considered in this paper the specification of prior distributions for the model parameters was done along the lines of the work described by Alexander *et al.* (2000) and Diggle *et al.* (2002). Specifically, the prior distribution for k was chosen to be gamma with density proportional to $k^{(1.5-1)} \exp(-0.01k)$ so that, *a priori*, k had a high variance. We chose independent improper uniform prior distributions for all the regression coefficients associated with the covariates and the constant term. For the parameters σ^2 and α we adopted the following vague priors: $f(\sigma^2) \propto 1/\sigma^2$ and $f(\alpha) \propto 1/\alpha^2$. In the absence of any prior knowledge about the model parameters, the choice of non-informative improper

Table 2. Coefficients of the Bayesian negative binomial geostatistical model based on overdispersed individual *Schistosoma mansoni* egg count data, obtained from 31 458 school-age children and young adults in East Africa during 1999–2004

Parameter	Spatial model			Non-spatial model				
	Posterior Mean	95% Bayesian CI		Posterior Median	Posterior Mean	95% Bayesian CI		Posterior Median
		Lower 2.5%	Upper 97.5%			Lower 2.5%	Upper 97.5%	
Intercept	9.941	5.768	13.223	10.055	10.295	9.826	10.770	10.292
Sex (Female)*	-0.413	-0.721	-0.106	-0.414	-0.463	-0.590	-0.336	-0.463
Elevation	-0.007	-0.010	-0.004	-0.007	-0.005	-0.005	-0.004	-0.005
Overdispersion parameter	0.060	0.058	0.062	0.060	0.031	0.030	0.032	0.031
Distance to nearest perennial water body†	-5.358	-7.512	-3.295	-5.364	-4.187	-4.334	-4.041	-4.188
Variance of the spatial random-effect	24.362	19.062	32.070	23.955				
Range of spatial correlation†	0.137	0.091	0.198	0.134				

* Reference category: Sex (Male).

† Unit: decimal degrees. 1 decimal degree = approximately 122 km at the equator.

priors was dictated by practical considerations, e.g. convergence issues and allowing the data to have a greater influence in determining the posterior.

We simulated realizations from the posterior distributions by means of a single-component Metropolis-Hastings algorithm. The parameters were updated by using a random-walk Metropolis algorithm with a Gaussian proposal density for the regression parameters, $\log(\sigma^2)$, $\log(\alpha)$ and $\log(k)$. At each iteration, the updating of σ^2 and α required the inversion of the $n \times n$ covariance matrix of the spatial effect, where n is the number of locations. In our example, $n=459$. The Markov chain was run for 110 000 iterations. After an initial burn-in of 10 000 iterations, the chain was sampled every 20th iteration to yield a sample of 5000 values from the posterior distribution of each of the model parameters. The convergence was judged by looking at the MCMC traces of the model parameters and also by computing the Monte Carlo errors. The traces differed in the extent to which they showed good mixing, but they all exhibited a reasonable degree of convergence to a stationary distribution. Spatial prediction was undertaken on a fine grid of 3304 locations with a grid-spacing of 12 km. At each prediction location 5000 values were simulated using the method outlined by Diggle *et al.* (1998). The Bayesian inference and prediction were undertaken using a program custom-written in the C language.

RESULTS

Intensity data were available for 31 458 school children (15 904 boys and 15 554 girls) aged 5–22 years (mean age 11.5 years, median age 12 years, 90% aged between 6 and 16 years). The median prevalence of *S. mansoni* infection was 3.3% ranging from 0 to

97.2% by school. The overall mean egg count was 116 epg (ranging from 0 to 3234 epg faeces by school) (Table 1). Figure 1 displays the spatial variation in infection intensity and shows that intensity was highest along the shores of the major lakes and in Uganda, along the river Nile.

