Acta Tropica 128 (2013) 365–377

Contents lists available at ScienceDirect

Acta Tropica

journal homepage: www.elsevier.com/locate/actatropica

Spatially explicit *Schistosoma* infection risk in eastern Africa using Bayesian geostatistical modelling

Nadine Schur[a,b], Eveline Hürlimann[a,b,c], Anna-Sofie Stensgaard[d,e], Kingford Chimfwembe[f], Gabriel Mushinge[f], Christopher Simoonga[g], Narcis B. Kabatereine[h], Thomas K. Kristensen[e], Jürg Utzinger[a,b], Penelope Vounatsou[a,b,*]

[a] *Department of Epidemiology and Public Health, Swiss Tropical and Public Health Institute, P.O. Box, CH-4002 Basel, Switzerland*
[b] *University of Basel, P.O. Box, CH-4003 Basel, Switzerland*
[c] *Centre Suisse de Recherches Scientifiques en Côte d'Ivoire, 01 BP 1303, Abidjan 01, Côte d'Ivoire*
[d] *Center for Macroecology and Evolution, Department of Biology, University of Copenhagen, Universitetsparken 15, DK-2100 Copenhagen, Denmark*
[e] *DBL, Department of Veterinary Disease Biology, University of Copenhagen, Thorvaldsensvej 57, DK-1871 Frederiksberg C, Denmark*
[f] *Department of Community Medicine, University of Zambia, P.O. Box 50110, Lusaka, Zambia*
[g] *Ministry of Health, P.O. Box 30205, 10101 Lusaka, Zambia*
[h] *Vector Control Division, Ministry of Health, P.O. Box 1661, Kampala, Uganda*

ARTICLE INFO

Article history:
Available online 14 October 2011

Keywords:
Bayesian geostatistics
East Africa
Markov chain Monte Carlo simulation
Risk mapping and prediction
Schistosoma haematobium
Schistosoma mansoni
Schistosomiasis

ABSTRACT

Schistosomiasis remains one of the most prevalent parasitic diseases in the tropics and subtropics, but current statistics are outdated due to demographic and ecological transformations and ongoing control efforts. Reliable risk estimates are important to plan and evaluate interventions in a spatially explicit and cost-effective manner. We analysed a large ensemble of georeferenced survey data derived from an open-access neglected tropical diseases database to create smooth empirical prevalence maps for *Schistosoma mansoni* and *Schistosoma haematobium* for a total of 13 countries of eastern Africa. Bayesian geostatistical models based on climatic and other environmental data were used to account for potential spatial clustering in spatially structured exposures. Geostatistical variable selection was employed to reduce the set of covariates. Alignment factors were implemented to combine surveys on different age-groups and to acquire separate estimates for individuals aged ≤20 years and entire communities. Prevalence estimates were combined with population statistics to obtain country-specific numbers of *Schistosoma* infections. We estimate that 122 million individuals in eastern Africa are currently infected with either *S. mansoni*, or *S. haematobium*, or both species concurrently. Country-specific population-adjusted prevalence estimates range between 12.9% (Uganda) and 34.5% (Mozambique) for *S. mansoni* and between 11.9% (Djibouti) and 40.9% (Mozambique) for *S. haematobium*. Our models revealed that infection risk in Burundi, Eritrea, Ethiopia, Kenya, Rwanda, Somalia and Sudan might be considerably higher than previously reported, while in Mozambique and Tanzania, the risk might be lower than current estimates suggest. Our empirical, large-scale, high-resolution infection risk estimates for *S. mansoni* and *S. haematobium* in eastern Africa can guide future control interventions and provide a benchmark for subsequent monitoring and evaluation activities.

1. Introduction

Schistosomiasis remains one of the most prevalent parasitic diseases in tropical and subtropical areas, particularly in sub-Saharan Africa (Steinmann et al., 2006; Utzinger et al., 2009). After many years of neglect, there is growing interest in the control of schistosomiasis and other neglected tropical diseases (Hotez et al., 2007; Fenwick et al., 2009; Utzinger et al., 2009). Reliable baseline maps of the geographical distribution of at-risk areas and estimates of the number of infected individuals are important tools to plan and evaluate control interventions in a cost-effective manner.

Abbreviations: BCI, Bayesian credible interval; CI, Confidence interval; EROS, Earth Resources Observation; GNTD database, Global Neglected Tropical Disease database; HII, human influence index; ISRIC, International Soil Reference and Information Center; LST, land surface temperature; MCMC, Markov chain Monte Carlo; MAE, mean absolute error; ME, mean error; MODIS, Moderate Resolution Imaging Spectroradiometer; NDVI, normalized difference vegetation index; OR, odds ratio; SD, standard deviation; SEDAC, Socioeconomic Data and Applications Center.

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

0001-706X/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.actatropica.2011.10.006

Most empirical mapping efforts for schistosomiasis only cover small geographical areas, e.g. a single village (Pinot de Moira et al., 2007), a district (Raso et al., 2005) or an entire country (Clements et al., 2006). Indeed, besides a few exceptions (Clements et al., 2008, 2010; Schur et al., 2011a), there is a paucity of large-scale mapping efforts. As part of the European Union (EU)-funded CONTRAST project, an up-to-date, open-access database of historical and contemporary prevalence surveys on schistosomiasis in Africa was developed (http://www.gntd.org) (Hürlimann et al., in press; Schur et al., 2011b; Stensgaard et al., 2013). Recently, we presented the first empirical schistosomiasis prevalence estimates for West Africa, based on the aforementioned database and a Bayesian-based geostatistical modelling approach using climatic and other environmental predictors (Schur et al., 2011a). We also presented population-adjusted risk estimates at country level and noted considerable differences from the widely cited statistics put forth by Chitsulo et al. for the mid-1990s (Chitsulo et al., 2000) and extrapolated estimates for mid-2003 (Steinmann et al., 2006). These previous estimates were based on population-adjusted statistics originally published by Utroska et al. (1989) and lack empirical modelling. Moreover, the estimates are likely outdated due to demographic and ecological transformations (e.g. water resources development and management), socio-economic development (e.g. improved access to clean water and sanitation) and implementation of large-scale control interventions, most notably regular deworming of school-aged children (Fenwick, 2006; Steinmann et al., 2006; Fenwick et al., 2009; Utzinger et al., 2009; WHO, 2010).

Bayesian geostatistical models fitted by Markov chain Monte Carlo (MCMC) simulation methods are increasingly utilised in disease risk mapping and prediction (Diggle et al., 1998; Banerjee et al., 2003). Such models have been employed for schistosomiasis risk profiling, i.e. mapping the distribution of *Schistosoma mansoni* in Burundi, Uganda and parts of Kenya and Tanzania (Clements et al., 2010). However, to our knowledge, large-scale model-based high-resolution *S. haematobium* and *S. mansoni* infection risk maps, including the number of infected individuals for the entire eastern African region, do not exist.

To fill this gap, we developed Bayesian geostatistical models based on climatic and other environmental risk factors, including different soil characteristics, to obtain empirical schistosomiasis risk maps and population-adjusted country prevalence estimates for an ensemble of 13 countries in eastern Africa. We analysed readily available survey data and implemented alignment factor models to account for the age-heterogeneity in the compiled survey data (Schur et al., 2011b). Geostatistical variable selection was applied to reduce the set of covariates to the most important predictors (George and McCulloch, 1993). Here, we report prevalence maps at 5 km × 5 km spatial resolution for *S. haematobium* and *S. mansoni* and estimated numbers of infected individuals at country level.

2. Data and methods

2.1. Disease data

Prevalence data on schistosomiasis for eastern Africa were extracted from the 'Global Neglected Tropical Disease' (GNTD) database (version: 5 October 2010) for all available survey years. The database assembles general information from the included publications, as well as study-specific information on survey population, time of the study, *Schistosoma* species, diagnostic test employed, and the number of infected individuals among those examined, stratified by age and sex (if available) (Hürlimann et al., in press). Data currently lacking geographical reference information were excluded. We also excluded entries based on non-direct diagnostic tests (e.g. immunofluorescence tests, antigen detections or questionnaire data) due to lower diagnostic sensitivities compared to direct diagnostic tests (e.g. schistosome egg detection in urine or stool). The proportion of rejected diagnostic techniques was low: 2.5% for *S. mansoni* and 0.6% for *S. haematobium*. Entries with missing information on the diagnostic technique (*S. mansoni*: 4.6% missing, *S. haematobium*: 8.4% missing) were assumed to be also largely based on direct examination techniques. We considered the bias that would arise from ignoring the missing data, as larger than the bias from potentially rejected diagnostic techniques among the data with missing information on the examination technique.

2.2. Climatic, demographic and environmental data

Climatic, demographic and environmental data were obtained from different freely accessible remote sensing data sources, as summarised in Table 1. Land surface temperature (LST) data were used as a proxy for day and night temperature, the normalized difference vegetation index (NDVI) as a proxy for moisture (Huete et al., 2002) and the human influence index (HII) for changes in the environment due to anthropometric activities (Sanderson et al., 2002). The land cover categories were re-grouped into six categories, as follows: (i) savannah and shrublands; (ii) forests; (iii) grasslands and sparsely vegetated areas; (iv) croplands; (v) urban areas; and (vi) wet areas. The soil parameters used were the following: bulk density (in kg/dm^3), available water capacity (in cm/m), pH and texture class (fine, medium and coarse). Furthermore, digitised maps on water body sources (rivers and lakes) in eastern Africa were combined and distance to the nearest water body source was calculated. LST, NDVI and rainfall data were summarised as overall averages over the period of 2000–2009 based on the mean. Land cover data from 2001 to 2004 were combined based on the most frequent category. Estimates for the percentage of individuals aged ≤20 years among the total population, stratified by country, were extracted from the U.S. Census Bureau International Database for the year 2010.

