Parassitologia 47: 115-125, 2005

Modeling the distribution of *Schistosoma mansoni* and host snails in Uganda using satellite sensor data and Geographical Information Systems

A. Stensgaard[1], A. Jørgensen[1], N.B. Kabatereine[2] , J.B. Malone[3], T.K. Kristensen[1]

[1]*Mandahl-Bart Research Centre for Biodiversity and Health in Developing Countries, DBL - Institute for Health Research and Development, Charlottenlund, Denmark;* [2]*Vector Control Division, Ministry of Health, Kampala, Uganda;* [3]*Pathobiological Sciences, School of Veterinary Medicine, Louisiana State University, Baton Rouge, Louisiana, USA.*

Abstract. The potential value of MODIS satellite sensor data on Normalized Difference Vegetation Index (NDVI) and land surface temperatures (LST) for describing the distribution of the *Schistosoma mansoni-"Biomphalaria pfeifferi"/Biomphalaria sudanica* parasite-snail system in inland Uganda, were tested by developing annual and seasonal composite models, and iteratively analysing for their relationship with parasite and snail distribution. The dry season composite model predicted an endemic area that produced the best fit with the distribution of schools with 5% prevalence. NDVI values of 151-174, day temperatures of 26-36 °C, and night temperatures of 15-20 °C were used as criteria for the prediction model. Using the same approach with host snail data indicated that most of Uganda is suitable "*B. pfeifferi*"/*B. sudanica* habitat, except for possibly the north-eastern region of the country. The parasite, however, appears to be restricted in its distribution in both the north-eastern and the south-western regions of Uganda. The absence of disease in the south-west can not be attributed to the absence of snail hosts. Results suggest a combination of satellite sensor data on temperature and standard climate data on precipitation, as the best ecological determinants of the *S. mansoni-"B. pfeifferi"/B. sudanica* system. Satellite composite models and logistic regression analysis, suggest low night time temperature as one of the significant factors inhibiting *S. mansoni* transmission in the south-western highland areas of Uganda. The developed models are, however, unique, representing species-specific ecologic preferences of the *S. mansoni* - "*B. Pfeifferi*"/*B. sudanica* system in inland Uganda. Further validation studies are needed to test the value of the model in other countries in East Africa.

Key words: *Schistosoma mansoni*, *Biomphalaria*, Geographical Information Systems, MODIS satellite, ecology, Uganda.

Studies on *Schistosoma mansoni* and its intermediate snail host *Biomphalaria* spp., around the world, have demonstrated important links between a number of environmental and climatic factors and the distribution and prevalence of the disease and disease vector. The parasite-host system is constrained in space and time by a number of biotic and abiotic factors, many of which are easily incorporated in a Geographical Information System (GIS). Data on climate and environment, such as land surface temperatures and vegetation indices can be easily obtained from various earth orbiting satellites, and represents a rapid way to develop spatial models of disease risk. These can, when resources are scarce, assist health authorities in making appropriate control and intervention measures in predicted high risk areas, which is particularly important in remote areas where ground-based meteorological and environmental data frequently are unavailable. Studies in a number of African countries have demonstrated that environmental features such as temperature and humidity, which affect the distribution of the parasite-snail system, may be determined from remotely sensed (RS) data (obtained mainly from the Advanced Very High Resolution Radiometer (AVHRR) sensor), and used within a GIS environment to provide increased understanding of parasite transmission (Hay *et al.*, 2000; Brooker *et al.*, 2001, 2002; Brooker and Michael, 2000; Malone *et al.*, 2001a; McNally, 2003).

In this paper we review results of a study of the associations between satellite sensor data from a new generation of satellite sensors, the Moderate Resolution Imaging Spectro-radiometer (MODIS) satellite and the distribution and prevalence of *S. mansoni* and its intermediate host snail species in Uganda.

Intestinal schistosomiasis in Uganda is a major health problem and has long been known to occur around the Greater Lakes (Prentice, 1972). A recent study on the epidemiology of intestinal schistosomiasis in Uganda found both a widespread occurrence of disease and a marked variability in infection prevalence, with some of the highest prevalence found in schools located near Lake Victoria and Lake Albert. No significant correlations were found between prevalence rates and AVHRR

Correspondence: Anna-sofie Stensgaard, Mandahl-Bart Research Centre for Biodiversity and Health in Developing Countries, DBL - Institute for Health Research and Development, Jaegersborg Allé 1D, 2920 Charlottenlund, Denmark, Tel +45 77 32 77 32, Fax +45 77 32 77 33, e-mail: asstensgaard@dblnet.dk

vegetation- or temperature derived variables, separately. The authors instead found that absence of transmission was associated with total annual precipitation <900 m and altitudes below 1400 m. Finally, it was established that proximity to larger lakes is a major risk factor for transmission in Uganda, and that prevalence consistently exceeded 50% (WHO's recommended threshold for mass treatment) in areas within 5 km of Lakes Victoria and Albert (Kabatereine *et al.*, 2004; Brooker *et al.*, 2004).

There are however, also other areas further inland and not immediately associated with larger permanent water bodies in Uganda, where high school prevalences have been observed, supposedly reflecting other important determining environmental factors, possibly identifiable with RS/GIS approaches. Presumably the responsible intermediate host snail in these areas is what is known as "*Biomphalaria pfeifferi*" in Uganda, whereas the transmission at the larger lakes mainly takes place through other *Biomphalaria* species associated with these waters. Climatic conditions, primarily rainfall and temperature, influence the distribution and density of snails and the rate of schistosomal development in the snail host (Appleton 1978; Sturrock 1993), and probably influence the distribution of schistosomiasis in Uganda. In order to examine possible preferable environmental and climatic ranges for the parasite-snail system in inland Uganda we look only at schools and snail that are not situated in the immediate vicinity of larger, permanent water bodies. We do this 1) to evaluate the use of MODIS satellite sensor vegetation- and temperature derived variables for mapping *S. mansoni* endemic areas in inland Uganda, 2) to identify the most significant explanatory climatic and environmental variables associated with the *S. mansoni-Biomphalaria* parasite-snail system in inland Uganda, and 3) to define environmental ranges of the main responsible intermediate host snail species in inland Uganda, and compare their distribution to that of the disease.

