Application of geo-spatial technology in schistosomiasis modelling in Africa: a review

Tawanda Manyangadze,1 Moses John Chimbari,1 Michael Gebreslassie,2 Samson Mukaratirwa3
1School of Nursing and Public Health, Department of Public Health Medicine, University of KwaZulu-Natal, Durban; 2School of Agriculture, Earth and Environmental Sciences, University of KwaZulu-Natal, Durban; 3School of Life Sciences, University of KwaZulu-Natal, Durban, South Africa

Abstract

Schistosomiasis continues to impact socio-economic development negatively in sub-Saharan Africa. The advent of spatial technologies, including geographic information systems (GIS), Earth observation (EO) and global positioning systems (GPS) assist modelling efforts. However, there is increasing concern regarding the accuracy and precision of the current spatial models. This paper reviews the literature regarding the progress and challenges in the development and utilization of spatial technology with special reference to predictive models for schistosomiasis in Africa. Peer-reviewed papers identified through a PubMed search using the following keywords: geo-spatial analysis OR remote sensing OR modelling OR earth observation OR geographic information systems OR prediction OR mapping AND schistosomiasis AND Africa were used. Statistical uncertainty, low spatial and temporal resolution satellite data and poor validation were identified as some of the factors that compromise the precision and accuracy of the existing predictive models. The need for high spatial resolution of remote sensing data in conjunction with ancillary data viz. ground-measured climatic and environmental information, local presence/absence intermediate host snail surveys as well as prevalence and intensity of human infection for model calibration and validation are discussed. The importance of a multidisciplinary approach in developing robust, spatial data capturing, modelling techniques and products applicable in epidemiology is highlighted.

Introduction

Schistosomiasis, caused by Schistosoma haematobium and S. mansoni, is a disease that mainly affects under-resourced communities and is often not prioritized in national budgets in sub-Saharan Africa (WHO, 2014). Approximately 600 million people are at risk worldwide and over 200 million people are infected (Chitsulo et al., 2000; King, 2009). The Schistosomiasis Control Initiative (SCI) works with national governments in sub-Saharan Africa to control and eliminate schistosomiasis. A recent national schistosomiasis survey contributed to the development of the schistosomiasis and soil transmitted helminths (STH) National Control Program in Zimbabwe that involves mass drug administration (MDA) (Midzi et al., 2014). Despite the efforts to control the disease in sub-Saharan Africa, 80-95% of the global, total number of schistosomiasis infected individuals still live in Africa (WHO, 2002; Steinmann et al., 2006; Utzinger et al., 2009; Hurlimann et al., 2011). It is estimated that about 200 000 deaths per year are caused by schistosomiasis in sub-Saharan Africa (WHO, 2014). The distribution of schistosomiasis is reflected by the geographical distribution of the intermediate host snail species (Despommier et al., 1994), which is a well-known phenomenon but difficult to predict and monitor. It is envisaged that knowing the drivers of the current distribution of schistosomiasis could improve our understanding on how climatic and environmental changes may influence the distribution of schistosomiasis in the future. This is critical to effectively and efficiently manage surveillance, control and prevention of the disease (Stensgaard et al., 2005). King et al. (2006) emphasize that the next generation of schistosomiasis control will be optimized using new monitoring tools and effective transmission containment. Enhanced by the advancements in statistical ecological modelling, spatial technologies which provide spatial data and tools for spatial analysis and predictive modelling have opened a new way for developing such monitoring tools. The development of geo-spatial technology such as geographical information systems (GIS) and global positioning system (GPS) have facilitated the integration of Earth observation (EO) driven environmental parameters with health data for the development of disease surveillance and control models (Beck et al., 2000). GIS applications in public health include the estimation of spatial variation of disease, determination of risk factors of disease, and improved deliv-
er of health services (Tanser et al., 2003). This review looks at the progress and challenges in the application of spatial technologies in mapping and modelling schistosomiasis in Africa. It was inspired by the works of Brooker et al. (2002a) and Brooker (2007) and Simoonga et al. (2009), who published detailed reviews of the past developments and use of GIS and remote sensing in schistosomiasis mapping and modelling in Africa. The potential future research priorities with emphasis on application of spatial technology in schistosomiasis modelling at local levels in Africa are discussed.

Selection criteria for the literature search undertaken

This review is based on a systematic search for relevant literature in the PubMed electronic search engine (http://www.ncbi.nlm.nih.gov/pubmed) following the method used by Simoonga et al. (2009). This search considered the studies using geo-spatial technologies for schistosomiasis prediction, modelling and mapping in Africa based on the following combination of terms and Boolean operations: geo-spatial analysis OR remote sensing OR earth observation OR geographic information systems OR prediction OR mapping OR modelling AND schistosomiasis AND Africa. Any literature which did not satisfy these criteria was excluded. The snowballing technique was used to obtain more literature based on the bibliography or reference list of previous reviews obtained by the search strategy described above. Literature on schistosomiasis without the element of geo-spatial analysis were used to strengthen the discussion and understanding why geo-spatial technology has been and/or can be used to understand the schistosomiasis transmission. The relevant literature was used to determine the levels of appreciation and use of geospatial technology in schistosomiasis modelling and/or mapping by extending the yearly publication graph by Simoonga et al. (2009). These authors accessed the PubMed database in early 2009 and gave a snapshot of the number of publications on GIS and remote sensing applications used with reference to schistosomiasis in Africa between 1996 and 2008. In this review, we extended the snapshot up to 2013 using the methodology mentioned above.