The MODIS/Terra data were processed using the 'MODIS Reprojection Tool' (Land Processes DAAC, USGS EROS). Rainfall estimates were converted in IDRISI 32 (Worcester, Clark University). Processing of the remaining data, distance calculations, and displaying of data and results were performed in ArcMap version 9.2 (ESRI). Further data processing was performed in Fortran 90 codes written by the authors. Remote sensing data were aligned to a common resolution of 5 km × 5 km. In particular, for data with high initial resolution (1 km × 1 km), the 'Aggregate' function of ArcMap was used to calculate the mean of all valid input cells that encompass the output cells (5 km × 5 km resolution). For data with low initial resolution (8 km × 8 km), the value of the input cell with the centroid closest to the centroid of the output cell was taken.

2.3. Statistical analysis

Bivariate logistic regressions were carried out to determine the relationship between the risk of *Schistosoma* infection and the potential covariates for each *Schistosoma* species separately. Covariates with non-linear outcome-predictor relations were treated as categorical.

Bayesian geostatistical logistic regression models with location-specific random effects and age-alignment factors (for details, see Appendix A) were fitted to identify the most significant predictors and to obtain spatially explicit schistosomiasis risk estimates. The random effects were considered as latent observations of a Gaussian process with variance–covariance matrix related to an exponential correlation function between any pair of locations. The variance–covariance is a matrix of $n \times n$, where n is the number of

Table 1
Data sources and properties of the climatic and other environmental covariates used to model schistosomiasis prevalence in eastern Africa.[a]

Source	Data type	Data period	Temporal resolution	Spatial resolution
Moderate resolution imaging spectroradiometer (MODIS)/Terra[b]	Land surface temperature (LST) for day and night	2000–2009	8 days	1 km
	Normalized difference vegetation index (NDVI)	2000–2009	16 days	1 km
	Land cover	2001–2004	Yearly	1 km
African Data Dissemination Service (ADDS)[c]	Rainfall	2000–2009	10 days	8 km
Earth Resources Observation (EROS) Center[d]	Altitude, slope and aspect	–	–	1 km
International Soil Reference and Information Center (ISRIC)[e]	Soil parameters	–	–	8 km
HealthMapper database[f]	Water bodies	–	–	Unknown
Socioeconomic Data and Applications Center (SEDAC)[g]	Human influence index (HII)	–	–	1 km
LandScan™ Global Population Database[h]	Population counts	2008	–	1 km

[a] All data accessed on 3 February 2011.
[b] Available at: https://lpdaac.usgs.gov/lpdaac/products/modis_products_table.
[c] Available at: http://earlywarning.usgs.gov/fews/africa/index.php.
[d] Available at: http://edc.usgs.gov/#/Find_Data/Products_and_Data_Available/gtopo30/hydro/.
[e] Available at: http://www.isric.org/data/isric-wise-derived-soil-properties-5-5-arc-minutes-global-grid-version-11.
[f] Available at: http://gis.emro.who.int/PublicHealthMappingGIS/HealthMapper.aspx.
[g] Available at: http://sedac.ciesin.columbia.edu/wildareas/.
[h] Available at: http://www.ornl.gov/landscan/.

survey locations. Model fit requires the inversion of this matrix. The datasets used for this study contain large numbers of survey locations and parameter estimation becomes unfeasible. Therefore, an approximation of the spatial process was used in the current application (see Appendix A).

The best set of covariates was determined using Gibbs variable selection (George and McCulloch, 1993). Indicator variables were linked to the regression coefficients to specify presence or absence of the corresponding covariate. In this study, variable selection was based on the estimation of the posterior inclusion probability with prior probability of 0.25. All covariates with a posterior inclusion probability larger than 0.5 were employed in the final model.

We employed MCMC simulation to estimate model parameters. Infection risk at unobserved locations was predicted via joint Bayesian kriging. A grid of prediction locations with a spatial resolution of $0.05^\circ \times 0.05^\circ$ (approximately 5 km × 5 km) was used, resulting in approximately 260,000 pixels. Population count estimates were linked to the grid to calculate the number of individuals aged ≤20 years and above per pixel. The number of individuals was merged with the model-based schistosomiasis risk predictions at the same locations to estimate the averaged number of infected individuals. A combined estimate for schistosomiasis risk of *S. mansoni* and *S. haematobium* was calculated on the assumption of independence between the two species, i.e. prevalence of *Schistosoma* spp. = prevalence of *S. mansoni* + prevalence of *S. haematobium* − (prevalence of *S. mansoni* × prevalence of *S. haematobium*).

2.4. Model validation

The performance of the models was assessed using a suite of model validations. In a first step, a sample of 80% of the survey locations was employed as training set for model fit, while the remaining 20% of the locations (test locations) were kept for model validation. Second, the predicted outcomes at the k test locations were compared to the observed outcomes via three different approaches: mean errors (ME), mean absolute errors (MAE) and Bayesian credible interval (BCI) comparisons (Gosoniu et al., 2006). The ME shows the overall tendency of a model to over- or underestimate the risk, and it is calculated by $\mathrm{ME} = 1/k\sum_{i=1}^{k}(p_i - \hat{p}_i)$, where p_i is the observed outcome and $\hat{p}_i$ the median of the predictions at test location i. The MAE provides information on the accuracy of a model based on the absolute distances between predictions and observations, $\mathrm{MAE} = 1/k\sum_{i=1}^{k}|p_i - \hat{p}_i|$. The proportion of test locations being correctly predicted within the q-th BCI of the posterior predictive distribution (restricted by the lower centiles c_i^l and upper centiles c_i^u) is the outcome of the BCI approach, i.e. $\mathrm{BCI}_q = 1/k\sum_{i=1}^{k}\min(I(c_i^l < p_i), I(c_i^u > p_i))$.

3. Results

3.1. Final datasets and preliminary statistics

The final datasets consisted of 1406 and 1851 survey locations for *S. haematobium* and *S. mansoni*, respectively. Among these, there were 1208 and 1558 unique locations, respectively. The number of surveys per country, year and age group are listed in Table 2. Prevalence in individuals aged ≤20 years ranged from 0% to 100% for both *Schistosoma* species with mean prevalence of 34.0% (median 26.7%, standard deviation (SD) 34.0%) for *S. haematobium* and 21.5% (median 8.6%, SD 27.4%) for *S. mansoni*. The distribution and the observed prevalence of the survey locations are shown in Figs. 1 and 2.

Data on the spatial distribution of the potential covariates influencing the distribution of schistosomiasis are presented in Appendix A. All considered covariates were significant in the bivariate logistic regressions. Typically, categorised covariates showed better predictive ability based on BIC than linear covariates, except for day temperature and altitude for *S. haematobium* and HII for *S. mansoni*. The implementation of the geostatistical variable selection approach led to a reduction in the final set of covariates. Distance to the closest freshwater body was excluded from the *S. haematobium* model, while NDVI was removed in the final *S. mansoni* model.

3.2. Spatial modelling results

Parameter estimates of the geostatistical model for *S. haematobium* are shown in Table 3. The results indicate that there was

Table 2
Overview of schistosomiasis data, stratified by survey year and age group given by country.

	Total (unique locations)	Survey year				Age group		All
		<1980	1980–1989	1990–1999	2000–2009	≤20 years	>20 years	
S. haematobium								
Burundi	0 (0)	0	0	0	0	0	0	0
Djibouti	0 (0)	0	0	0	0	0	0	0
Eritrea	7 (7)	7	0	0	0	4	1	2
Ethiopia	82 (56)	27	24	31	0	59	13	10
Kenya	172 (136)	76	50	33	13	123	18	31
Malawi	67 (62)	9	8	43	7	58	9	0
Mozambique	105 (103)	93	0	4	8	103	2	0
Rwanda	0 (0)	0	0	0	0	0	0	0
Somalia	73 (60)	48	25	0	0	40	18	15
Sudan	152 (135)	7	142	0	3	124	27	1
Tanzania	421 (351)	93	185	29	114	375	39	7
Uganda	57 (50)	16	2	0	39	50	7	0
Zambia	270 (248)	56	25	45	144	187	78	5
Total	1406 (1208)	432	461	185	328	1123	212	71
S. mansoni								
Burundi	85 (35)	0	67	18	0	47	38	0
Djibouti	0 (0)	0	0	0	0	0	0	0
Eritrea	11 (10)	8	3	0	0	7	1	3
Ethiopia	528 (438)	94	249	137	48	373	132	23
Kenya	142 (109)	68	31	22	21	119	15	8
Malawi	21 (21)	7	8	6	0	14	7	0
Mozambique	101 (96)	93	0	6	2	96	5	0
Rwanda	4 (4)	0	4	0	0	0	4	0
Somalia	10 (9)	8	2	0	0	4	1	5
Sudan	183 (156)	47	128	3	5	149	30	4
Tanzania	151 (129)	52	5	12	82	125	23	3
Uganda	432 (383)	28	21	21	362	381	38	13
Zambia	183 (168)	47	15	9	112	118	63	2
Total	1851 (1558)	452	533	234	632	1433	357	61

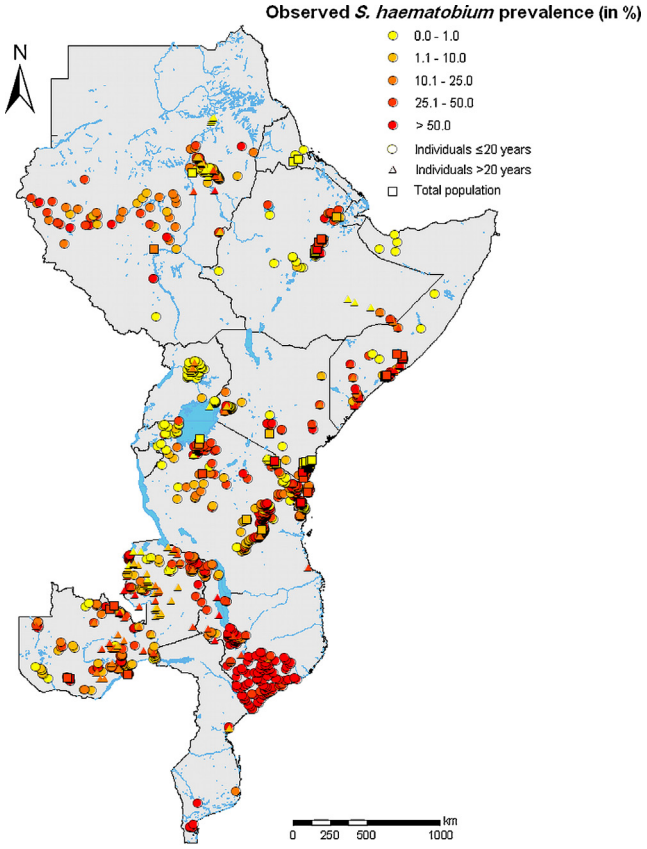

Fig. 1. Observed prevalence of *S. haematobium* across eastern Africa obtained from the GNTD database, including data until October 2010.