The results of the Bayesian geostatistical model are presented in Table 2. Model coefficients indicate that females had significantly lower egg counts than males and that higher egg counts are found in areas at low elevations and close to perennial water. The posterior mean of $0.137 \times 122 \approx 16.7$ km for α , the range of spatial correlation, corresponds to a 'half-distance' for correlation of $16.7 \times \log 2 \approx 11.6$ km, with $3 \times 16.7 \approx 50.1$ km being the distance at which the correlation reduces to 0.05. The 95% posterior interval for α of (11.1, 24.2) km indicates that the residual spatial variation operates on a small scale. This is in agreement with the empirical variogram of the mean number of eggs in each location in Fig. 2A which exhibits a rising trend up to a distance of around 40 km. The posterior mean value of the overdispersion parameter was 0.060, confirming that the data were highly overdispersed even after accounting for spatial correlation. In addition, the estimate of the overdispersion parameter for the non-spatial model was 0.031, suggesting that inclusion of the spatial random-effect accounted for some of the overdispersion and, despite similar posterior mean values, the Bayesian credible intervals were narrower for the non-spatial model than the spatial model. All the parameters have mostly symmetric posterior distributions excepting σ^2 and α which are slightly positively skewed. The empirical variogram of the Pearson residuals from the fitted non-spatial regression model in Fig. 2B is quite similar to the variogram in Fig. 2A. This indicates that the

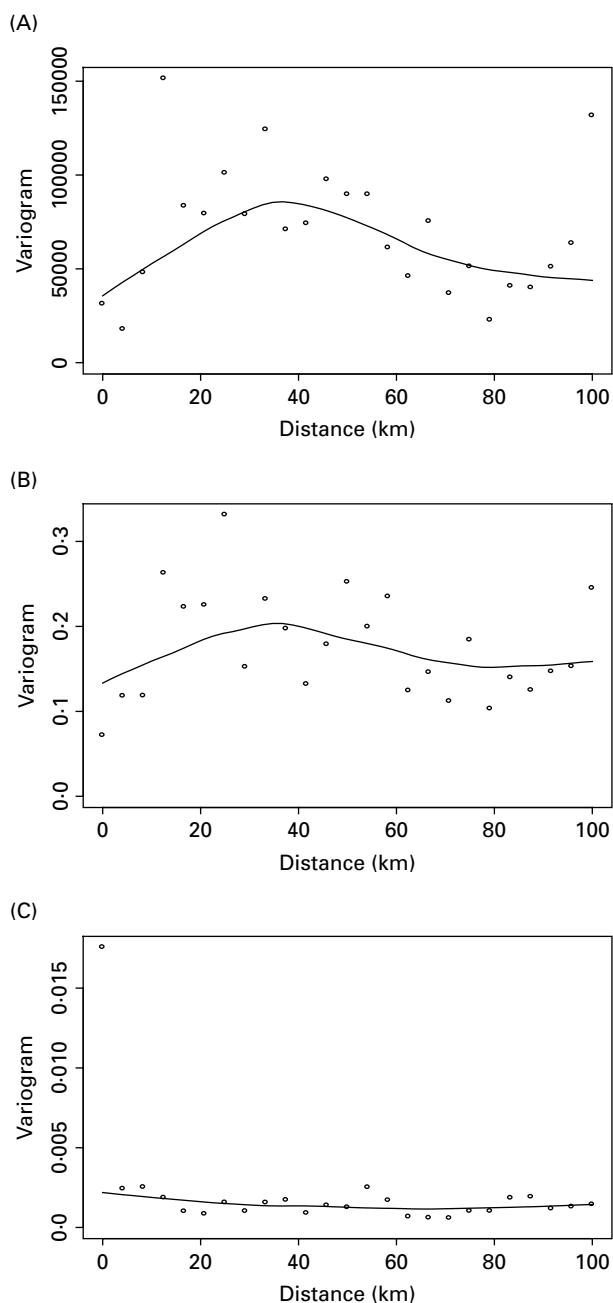


Fig. 2. (A) Empirical variogram of the mean number of eggs in each location. (B) Empirical variogram of the average of the Pearson residuals for each location from the non-spatial model. (C) Empirical variogram of the average of the Pearson residuals for each location from the spatial model. The curves were produced using a Loess smoother.

non-spatial regression model is inadequate in explaining the short-range spatial structure, whereas the variogram of the residuals from the spatial model shown in Fig. 2C is essentially flat.

The map of the posterior mean predicted intensity value is presented in Fig. 3. Model predictions of infection intensity indicate a high intensity in areas adjacent to Lakes Victoria, Kyoga, Albert and the northern part of Lake Edward as well as along the

Nile River valley in Uganda. Low intensity areas were predicted in northeast and southwest Uganda and much of northwest Tanzania away from the lakeshore. Fig. 4 presents the standard deviations of the modelled intensities. As expected, a lower level of uncertainty was apparent in locations close to observed data points and a higher level of uncertainty was apparent in locations distant from observed data points. Since observed data were well distributed across the region, most areas were characterized by a reasonable level of prediction certainty. An exception to this was northeast Uganda and areas between Lakes Victoria, Albert and Edward in western Uganda, areas which are too arid and too high, respectively to support transmission of *S. mansoni*, and therefore were not sampled.