Search results

The search for literature conducted for the period 2009 - 2013 revealed 59 hits of which 36 were considered relevant. The relevant publications (36) were further categorized by year of publication and combined with Simoonga et al. (2009) results for the period 1996-2008 as presented in Figure 1. The annual number of publications on remote sensing and GIS with application to schistosomiasis in Africa has generally increased over the years as shown in Figure 1. This indicates the increased appreciation and usefulness of geospatial technology for schistosomiasis control and management through mapping, modelling or prediction. The key publications from 2009 to 2013 and 2 from 2008 which were not captured by Simoonga et al. (2009) as well as one for 2014 were categorized into the following three groups: modelling intermediate snail hosts (Table 1), modelling schistosomiasis (Table 2) and those modelling co-infection or co-endemicity of schistosomiasis and STHs (Table 3). The subsequent sections give a detailed discussion based on these publications as well as those mentioned elsewhere for example by Simoonga et al. (2009).

<table>
<thead>
<tr>
<th>Objective</th>
<th>Scale/Study area</th>
<th>Methodology/subject</th>
<th>Published Information and/or data</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Effect of climate change on the distribution of intermediate host snails</td>
<td>National/Zimbabwe</td>
<td>Maxent/species distribution</td>
<td>Temperature, DEM, NDI, PET, wetlands, soil pH, snail survey data</td>
<td>Pedersen et al. (2014)</td>
</tr>
<tr>
<td>Large-scale environmental and climate determinants of the distribution</td>
<td>Continental/Africa</td>
<td>Bayesian logistic regression/spatial analysis</td>
<td>Temperature, rainfall, anthropogenic biomes, percentage tree cover, distance to closest water bodies, soil pH</td>
<td>Stensgaard et al. (2013)</td>
</tr>
<tr>
<td>Analysis of variables associated with the distribution of snails;</td>
<td>Regional/countries bordering Lake Victoria</td>
<td>Multivariate logistical regression/spatial analysis</td>
<td>Rainfall, habitat types, water depth, turbidity, water chemistry, malacological survey data, snail infection rates and substrate type</td>
<td>Standley et al. (2012)</td>
</tr>
</tbody>
</table>
Progress in geo-spatial technology application in schistosomiasis modelling

Spatial technologies have provided an invaluable analytical tool to better understand the determinants and distribution of schistosome infections in Africa (Figure 1 and Tables 1-3). It is well known that climate and environmental factors determine the distribution of schistosomiasis (Appleton, 1978; Brown, 1994; Brooker, 2002) hence it is restricted in space and time by environmental factors (Rollinson et al., 2001; Malone, 2005). In this regard, spatial technologies are useful in understanding the distribution of parasites and their hosts as depicted in the maps showing the interplay with spatial and temporal features of the environment. The general objective of the models is to link the variables related to schistosomiasis transmission or snails with spatial parameters (Simoonga et al., 2009). Most studies use spatial modelling techniques such as logistic regression (Brooker et al., 2001, 2002b), Maxent (Stensgaard et al., 2013; Pedersen et al., 2014), genetic algorithm for rule-set prediction (GARP) (Stensgaard et al., 2006), generalized linear models (GLMs) and generalised additive models (GAMs) (Pfukenyi et al., 2006). There are also non-regression models such as the Bayesian geostatistical approach for modelling intermediate host snails distribution and prevalence of schistosomiasis (Raso et al., 2005; Vaunatsou et al., 2009; Schur et al., 2013). The use of GIS and remote sensing in these models has contributed towards optimized schistosomiasis control efforts at different spatial scales through identifying vulnerable populations for mass treatment (Brooker, 2007) and permitting more rational allocation of resources for cost-effective control (Beck et al., 1997, 2000). However, there are limitations with regards the applicability and robustness of these models that compromise their effectiveness in promoting community public health especially at local levels.

Geo-spatial technology in schistosomiasis modelling: pros and cons

Earth observation has been providing spatial data (and will continue to do so) for developing GIS models for predicting and mapping the risk of schistosomiasis, mostly in inaccessible regions of Africa and in scenarios involving environmental or climate change. However, there are still challenges which warrant further research and refinements/improvements (Herbreteau et al., 2007; Simoonga et al., 2009). In the early stages of the development of GIS and remote sensing technology, the main challenge was the possible resistance to the uptake of EO technology mainly attributed to costs of image processing equipment, expertise and subsequent ground validation (Hay, 1997). To date, it might be true that remote sensing has not become the wonder tool as it was expected to be, to echo Herbreteau et al. (2007), mainly because of limited capacity in processing and use of remote sensing data especially in Africa. EO data requires processing and understanding of the purpose for which it is intended without which the output may be as meaningless as the raw data. Therefore the existing schistosomiasis predictive models are weakened by several factors, including statistical uncertainty in variable selection criteria and methods used, low spatial resolution, failure to utilize the temporal aspect of EO data for spatio-temporal prediction of schistosomiasis, limited application of the developed models in different areas as well as uncertainty and lack of vigorous validation as discussed below.