Materials and methods

Parasitological data

A recent survey of schools in Uganda revealed an overall prevalence rate of 20.4% among examined schoolchildren with some schools having more than 90% prevalence (Kabatereine *et al.*, 2004). The survey was conducted by the Vector Control Division, Ministry of Health, Uganda, in the period 1998-2002 and constitutes the parasitological data used in the present study. Data is available from 201 georeferenced schools across Uganda, and includes a total of 13,798 schoolchildren (Kabatereine *et al.*, 2004).

For the purpose of the present study, schools situated within a 5 kilometer buffer around all the major permanent water bodies were eliminated, and the 153 remaining "inland" schools were used for the following analyses.

Snail data

The main host snails of *S. mansoni* in habitats associated with swampy lake margins and rivers in Uganda are *Biomphalaria sudanica*, and further away from permanent waters seasonal transmission is believed to take place through "*B. pfeifferi*", which is known also to inhabit temporary water bodies (Prentice, 1972; Brown, 1994). It is assumed that transmission of *S. mansoni* in these areas must take place through "*B. pfeifferi*" and/or *B. sudanica*, since the other Ugandan host snail species *Biomphalaria stanleyi* (in Lake Albert) and *Biomphalaria choanomphala* (in Lake Victoria), only exists in habitats very closely associated with these lakes (Prentice, 1972). However, a recent study on the phylogeny of the Ugandan freshwater snail species shows that the species termed "*B. pfeifferi*" according to morphological characters actually have a closer phylogenetic relationship with *B. sudanica* from Uganda, than with *B. pfeifferi* from other African countries (Jørgensen, 2003). Therefore, this study includes both *B. sudanica* and "*B. pfeifferi*" localities to investigate the inland *S. mansoni*-host snail system, as it is currently not clear which of these constitutes the main intermediate host in the inland areas of Uganda.

Freshwater snail specimens were collected in Uganda in the period 2000-2003, mainly by A. Jørgensen. In total, more than 20,000 specimens have been collected from 89 different sampling sites. A variety of sampling methods were employed on each locality depending on water depth, sediment type and vegetation. All sampling localities were georeferenced using Geographical Positioning System (GPS). No sampling was conducted north of Lake Albert due to the general insecurity of the area. Species identification was conducted by referencing Mandahl-Barth (1954, 1957a,b), Brown (1994), Kristensen (1987), and the DBL - Institute for Health Research and Development reference collection. A list of species for each locality were constructed and imported to ArcView® 8.3 (ESRI, Redlands, CA). Furthermore, records of relevant *Biomphalaria* snail species recorded by Mandahl-Barth (1954) (corrected to currently accepted species) and geo-referenced within two hundredths of a decimal degree latitude and longitude were added to the data set to get as many snail reference points as possible, giving a total of 131 localities. Snail sampling localities more than 5 km away from larger water bodies, were isolated for further spatial analysis, to evaluate climatic and environmental ranges for this inland "snail species complex" alone, and to compare the distribution of the host snails to that of the parasite. Of the original 131 snail localities, 57 were found further than 5 km away from larger permanent water bodies. Of these, *B. sudanica* and/or "*B. pfeifferi*" were found at 22 of the localities.

GIS and satellite sensor data

The GIS databases were constructed using GIS software packages ArcView® 3.3 and ArcGIS® 8.3. The basic GIS was developed using a compilation of resource data on administrative zones, infrastructure, elevation and climate from the Minimum Medical Database Spatial Decision Support System for the Intergovernmental Authority on Development-Nile Basin Region (IGAD/Nile MMDb) (Malone *et al.*, 2001b). Land cover, vegetation maps and mapped water bodies were obtained from the Uganda National Biomass Study (Forest Department, Uganda, 1996). All GIS data were georeferenced to the geographic latitude-longitude decimal-degree projection, WGS 84 spheroid and WGS 84 datum format. Satellite sensor land surface temperature (LST) and Normalized Difference Vegetation Index (NDVI) from the MODIS satellite onboard the NASA Terra, were obtained from the United States Geological Survey's (USGS) EROS datacenter and image processing was performed using ERDAS Imagine 8.4™ (ERDAS Inc., Atlanta, GA, USA). The data from sensors like MODIS are advanced in spectral, spatial, radiometric terms and temporal resolutions and are further complimented by advances in cloud/haze removal algorithms, time composing, and normalization of data into reflectance. MODIS LST products include day time (T_{day}) and night time (T_{night}) global 1 km products, and a global 250 m NDVI product. NDVI is used as a surrogate for soil moisture (Malone *et al.*, 2001a). These products are available for every 8 – 16 days during the period 2000 onwards (for more information on the MODIS satellite sensor products see http://modis.gsfc.nasa.gov/).

Average annual composite NDVI maps were produced by combining all 16-day interval NDVI maps in the period 2000-2003. Composite land suface temperature (LST) maps were produced by combining all 8-day interval LST maps, to reduce the variation between climate years. Based on Uganda's two principal rainy seasons, March to May, and August to November, average wet season and dry season composite maps were similarly produced. NDVI was rescaled to positive values (0-200). The temperature measures used in subsequent analyses were T_{day}, T_{night} and T_{diff} (T_{day} minus T_{night}).

Larger permanent water bodies, surrounding wetlands and their boundaries were identified through the National Biomass Study (Uganda Forest Department, 1996), a major study aimed at classifying vegetation types and water bodies in Uganda based on 30 m resolution SPOT satellite images, combined with field validation studies. Water bodies were also identified using the 2000-2003 MODIS satellite products as additional validation, using standard classification procedures.

Climate grid data

A climate surface grid (5 km^2 cell) consisting of monthly long-term normal climate data was obtained from the IGAD/Nile MMDb. It includes GIS vector data on monthly rainfall in mm and monthly potential and actual evapotranspiration, for the climatic normal period 1931-1960 (Malone *et al.*, 2001a). An average based on wet season, dry season and annual months respectively, were calculated for both precipitation (PRE) and the ratio of rain/potential evapotranspiration (PPE). The latter is an index that represents not just amount of rainfall, but the availability of water in a given period.

Data extraction and model developments

Elimination of schools and snail localities within buffer zones of 5 km around all identified permanent water bodies were undertaken using the Spatial Analyst extension of ArcGIS® 8.3. Buffer zones of 5 km diameter centered on each of remaining 153 schools and 57 snail localities were created and used to extract mean values for all MODIS surrogate climate data, as well as mean values for other environmental variables.