DISCUSSION

This study presents a novel application of Bayesian geostatistical modelling to predict intensity of infection with *S. mansoni*, a highly aggregated parasitological index. Results identify the role of environmental risk factors in explaining geographical heterogeneity in infection intensity and show how these factors can be used to develop a predictive map. Such a map provides an empirical basis for identifying priority areas when implementing control and for predicting the potential impact of control.

The observed risk factors are already well known and are consistent and interpretable with the epidemiology of infection and known biology of freshwater snails, the intermediate hosts for *S. mansoni*. Malacological studies confirm that *Biomphalaria* snail species in East Africa tend to prefer large water body habitats (Prentice *et al.* 1972; Lwambo *et al.* 1999). Furthermore, individuals living on or near the shore are more likely to undertake risky water contact behaviour, increasing their exposure to infection. Observed differences among males and females may also be exposure-related, although it is also suggested that hormonal differences between males and females may account for higher infection rates in males than females (Webster *et al.* 1997). The role of other host-specific factors, such as genetics, immunity, nutrition, hygienic behaviour and education, as risk factors is also recognized. Climatic conditions, primarily temperature and rainfall, have been shown to influence the distribution and density of snails and the rate of schistosomal development in the snail host (Appleton, 1978; Sturrock, 1993). At high altitudes, low temperatures cause snails to die before the cercariae mature from sporocysts, limiting transmission (Pflüger, 1980).

Our predictive map of schistosome transmission is less affected by seasonality than other, more seasonally dynamic, parasitic diseases, such as malaria. Although seasonal dynamics in schistosome transmission may occur, such fluctuations may be of

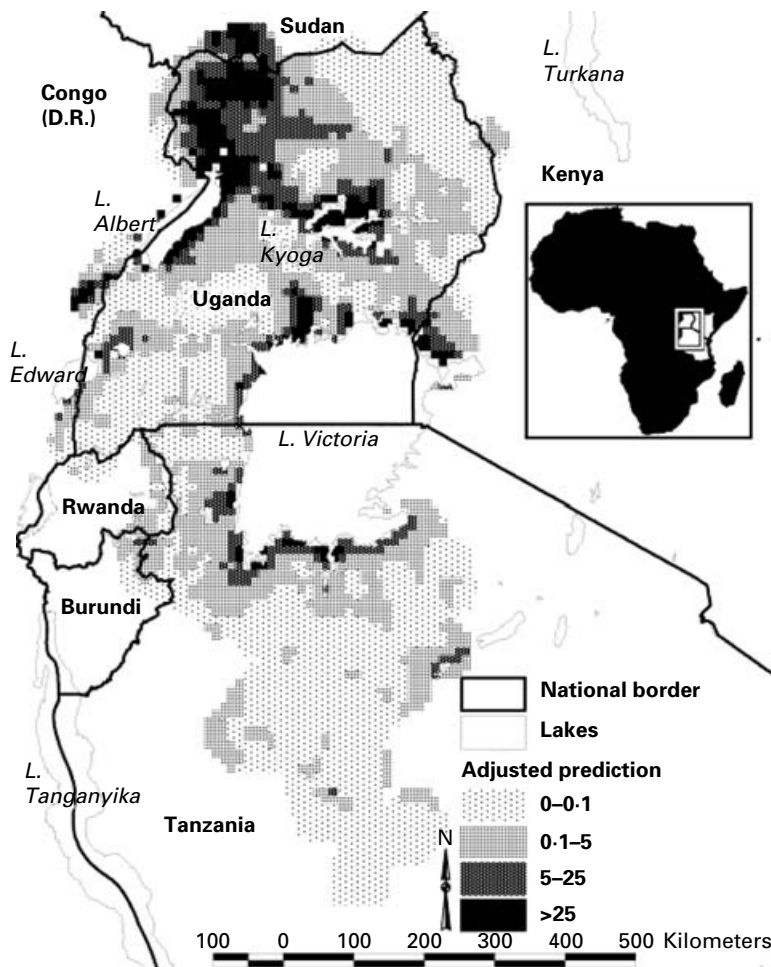


Fig. 3. Predicted intensity of infection (epg) with *Schistosoma mansoni* in East Africa, adjusted for covariates (distance to perennial water body and elevation).

little significance to the overall parasite equilibrium within communities. This is because the life-span of adult worms is typically much longer (1–10 years) than the periods in the year during which the basic reproductive number (R_0) is less than unity, and R_0 will on average be greater than one, maintaining overall endemicity (Anderson, 1982).