Unjustified variable selection criteria

The role of non-climatic factors such as topography, distance to water and soil types have been considered in large-scale studies (Tables 1-3). However they have not been given considerable attention compared to climatic factors as determinants of the spatial distribution of schistosomiasis particularly at the local level. These factors could easily be mapped using GIS to determine their influence on schistosomiasis transmission. For example, the distance to water determines

Figure 1. Number of publications pertaining to remote sensing and geographic information system with application to schistosomiasis in Africa from 1996 to 2008 (Simoonga et al., 2009) and 2009 to 2013 (this review).

Figure 2. Theoretical framework for a schistosomiasis predictive modelling.
<table>
<thead>
<tr>
<th>Objective</th>
<th>Scale/study area</th>
<th>Methodology/subject</th>
<th>Published information and/or data</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Generation of point prevalence maps</td>
<td>Country level/Nigeria</td>
<td>Bivariate logistic regressions, Bayesian geostatistical regression</td>
<td>Annual precipitation and mean temperature, soil acidity, georeferenced prevalence data on <em>S. haematobium</em> from journals and reports</td>
<td>Ekpo et al. (2013)</td>
</tr>
<tr>
<td>Overview of major statistical challenges encountered in mapping historical schistosomiasis</td>
<td>Regional/East Africa</td>
<td>Bayesian geostatistical variable selection (Gibbs selection and parameter-expanded normal mixture of inverse-gamma), step-wise logistic regression</td>
<td>Precipitation, DEM temperature, NDVI, land cover, soil moisture and acidity, distance to freshwater bodies</td>
<td>Chammartin et al. (2013)</td>
</tr>
<tr>
<td>Smoothed, empirical independent prevalence maps for <em>S. mansoni</em> and <em>S. haematobium</em></td>
<td>Coast line/Kenya</td>
<td>MCMC simulation, Gibbs selection, bivariate logistic regression</td>
<td>DEM, aspect, precipitation, soil parameters, water bodies, HII, population counts and densities, georeferenced disease survey data</td>
<td>Schur et al. (2013)</td>
</tr>
<tr>
<td>Prevalence rates in schools from available, multiple data sources</td>
<td>Local/Uganda</td>
<td>Bayesian bivariate spatial modelling</td>
<td>LST, precipitation, altitude, land cover, EVI, distance to the nearest water body, questionnaire and urine filtration data</td>
<td>Sturrock et al. (2013)</td>
</tr>
<tr>
<td>Infection and water-contact and among mothers and preschool children</td>
<td>Regional/East Africa</td>
<td>GPS data logging, kernel densities for visualisation, univariate regression models</td>
<td>GPS data, visual satellite data, such as Landsat imagery</td>
<td>Seto et al. (2012)</td>
</tr>
<tr>
<td>Mapping distribution of schistosomiasis with regard to age and heterogeneity</td>
<td>Regional/West Africa</td>
<td>Bayesian geostatistical model with age- and country-specific alignment</td>
<td>DEM, NDVI, LST, land cover, precipitation, historical survey data</td>
<td>Schur et al. (2011b)</td>
</tr>
<tr>
<td>Production of reliable prevalence maps</td>
<td>Regional/Ethiopia</td>
<td>Bayesian bivariate logistic regression, multi-variate Gaussian spatial modelling and MCMC simulation</td>
<td>DEM, NDVI, LST, land cover, precipitation, freshwater bodies, population density, disease data and ecology</td>
<td>Schur et al. (2011c)</td>
</tr>
<tr>
<td>Rainfall and temperature datasets for assessing <em>S. mansoni</em> transmission in an irrigated agricultural area</td>
<td>Regional/Burkina Faso, Mali and Niger</td>
<td>Spearman rank correlation</td>
<td>Precipitation and LST, monthly air temperature and monthly total precipitation (from local weather station), parasitological data (from local hospital)</td>
<td>Xue et al. (2011)</td>
</tr>
<tr>
<td>Prediction of spatial variation of people infected with schistosomiasis</td>
<td>Regional, provincial/Country</td>
<td>Bayesian multinomial regression</td>
<td>NDVI, (as proxy for rainfall), LST, field survey data, distance to perennial inland water body, population density</td>
<td>Clements et al. (2009)</td>
</tr>
<tr>
<td>Study if relaxation of assumption of stationarity and negative binomial mislead because of excess people without infection</td>
<td>Ogun State/Nigeria</td>
<td>Bayesian geo-statistical ZI regression model</td>
<td>DEM, NDVI, LST and rainfall, household characteristics, such as age, sex, asset ownerships, parasitology data, GPS coordinates, rivers</td>
<td>Vounatsou et al. (2009)</td>
</tr>
<tr>
<td>Quantification of infection risk by predictive maps of the probability of infection</td>
<td>Lusaka Province/Zambia</td>
<td>Binary logistic regression models</td>
<td>Rainfall vegetation and temperature, latitude, land cover, soil types and infection data (from morbidity questionnaire)</td>
<td>Ekpo et al. (2008)</td>
</tr>
<tr>
<td>Epidemiological studies of schistosomiasis to enhance spatial targeting of control</td>
<td>Three Bayesian logistic regression models (ordinary, with random effects, and with assumed spatially correlated random effects via MCMC simulation)</td>
<td>DEM, NDM and temperature, prevalence data, individual-level demographic covariates (age and sex), georeferenced potential transmission sites near primary schools</td>
<td>Simoonga et al. (2008)</td>
<td></td>
</tr>
</tbody>
</table>

**Notes:**
- DEM, digital elevation model (altitude);
- NDVI, normalized difference vegetation index;
- MCMC, Markov Chain Monte Carlo;
- HII, human influence index;
- LST, land surface temperature;
- EVI, enhanced vegetation index;
- ZI, zero-inflated;
- GPS, global positioning system.
Table 3. Studies on modelling of schistosomiasis co-infection and/or co-endemicity in Africa.