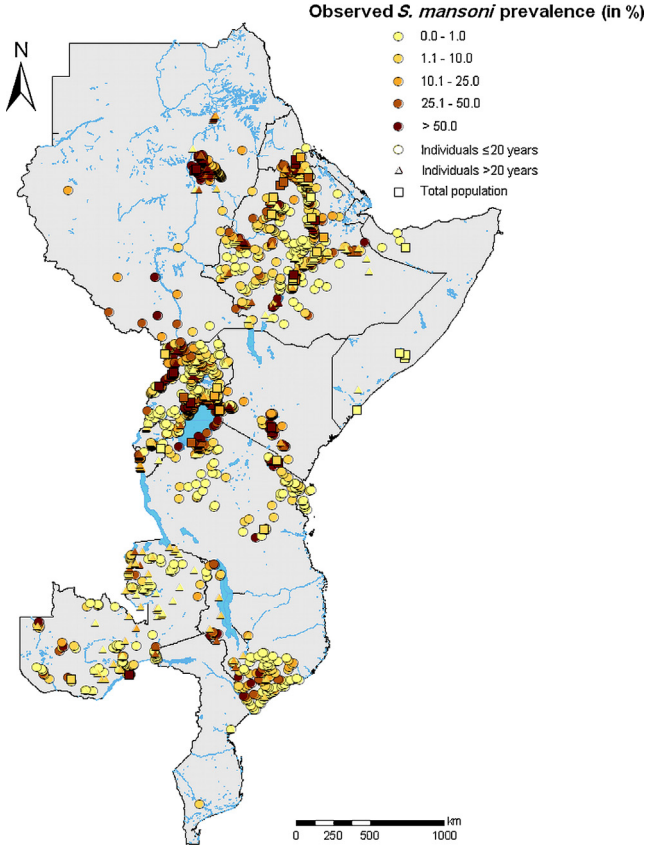

Fig. 2. Observed prevalence of *S. mansoni* across eastern Africa obtained from the GNTD database, including data until October 2010.

Table 3
Logistic regression parameter estimates for *S. haematobium* summarised by odds ratios (OR), 95% confidence intervals (CI) and 95% Bayesian credible intervals (BCI).

	Bivariate non-spatial OR (95% CI)	Multivariate non-spatial OR (95% CI)	Multivariate spatial OR (95% BCI)
Survey year			
<1980	1.00	1.00	1.00
1980–1989	0.58 (0.57, 0.59)*	1.07 (1.04, 1.10)*	1.02 (0.99, 1.08)
1990–1999	0.81 (0.79, 0.84)*	0.94 (0.90, 0.97)*	0.70 (0.67, 0.74)*
2000–2009	0.52 (0.51, 0.54)*	0.59 (0.57, 0.61)*	0.47 (0.45, 0.50)*
Latitude (in°)			
<−11.1	1.00	1.00	1.00
−11.1 to −5.5	0.97 (0.94, 0.99)*	0.37 (0.35, 0.39)*	0.74 (0.64, 0.83)*
−5.6 to 1.7	1.49 (1.45, 1.53)*	0.39 (0.37, 0.41)*	1.37 (1.22, 1.57)*
>1.7	0.60 (0.59, 0.62)*	0.34 (0.31, 0.35)*	0.61 (0.57, 0.65)*
Longitude (in°)			
<32.7	1.00	1.00	1.00
32.7–35.5	1.18 (1.14, 1.21)*	0.90 (0.86, 0.94)*	0.59 (0.54, 0.63)*
35.6–39.1	1.69 (1.64, 1.75)*	1.26 (1.18, 1.34)*	1.00 (0.96, 1.05)
>39.1	2.56 (2.49, 2.63)*	1.57 (1.47, 1.68)*	0.67 (0.63, 0.72)*
Altitude (in m)			
<200	1.00	1.00	1.00
200–559	0.27 (0.26, 0.28)*	0.48 (0.46, 0.51)*	1.00 (0.96, 1.05)
560–1102	0.49 (0.47, 0.50)*	0.66 (0.62, 0.70)*	0.57 (0.54, 0.60)*
>1102	0.54 (0.52, 0.55)*	0.57 (0.53, 0.61)*	0.50 (0.47, 0.53)*
Day LST (in °C)	0.92 (0.92, 0.92)*	0.97 (0.97, 0.98)*	0.92 (0.92, 0.92)*
Night LST (°C)			
<17.7	1.00	1.00	1.00
17.7–19.8	2.34 (2.27, 2.41)*	2.27 (2.16, 2.38)*	2.20 (2.09, 2.34)*
19.9–21.7	1.53 (1.49, 1.58)*	1.91 (1.81, 2.03)*	2.26 (2.09, 2.43)*
>21.7	2.09 (2.03, 2.15)*	1.84 (1.73, 1.97)*	1.85 (1.70, 2.00)*
Rainfall (in mm)			
<216	1.00	1.00	1.00
216–276	1.71 (1.68, 1.75)*	1.49 (1.44, 1.54)*	0.63 (0.60, 0.67)*
>276	0.85 (0.83, 0.87)*	1.12 (1.08, 1.17)*	0.77 (0.72, 0.83)*
NDVI			
<0.40	1.00	1.00	1.00
0.40–0.49	2.74 (2.67, 2.80)*	1.23 (1.19, 1.28)*	1.67 (1.61, 1.76)*
0.50–0.59	2.08 (2.03, 2.14)*	1.15 (1.10, 1.20)*	1.53 (1.45, 1.64)*
>0.59	2.32 (2.26, 2.38)*	0.89 (0.85, 0.94)*	1.23 (1.15, 1.33)*
Land cover			
Savannah and shrublands	1.00	1.00	1.00
Forests	0.98 (0.96, 1.01)	0.94 (0.90, 0.97)*	1.00 (0.99, 1.03)
Grasslands and sparsely vegetated	0.79 (0.77, 0.81)*	0.81 (0.78, 0.84)*	0.67 (0.64, 0.70)*
Croplands	1.25 (1.22, 1.28)*	1.24 (1.20, 1.28)*	1.07 (1.01, 1.12)*
Urban	0.86 (0.83, 0.88)*	0.63 (0.60, 0.66)*	1.00 (0.97, 1.02)
Wet areas	0.87 (0.83, 0.92)*	0.65 (0.60, 0.70)*	0.72 (0.66, 0.78)*
Slope (in°)			
<0.16	1.00	1.00	1.00
0.16–0.43	0.62 (0.60, 0.64)*	0.74 (0.71, 0.76)*	0.87 (0.83, 0.90)*
0.44–1.15	0.52 (0.51, 0.54)*	0.78 (0.75, 0.81)*	0.71 (0.67, 0.75)*
>1.15	0.74 (0.72, 0.76)*	1.08 (1.04, 1.12)*	1.17 (1.12, 1.23)*
Aspect (in°)			
<48.8	1.00	1.00	1.00
48.8–105.0	1.40 (1.36, 1.43)*	1.40 (1.35, 1.45)*	1.43 (1.39, 1.48)*
105.1–202.4	1.05 (1.03, 1.08)*	1.04 (1.01, 1.07)*	1.00 (0.99, 1.02)
>202.4	0.60 (0.58, 0.61)*	0.95 (0.92, 0.99)*	0.93 (0.89, 0.97)*
Human influence index			
<17	1.00	1.00	1.00
17–19	1.19 (1.15, 1.23)*	1.15 (1.10, 1.19)*	1.00 (0.97, 1.04)
20–24	1.48 (1.43, 1.52)*	1.40 (1.35, 1.45)*	1.07 (1.01, 1.13)*
>24	1.20 (1.17, 1.24)*	1.31 (1.26, 1.36)*	0.93 (0.89, 0.98)*
Bulk density (in kg/dm^3)			
<1.32	1.00	1.00	1.00
1.32–1.34	1.25 (1.21, 1.29)*	0.65 (0.62, 0.68)*	0.81 (0.75, 0.91)*
1.35–1.50	1.35 (1.31, 1.39)*	0.49 (0.47, 0.52)*	0.55 (0.50, 0.61)*
>1.50	1.98 (1.93, 2.03)*	1.43 (1.35, 1.51)*	1.02 (0.99, 1.04)
Available water capacity (in cm/m)			
<8	1.00	1.00	1.00
8–9	0.78 (0.76, 0.80)*	2.43 (2.25, 2.63)*	1.69 (1.55, 1.84)*
10–12	0.31 (0.30, 0.32)*	0.95 (0.88, 1.02)	2.06 (1.90, 2.22)*
>12	0.56 (0.55, 0.57)*	1.04 (0.95, 1.12)	1.43 (1.33, 1.52)*
pH in water			
<5.9	1.00	1.00	1.00
5.9–6.8	1.11 (1.09, 1.13)*	0.85 (0.82, 0.89)*	0.76 (0.72, 0.81)*
>6.8	0.67 (0.66, 0.69)*	1.08 (1.01, 1.15)*	1.17 (1.07, 1.26)*
Texture class			
Medium	1.00	1.00	1.00
Coarse	1.76 (1.72, 1.80)*	0.49 (0.44, 0.53)*	1.30 (1.10, 1.49)*
Fine	0.79 (0.77, 0.80)*	0.43 (0.41, 0.45)*	0.90 (0.80, 1.00)
			Mean (95% BCI)
Sigma^2	–	–	3.83 (3.15, 4.49)
Range (in km)	–	–	355.3 (341.0, 372.0)

* Significant correlation based on 95% CI or 95% BCI.

a significant decrease in *S. haematobium* infection risk from the beginning of the 1990s onwards. Latitude and longitude effects suggested an elevated risk around Lake Victoria and the countries south of Tanzania. Higher day LST averages and yearly rainfall estimates above 216 mm were associated with low risk of infection. There was a non-linear relation between *S. haematobium* with night LST and NDVI indicating a positive relation and a decrease in risk at the lowest and the highest values of those two covariates. Areas covered with croplands showed a higher schistosomiasis risk than those with savannahs and shrublands. Altitude showed a negative relation, while topological parameters showed high and low risk in East and West direction, respectively. Most of our locations were in low flow velocity flat areas, and therefore the observed positive effect of high slope regions is rather misleading. Soil-related parameters showed that schistosomiasis is more common in areas where the soils have a pH that is neutral to acid, high water capacity and coarse texture. The influence of anthropometric activities to the environment appeared to be positively associated with schistosomiasis at intermediate levels of the covariate.