Previous studies predicting parasitic diseases in Africa on the basis of published data suffer from certain methodological problems including dissimilar age groupings of surveys and use of different diagnostic techniques to assess infection (e.g. Kleinschmidt *et al.* 2001). The inclusion of survey data from comparable age groups collected using the same diagnostic method (Kato-Katz quantitative smear) reduces this inherent variability. However, methodological problems related to the limited sensitivity and specificity of single stool samples to detect light infection (de Vlas *et al.* 1993) constitutes a limitation of the current analysis and may widen the prediction errors by introducing greater variability into the estimates of infection intensity.

Our approach highlights a number of key methodological limitations and areas for further work.

First, the current model does not account for anisotropic and non-stationary spatial variation. Data are currently being collected in other countries with SCI-supported control programmes, including Zambia in southern Africa and Burkina Faso, Mali and Niger in west Africa. These data cover wide areas with marked environmental diversity. Hence, there is scope of investigating the possible presence of anisotropy and non-stationarity in observed spatial variation. Second, there is the possibility that the degree of parasite aggregation may vary between endemic populations, reflecting area-specific differences in host exposure and immunity (Guyatt *et al.* 1994). We did not observe any variation in aggregation patterns with host age, presumably reflective of the relatively narrow age range of individuals included in the analysis, or according to environmental covariates. Third, zero-truncated negative binomial distribution has previously been used to describe patterns of *Wuchereria bancrofti* (Grenfell *et al.* 1990), *Onchocerca volvulus* (Filipe *et al.* 2005) and *Loa loa* (Pion *et al.* 2006) and may provide an alternative approach to modelling aggregated distributions. This would involve using a mixture