<table>
<thead>
<tr>
<th>Objective</th>
<th>Scale/study area</th>
<th>Methodology/subject</th>
<th>Published information and/or data</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mapping of schistosomiasis and STH</td>
<td>Regional/Southwest Cameroon</td>
<td>Simple GIS operations for mapping and visualization</td>
<td>Parasitological data</td>
<td>Tchuenté et al. (2013)</td>
</tr>
<tr>
<td>Determination of spatial patterns of co-endemicity of <em>S. mansoni</em> and STHs</td>
<td>Regional/Great lakes East Africa</td>
<td>Bayesian multivariable geostatistics, fixed effects logistic regression models for variable selection</td>
<td>NDVI, LST, (as proxy for rainfall), field survey data, distance to perennial inland water bodies, population density</td>
<td>Clements et al. (2010)</td>
</tr>
<tr>
<td>Mapping of schistosomiasis and STH</td>
<td>Regional/Cameroon</td>
<td>Simple GIS operations for mapping and visualization</td>
<td>Parasitological data</td>
<td>Tchuenté et al. (2012)</td>
</tr>
<tr>
<td>Rapid mapping of schistosomiasis and STH</td>
<td>Regional/South Sudan</td>
<td>GIS mapping</td>
<td>Prevalence data</td>
<td>Finn et al. (2012)</td>
</tr>
<tr>
<td>Co-endemicity of <em>S. mansoni</em>, <em>S. haematobium</em> and STH</td>
<td>Country/Sierra Leone</td>
<td>Binomial logistic regression and probabilistic approach</td>
<td>NDVI, LST, DEM, perennial inland water bodies, population density, prevalence data</td>
<td>Hodges et al. (2012)</td>
</tr>
<tr>
<td>Nationwide, integrated mapping of NTDs</td>
<td>Country/Togo</td>
<td>GIS mapping</td>
<td>Prevalence data</td>
<td>Dorkenoo et al. (2012)</td>
</tr>
<tr>
<td>Mapping of schistosomiasis and STH prevalence</td>
<td>Country/Sierra Leone</td>
<td>GIS mapping</td>
<td>Prevalence data</td>
<td>Hodges et al. (2012)</td>
</tr>
<tr>
<td>Prevalence and distribution of schistosomiasis and STH</td>
<td>Local/Kenya</td>
<td>GIS mapping and spatial analysis</td>
<td>Prevalence data, distance to water bodies</td>
<td>Odiere et al. (2011)</td>
</tr>
<tr>
<td>Mapping of the risk for co-infection of schistosomiasis and hookworm</td>
<td>Regional/Country</td>
<td>Bayesian geostatistical shared</td>
<td>NDVI, LST, DEM, Land cover, precipitation, historical survey data</td>
<td>Schur et al. (2011a)</td>
</tr>
<tr>
<td>Mapping of co-infection of <em>S. haematobium</em> and hookworm to assist planning of national parasitic disease control</td>
<td>Country/Ghana</td>
<td>Bayesian geostatistical methods: multi-nomial regression and zero-inflated Poison regression models</td>
<td>NDVI, LST, distance to perennial inland water bodies, parasitological data</td>
<td>Magalhães et al. (2011)</td>
</tr>
<tr>
<td>To perform a national mapping of schistosomiasis and STH spatial helminth epidemiology of mono- and co-infections with the aim to predict infection risk</td>
<td>Country/Sierra Leone</td>
<td>Bayesian based geostatistical spatial prediction model</td>
<td>NDVI, LST, DEM, location of nearest perennial water bodies, prevalence data, including sex and age, population density</td>
<td>Koroma et al. (2010)</td>
</tr>
<tr>
<td></td>
<td>Regional/East Africa</td>
<td>Bayesian spatial multi-nomial logistic regression model</td>
<td>NDVI, LST, DEM, distance to large, perennial water bodies, data from standardised school survey of <em>S. mansoni</em> and hookworm, degree of urbanisation (urban, peri-urban, rural and extreme rural)</td>
<td>Brooker and Clements (2009)</td>
</tr>
</tbody>
</table>