Initial evaluation of the extracted satellite sensor data was done to define T_{day}, T_{night} and NDVI value ranges consistent with 1) the *S. mansoni* endemic areas, and 2) the "*B. pfeifferi*" and *B. sudanica* distribution areas. Analysis of scatter diagrams made by plotting extracted mean values against "*B. pfeifferi*"/*B. sudanica* presence localities, and school localities ≥5% prevalence, allowed definition of the ranges of NDVI, T_{day} and T_{night}. For the parasitological data set on *S. mansoni* prevalence, these initially identified range values were then used for GIS queries to produce overlay maps of areas suggested to be suitable for *S. mansoni* inland in Uganda. Queries were then repeated iteratively using increments of NDVI and LST measures separately, to derive best-fit values by matching query-defined endemic areas to the mapped schools with ≥5% prevalence. The narrowest ranges of T_{day}, T_{night} and NDVI values were defined for which the maximum number of schools with ≥5% prevalence were contained within the GIS query map area. Using these separately defined value ranges for NDVI and LST, the satellite composite maps were then queried to produce a GIS map showing areas, where NDVI, T_{day} and T_{night} criteria were met simultaneously for each of the 3 criteria ranges. The resulting overlay map represented the model-predicted endemic area.

To determine the relative suitability of Uganda's wet and dry season for the *S. mansoni* - "*B. pfeifferi*"/*B. sudanica* system, as well as the relative influence of each season within the annual composite model, composite models were also developed for the dry season and wet season and evaluated with regards to their positive and negative predictive values (Table 1 and Table 2), using same procedures as that described for the development of the annual composite model. The described procedures are largely the same as those described by Malone *et al.* (2001a) for defining climatic and environmental preferences for the *S. mansoni*-*B. pfeifferi* system in Ethiopia.

For the snail data set, the same procedures were repeated using "*B. pfeifferi*"/*B. sudanica* presence points, to define NDVI, T_{day} and T_{night} criteria ranges, and to produce annual, wet season and dry season model predicted presence areas, representing "*B. pfefferi*"/*B. sudanica* suitable habitats. Due to the lack of snail sampling localities north of Lake Albert, schools further than 5 km away from boundaries of the larger lakes, with ≥5% prevalence were hypothesized to represent *"B. pfeifferi"/B. sudanica* presence points.

School prevalence rates were then classified into the categories 0%, 0-5% and >5%, representing non-endemic, marginally endemic and endemic areas (Malone *et al.*, 2001a). These prevalence categories were used to evaluate the *S. mansoni* predictive models positive and negative predictive values (Table 1).

Statistical evaluations

Initially, the ranges for schools with more than 5% prevalence were used to identify potentially endemic areas in Uganda, and the resultant models tested with regards to their positive and negative predictive value. The same was done for "*B. pfeifferi*" and *B. sudanica* presence localities.

Spearman rank correlation and logistic regression analysis were performed to identify significant MODIS, environmental and climatic features associated with *S. mansoni* prevalence patterns (dichotomized to "endemic" and "non-endemic" schools) and host snail sites. The logistic regression models, using all available environmental, MODIS and climatic variables, were tested to find the most significant explanatory factors for the variation in the *S. mansoni* prevalence patterns in Uganda, as well as that combination of factors that best describes the geographical variation in endemicity. Model reduction and identification of significant variables were done using backwards elimination procedures. The schools were considered comparable for this type of analysis, as the prevalences are based on examinations on fairly equal number of boys and girls in approximately the same age-groups. All the independent variables were initially tested for linearity and correlation with each other. Interactions between the various variables were tested for after adjusting for the main effects. Statistical significance was considered acceptable if $P < 0.05$. Analyses were performed using SAS software, Version 8 of the SAS system for Windows (SAS Institute Inc, Cary, NC, USA).

Results

Evaluation of the MODIS composite models for Schistosoma mansoni *endemic areas in Uganda*

To develop optimum climatic criteria ranges for both annual, wet season and dry season, mean values from 5 km diameter buffer zones centered on each school were extracted from the NDVI and LST composite maps for each season, by iteratively fitting query defined areas, to that of the spatial distribution of schools ≥5% prevalence. The narrowest, best fit query results that included a maximum of schools with ≥5% prevalence for the annual composite maps, were 156-177 for NDVI, 27-35 ºC for T_{day} and 15-20 ºC for T_{night}. This combined query result represented the annual composite model, which had a positive predictive value of 98.2% (Table 1). Similar analysis of values extracted from the wet season composite maps versus school prevalence data, produced best-fit value ranges for the wet season of 160 - 176 for NDVI, 27-35ºC for T_{day} and 16-20ºC for T_{night}, with a positive predictive value of 94.7% for schools with ≥5% prevalence. This combined query result constituted the wet season model. Repeating analyses for the dry season, gave best-fit value dry season ranges in relation to schools of ≥5% prevalence, of

Table 1. Predictive values of the MODIS NDVI, T_{day} and T_{night} composite annual and seasonal models, based on the narrowest identifiable query ranges for schools with 5% prevalence, as compared to schools with no prevalence, <5% and >5% prevalence.

	No. of schools within predicted area	No. of schools outside predicted area	Positive predictive values (%)	Negative predictive values (%)
Annual composite map (NDVI 156-177 T_{day} 27-35°C, T_{night} 15-20 °C)				
0% prevalence	36	27		42.9
< 5 % prevalence	29	4	87.9	
> 5 % prevalence	56	1	98.2	
Wet season composite map (NDVI 160-176 T_{day} 27-35°C, T_{night} 16-20 °C)				
0% prevalence	35	28		44.4
< 5% prevalence	29	4	87.9	
> 5% prevalence	54	3	94.7	
Dry season composite map (NDVI 151-174, T_{day} 26-36°C, T_{night} 15-20 °C)				
0% prevalence	38	25		39.7
< 5% prevalence	31	2	93.9	
> 5% prevalence	57	0	100	