The estimated odds ratios (ORs) of the predictors for *S. mansoni* infection risk are given in Table 4. The risk increased during the 1980s and 1990s, but decreased slightly from 2000 onwards. A South-to-North and West-to-East trend of increasing prevalence is indicated by latitude and longitude. Low risk of infection is associated with high average day LST and low night LST values and rainfall showed a significant positive relation. Forested areas were found to be related to highest risk, while low risk estimates were found in built up environments, such as urbanised settings. An elevated risk was suggested at very low and very high levels of altitude in flat regions and areas pointing towards West, as indicated by the aspect parameter. Locations with distances of more than 1.6 km to the nearest freshwater body did not appear to be different from locations in close proximity (within 0.6 km) to freshwater bodies. However, intermediate distances showed a positive relation to *S. mansoni* infections. Higher risks were associated with soils of medium texture, high basicity, low bulk density and low water capacity. The HII showed no relation with *S. mansoni*.

The estimated spatial parameters were similar for both *Schistosoma* species. Spatial ranges of 355 km (95% BCI: 341–372 km) and 357 km (95% BCI: 337–379 km) were observed for *S. haematobium* and *S. mansoni*, respectively, and respective spatial variation of 3.83 (95% BCI: 3.15–4.49) and 3.79 (95% BCI: 3.16–4.50).

3.3. *Schistosoma* infection risk maps

The spatial distribution of *S. haematobium* risk throughout eastern Africa is shown in Fig. 3A. Large areas of high infection risk (>50%) were predicted for central Mozambique, the south of Lake Victoria and around the Sudanese and Eritrean border. Low risk areas with predicted infection risks <10% were mainly located in mid/northern Zambia, around Mount Kilimanjaro, in the north of Lake Victoria, northern Sudan and in Ethiopia. The map of the SD of the prediction error for *S. haematobium* (Fig. 3B) demonstrates that areas of relatively high uncertainty (above 30%) are mainly found in areas of high infection risk and far away from sampled survey locations. We found a mean SD of about 23%, varying from 0% to 32.5%, with areas of low uncertainty typically in close proximity to sub-sampled locations.

Fig. 4A displays the *S. mansoni* infection risk map, and Fig. 4B shows the corresponding map of the SD of the prediction error. Low risk areas (predicted infection risk <10%) occur in large parts of Zambia, central Tanzania, around the Ugandan and Kenyan border and the Ethiopian highlands. High risk areas (predicted infection risk >50%) are located in northern Mozambique, southern Tanzania, around Lake Victoria, south-eastern Kenya and small areas in Ethiopia and Sudan. Main areas of high uncertainty are found in Mozambique, south-western Sudan, northern Eritrea, and parts of Somalia and Kenya, while areas of low uncertainty are located around the sampled survey locations and low-risk settings. We calculated a mean SD of 23.5%, varying from 0% to 32.5%.

3.4. Country prevalence estimates and numbers of infected individuals

Population-adjusted country prevalence estimates are summarised in Table 5 for individuals aged ≤20 years and the total population. For *S. haematobium*, prevalence estimates for the total population vary from 11.9% (Djibouti) to 40.9% (Mozambique), whereas for *S. mansoni* they vary between 12.9% (Uganda) and 34.5% (Mozambique). In Burundi, Malawi, Mozambique, Rwanda and Zambia, *S. haematobium* was the predominant species, while in Djibouti, Eritrea, Kenya, Somalia and Sudan, *S. mansoni* was the primary *Schistosoma* species. Both species were estimated to have similar country prevalence in Ethiopia, Tanzania and Uganda. Combined schistosomiasis prevalence estimates, assuming independence in the occurrence of the two species, ranged between 25.3% (Uganda) and 55.6% (Mozambique) for the total population.

The number of infected individuals per country, stratified by individuals aged ≤20 years and the total population, is given in Table 6. High numbers of infected individuals among the total population (>5 million) were predicted for Ethiopia, Malawi, Mozambique, Sudan, Tanzania and Uganda for *S. haematobium*, and in Ethiopia, Kenya, Mozambique, Sudan and Tanzania for *S. mansoni*. Low numbers (<1 million) were only observed in Djibouti for *S. haematobium* and *S. mansoni*. The combined number of infected individuals vary from 147,000 (Djibouti) to 29.1 million (Ethiopia) with a total of approximately 122 million infections in the 13 countries considered here in eastern Africa.

3.5. Model validation results

Model validation based on 80% of the survey locations resulted in MEs of 0.6 for *S. haematobium* and −1.7 for *S. mansoni*, and MAEs of 15.3 and 14.2, respectively. The percentage of test locations correctly predicted by 95% BCIs is 78.3% for *S. haematobium* and 71.1% for *S. mansoni*.

4. Discussion

To our knowledge, we present the first smooth empirical schistosomiasis prevalence maps at a spatial resolution of 5 km × 5 km for an ensemble of 13 countries in eastern Africa. The maps are stratified by the two main *Schistosoma* species, *S. haematobium* and *S. mansoni*. Bayesian geostatistical models with an approximation of the spatial process were employed to handle the large amount of unique survey locations extracted from a readily available open-access GNTD database (Hürlimann et al., in press; Schur et al., 2011a,b; Stensgaard et al., 2013). Our prevalence maps are accompanied by contemporary population-adjusted prevalence estimates and number of infected individuals on a country-by-country basis. An attempt was made to employ factors to align surveys arising from different risk groups, namely individuals aged ≤20 years and entire communities. This enabled us to obtain age-adjusted risk estimates for individuals aged ≤20 years, who are known to carry the highest schistosomiasis risk (WHO, 2002), as well as entire populations. The spatial resolution of 5 km × 5 km is a compromise between computational burden and estimation accuracy. A map of 1 km × 1 km resolution would result in a total of more than 6 million pixels in the study area, as compared to 260,000 pixels when a 5 km × 5 km resolution is chosen (a 25-fold difference). In addition, schistosomiasis risk is influenced by local

Table 4
Logistic regression parameter estimates for *S. mansoni* summarised by odds ratios (OR), 95% confidence intervals (CI) and 95% Bayesian credible intervals (BCI).