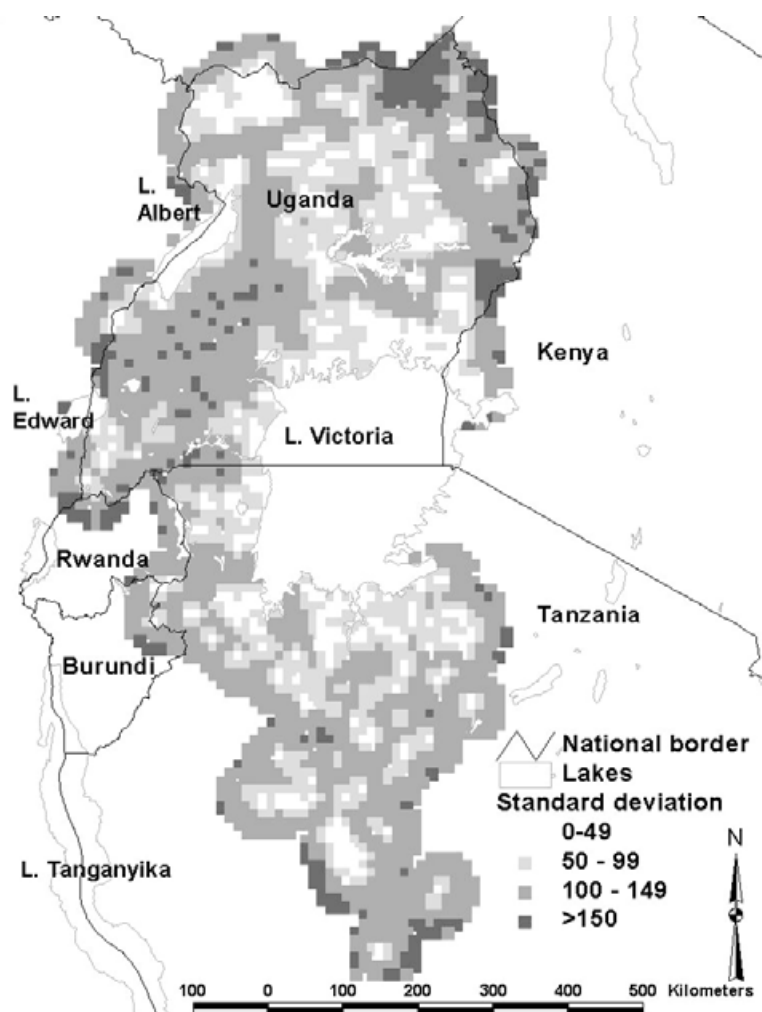


Fig. 4. Posterior standard deviations of the model predictions of the intensity of infection with *Schistosoma mansoni* in East Africa.

model, i.e. a negative binomial distribution with an extra point mass at zero. Over the next few years improved approaches to modelling geostatistical processes and transmission dynamics will iteratively improve our ability to predict disease patterns for geographical targeting of interventions.

As well as identifying areas where population-based morbidity control, using praziquantel, is most warranted (Brooker and Michael, 2000; Brooker, 2002), spatial predictions of infection intensity can also help define the frequency of treatment required to reduce morbidity and to model the potential impact of treatment. Mathematical models of transmission dynamics (Chan *et al.* 1995) use mean intensity of infection to predict rates of re-infection over fixed time-intervals, and provide estimates of infections and heavy infections prevented each year and the number of morbidity cases/year prevented. The integration of Bayesian geostatistical modelling with transmission dynamics models will provide an important analytical platform to help better understand the epidemiology of schistosomiasis and provide a more robust evidence-based approach to implementing and evaluating control.

This project was supported by the Schistosomiasis Control Initiative which is generously supported by the Bill and Melinda Gates Foundation. S. B. is supported by a Wellcome Trust Advanced Training Fellowship (073656). We thank the many field workers who carried out the field surveys. We also thank Drs Narcis Kabateriene, Nicholas Lwambo, Sian Clarke, Benson Estambale and Njagi Kiambo for providing access to the survey data, Dr Simon Hay for providing the satellite data and Professor Alan Fenwick for his support and contributions.

REFERENCES

- Alexander, N., Moyeed, R. and Stander, J. (2000). Spatial modelling of individual-level parasite counts using the negative binomial distribution. *Biostatistics* **1**, 453–463.
- Alexander, N. D., Moyeed, R. A., Hyun, P. J., Dimber, Z. B., Bockarie, M. J., Stander, J., Grenfell, B. T., Kazura, J. W. and Alpers, M. P. (2003). Spatial variation of Anopheles-transmitted *Wuchereria bancrofti* and *Plasmodium falciparum* infection densities in Papua New Guinea. *Filaria Journal* **2**, 14.
- Anderson, R. M. (1982). The population dynamics and control of hookworm and roundworm infection. In *Population Dynamics of Infectious Diseases: Theory and*