STH, soil-transmitted helminths; NDVI, normalized difference vegetation index; LST, land surface temperature; GIS, geographic information system; DEM, digital elevation model (altitude); NTDs, neglected diseases.
the human water-contact behaviour, which has a strong influence on
the prevalence and intensity of schistosomiasis (Stensgaard et al.,
2013; Chimbari et al., 2003). Furthermore, studies have shown that
Biomphalaria pfeifferi can tolerate a maximum flow speed of up to 0.3
m/s (Appleton, 1978; Kloos et al., 2001) and soil types determine the
water-soluble calcium (Ca) and pH levels. Together, these factors have
an effect on the presence and density of snails that ultimately decides
the risk for infection. Further, Raso et al. (2005) observed that distance
to permanent rivers was significantly associated with S. mansoni infec-
tion in the non-spatial logistic regression but showed no significant
association in spatially explicit models. Saathof et al. (2002) also con-
cluded that most of the environmental factors (slope, distance to water
and vegetation cover) inadequately explain the spatial pattern of schis-
tosomiasis infection at the sub-district level except with regard to alti-
tude. Although this is probably true, this assertion still requires further
investigation. Moreover, rainfall and temperature were not used in this
particular study as the necessary data for that spatial resolution were
not available and it was anticipated that the variation within such a
small area would have been small (Moodley et al., 2003). Rainfall and
normalized difference vegetation index (NDVI) have been widely used
as proxies for water availability. In their study, Standley et al. (2012)
noted that water availability may be insignificant in the context of per-
manent large water bodies. However, the significance of these environ-
mental factors may still need to be investigated as they may be proxies
for water availability in temporal and spatial modelling of schistosomi-
asis at the local level.

In their large review, Simoonga et al. (2009) noted that a few
researchers have attempted to relate disease prevalence and socio-eco-
nomic profiles of the local population using GIS applications. Such
analyses may be linked to intermediate host snails distribution datasets explained by EO environmental and climatic factors.
Schistosomiasis is heterogeneous, so there is need to undertake
localized studies to establish exposure risk factors and link water con-
tact patterns with malacological surveys (Simoonga et al., 2008). Schur
et al. (2013) used Bayesian modelling and they believe that their model
could have performed better if they had included intermediate host
snail data. Sturrock et al. (2013) highlighted the need to consider loca-
tions of transmission sites and not just distance to water bodies (a
river in their case) to improve the performance of their prediction
model. Simoonga et al. (2009) reported that socio-economic factors,
such as availability of sanitary facilities and safe water supply may bet-
er explain different levels of transmission at local-scale than for
instance, poverty quintiles as seen in a the micro-level study in Côte
d’Ivoire by Raso et al. (2005). Chimbari et al. (2003) attributed differen-
ces in schistosomiasis distribution and intensity between Lake
Kariba (Zimbabwe) and Siavonga (Zambia) (10 km apart) to different
sanitation (better in Lake Kariba) and access to water from the lake
(easy access for Siavonga compared to Kariba). Seto et al. (2012) used
wearable GPS data-loggers for mapping and assessing the exposure of
women and children to risk factors such as access to water. This, how-
ever may have influenced behaviour and affecting the performance of
the prediction model. This opens avenues for further research as the
inclusion of these factors may improve the performance of the model.

**Statistical uncertainty**

The use of simple threshold analysis (Malone et al., 2001) or logistic
regression (Brooker et al., 2001, 2002a; Ekpo et al., 2008) modelling to
predict infection risk are both limited by the inability of these methods to con-
sider spatial correlation of infection and environmental variables. This
leads to underestimation of standard errors of covariate coefficients result-
ing in erroneous inference and justifies the need for assessment of uncer-
tainties inherent in data and modelling techniques (Brooker, 2007).
Variable selection in predictive modelling is a major challenge due to ana-
tycal problems caused by over-fitting, confounding and non-indepen-
dence of data (Craig et al., 2007). Schistosomiasis transmission predictive
models mostly derived from regression models are conservative, because
they use only a few (two or three) climate variables to model the disease.
While this is simple to understand as required in most models, it could be
more reasonable and realistic to include more factors such as the effects
of sanitation, water-related activities (Simoonga et al., 2009; Stensgaard
et al., 2013), snail presence and absence data (Moodley et al., 2003; Schur
et al., 2013). The mapping of infected snails could also help clarify the com-
plex interaction between snail, parasites and the environmental factors
that are usually used to predict the distributions (Simoonga et al., 2009).
However, the approach requires systematic and repeatable staged variable
selection procedures, including spatial analysis to achieve a parsimonious
model with the desired level of internal and external validity (Craig et al.,
2007). This could be resolved by using the machine-learning algorithms
such as Maxent (Stensgaard et al., 2013; Pedersen et al., 2014) and GARP
(Stensgaard et al., 2006), which can handle larger number of variables
(continuous and categorical) than regression models. Bayesian geostatis-
tical approach has been widely used in schistosomiasis modelling over the
seven latest years (Raso et al., 2005; Clements et al., 2006; Vaunatsou et
al., 2009; Schur et al., 2011a, 2011b, 2011c, 2013). This approach is regarded
as a flexible and robust (Brooker, 2007) that takes into account spatial vari-
ability of epidemiological and environmental data as it uses the semivari-
ogram of the spatial process (Chiles and Definier, 1999). However, only a
few aspects of geostatistical methods have been explored in schistosomia-
sis modelling. For example isotropy is more often used than anisotropy
(Simoonga et al., 2008; Vaunatsou et al., 2009). Isotropy is independent of
location and direction and assumes that spatial correlation is only a func-
tion of distance (Gosoniu et al., 2009). This is contrary to anisotropy, which
assumes that spatial correlation is a function of distance in relation to both
location and direction. It is likely that there is high correlation of intensity
and prevalence of schistosomiasis towards transmission sites
(Vaunatsou et al., 2009) and linked to the main flow direction of the river
(Beck-Wörner et al., 2007). Schur et al. (2013) applied anisotropy at the
regional scale in East Africa and recommend the application of this method
at the local scale. These authors emphasise that ignoring anisotropy could
influence the strength of association and thus also the spatial range-
parameter estimates, which might reduce model ability, especially in the
presence of strong anisotropy. Most of the studies on geospatial modeling
and/or mapping of schistosomiasis at the local scale do not consider co-
demicity or co-infection of S. mansoni and S. haematobium (Gryseels,
1996; Raso et al., 2005; Brooker and Clements, 2009). Advanced geostatis-
tical capabilities, such as shared component modelling (Schur, 2011a) and
Bayesian geostatistical multinomial regression modelling (Magalhães et
al., 2011), which can model co-infections, have not been fully exploited
with reagrd to schistosomiasis co-infections and/or co-endemicity model-
ing. These techniques allow simultaneous modelling of co-endemicity or
co-infection of the two common schistosome species in Africa (S. mansoni
and S. haematobium) and investigate the independence between the two
and how they respond to different climatic and environmental factors.
However, the results are affected by inconsistency in data that can be due
to different samples, different sample sizes and time of sampling, which
may compromise the quality of model outputs (Hodges et al., 2012). This
type of inconsistencies can be handled well using geostatistical methods
such as shared component modelling as compared to Bayesian geostatis-
tical multinomial regression modeling which can only use data from sur-