	Bivariate non-spatial OR (95% CI)	Multivariate non-spatial OR (95% CI)	Multivariate spatial OR (95% BCI)
Study year			
<1980	1.00	1.00	1.00
1980–1989	1.30 (1.26, 1.33)*	1.21 (1.17, 1.25)*	1.86 (1.79, 1.91)*
1990–1999	1.84 (1.79, 1.89)*	1.98 (1.92, 2.05)*	1.99 (1.89, 2.08)*
2000–2009	1.11 (1.08, 1.14)*	1.13 (1.09, 1.17)*	1.48 (1.41, 1.55)*
Latitude (in°)			
<−2.8	1.00	1.00	1.00
−2.8–1.0	2.21 (2.15, 2.27)*	3.67 (3.53, 3.81)*	1.80 (1.57, 2.08)*
1.1–9.8	2.42 (2.34, 2.49)*	4.21 (4.04, 4.38)*	1.41 (1.29, 1.55)*
>9.8	2.22 (2.16, 2.29)*	5.39 (5.12, 5.66)*	2.44 (2.28, 2.61)*
Longitude (in°)			
<32.9	1.00	1.00	1.00
32.9–37.2	1.30 (1.28, 1.33)*	1.00 (0.97, 1.03)	2.21 (2.03, 2.39)*
>37.2	1.45 (1.42, 1.48)*	1.63 (1.55, 1.70)*	2.82 (2.62, 2.97)*
Altitude (in m)			
<801	1.00	1.00	1.00
801–1154	0.80 (0.78, 0.82)*	0.57 (0.54, 0.60)*	0.53 (0.50, 0.57)*
1155–1513	0.88 (0.86, 0.90)*	0.59 (0.56, 0.63)*	0.50 (0.46, 0.55)*
>1513	0.59 (0.58, 0.60)*	0.59 (0.55, 0.63)*	1.04 (0.99, 1.18)
Day LST (in °C)			
<27.8	1.00	1.00	1.00
27.8–30.1	1.16 (1.13, 1.19)*	0.96 (0.93, 1.00)	1.23 (1.19, 1.27)*
30.2–33.1	0.88 (0.86, 0.90)*	0.75 (0.72, 0.78)*	1.00 (0.99, 1.02)
>33.1	1.18 (1.15, 1.20)*	0.40 (0.38, 0.42)*	0.75 (0.72, 0.79)*
Night LST (in °C)			
<15.7	1.00	1.00	1.00
15.7–18.0	2.68 (2.60, 2.76)*	3.82 (3.67, 3.98)*	1.67 (1.59, 1.75)*
18.1–20.1	2.35 (2.28, 2.42)*	5.10 (4.85, 5.37)*	2.27 (2.09, 2.46)*
>20.1	2.57 (2.50, 2.64)*	4.01 (3.74, 4.30)*	1.61 (1.50, 1.80)*
Rainfall (in mm)			
<216	1.00	1.00	1.00
216–312	0.49 (0.48, 0.50)*	0.77 (0.74, 0.80)*	1.56 (1.48, 1.66)*
>312	0.80 (0.79, 0.82)*	0.91 (0.87, 0.95)*	3.19 (3.02, 3.46)*
Land cover			
Savannah and shrublands	1.00	1.00	1.00
Forests	0.87 (0.84, 0.89)*	0.78 (0.75, 0.81)*	1.15 (1.10, 1.19)*
Grasslands and sparsely vegetated	1.15 (1.12, 1.18)*	1.03 (1.00, 1.07)	1.00 (0.99, 1.02)
Croplands	0.82 (0.80, 0.84)*	0.94 (0.91, 0.97)*	1.00 (0.97, 1.02)
Urban	0.39 (0.38, 0.40)*	0.74 (0.71, 0.78)*	0.58 (0.54, 0.61)*
Wet areas	1.13 (1.09, 1.17)*	0.79 (0.75, 0.84)*	0.97 (0.90, 1.00)
Slope (in°)			
<0.28	1.00	1.00	1.00
0.28–0.92	1.35 (1.32, 1.38)*	1.29 (1.26, 1.33)*	1.01 (0.99, 1.05)
>0.92	1.10 (1.08, 0.13)*	1.03 (1.01, 1.06)*	0.96 (0.92, 0.99)*
Aspect (in°)			
<73.1	1.00	1.00	1.00
73.1–182.2	1.06 (1.03, 1.08)*	1.28 (1.24, 1.31)*	1.03 (1.00, 1.08)
>182.2	1.27 (1.24, 1.29)*	1.34 (1.30, 1.38)*	1.13 (1.09, 1.18)*
Human influence index	0.96 (0.96, 0.96)*	0.96 (0.96, 0.97)*	1.00 (0.99, 1.00)
Distance to closest water body (km)			
<0.59	1.00	1.00	1.00
0.59–1.56	1.23 (1.20, 1.26)*	1.14 (1.11, 1.18)*	1.16 (1.11, 1.21)*
1.57–3.70	1.04 (1.02, 1.07)*	0.85 (0.83, 0.88)*	0.99 (0.96, 1.01)
>3.70	0.72 (0.71, 0.74)*	0.72 (0.70, 0.75)*	0.99 (0.96, 1.05)
Bulk density (in kg/dm^3)			
<1.30	1.00	1.00	1.00
1.30–1.31	1.49 (1.45, 1.53)*	1.69 (1.61, 1.78)*	1.31 (1.21, 1.39)*
1.32–1.40	0.73 (0.71, 0.75)*	0.57 (0.54, 0.59)*	0.59 (0.56, 0.62)*
>1.40	0.57 (0.56, 0.59)*	0.39 (0.37, 0.41)*	0.84 (0.80, 0.88)*
Available water capacity (in cm/m)			
<8	1.00	1.00	1.00
8–10	0.72 (0.70, 0.75)*	0.47 (0.44, 0.50)*	0.32 (0.28, 0.36)*
10–12	1.00 (0.96, 1.04)	0.35 (0.33, 0.38)*	0.59 (0.54, 0.65)*
>12	1.05 (1.02, 1.09)*	0.23 (0.21, 0.25)*	0.65 (0.60, 0.72)*
pH in water			
<5.2	1.00	1.00	1.00
5.2–6.7	0.79 (0.77, 0.81)*	1.16 (1.10, 1.21)*	0.77 (0.72, 0.85)*
>6.7	1.28 (1.25, 1.31)*	1.36 (1.29, 1.43)*	0.47 (0.44, 0.49)*
Texture class			
Medium	1.00	1.00	1.00
Coarse	0.56 (0.53, 0.59)*	0.36 (0.33, 0.39)*	0.48 (0.44, 0.54)*
Fine	0.92 (0.91, 0.94)*	0.58 (0.56, 0.60)*	0.69 (0.67, 0.72)*
			Mean (95% BCI)
Sigma^2	–	–	3.79 (3.16, 4.50)
Range (in km)	–	–	356.9 (336.9, 379.0)

* Significant correlation based on 95% CI or 95% BCI.

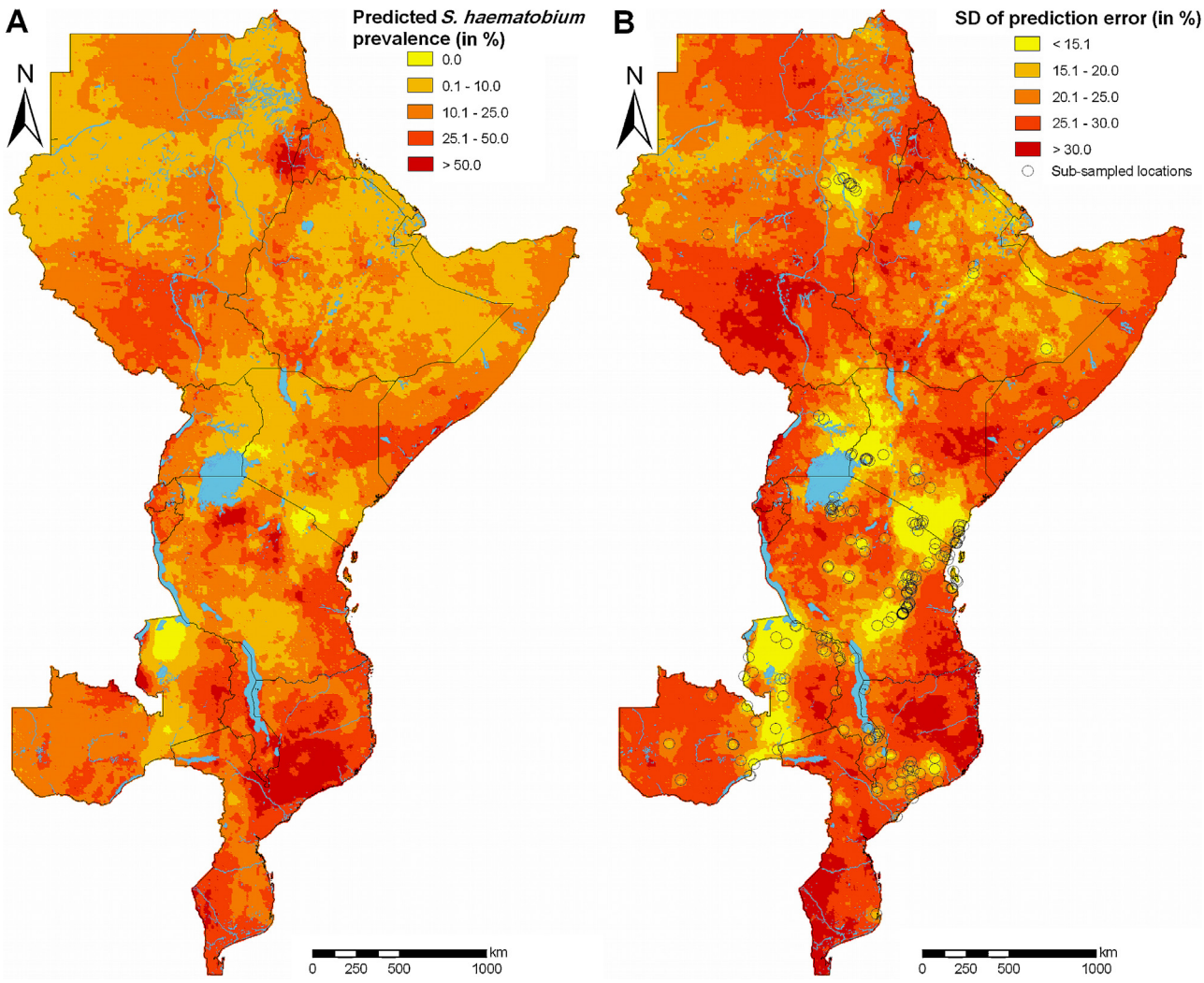

Fig. 3. Predicted median of infection risk for individuals aged ≤20 years for *S. haematobium* during the period of 2000–2009 based on joint Bayesian kriging (A) and standard deviation (SD) of the prediction error with sub-sampled survey locations (B).

factors (e.g. people's movements, behaviour, socio-economic factors), which are unknown and therefore prediction at very high spatial resolution may not be rational.

The GNTD database represents an important output of the EU-funded CONTRAST project. As of early October 2010, the database contained over 4000 survey locations across eastern Africa that have been obtained through a systematic search of published and unpublished sources. Importantly, various remotely sensed parameters were incorporated into our models to evaluate the effect of climate and other environmental factors on *Schistosoma* infection risk. For the first time in large-scale schistosomiasis risk profiling, different soil characteristics were also included, such as pH and available water capacity, which might have an effect on the intermediate host snails, and hence potentially influence disease risk. Another initiative on mapping helminthic infections has been taken by the Global Atlas of Helminth Infections (GAHI; http://www.thiswormyworld.org) project (Brooker et al., 2010). It would be interesting to compare our estimates with the ones from the GAHI project once they become available for schistosomiasis.