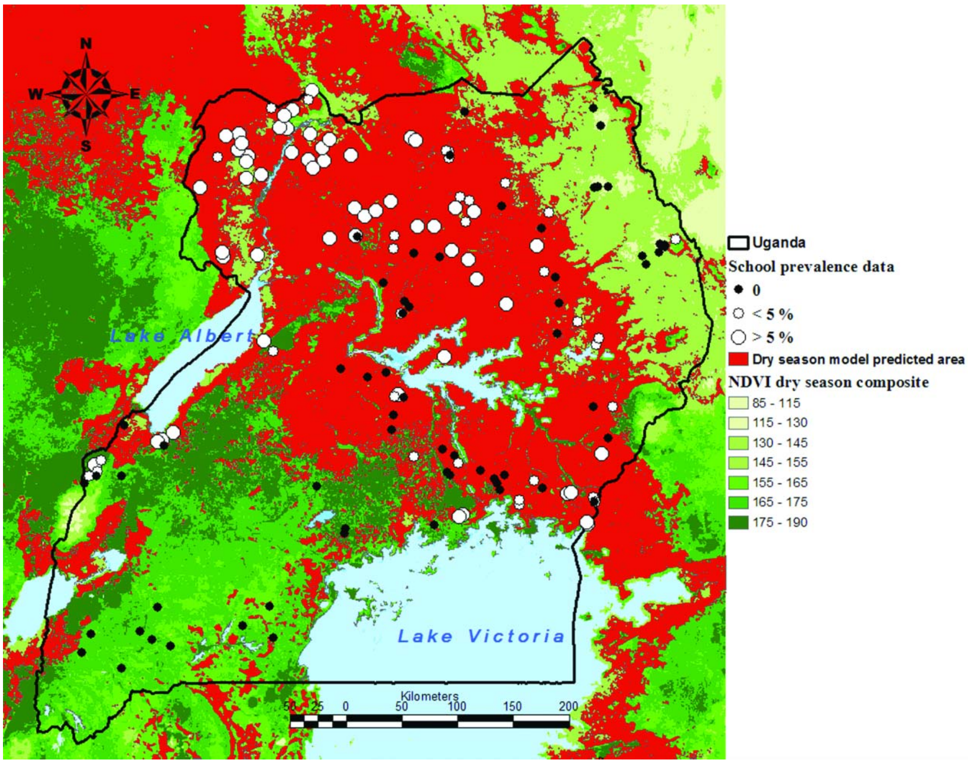

Fig. 1. GIS dry season composite model predicted endemic areas (red), based on climatic range value query for NDVI (151-174), T_{day} (26-36 °C) and T_{night} (15-20 °C). The model-predicted endemic areas are overlaid on a dry season composite NDVI map.

151 - 174 for NDVI, 26 - 36 ºC for T_{day} and 15 - 20 ºC for T_{night}, and a positive predictive value of 100% for schools of ≥5% prevalence. Each of the three models best-fit query ranges for NDVI and LST measures, along with positive and negative predictive values are listed in Table 1. An illustration of the dry season composite model-predicted area, in a combined GIS query map can be seen in Fig. 1.

Investigation of the GIS query ranges on each of the composite NDVI, T_{day} and T_{night} maps separately revealed that the exclusion of the north-eastern part of Uganda from the endemic predicted area was due largely to NDVI values below the range query limits identified for schools with ≥ 5% prevalence. The exclusion of the large area in the south-western part of Uganda, was due mainly to lower night and day temperatures than the criteria ranges used for the model GIS query.

Evaluation of MODIS composite data models for the distribution of the intermediate host snails

The climatic ranges identified through scatter diagram analysis of *"B. pfeifferi"* and *B. sudanica* presence points plotted against extracted annual, wet season and dry season composite NDVI and temperature values, were used for initial GIS queries of the composite MODIS satellite maps, using the same approach as with *S. mansoni*. The previously identified ranges found for non-lakeshore schools ≥5% prevalence (sites outside 5 km buffer around larger water bodies), were also used where these went beyond ranges otherwise identified for snail presence points, due to the lack of snail sampling in the area north of Lake Albert. The resultant annual, wet season and dry season composite model predicted areas, representing suitable snail habitat, captured all *"B. pfeifferi"/B. sudanica* presence sites. The annual composite model had a positive predictive value for schools ≥ 5% prevalence of 98.5 %, the wet season composite model one of 96.5%, and the dry season composite model a positive predictive value of 100% for ≥ 5% prevalence non-lakeshore schools. These, and other results for the MODIS composite model predicted snail habitats, can be seen in Table 2.

Analysis of the MODIS composite satellite maps, using the individual MODIS NDVI and temperature range values separately, revealed that the exclusion of the northern-eastern area is due to NDVI values below the lower limits found for both the *Biomphalaria* snails sites and the ≥ 5% *S. mansoni* prevalence schools. For night time and day time LST, lower query ranges contributed to the exclusion only of the Rwenzori mountain area in the east, the Mount Elgon areas in the west at the Kenyan border, and small patches of highland areas at the very south-east corner of Uganda. The combined dry season query map can be seen as an overlay on the dry season NDVI composite map in Fig. 2.

Table 2. Predictive values of the MODIS NDVI, T_{day} and T_{night} seasonal and annual composite models, based on the narrowest identifiable climatic ranges for "*B. pfeifferi*" and *B. sudanica*. Schools with 5% prevalence are included for evaluation of the model, as they are hypothesized to represent presence of *"B. pfeifferi"/B. sudanica*.

Composite model	NDVI, T_{day} and T_{night} (°C)	No. of positive sites within the model area	Positive predictive value (%)	No. of negative sites outside the model area	Negative predictive values (%)
"B. pfeifferi"/B. sudanica					
Annual	156-177, 25-35, 12-20	22/22	100	2/36	5.6
Wet season	160-176, 25-35, 12-20	22/22	100	3/36	8.3
Dry season	151-175, 24-36, 12-20	22/22	100	1/36	2.8
Schools *S. mansoni*					
Annual		56/57	98.5	12/63	19.0
Wet season		55/57	96.5	12/63	19.0
Dry season		57/57	100	11/63	17.5