[page 104]
veys screening for multiple infections simultaneously (Schur et al., 2011a). Generally, there is paucity of information on the geographical distribution of both species within co-endemic regions and knowledge of micro-geographical variation of single and mixed schistosoma infections and morbidity. Thus, modelling the two parasites could provide important insights into the drivers of infection and disease, which could help tailor schistosomiasis control and elimination efforts (Meurs et al., 2013). It is believed that chronic infections cause adverse morbidity-related effects that are exacerbated by infections by multiple species and high parasite loads (Pullan and Brooker, 2008). Modelling schistosome co-infections could make it easier to determine whether these infections impact species-specific morbidity compared to single species infection (Gouvras et al., 2012). To date, the effects of schistosome co-infection on morbidity are not clear (Sang et al., 2014; Meurs et al., 2012) and modelling these co-infections could help targeting specific micro-geographical locations for further research on design of superior intervention strategies.

**Low spatial resolution — remote sensing data**

The increasing use of spatial low-resolution imagery (500 m or less) has provided an opportunity to explore the distribution of schistosomiasis at broad scales (country, regional and continental) (Brooker and Michael, 2000; Brooker, 2002; Clennon et al., 2006). Still, however, ineffective schistosomiasis prediction remains a public health concern in geographically restricted areas as low-resolution investigations do not consider local heterogeneity of snails and schistosomiasis (Kitron et al., 2006). The small-scale fociality of schistosomiasis is well recognised and the causes of heterogeneity are varied and reflect many human and ecological factors (Kloos et al., 1997, 1998; Woolhouse and Chandiwana, 1989). This makes the small-scale distribution difficult to predict (Brooker, 2002) as the effects of local heterogeneity are averaged out at broader scales and ecological patterns often appear more regular (Wiens, 1989; Levin, 1992). Brooker (2002), emphasized that the use of remote sensing has not been very successful in capturing the well-known local variation of schistosomiasis transmission, for example, the 8-km advanced very high resolution radiometer (AVHRR) used by Brooker et al. (2002a, 2002b) may generalize the spatial variation of schistosomiasis. Although satellite instruments can deliver higher resolution imagery, such as SPOT 5 and 6 (http://www.geoairbusds.com/en/147-spot-6-7-satellite-imagery) and Landsat 8 (http://landsat.usgs.gov/landsat8.php) even most recent African studies are based on low-resolution satellite products from the moderate resolution imaging spectroradiometer (MODIS) (http://modis.gsfc.nasa.gov) and American National Oceanic and Atmospheric Administration (NOAA) (http://www.noaa.gov) AVHRR (Tables 1-3).

There is need for the use of high-resolution instruments to capture the local ecological spatial variation of intermediate host snails or prevalence of schistosomiasis (Brooker et al., 2001, 2002a). Indeed, high-resolution imagery is available but has only been used occasionally due to the current high cost. For example, the very recent study of De Roeck et al. (2014) on Fasciola hepatica in Belgium using drones and very high resolution (VHR) imagery from the commercial WorldView2 satellite (https://www.digitalglobe.com/sites/default/files/DG_WorldView2_DS_PROD.pdf) is a case in point. Fine-scale monitoring is of key importance to refine currently existing broad-scale infection risk models, and costs might diminish with time allowing a more widely use of the technology. Generally, however, there is no single natural scale, at which ecological patterns are studied (Levin, 1992). In each case, the appropriate scale is dictated by the goals of the study, system and available data. Hence, there is need for developing remote sensing predictive models for targeting schistosomiasis control at local levels. There is also need to focus on the household as spatial points in high endemic areas instead of schools, as this will help to avoid spatial aggregation and allow appropriate finite scale spatial mapping and give insight into the micro-epidemiology of schistosomiasis (Simoonga et al., 2008).