Clements et al. (2010) previously presented a *S. mansoni* risk map using Bayesian geostatistical modelling for Burundi, Uganda and parts of Kenya and Tanzania based on geo-referenced school surveys carried out between 1998 and 2007. Their map shows similar schistosomiasis risk patterns than the one presented here, yet discrepancies are evident in the area north of Lake Albert and areas in proximity to the Kyoga, Edward and George lakes. Importantly though, the prevalence estimates in both maps are similar. One decade ago, another risk map using non-spatial logistic regression was published for Tanzania, focussing on *S. haematobium* (Brooker et al., 2001). However, this map does not show the actual level of schistosomiasis risk, but rather probabilities that the predicted risk is above a certain cut-off fixed at 50%. This cut-off has been proposed by the World Health Organization (WHO); areas where >50% of school-aged children are infected warrant yearly preventive chemotherapy to entire communities (WHO, 2002, 2006). However, such maps do not provide detailed information for lower risk areas or the number of infected individuals, which is important for operational and programmatic reasons. Additionally, such maps cannot be used for monitoring and evaluation of interventions. A smaller Bayesian geostatistical risk map covering areas of north-western Tanzania for *S. haematobium* (Clements et al., 2006) revealed similar patterns of risk as predicted in our map. Nonetheless, differences, especially in the estimated prevalence level, can be found in areas further away from Lake Victoria where we predicted higher risk of infections than Clements and colleagues. In the 1950s and 1970s, higher prevalence of *S. haematobium* compared to *S. mansoni* was found in the Lango region of Central Northern Uganda (Schwetz, 1951; Bradley et al., 1967), while recent

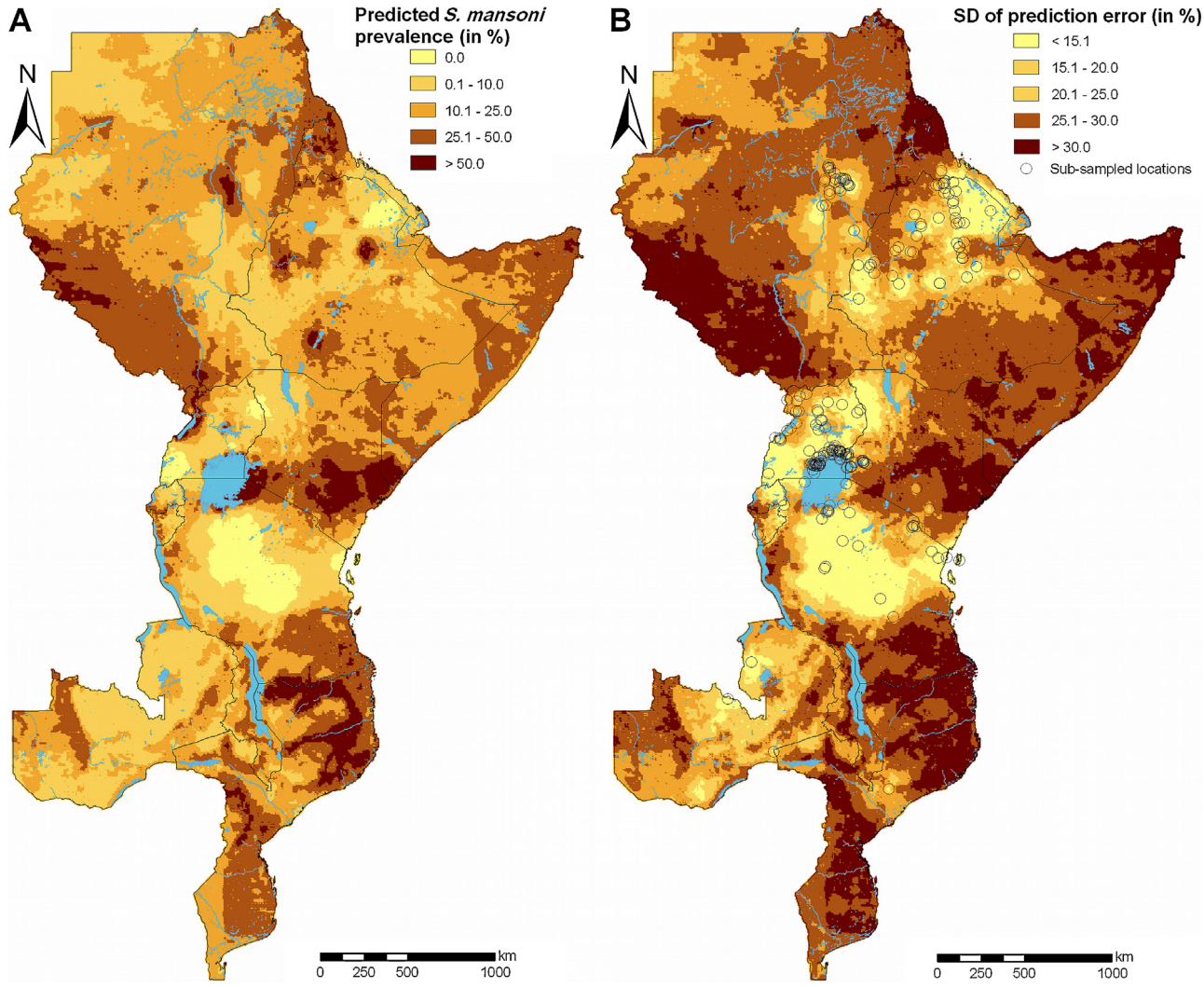

Fig. 4. Predicted median of infection risk for individuals aged ≤20 years for *S. mansoni* during the period of 2000–2009 based on joint Bayesian kriging (A) and standard deviation (SD) of the prediction error with sub-sampled survey locations (B).

investigations in the same area only detect few *S. haematobium* cases. The underlying reasons for this decline remain to be determined. Therefore, we are likely to overestimate the current *S. haematobium* and schistosomiasis risk in this region.

We obtained estimates of the number of infected individuals aged ≤20 years, and for all age groups, on a country basis by overlaying population data adjusted for 2010 on the predicted risk surfaces for the two *Schistosoma* species. These estimates are empirical model-based, while previous country estimates presented by Chitsulo et al. (2000), Steinmann et al. (2006) and Utzinger et al. (2009), are interpolations of limited survey data for a whole country. Chitsulo et al. reported 57.5 million infected individuals in eastern Africa, which is less than half of our combined schistosomiasis prevalence estimate (122 million). We observed at least three-fold more infected individuals in Burundi, Eritrea, Ethiopia, Rwanda and Sudan and similar numbers in Mozambique and Tanzania.

How can these differences be explained? First and foremost, populations have grown. The Chitsulo et al. estimates are calculated for the mid-1990s (estimated population in the 13 countries: 219.4 million) compared to our estimates for the year 2010 (328.4 million). Second, individuals might have moved into areas in close proximity to freshwater bodies or newly established irrigation systems. These areas are likely to be linked to higher *Schistosoma* infection risks (Steinmann et al., 2006), even though no significant effect for the distance to the nearest freshwater body was observed for *S. haematobium*. Third, large-scale preventive chemotherapy programmes (Fenwick et al., 2009; WHO, 2010), improved sanitation (WHO/UNICEF, 2010), water resources development and management (Fenwick, 2006; Steinmann et al., 2006), urban–rural movements and socio-economic development are important underlying determinants of changing schistosomiasis risk patterns. Fourth, discrepancies might be related to interpolations of few data points over large areas without taking into account model-based predictions on the basis of climate, environment and disease data. Fifth, we might also underestimate country-specific schistosomiasis prevalence due to the assumption of independence between the occurrence of *S. haematobium* and *S. mansoni*. Simultaneous infections with both species in areas where the species co-exist might be more frequent (e.g. due to similar and highly behavioural infection pathways) or less frequent (e.g. due to protective factors) than expected by chance.

Model validation at 20% of the original survey locations included in our models showed that we are able to correctly predict more than 70% of the locations when considering 95% BCIs. In general, our predictions are approximately 14–15% away from the observed prevalence with a small tendency for *S. mansoni* to overestimate the risk. We are encouraged by these results due to the complexity of

Table 5
Population-adjusted prevalence of *S. haematobium* and *S. mansoni* in individuals (≤20 years) and in the total population, stratified by country in eastern Africa (predicted for the period 2000–2009) based on 2010 population estimates with 95% Bayesian credible interval (BCI).

Country	*S. haematobium*		*S. mansoni*		Schistosomiasis[a]		Prevalence[b] (%)
	Prevalence (%) Age ≤20 years (95% BCI)	Prevalence (%) Entire population (95% BCI)	Prevalence (%) Age ≤20 years (95% BCI)	Prevalence (%) Entire population (95% BCI)	Prevalence (%) Age ≤20 years (95% BCI)	Prevalence (%) Entire population (95% BCI)	
Burundi	32.2 (5.1, 76.9)	29.9 (4.7, 71.2)	21.0 (6.3, 49.5)	20.2 (6.1, 47.7)	42.3 (11.7, 83.9)	40.3 (11.1, 80.7)	13.3
Djibouti	12.8 (0.9, 69.0)	11.9 (0.8, 63.9)	21.6 (2.0, 77.6)	20.8 (1.9, 74.9)	30.0 (2.7, 87.8)	28.8 (2.6, 86.0)	–
Eritrea	24.0 (6.6, 64.3)	22.2 (6.1, 59.6)	32.4 (11.3, 61.6)	31.2 (10.9, 59.3)	44.1 (16.1, 79.4)	42.5 (15.4, 77.6)	7.2
Ethiopia	19.5 (9.8, 31.0)	18.1 (9.1, 28.7)	22.9 (15.9, 30.9)	22.1 (15.4, 29.8)	33.8 (22.4, 45.0)	32.5 (21.6, 43.4)	7.1
Kenya	16.6 (10.1, 25.5)	15.4 (9.4, 23.7)	35.6 (21.6, 51.6)	34.3 (20.8, 49.8)	42.9 (27.8, 58.6)	41.4 (26.7, 56.8)	30.0
Malawi	37.8 (27.0, 50.5)	35.0 (25.1, 46.8)	27.1 (9.3, 56.6)	26.1 (9.0, 54.6)	50.0 (32.5, 72.2)	47.7 (30.8, 70.1)	42.9
Mozambique	44.2 (32.4, 55.2)	40.9 (30.0, 51.2)	35.8 (21.4, 49.4)	34.5 (20.6, 47.6)	57.7 (42.7, 69.7)	55.6 (40.9, 67.5)	69.8
Rwanda	31.3 (5.8, 75.0)	29.0 (5.3, 69.6)	15.8 (3.8, 41.6)	15.3 (3.6, 40.1)	38.6 (9.2, 81.4)	36.6 (8.8, 78.1)	5.9
Somalia	26.3 (17.5, 37.2)	24.4 (16.3, 34.5)	32.3 (18.6, 50.1)	31.2 (17.9, 48.3)	44.0 (29.4, 59.2)	42.5 (28.2, 57.8)	18.0
Sudan	23.4 (16.2, 33.8)	21.7 (15.1, 31.3)	29.5 (21.8, 39.0)	28.4 (21.1, 37.6)	40.5 (30.7, 52.2)	39.0 (29.5, 50.5)	18.2
Tanzania	24.8 (19.2, 32.0)	23.0 (17.8, 29.7)	20.0 (13.9, 28.7)	19.3 (13.4, 27.6)	38.2 (30.6, 48.0)	36.4 (29.1, 45.8)	51.5
Uganda	17.5 (7.7, 33.5)	16.2 (7.1, 31.1)	13.4 (10.0, 18.0)	12.9 (9.6, 17.3)	26.6 (16.8, 42.2)	25.3 (16.1, 39.9)	32.0
Zambia	26.1 (17.9, 34.3)	24.2 (16.6, 31.8)	16.2 (8.0, 28.0)	15.6 (7.7, 27.0)	34.4 (23.4, 46.1)	32.8 (22.1, 44.2)	26.6