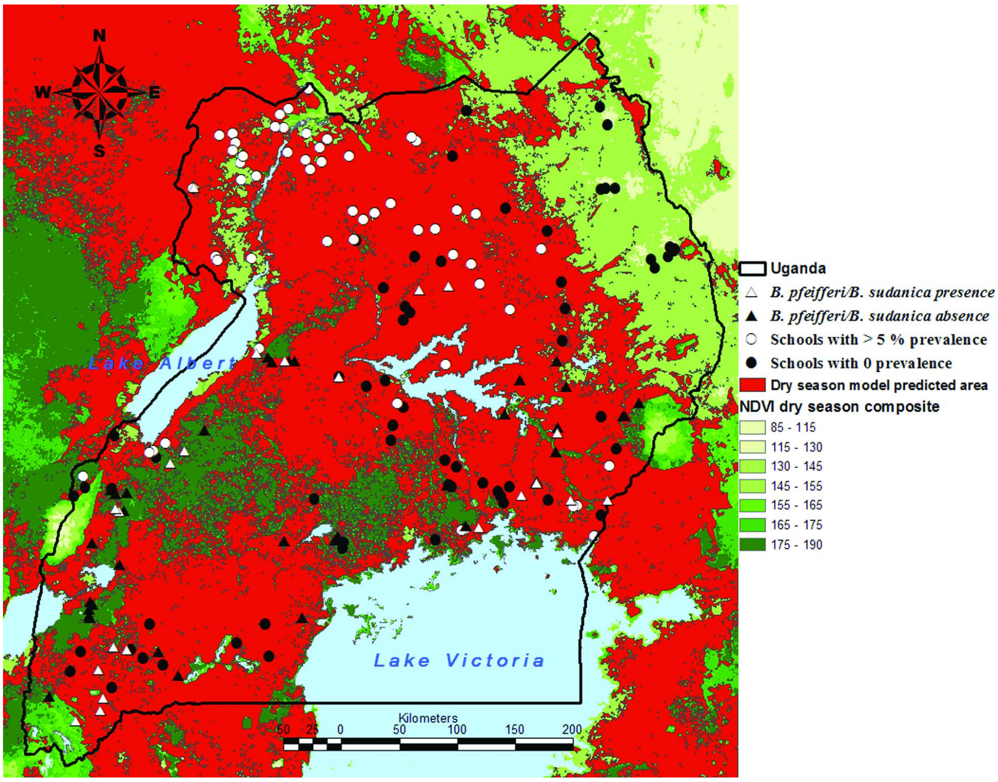

Fig. 2. The GIS query dry season MODIS composite model predicted area (red) for potential "*B. pfeifferi*" and *B. sudanica* habitats in inland (non-lakeshore) Uganda. The predicted area represents the narrowest, best-fit GIS query map, using NDVI ranges (151-175), T_{day} (24-36 °C) and T_{night} (12-20 °C), that included a maximum of snail presence sites, and non-lakeshore schools with 5%. The schools were included to compensate for the lack of snail sampling localities in the northern parts of Uganda.

Identifying significant environmental explanatory variables for S. mansoni

Initial Spearman's rank correlation analysis between disease prevalence and all independent variables separately, showed correlations with the majority of the climatic and environmental variables used in the present study. No significant correlation were however found with MODIS composite annual and wet season NDVI, and only a very weak negative correlation observed with dry season NDVI (*r*=-0.188; *P*<0.038). Furthermore, no correlation was demonstrated with wet season PPE, or T_{diff}. All other MODIS satellite sensor and climatic variables were positively correlated with disease prevalence (*r*=0.200 to 0.443; *P* <0.01), the strongest correlations being with wet season night-time temperature ($T_{nightwet}$) (*r*=0.4434; *P* <0.0001). A negative correlation was furthermore observed with elevation (*r*= -0.553; *P* <0.0001).

Multi-colinearity was initially observed between several of the explanatory variables, most notably between MODIS satellite sensor NDVI and several LST products (*r*=0.87 to 0.90; *P* <0.0001). Also, precipitation (PRE) and rain minus potential evapotranspiration (PPE) variables were highly correlated (*r*=0.87 to 0.92; *P* <0.0001). Due to this high correlation between several of the independent variables, several models including different combinations of non-colinear independent variables had to be developed separately. All models were initially developed using backwards elimination, with subsequent tests for interactions. A total of eight logistic regression models were fitted to the school *S. mansoni* prevalence data ≥5%, all with two or three significant explanatory variables, most of them based on wet season variables. None of the dry season composite variables showed any statistically significant association with *S. mansoni* prevalence when tested in a logistic regression model. Seven of the models included the variable elevation in combination with T_{day} and precipitation (PRE) *or* precipitation minus evapotranspiration (PPE), and one model included a combination of variables T_{night} and precipitation**.** The models were compared using a combination of the Akaike Information Criteria (AIC) (Bozdogan, 1987), coefficients of determination (max-rescaled R-square), Hosmer-Lemeshow goodness-of-fit statistic, and a measure of prediction error (Sp) (Hosmer and Lemeshow, 1989). AIC, is a measure of goodness of fit, penalized by model complexity (Ripley, 1996), and is quoted to give a fair comparison between models with an uneven number of parameters. The lower the AIC, the better the model (Burnham and Anderson, 2002). The fit of all models was assessed by the Hosmer-Lemeshow goodness-of-fit test, which compares predicted outcomes with observed outcomes by decile. A larger *p*-value implies better fit. Following the methodology proposed by Nagelkerke (1991), we also utilized a maximum rescaled R-square to determine the absolute percentage of variation explained by each model. The two best performing models with regards to these criteria can be seen in Table 3. The resultant regression equations for these models were applied in the GIS, to produce *S. mansoni* predicted probability risk maps, showing the model predicted probabilities of an area being "endemic" (having schools ≥5% prevalence). The risk map for the logistic regression model with significant variables $T_{nightwet}$ and PRE_{wet} (model 2) is shown in Fig. 3.

Table 3. Statistical results from logistic regression analysis of the association between endemicity (5% prevalence) and the significant variables, wet season day time temperature (T_{daywet}) average monthly precipitation in the wet season (PRE_{wet}) and wet season average night time temperatures ($T_{nightwet}$). Adjusted odds ratios for increases in factors of one and several units are also shown.

Parameter	Estimate	Error	*p*-value	Wald 95% CF limits	AIC	R-square	Hosmer & Lemeshow Goodness of fit (*p*-value)	Prediction error (S_p)
Model 1:					150.7	0.4384	0.1134	0.3890
Intercept	-14.234	6.1613	0.0236					
Elevation	-0.006	0.0016	0.0004	-0.009 – -0.003				
T_{daywet}	0.293	0.1168	0.0123	0.063 – 0.521				
PRE_{wet}	0.084	0.0223	0.0002	0.040 – 0.129				
Model 2:					152.5	0.4155	0.8448	0.3975
Intercept	- 26.272	5.0038	< 0.0001					
$T_{nightwet}$	0.946	0.2009	< 0.0001	0.552 – 1.340				
PRE_{wet}	0.073	0.0192	0.0001	0.035 – 0.110				

Adjusted odds ratios for different increases in units in the significant factors:

	1 unit	5 units	10 units	100 units	500 units
Elevation (m)	0.994	-	-	0.570	0.060
T_{daywet}	1.339	4.312	-	-	-
PRE_{wet}	1.076	1.439	2.071	-	-
$T_{nightwet}$	2.576	113.41	-	-	-

Odds ratios are shown for a change of one unit for each significant variable. Adjusted odds ratios were calculat-