**Failure to utilize the temporal domain of remote sensing**

The temporal characteristic of remote sensing has not been fully utilized in modelling the temporal variation of schistosomiasis in Africa. Brooker (2002) has highlighted that the temporal variation of schistosomiasis has received far too little attention despite its relevance in understanding the spatial distribution of infection. Seasonal and intra-seasonal modelling of schistosomiasis may capture the variation of snail density and occurrence as some snails maybe washed away during the rainy season (Appleton, 1978; Kloos et al., 2001) and some may die due to desiccation during the dry season (Rollinson et al., 2001). This may also explain the spatial and temporal variability of point prevalence, infection rate and intensity of schistosomiasis especially at local levels. This emphasizes the relevance of intra-seasonal and seasonal modelling of schistosomiasis in timely allocation of resources as well as targeting of control programs. Most of the schistosomiasis predictive models consider annual distribution of schistosomiasis (Stensgaard et al., 2013; Pedersen et al., 2014) but not the component of seasonality. Considering the temporal resolution of satellite data could help to synchronise the temporal differences between data collection of variables and disease or parasitological data as highlighted by Sturrock et al. (2013). There is need for ecological niche modelling of seasonal vector population dynamics combining ecological niche models with purpose-built, temporal high-resolution satellite remote sensing data (Kulkarni et al., 2010). On the other hand, remote sensing technology has been available for more than 30 years, but only a few studies such as that by Pedersen et al. (2014) have taken up the challenge to model the changes in snail habitats over the past years to estimate the possible distribution of schistosomiasis in relation to climate and environmental changes.

**Application limitations**

Brooker (2002) emphasized the need to develop separate models for each snail-schistosome system due to their different habitat types and environmental suitability. Thus, these models are not transferable to other regions or places. This is evident from the work of Malone et al. (2001), who developed a region-specific schistosomiasis predictive model with limited application elsewhere where different ecological conditions and snail species prevail. This compromises the validity of the continental- and regional-level models, advocating the local scale. Until 2009, there was no schistosomiasis model at a continental level (Simoonga et al., 2009) and, to our knowledge, Stensgaard et al. (2013) is the only study investigating schistosomiasis distribution in Africa as a whole. Simoonga et al. (2009) highlight several challenges that must be overcome in order to further improve the GIS-based mapping of intermediate host snails at the continental scale. Consequently, there is a lack of large-scale, geo-referenced quality data on the presence/
absence of snails as well as parasite-snail compatibilities. There is also need for a more complete understanding of snail species identities and their efficiencies as intermediate hosts as prescribed by Stothard et al. (2002). The challenge is to develop a composite risk map of schistosomiasis (Brooker, 2002), which could be achieved through shared component modelling technique as highlighted above.

**Lack of vigorous validation or accuracy assessment**

An important but often difficult part of a disease model is the assessment of applicability and validity, especially if outputs are to be used for disease control. The quality of predictive (presence or absence) models of a species is normally judged by the number of prediction errors or its accuracy (Fielding and Bell, 1997). Accuracy refers to the correctness of remotely sensed data or model outputs, which measures the agreement between a standard assumed to be the correct and classified image or the result of a model of unknown quality (Foody, 2001; Campbell, 2006). Despite the wide use of predictive models, most applications do not give sufficient consideration to model error and uncertainty (Barry and Elith, 2006). Datasets used to statistically develop the models are often of uncertain accuracy and are not always easily producible as the results vary with training data and methods used (Tanser et al., 2003). These generic disadvantages vary from worldwide or continent-wide (Stensgaard et al., 2013) to regional (Schur et al., 2011a, 2011b, 2011c) statistically-driven models; the models are too coarse to guide intervention efforts and their capacity to predict prevalence remains uncertain (Kulkarni et al., 2010).

Evaluating the predictive models is a crucial step for determining its suitability for specific applications (Guisan and Hofer, 2003; Allouche et al., 2006). In this case, the focus is on schistosomiasis control and the possibility of comparing with other models and classification techniques (Powell et al., 2004). Any approach to ecological modelling has little merit if predictions cannot be assessed (Verbyla and Litvaitis, 1989) and any maps or satellite products or models without associated accuracy remain untested hypotheses (Strahler et al., 2006). The most commonly used statistical measures of error of predictive models include error matrix (Morissette et al., 2005), the Cohen Kappa statistics, the threshold-independent receiver operating characteristic (ROC) approach (Fielding and Bell, 1997) achieved by calculating the area under the ROC curve (AUC), a Gini coefficient AUC (Copas, 1999), the true skill statistic (TSS) (McPherson et al., 2004), Cohen’s Kappa z-test and MacNemar’s test. Most of the studies use EO and disease data for schistosomiasis predictive modelling from electronic databases and only a few models for example Schur et al. (2011a, 2011b, 2011c, 2013) have been adequately evaluated and provided with proper statistical quantification of error. The main constraint in validation is lack of updated, comprehensive, good quality empirical data (Moodley et al., 2003; Tanser et al., 2003). This has compromised the quality, applicability and reliability of the developed models. The maps representing the world-wide or country-wide burden of schistosomiasis generally reflect the reported distribution of clinical episodes of this disease. However, the scope and accuracy of such reports are limited by the extent of health care coverage, the efficacy of surveillance and also by the quality of the reporting systems (Kiszewski et al., 2004). This paucity of epidemiological data hinders large-scale quantification of the burden of a disease (Brooker et al., 2002a, 2002b).