[a] Both *S. haematobium* and *S. mansoni* combined, assuming independence between the two species.
[b] Estimated country prevalence of infected individuals with schistosomiasis over all age groups in 1995, as presented by Chitsulo et al. (2000).

schistosomiasis disease transmission in reality (Stensgaard et al., 2013). Nevertheless, certain modelling assumptions might have influenced model performance. For example, overall prevalence and infection intensity depend on the sensitivity and specificity of the diagnostic technique (Bergquist et al., 2009). However, we assumed that the different diagnostic techniques in our dataset have similar ability to detect a *Schistosoma* infection, which might bias the results. Spatial models accounting for sensitivity and specificity could be incorporated in the models, as demonstrated by Wang et al. (2008). However, due to a large number of missing or incomplete information in our underlying data, assumptions on the diagnostic techniques and the sampling effort would be required, which may introduce considerable bias. Another concern is the amount of zero outcomes (i.e. none of the study participants found to be infected), especially for *S. mansoni* (*S. mansoni*: 30.0%; *S. haematobium*: 14.3%). Zero-inflated models could be implemented instead. Such models modify the likelihood function and add an additional model parameter capturing the over-dispersion arising by the zeros (Vounatsou et al., 2009). Furthermore, our models are assumed to be isotropic stationary, which implies that the spatial random effect is stable throughout the study area (Gosoniu et al., 2011) and that the spatial correlation is the same within the same distance irrespective of direction (Ecker and Gelfand, 2003). This is a potentially inappropriate assumption because dry regions might be less suitable for the disease to spread than humid region and therefore the spatial range could be smaller. Additionally, intermediate host snails spread along rivers and lakeshores, and hence correlation is likely to be attributed to directions.

School-aged children are known to carry the highest risk of *Schistosoma* infection, and hence are the key target group for preventive chemotherapy (WHO, 2002, 2006). However, large amounts of surveys included in the database are either community-based or involved adults only. It follows that these surveys are related to lower risks of infection. Hence, unadjusted combination of all surveys in one model, irrespective of age-group, would result in inaccurate community risk estimates. We incorporated age-alignment factors to merge studies based on the three main age groups (individuals aged ≤20 years, individuals aged >20 years, entire communities) present in our data. These factors are expected to increase model performance compared to models considering only one age group or models without any alignment between the different age groups (Schur et al., 2011b). However, age adjustment could be refined and adopted to different disease transmission settings in order to further enhance model performance.

Temporal trends included in the risk estimation highlighted the differences in schistosomiasis prevalence levels between the 1980s, 1990s and the 2000s. The risk of infection for *S. haematobium* has been lowered during the past two decades, while *S. mansoni* risk increased during the 1980s and 1990s and dropped slightly during the present decade. Major water resources development and management activities might explain the observed increase in *S. mansoni* risk over the past decades. Human-altered habitats has been shown to be an important determinant of the distribution of the major intermediate host snail species at the African continent (Stensgaard et al., 2013). Indeed, there is evidence that urinary schistosomiasis is replaced by intestinal schistosomiasis in face of irrigation schemes and large dams (Abdel-Wahab et al., 1979; Steinmann et al., 2006). This phenomenon is also referred to as "Nile shift", as it has been documented first in the Nile delta of Egypt after the completion of the Aswan dam. On the one hand, several African countries have (re-) established national schistosomiasis control programmes emphasizing preventive chemotherapy to school-aged children (Fenwick et al., 2009; WHO, 2010), which reduced the community prevalence of both *S. haematobium* and *S. mansoni*.

Table 6
Estimated number of infected in individuals (≤20 years) and in the total population, stratified by country in eastern Africa (predicted for the period 2000–2009) based on 2010 population estimates with 95% Bayesian credible interval (BCI).

Country	Age ≤20 years (million)	Entire population (million)	*S. haematobium*		*S. mansoni*		Schistosomiasis[a]		Infected[b] (million)
			Infected (million) Age ≤20 years (95% BCI)	Infected (million) Entire population (95% BCI)	Infected (million) Age ≤20 years (95% BCI)	Infected (million) Entire population (95% BCI)	Infected (million) Age ≤20 years (95% BCI)	Infected (million) Entire population (95% BCI)	
Burundi	5.528	9.445	1.780 (0.281, 4.248)	2.820 (0.445, 6.728)	1.159 (0.347, 2.736)	1.908 (0.571, 4.506)	2.340 (0.645, 4.640)	3.806 (1.046, 7.620)	0.84
Djibouti	0.251	0.512	0.032 (0.002, 0.173)	0.061 (0.004, 0.327)	0.054 (0.005, 0.195)	0.107 (0.010, 0.383)	0.076 (0.007, 0.221)	0.147 (0.013, 0.440)	–
Eritrea	3.006	5.477	0.721 (0.197, 1.932)	1.218 (0.333, 3.262)	0.974 (0.340, 1.850)	1.710 (0.597, 3.249)	1.324 (0.484, 2.387)	2.329 (0.843, 4.252)	0.26
Ethiopia	52.200	89.500	10.165 (5.096, 16.180)	16.157 (8.100, 25.718)	11.946 (8.313, 16.132)	19.746 (13.740, 26.665)	17.656 (11.712, 23.500)	29.095 (19.320, 38.803)	4.00
Kenya	21.900	40.300	3.647 (2.217, 5.606)	6.209 (3.774, 9.543)	7.813 (4.730, 11.326)	13.833 (8.373, 20.051)	9.420 (6.105, 12.867)	16.693 (10.775, 22.899)	6.14
Malawi	8.390	14.400	3.168 (2.267, 4.233)	5.047 (3.612, 6.745)	2.272 (0.782, 4.753)	3.764 (1.295, 7.875)	4.194 (2.726, 6.055)	6.883 (4.435, 10.114)	4.20
Mozambique	11.900	20.200	5.273 (3.867, 6.594)	8.263 (6.060, 10.334)	4.271 (2.556, 5.899)	6.960 (4.166, 9.613)	6.895 (5.100, 8.317)	11.224 (8.253, 13.624)	11.30
Rwanda	5.868	10.700	1.834 (0.337, 4.403)	3.113 (0.573, 7.473)	0.928 (0.221, 2.439)	1.639 (0.390, 4.306)	2.263 (0.541, 4.776)	3.930 (0.943, 8.387)	0.38
Somalia	5.149	9.150	1.354 (0.902, 1.917)	2.230 (1.486, 3.158)	1.664 (0.957, 2.578)	2.851 (1.639, 4.415)	2.266 (1.513, 3.047)	3.890 (2.575, 5.289)	1.71
Sudan	23.400	42.100	5.482 (3.801, 7.900)	9.148 (6.343, 13.183)	6.902 (5.110, 9.119)	11.976 (8.867, 15.825)	9.465 (7.180, 12.218)	16.416 (12.421, 21.260)	4.85
Tanzania	23.600	42.100	5.840 (4.537, 7.544)	9.666 (7.510, 12.487)	4.717 (3.282, 6.759)	8.119 (5.650, 11.634)	8.998 (7.219, 11.309)	15.304 (12.229, 19.273)	15.24
Uganda	21.300	33.600	3.730 (1.634, 7.150)	5.450 (2.388, 10.447)	2.858 (2.131, 3.831)	4.343 (3.238, 5.821)	5.674 (3.589, 8.996)	8.511 (5.402, 13.434)	6.14
Zambia	6.473	10.900	1.688 (1.160, 2.221)	2.635 (1.810, 3.467)	1.051 (0.518, 1.809)	1.706 (0.841, 2.938)	2.229 (1.513, 2.983)	3.578 (2.408, 4.815)	2.39
TOTAL	188.965	328.384	44.714	72.017	46.609	78.662	72.800	121.806	57.45

[a] Both *S. haematobium* and *S. mansoni* combined, assuming independence between the two species.
[b] Estimated country prevalence of infected individuals with schistosomiasis over all age groups in 1995, as presented by Chitsulo et al. (2000).

In comparison to a similar analysis done for West Africa, including Cameroon (Schur et al., 2011a), we implemented further potential covariates in our models such as latitude, longitude, slope, aspect and soil parameters. To our knowledge, we have now implemented soil parameters for the first time in large-scale geostatistical schistosomiasis risk mapping. Importantly, soil parameters were indeed related to the risk of schistosomiasis transmission, and hence improved outcome predictions. While, pH was a predictor for both schistosome species, available water capacity was associated with *S. haematobium*, whereas bulk density, and texture class showed an association with *S. mansoni*. These soil factors directly influence snail habitats and larval survival in the environment (Madsen, 1985a,b; Bavia et al., 1999), and hence are important predictors of schistosomiasis.