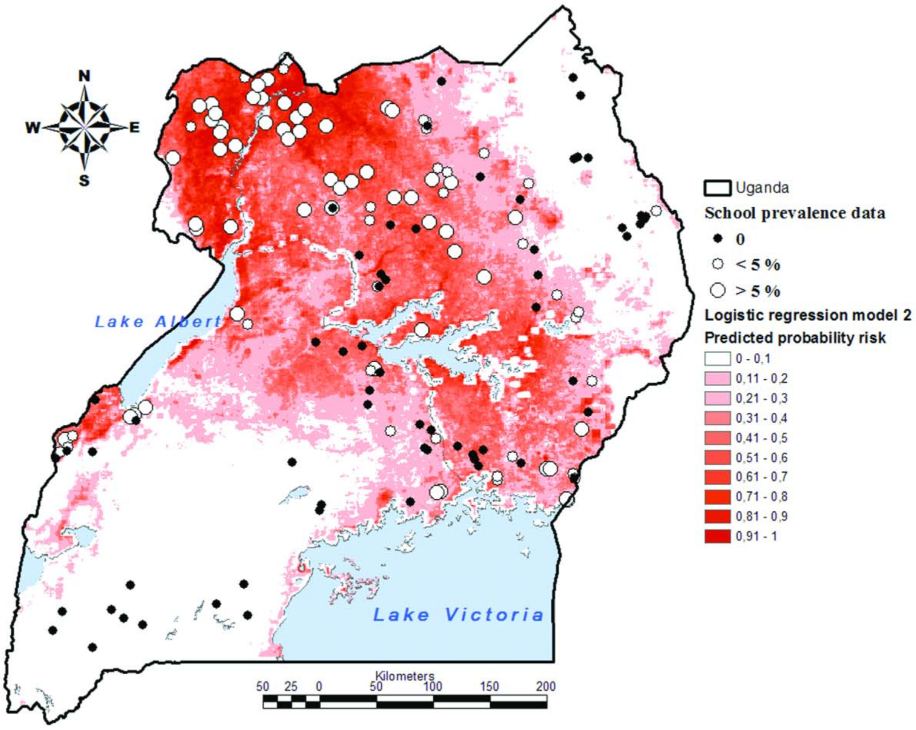

Fig. 3. GIS schistosomiasis risk prediction map for Uganda, based on 153 schools not associated with larger permanent water bodies. The map shows the predicted probability of an area being "endemic" (having more than 5% prevalence). Probabilities are calculated from the logistic regression model 2 equation and includes significant variables $T_{nightwet}$ and PRE_{wet}.

ed to give more meaningful interpretation. They show that an increase in 100 meters altitude, according to model 1, gives approximately 1/2 the odds for an area being endemic (or twice the odds for an area being non-endemic), and that an increase of 10 mm average monthly rainfall in the wet season gives twice the odds of an area being endemic.

Identifying significant explanatory variables for intermediate host snail presence

Spearman rank correlation analysis revealed that presence of "*B. pfeifferi*" and/or *B. sudanica* were weakly positively correlated with annual and seasonal T_{day}, (p=0.208 to 0.211; P <0,01), but not with T_{night} measures. Furthermore, presence was weakly correlated with several of the precipitation measures (r= 0.1963 to 0.2240; P <0.01). No correlation could be shown with elevation using the present number of snail sampling localities.

Discussion

Evaluating the suitability of MODIS surrogate climate data for defining S. mansoni *endemic areas in Uganda*

The suitability of MODIS satellite sensor data for identifying and modeling *S. mansoni* -"*B.pfeifferi*"/*B. sudanica* system in areas in inland Uganda was initially evaluated. By analyzing the MODIS NDVI and day time and night time composite LST query map ranges based on values extracted from school sites in a GIS, annual and seasonal model- predicted endemic map areas that included 94.7-100% of schools ≥5% prevalence could be defined (Table 1 and Fig. 1). Optimum criteria ranges used for queries were NDVI values 156-177, T_{day} values of 27-35 °C and T_{night} values of 15-20 °C for the annual composite model. Only small variations were observed between the seasonal developed composite models and the annual composite models, although the dry season had the highest positive predictive values. The dry season models NDVI values (151-174) were slightly lower and with a wider range, and with T_{day} values of 26-36 °C and T_{night} values of 15-20 °C, closely resembling the annual and wet season model values. In general, seasonal variation in climate is less extreme in Uganda, than in many other highland areas, with mean monthly temperatures highest in February and lowest in July, both dry season months (Cox *et al.*, 1999), explaining the low variation in temperature measures inter-seasonally. The annual and seasonal model defined endemic risk areas for *S. mansoni* (Table 1) and suitable host snail habitat areas (Table 2) had lower negative predictive values. However, this is not unexpected, since large variation in schistosomiasis prevalence between closely situated schools and communities is a well known phenomena, due to the multifactorial nature of the disease. The lower negative predictive values for the annual and seasonal model for suitable snail habitat areas, are not surprising either, since the absence of snails at a particular site can be a result of sampling effort and the highly temporal variation in snail-microhabitats.

Analysis of MODIS satellite NDVI and LST maps individually, showed that the exclusion of the north-eastern region of Uganda from the model-predicted endemic area was due to low NDVI values, indicating a moisture level below what can sustain a *S. mansoni* -*"B. pfeifferi"/B. sudanica* system. This corresponds well to areas where it was considered unlikely to find *S. mansoni* by Kabatereine *et al.* (2004), due to low total annual precipitation. The exclusion of the region in the south-western region of Uganda from the model-predicted endemic area, was shown to be mainly due to low night time temperatures (below the query range value of 15 ºC) with some influence also from low day time temperatures (below the query range value of 26 ºC). These areas correspond to areas in the south-west regions that were considered unlikely to have *S. mansoni* transmission by Kabatereine *et al.* (2004) due to high altitudes with some contribution from total annual precipitation below 900 mm.