- Applications* (ed. Anderson, R. M.), pp. 67–109. Chapman and Hall, London.
- Anderson, R. M.** (1993). Epidemiology. In *Modern Parasitology* (ed. Cox, F. E. G.), pp. 75–166. Blackwell, Oxford.
- Anderson, R. M. and May, R. M.** (1991). *Infectious Diseases of Humans: Dynamics and Control*. Oxford University Press, Oxford, UK.
- Appleton, C. C.** (1978). Review of literature on abiotic factors influencing the distribution and life-cycles of Bilharziasis intermediate host snails. *Malacological Review* **11**, 1–25.
- Basáñez, M.-G. and Boussinesq, M.** (1999). Population biology of human onchocerciasis. *Philosophical Transactions of the Royal Society, B* **354**, 809–826.
- Brooker, S.** (2002). Schistosomes, snails and satellites. *Acta Tropica* **82**, 209–216.
- Brooker, S., Clements, A. C. A. and Bundy, D. A. P.** (2006). Global epidemiology, ecology and control of soil-transmitted helminth infections. *Advances in Parasitology* **62**, 223–265.
- Brooker, S., May, S. I., Issae, W., Hall, A., Kihamia, C. M., Lwambo, N. J. S., Wint, W., Rogers, D. J. and Bundy, D. A. P.** (2001). Predicting the distribution of urinary schistosomiasis in Tanzania using satellite sensor data. *Tropical Medicine and International Health* **6**, 998–1007.
- Brooker, S. and Michael, E.** (2000). The potential of geographical information systems and remote sensing in the epidemiology and control of human helminth infections. *Advances in Parasitology* **47**, 245–288.
- Brooker, S., Miguel, E. A., Moulin, S., Waswa, P., Namunyu, R., Guyatt, H. and Bundy, D. A. P.** (2001). The potential of rapid screening methods for *Schistosoma mansoni* in Western Kenya. *Annals of Tropical Medicine and Parasitology* **95**, 343–351.
- Carabin, H., Escalona, M., Marshall, C., Vivas-Martinez, S., Botto, C., Joseph, L. and Basáñez, M.-G.** (2003). Prediction of community prevalence of human onchocerciasis in the Amazonian onchocerciasis focus: Bayesian approach. *Bulletin of the World Health Organization* **81**, 482–490.
- Clements, A., Lwambo, N., Blair, L., Nyandindi, U., Kaatano, G., Kinunghi, S., Webster, J., Fenwick, A. and Brooker, S.** (2006). 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.
- Chan, M. S., Guyatt, H. L., Bundy, D. A., Booth, M., Fulford, A. J. and Medley, G. F.** (1995). The development of an age structured model for schistosomiasis transmission dynamics and control and its validation for *Schistosoma mansoni*. *Epidemiology and Infection* **115**, 325–344.
- Clarke, S., Njagi, K., Jukes, M. C. H., Estambale, B., Khasakhala, L., Ajanga, A., Otodo, J., Ochola, S., Magnussen, P. and Brooker, S.** (2005). Intermittent preventive treatment in schools: malaria parasitaemia, anaemia and school performance [MIM-KN-409248]. Meeting abstract. Fourth MIM Pan-African Malaria Conference. *Acta Tropica* **95S**, S133–S134.
- de Vlas, S. J., Nagelkerke, N. J., Habbema, J. D. and van Oortmarssen, G. J.** (1993). Statistical models for estimating prevalence and incidence of parasitic diseases. *Statistical Methods in Medical Research* **2**, 3–21.
- Diggle, P., Moyeed, R., Rowlingson, B. and Thompson, M.** (2002). Childhood malaria in the Gambia: a case-study in model-based geostatistics. *Applied Statistics* **51**, 493–506.
- Diggle, P., Tawn, J. A. and Moyeed, R.** (1998). Model-based geostatistics. *Applied Statistics* **47**, 299–350.
- Filipe, J. A. N., Boussinesq, M., Rez, A., Collins, A. C., Vivas-Martinez, S., Grillet, M.-G., Little, M. P. and Basáñez, M.-G.** (2005). Human infection patterns and heterogeneous exposure in river blindness. *Proceedings of the National Academy of Sciences, USA* **102**, 15265–15270.
- 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.
- Gemperli, A., Vounatsou, P., Sogoba, N. and Smith, T.** (2006). Malaria mapping using transmission models: application to survey data from Mali. *American Journal of Epidemiology* **163**, 289–297.
- Grenfell, B. T., Das, P. K., Rajagopalan, P. K. and Bundy, D. A. P.** (1990). Frequency distribution of lymphatic filariasis microfilariae in human population: population processes and statistical estimation. *Parasitology* **101**, 417–427.
- Guyatt, H. L., Smith, T., Gryseels, B., Lengeler, C., Mshinda, H., Siziya, S., Salanave, B., Mohome, N., Makwala, J., Ngimbi, K. P. and Tanner, M.** (1994). Aggregation in schistosomiasis: comparison of the relationships between prevalence and intensity in different endemic areas. *Parasitology* **109**, 45–55.
- Hay, S. I., Graham, A. J. and Rogers, D. J.** (2006). Global mapping of infectious diseases: methods, examples and emerging applications. *Advances in Parasitology* **62**, 1–381.
- Hay, S. I., Randolph, S. E. and Rogers, D. J.** (2000). Remote sensing and geographic information systems in epidemiology. *Advances in Parasitology* **47**, 1–344.
- Kabatereine, N. B., Brooker, S., Tukahebwa, E. M., Kazibwe, F. and Onapa, A.** (2004). Epidemiology and geography of *Schistosoma mansoni* in Uganda: implications for planning control. *Tropical Medicine and International Health* **9**, 372–380.
- Kleinschmidt, I., Omumbo, J., Briet, O., van de Giesen, N., Sogoba, N., Mensah, N. K., Windmeijer, P., Moussa, M. and Teuscher, T.** (2001). An empirical malaria distribution map for West Africa. *Tropical Medicine and International Health* **6**, 779–786.
- Lwambo, N. J. S., Siza, J. E., Brooker, S., Bundy, D. A. P. and Guyatt, H.** (1999). Patterns of concurrent infection with hookworm and schistosomiasis in school children in Tanzania. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **93**, 497–502.
- Nodtvedt, A., Dohoo, I., Sanchez, J., Conboy, G., DesCjeteaux, L., Keefe, G., Leslie, K. and Campbell, J.** (2002). The use of negative binomial modelling in a longitudinal study of gastrointestinal parasite burdens in