**Potential future research priorities**

Availability of affordable treatment of schistosomiasis (praziquantel) has led to increased interest and commitment to effective and efficient control of this disease. However, control resources are inevitably limited, necessitating predictive models that can rapidly and accurately identify and map high-risk communities so that interventions can be targeted in a spatially-explicit and cost-effective manner (Brooker, 2009). Geospatial technologies are promising with respect to meeting this objective. However, disease data, climatic and environmental data must not only be reliable, but also be possible to collect at suitable spatial and temporal resolutions. Figure 2 illustrates the proposed theoretical framework for schistosomiasis predictive modeling considering four phases; data collection, snail prediction modeling, schistosomiasis predictive modeling and validation which are based on data quality and availability as well as model accuracy. The whole system is sustained by health research institutional capacity in terms of skills, funds and equipment to generate high quality data and achieve desirable predictive model accuracy (Figure 2).

Although remote sensing has proved to be a reliable source of climatic and environmental data, there is need to consider satellites with higher spatial resolution such as SPOT-6 as opposed to low spatial resolution imagery of 1 and 8 km for MODIS and AVHRR, respectively, while ultrahigh-resolution imagery will have to wait until costs diminishes. The EO technology could offer higher spatial and spectral, more frequent coverage and lower cost data as suggested by Lleo et al. (2008), however, specialized skills and expertise are a pre-requisite so as to realize the full advantages of these developments. Mostly, EO data are developed for a wide range of applications and epidemiologists have to develop specific products or applications for specific purposes rather than relying on already processed or off-shelf products such as the normalized difference vegetation index (NDVI) and land surface temperature (LST) with no metadata. These products may have been developed for different purposes at different scales and may not serve the same purpose with same level of accuracy required in schistosomiasis modelling. This indicates the need for investing more time and resources in the development, application and use of the space technology in epidemiology.

Field studies will still be needed to generate high quality data including climatic and environmental factors for calibration of EO data as shown in Figure 1. The output predictive models should also be validated against field observations as argued by Bergquist et al. (2009) to realize their usefulness in community health and climate change decision making process especially at the local level in Africa. This will help to capture the local focality of schistosomiasis as the use of high quality and reliable data could help to refine the geostatistical techniques and adoption of ecological tools such as Maxent (Phillips et al., 2006), which are promising to produce highly performing schistosomiasis predictive models (Stensgaard et al., 2013). The field measured data could also be complemented by laboratory experiments as the behavior of snails and schistosome parasites could change due to climatic and environmental changes. For example Brown (1994) studied the distribution of freshwater host snails in Africa and divided them into two groups (tropical and temperate species) based on the climate where they occur. However the snail species tolerance ranges for temperature might have changed over time, which requires more field studies for verification and comparison. On the other hand, systematic field studies on the relationship between densities and infection rates in snails and those in humans would still be highly useful (Gryseels, 1996). The rational use of remote sensing data is dependent of the quality of infor-
mation from the field (De la Rocque et al., 2005), which requires sophisticated geo-spatial statistical methods for analysis and predictive modeling and dedicated fieldwork to validate the observations (Herbreteau et al., 2005). Hence, the need for interdisciplinary approach in which epidemiologists collaborate with software programmers, geographers and spatial analysts to create robust techniques and products for use with epidemiological data (Jacquez, 2000; Graham et al., 2004; Herbreteau et al., 2007).

Conclusions

Geo-spatial technologies are invaluable for schistosomiasis mapping and transmission prediction, particularly in Africa. However, more extensive applications of these tools have been hampered by lack of training, gaps in data (quality and quantity, particularly climatic, environmental, epidemiologic and parasitologic data) and inadequate tools for data gathering (Hay, 2000). These are clear indications of possible sources of errors and uncertainties that have propagated the schistosomiasis transmission modelling in Africa. Therefore, Herbreteau et al. (2007) is correct in saying that major elements of geospatial technologies have not yet met our current needs. However, in China, Yang et al. (2005) viewed future prospects of GIS and remote sensing application in disease mapping as bright and promising; hence, it might be too early to blame the technology. Instead, there is need to take advantage and sharpen ideas and skills to develop more and better methods through further research and refinement of the schistosomiasis predictive models to meet community needs. This could be achieved through collaboration between epidemiologists, geographers and software programmers, use of high-quality remote sensing and ground measured data and thorough validation protocols. It would also allow the adoption and use of some tools developed in ecology such as Maxent and GARP machine-learning algorithms and refining geostatistical techniques for predictive modelling of schistosomiasis at the local level.

References


of potential applications using the example of animal trypanosomosis in West Africa. Parasitology 47:97-104.


Malone JB, Yilma JM, McCarroll JC, Erko B, Mukaratirwa S, Zhou XY,


Woolhouse MEJ, Chandiwana SK, 1989. Spatial and temporal heterogeneity in the population dynamics of *Bulinus globosus* and *Biomphalaria pfeifferi* and the epidemiology of their infection with schistosomes. Parasitology 98:21-34.