Applying similar approaches for the *"B. pfeifferi"/B. sudanica* sites, results suggested most of Uganda - with the possible exception of the north-eastern region - as suitable habitat. Thus, the absence of suitable host snails was ruled out, as the primary reason for disease absence in the south-western region of Uganda, and supports the findings that climatic features such as low night temperatures and precipitation, limits the distribution and abundance of *S. mansoni.*

Using NDVI (125-145) and T_{day} temperature ranges (T_{max} 20-33 ºC) applied by Malone *et al.* (2001a) for Ethiopia, wrongly predicted a large area around the Albert Nile to be unsuitable for *S. mansoni* transmission, largely due to average day time temperatures above 33°C in a large part of this area. Using NDVI ranges 125-145 identified for Ethiopia in a GIS query for Uganda eliminates the north-eastern region again, but this is due to the fact that NDVI values in Uganda in all other areas but this are above 145. This indicates that Uganda in general has a lesser variation in moisture regime and thus NDVI than Ethiopia, and that satellite sensor data values of NDVI and LST should be defined for the specific parasite-snail species system (Malone *et al.*, 2001a), or similar ecological zones, as also proposed by Brooker *et al.* (2001). Also, very importantly, the differences in NDVI and temperature regimes observed between the countries, could be exaggerated due to both different applied satellite sensor products (MODIS versus AVHRR) and the different time period of the AVHRR (1992-1995) products compared to the MODIS product. The MODIS satellite sensor products are only available for 2000 onwards, and *S. mansoni* transmission, even though the data is on children, could just as well have taken place due to previous years climatic conditions. In general, the limited time range of the available satellite sensor data sets must be considered a limitation when trying to define ultimate climatic ranges for schistosomiasis.

Climatologists consider longer term (30-year-average) standard climate data minimally suitable for accurately defining climate patterns (Malone *et al.*, 2001a).

Even though unable to capture the well-known small-scale focal nature of schistosomiasis, climatic range identification models based on satellite remote sensing data, can supply regional scale information to health authorities as an indication of 'permissive' geographical areas where *S. mansoni* may be found. Climate-based models would properly be more accurate in defining areas in Uganda where *S. mansoni* is unlikely to occur, thuse helping to exclude areas and narrowing focus on priority areas.

Environmental and climatic variables significantly associated with the snail-parasite system

Many environmental and climatic variables are often highly correlated, which represents a challenge and a potential problem when developing models using a large number of such variables. It can make it difficult to separate the effects of the independent variables statistically (Morgenstern, 1998). However, using logistic regression analysis, we identified which of all the available climatic and environmental variables best described the geographical variation in endemic (≥ 5% prevalence) schools in Uganda. Initial tests for colinearity revealed that NDVI and water availability measures (PRE and PPE), which would be expected to be highly correlated, since NDVI is a surrogate for water availability, only were highly correlated in the wet season and then 'only' by 48%. Elevation which was expected to be highly correlated with temperature, was only significantly correlated with the average night time temperature, T_{night}, and then only with 38%. Surprisingly, all composite temperature measures were instead very highly correlated with all NDVI composite values (85-90%), making it difficult to separate their effects statistically. However, both Spearman rank correlation analysis and logistic regression, revealed that NDVI was not significantly correlated with *S. mansoni* prevalence in Uganda. Logistic regression analysis of all environmental variables in the database showed that either a combination of 1) wet season day time temperature (T_{daywet}), precipitation (PRE_{wet}) and elevation or 2) a combination of wet season night-time temperatures ($T_{nightwet}$) and precipitation (PRE_{wet}) alone, best described the variation in endemic and non-endemic schools (Table 3). These findings further indicate that although NDVI could be used as a moisture surrogate in the composite range-finding models, precipitation from a 30-year-average standard climate data performs better statistically. Also, logistic regression indicate that much of the same information can be captured by model 2 (night time temperatures in combination with precipitation), having a similar AIC value as model 1, but with a significantly higher Hosmer-Lemeshow value, indicating a better fit. T_{night} in model 2 has an odds ratio of 2.576, indicating that an increase in night temperature by 1ºC, leads to more than twice the odds of an area being endemic. This suggests a potential use of night time temperature as an important explanatory climatic variable in future studies of similar parasite-snail systems. The max-rescaled R-square (0.4155) indicates that model 2 explains a significant proportion (41.6%) of the variation in endemic and non-endemic schools.

On average, the MODIS composite night time LST in the excluded south-western region of Uganda are below 15 ºC. Experimental studies on the temperature limits and tolerances of the intra-molluscan stages of *S. mansoni*, indicates that extended periods of temperatures below 16ºC, causes the snail host to die before schistosome cercaria can mature (Joubert *et al.*, 1986; Pflüger, 1980). The upper day time temperature limits identified by the MODIS composite models (35ºC for the wet season composite model, and 36ºC for the dry season composite model), also corresponds well to the upper constant temperature limits of 35ºC of *S. mansoni* in *B. glabrata*, reported by Pflüger (1980). It is important to point out, however, that results reported from Ethiopia by Malone *et al.* (2001a) was due to an upper limiting temperature of >27°C during gonadal development of *B. pfeifferi* snail hosts, especially in coastal areas (Appleton and Eriksson, 1984). The lack of such an effect in Uganda is supportive of the unique identification of "*B. pfeifferi*" and its relationship to *B. sudanica* there.

Thus, night time temperatures below 15ºC could be the limiting factor, in combination with the relatively low precipitation in this region. However, it should be kept in mind that schistosomiasis is a multi factorial disease, and the logistic regression model, based on regional wet season T_{night} and precipitation, only explains 41% of the variation between endemic and non-endemic school sites. Further local scale studies in this part of the country to positively confirm the absence of the disease and investigations of alternative reasons for this absence, is required.

Ideally, in future work, the predictive performances (accuracy and usefulness) of the predictive models should be carried out, by applying them to other similar ecological areas in the great lakes region, and comparing with known *S. mansoni*-snail distributional patterns. However, as pointed out by Malone *et al.* (2001a) different *S. mansoni - Biomphalaria* spp. systems properly occupy unique spatial and temporal niches in the environment, which means that each major snail-parasite system should be modeled separately. This can, however, be complicated by the current difficulties with snail species identification.

Since phylogenetic investigations indicate that "*B. pfeifferi*" in Uganda, appears to be of another species than, *B. pfeifferi* found elsewhere in Africa (Jørgensen, 2003), and phylogenetically resembles Ugandan *B. sudanica*, the models developed in the present study for Uganda, might not be very accurate or useful for other regions - even if the ecological zones are similar, as suggested by Brooker *et al.* (2001). Further investigations into the phylogeny of the *Biomphalaria* species in Uganda, along with snail sampling in the northern regions, would contribute to clarify some of these issues. Likewise, more studies investigating the climatic and environmental limits and tolerances of different *S. mansoni-Biomphalaria* spp. systems, both experimentally and in the field, could help improve model development and validation - making them more accurate and useful for health authorities to prioritize effective control programs.