- Canadian dairy cows. *Canadian Journal of Veterinary Research* **66**, 249–257.
- Pflüger, W.** (1980). Experimental epidemiology of schistosomiasis I. The prepatent period and cercarial production of *Schistosoma mansoni* in *Biomphalaria* snails at various constant temperatures. *Zeitschrift für Parasitenkunde* **63**, 159–169.
- Pion, S. D. S., Filipe, J. A. N., Kamgno, J., Gardon, J., Basáñez, M.-G. and Boussinesq, M.** (2006). Microfilarial distribution of *Loa loa* in the human host: population dynamics and epidemiological implications. *Parasitology* **131**, 1–12.
- Prentice, M. A.** (1972). Distribution, prevalence and transmission of schistosomiasis in Uganda. *Uganda Medical Journal* **1**, 136–139.
- 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., Singer, B. H., N’Goran, E. K., Tanner, M. and Utzinger, J.** (2006). 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.
- Shaw, D. J., Grenfell, B. T. and Dobson, A. P.** (1998). Patterns of macroparasite aggregation in wildlife host populations. *Parasitology* **117**, 597–610.
- Sturrock, R. F.** (1993). The intermediate hosts and host-parasite relationships. In *Human Schistosomiasis* (ed. Jordan, P., Webb, G. and Sturrock, R. F.), pp. 33–85. CAB International, Wallingford.
- Tatem, A. J., Noor, A. M. and Hay, S. I.** (2005). Assessing the accuracy of satellite derived global and national urban maps in Kenya. *Remote Sensing of Environment* **96**, 87–97.
- Thogmartin, W. E., Sauer, J. R. and Knutson, M. G.** (2004). A hierarchical spatial model of avian abundance with application to cerulean warblers. *Ecological Applications* **14**, 1766–1779.
- Thomson, M. C., Connor, S. J., D’Alessandro, U., Rowlingson, B., Diggle, P., Cresswell, M. and Greenwood, B.** (1999). Predicting malaria infections in Gambian children from satellite data and bed net use surveys: the importance of spatial correlation in the interpretation of results. *American Journal of Tropical Medicine and Hygiene* **61**, 2–8.
- Webster, M., Correa-Oliveira, R., Gazzinelli, G., Viana, I. R., Fraga, L. A., Silveira, A. M. and Dunne, D. W.** (1997). Factors affecting high and low human IgE responses to schistosome worm antigens in an area of Brazil endemic for *Schistosoma mansoni* and hookworm. *American Journal of Tropical Medicine and Hygiene* **57**, 487–494.
- Woolhouse, M. E. J., Etard, J.-F., Dietz, K., Ndhlovu, P. D. and Chandiwana, S. K.** (1998). Heterogeneities in schistosome transmission dynamics and control. *Parasitology* **117**, 475–482.
- World Health Organization.** (1999). Report of the WHO informal consultation on schistosomiasis control. WHO/CDS/CPC/SIP/99.2. World Health Organization, Geneva.
- Yang, G. J., Vounatsou, P., Zhou, X. N., Tanner, M. and Utzinger, J.** (2005). A Bayesian-based approach for spatio-temporal modeling of county level prevalence of *Schistosoma japonicum* infection in Jiangsu province, China. *International Journal for Parasitology* **35**, 155–162.
- Yapi, Y. G., Briet, O. J., 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.
- Zheng, T., Salganik, M. J. and Gelman, A.** (2006). How many people do you know in prison? Using overdispersion in count data to estimate social structure in networks. *Journal of American Statistical Association* **101**, 409–423.