5. Conclusions and outlook

Our country-specific estimates on the number of schistosome-infected individuals in eastern Africa revealed considerable differences to previous and widely cited statistics. Our new estimates, together with the *Schistosoma* infection prevalence maps, are useful decision tools for disease control managers to efficiently guide interventions to high-risk areas, to plan the frequency of deworming campaigns, to estimate the required drug supplies at the operational unit of drug deployment (e.g. district) in order to reduce the burden of schistosomiasis, and to monitor progress of interventions to ultimately interrupt transmission. Regions of high model uncertainty need to be studied in greater detail to further validate our results and to deepen our knowledge on the spatial distribution of *Schistosoma* infection. In the future, we plan to further expand this work to obtain Africa-wide prevalence estimates and to study temporal trends. In addition, we will include the geographical distribution of key intermediate host snail species in our spatial models to improve model-based predictions and to study the importance of climatic and other environmental covariates on schistosomiasis risk, while accounting for the presence and absence of intermediate hosts. Finally, we will probe the assumption of independence between *S. haematobium* and *S. mansoni* by jointly modelling both species to enhance our combined risk estimates.

Acknowledgements

We thank the many collaborators who contributed georeferenced schistosomiasis survey data to our GNTD database. NS and EH are grateful for the financial support of the EU-funded CONTRAST project (FP6-STREP-2004-INCO-DEV project no. 032203). This investigation received further financial support from the Swiss National Science Foundation granted to PV (project no. 325200-118379). ASS was supported by a PhD fellowship at the University of Copenhagen, partly funded by DHI Denmark, and thanks the Danish National Research Foundation for its support of the Center for Macroecology, Evolution and Climate.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.actatropica.2011.10.006.

References

Abdel-Wahab, M.F., Strickland, G.T., El-Sahly, A., El-Kady, N., Zakaria, S., Ahmed, L., 1979. Changing pattern of schistosomiasis in Egypt 1935–79. Lancet 314, 242–244.

Banerjee, S., Gelfand, A.E., Carlin, B.P., 2003. Hierarchical Modeling and Analysis for Spatial Data, 1st ed. Chapman and Hall/CRC.

Bavia, M.E., Hale, L.F., Malone, J.B., Braud, D.H., Shane, S.M., 1999. Geographic information systems and the environmental risk of schistosomiasis in Bahia, Brazil. Am. J. Trop. Med. Hyg. 60, 566–572.

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

Bradley, D.J., Sturrock, R.F., Williams, P.N., 1967. The circumstantial epidemiology of *Schistosoma haematobium* in Lango district, Uganda. East Afr. Med. J. 44, 193–204.

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

Brooker, S., Hotez, P.J., Bundy, D.A.P., 2010. The global atlas of helminth infection: mapping the way forward in neglected tropical disease control. PLoS Negl. Trop. Dis. 4, e779.

Chitsulo, L., Engels, D., Montresor, A., Savioli, L., 2000. The global status of schistosomiasis and its control. Acta Trop. 77, 41–51.

Clements, A.C.A., Deville, M., Ndayishimiye, O., Brooker, S., Fenwick, A., 2010. Spatial co-distribution of neglected tropical diseases in the east African great lakes region: revisiting the justification for integrated control. Trop. Med. Int. Health 15, 198–207.

Clements, A.C.A., Garba, A., Sacko, M., Touré, S., Dembelé, R., Landouré, A., Bosque-Oliva, E., Gabrielli, A.F., Fenwick, A., 2008. Mapping the probability of schistosomiasis and associated uncertainty, West Africa. Emerg. Infect. Dis. 14, 1629–1632.

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

Diggle, P.J., Tawn, J.A., Moyeed, R.A., 1998. Model-based geostatistics. Appl. Stat. 47, 299–350.

Ecker, M.D., Gelfand, A.E., 2003. Spatial modeling and prediction under stationary non-geometric range anisotropy. Environ. Ecol. Stat. 10, 165–178.

Fenwick, A., 2006. Waterborne infectious diseases—could they be consigned to history? Science 313, 1077–1081.

Fenwick, A., Webster, J.P., Bosque-Oliva, E., Blair, L., Fleming, F.M., Zhang, Y., Garba, A., Stothard, J.R., Gabrielli, A.F., Clements, A.C.A., Kabatereine, N.B., Toure, S., Dembele, R., Nyandindi, U., Mwansa, J., Koukounari, A., 2009. The Schistosomiasis Control Initiative (SCI): rationale, development and implementation from 2002–2008. Parasitology 136, 1719–1730.

George, E.I., McCulloch, R.E., 1993. Variable selection via Gibbs sampling. J. Am. Stat. Assoc. 88, 881–889.

Gosoniu, L., Vounatsou, P., Sogoba, N., Smith, T., 2006. Bayesian modelling of geostatistical malaria risk data. Geospat. Health 1, 127–139.

Gosoniu, L., Vounatsou, P., Sogoba, N., Maire, N., Smith, T., 2011. Mapping malaria risk in West Africa using a Bayesian nonparametric non-stationary model. Comput. Stat. Data Anal. 53, 3358–3371.

Hotez, P.J., Molyneux, D.H., Fenwick, A., Kumaresan, J., Ehrlich Sachs, S., Sachs, J.D., Savioli, L., 2007. Control of neglected tropical diseases. N. Engl. J. Med 357, 1018–1027.

Huete, A., Didan, K., Miura, T., Rodriguez, E.P., Gao, X., Ferreira, L.G., 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. Remote Sens. Environ. 83, 195–213.

Hürlimann, E., Schur, N., Boutsika, K., Stensgaard, A.S., Laserna de Himpsl, M., Ziegelbauer, K., Laizer, N., Camenzind, L., Di Pasquale, A., Ekpo, U.F., Simoonga, C., Mushinge, G., Saarnak, C.F.L., Utzinger, J., Kristensen, T.K., Vounatsou, P. Toward an open-access global database for mapping, control, and surveillance of neglected tropical diseases. PLoS Negl. Trop. Dis. 5, e1404.

Madsen, H., 1985a. Ecology and Control of African Freshwater Pulmonate Snails. Part I: Life Cycle and Methodology. Danish Bilharziasis Laboratory, Copenhagen.

Madsen, H., 1985b. Ecology and Control of African Freshwater Pulmonate Snails. Part 2: Basic Principles in Ecology of Freshwater Snails. Danish Bilharziasis Laboratory, Copenhagen.

Pinot de Moira, A., Fulford, A.J.C., Kabatereine, N.B., Kazibwe, F., Ouma, J.H., Dunne, D.W., Booth, M., 2007. Microgeographical and tribal variations in water contact and *Schistosoma mansoni* exposure within a Ugandan fishing community. Trop. Med. Int. Health 12, 724–735.

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

Sanderson, E.W., Jaiteh, M., Levy, M.A., Redford, K.H., Wannebo, A.V., Woolmer, G., 2002. The human footprint and the last of the wild. Bioscience 52, 891–904.

Schur, N., Hürlimann, E., Garba, A., Traoré, M.S., Ndir, O., Ratard, R.C., Tchuem Tchuenté, L.A., Kristensen, T.K., Utzinger, J., Vounatsou, P., 2011a. Geostatistical model-based estimates of schistosomiasis prevalence among individuals ≤20 years in West Africa. PLoS Negl. Trop. Dis. 5, e1194.

Schur, N., Utzinger, J., Vounatsou, P., 2011b. Modelling age-heterogeneous *Schistosoma haematobium* and *S. mansoni* survey data via alignment factors. Parasit. Vectors 4, 142.

Schwetz, J., 1951. Communications on vesical bilharzia in the Lango district of Uganda. Trans. R. Soc. Trop. Med. Hyg. 44, 501–513.

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

Stensgaard, A., Utzinger, J., Vounatsou, P., Hürlimann, E., Schur, N., Saarnak, C.F.L., Mushinge, G., Simoonga, C., Kabatereine, N.B., Tchuem Tchuenté, L., Rahbek, C., Kristensen, T.K., 2013. Large-scale determinants of intestinal schistosomiasis and intermediate host snail distribution across Africa: does climate matter? Acta Trop. 128, 378–390.

Utroska, J.A., Chen, M., Dixon, H., Yoon, S., Helling-Borda, M., Hogerzeil, H.V., Mott, K.E., 1989. An Estimate of the Global Needs for Praziquantel Within Schistosomiasis Control Programmes. World Health Organization, Geneva.

Utzinger, J., Raso, G., Brooker, S., de Savigny, D., Tanner, M., Ørnbjerg, N., Singer, B.H., N'Goran, E.K., 2009. Schistosomiasis and neglected tropical diseases: towards integrated and sustainable control and a word of caution. Parasitology 136, 1859–1874.

Vounatsou, P., Raso, G., Tanner, M., N'Goran, E.K., Utzinger, J., 2009. Bayesian geostatistical modelling for mapping schistosomiasis transmission. Parasitology 136, 1695–1705.

Wang, X.H., Zhou, X.N., Vounatsou, P., Chen, Z., Utzinger, J., Yang, K., Steinmann, P., Wu, X.H., 2008. Bayesian spatio-temporal modeling of *Schistosoma japonicum* prevalence data in the absence of a diagnostic 'gold' standard. PLoS Negl. Trop. Dis. 2, e250.

WHO, 2002. Prevention and control of schistosomiasis and soil-transmitted helminthiasis: report of the WHO expert committee. WHO Tech. Rep. Ser. 912, 1–57.

WHO, 2006. Preventive Chemotherapy in Human Helminthiasis: Coordinated use of Anthelminthic Drugs in Control Interventions: A Manual for Health Professionals and Programme Managers. World Health Organization, Geneva, pp. 1–62.

WHO, 2010. Schistosomiasis. Wkly. Epidemiol. Rec. 85, 158–164.

WHO/UNICEF, 2010. Progress on Sanitation and Drinking-Water—2010 Update, WHO/UNICEF Joint Monitoring Programme for Water Supply and Sanitation. WHO, UNICEF.