Acknowledgements

The staff at the Vector Control Division in Uganda is warmly thanked for their logistic assistance during snail sampling. The study was supported by grants from "WWF Verdensnaturfonden / Novo Nordisk Biodiversitetslegatet" and the Danish Development Research Council. The Willum Kann Rasmussen Foundation is thanked for making automated sequencing facilities available at the DBL - Institute for Health Research and Development.

Karen Gruszynski is thanked for her technical assistance with the MODIS satellite imagery processing during a 1-month GIS training visit to Louisiana State University by the senior author.

References

Appleton CC (1978). Review of Literature on abiotic factors influencing the distribution and life-cycles of Bilharziasis intermediate host snails. Malacological Review 11: 1–25.

Appleton CC, Eriksson IM (1984). The influence of fluctuating above-optimal temperature regimes on the fecundity of *Biomphalaria pfeifferi* (Mollusca: Planorbidae). Trans R Soc Trop Med Hyg 78(1): 49-54.

Bozdogan H (1987). Model selection and Akaikes Information Criterion (AIC): the general theory and its analytical extensions. Psychometrika 52: 163-180.

Brooker S, Michael E (2000). The potential of geographical information systems and remote sensing in the epidemiology and control of human helminth infections. Adv Parasitol 47: 245–288.

Brooker S, Hay SI, Issae W, Hall A, Kihamia CM, Lwambo NJS, Wint W, Rogers DJ, Bundy DAP (2001). Predicting the distribution of urinary schistosomiasis in Tanzania using satellite sensor data. Trop Med Intern Health 6: 998-1007.

Brooker S (2002). Schistosomes, snails and satellites. Acta Trop 82: 207–214.

Brooker S, Kabatereine NDB, Clements ACA, Stothard JR (2004). Schistosomiasis control. Comment in The Lancet 363 (9409): 658-659.

Brown DS (1994). Freshwater snails of Africa and their medical importance, 2nd Ed. Taylor & Francis, London.

Burnham KP, Anderson DR (2002). Model Selection and Multimodel Inference: a practical information-theoretic approach, 2nd edition. Springer-Verlag, New York.

Cox J, Craig MH, le Sueur D, Sharp BL (1999). Mapping Malaria Risk in the Highlands of Africa (Durban, South Africa), Mapping Malaria Risk in Highlands of Africa, MARA/HIMAL Technical report, December 1999.

Hay SI, Randolph SE, Rogers DJ (Eds) (2000). Remote Sensing and Geographical Information Systems in Epidemiology. Academic Press, London. Advances in Parasitology 47.

Hosmer DW, Lemeshow S (1989). Applied Logistic Regression. New York: John Wiley & Sons, Inc.

Joubert PH, Pretorius SJ, de Knock KN, van Eeden JA (1986). Survival of *Bulinus africanus* (Krauss), *Bulinus globosus* (Morelet) and *Biomphalaria pfeifferi* (Krauss) at constant high temperatures. South African Journal of Zoology 21: 85–88.

Jørgensen A (2003). Diversity and phylogeny of African freshwater gastropods with special emphasis upon the molecular phylogeny of “ancyloplanorbidae” and *Biomphalaria* and snail biodiversity within the Great East African Lakes. Ph. D Thesis to the Faculty of Science, University of Copenhagen, December 2003.

Kabatereine NB, Brooker S, Tukahebwa EM, Kazibwe F, Onapa A (2004). Epidemiology and geography of *Schistosoma mansoni* in Uganda: implications for planning control. Tropical Medicine and International Health 9: 372-380.

Kristensen TK (1987). A field guide to African Freshwater Snails 2: East African Species. 2nd edition. Charlottenlund, Danish Bilharziasis Laboratory.

Mandahl-Barth G (1954). The freshwater mollusks of Uganda and adjacent territories. Annls. Mus. r. Congo belge, Sér. 8°. 32: 1-206.

Mandahl-Barth G (1957a). Intermediate hosts of *Schistosoma*. African *Biomphalaria* and *Bulinus*: 1. *Biomphalaria*. Bull Wld Htlh Org 16: 1103-1163.

Mandahl-Barth G (1957b). Intermediate hosts of *Schistosoma*. African *Biomphalaria* and *Bulinus*: 2. *Bulinus*. Bull Wld Hlth Org 17: 1-65.

Malone JB, Huh OK, Fehler DP, Wilson PA, Wilensky DE, Holmes RA, Elmagdoub AI (1994). Temperature data from satellite imagery and the distribution of schistosomiasis in Egypt. American Journal of Tropical Medicine and Hygiene 50: 714-722.

Malone JB, Yilma JM, McCarroll JC, Erko B, Mukaratirwa S, Zhou X (2001a). Satellite climatology and the environmental risk of *Schistosoma mansoni* Ethiopia and east Africa. Acta Trop 79: 59-72.

Malone JB, McCarroll JC, Kristensen TK, Yilma JM, Erko B, El Bahy MM, Corbett JD (2001b). Minimum Medical Database Spatial Decision Support System for the Inter-Governmental Authority on Development-Nile Basin Region. Manual and CDROM. (www. GnosisGIS.org): 49 p.

McNally K (2003). Developing risk assessment maps for *Schistosoma haematobium* based on climate grids and remotely sensed data. MS Thesis. Louisiana State University, 42 pp.

Morgenstern H (1998). Ecologic Studies. In: Rothman K, Greenland S (eds). Modern Epidemiology, 2nd edn. Philadelphia: Lippincott-Raven: 59–80 pp.

Nagelkerke NJD (1991). Miscellanea: A note on a general definition of the coefficient of determination. Biometrik 78: 691-692.

NBSP (National Biomass Study Project) (1995). Technical Report National Biomass Study. Forestry Department. Ministry of Natural Resources. Kampala, Uganda.

Prentice MA (1972). Distribution, prevalence and transmission of schistosomiasis in Uganda. Uganda Medical Journal 1: 136–139.

Pflüger W (1980). Experimental epidemiology of schistosomiasis I. The prepatent period and cercarial production of *Schistosoma mansoni* in *Biomphalaria* snails at various constant temperatures. Zeitschrift für Parasitenkunde 63: 159–169.

Ripley BD (1996). Pattern recognition and neural networks: Cambridge University Press, Cambridge, UK.

Sturrock RF (1993). The intermediate hosts and host–parasite relationships. In: Jordan P, Webbe G, Sturrock RF (eds) Human Schistosomiasis, CAB International, Wallingford, 33–85 pp.