Geospatial Health 2015; volume 10:398

Use of an ecologically relevant modelling approach to improve remote sensing-based schistosomiasis risk profiling

Yvonne Walz,[1,2] Martin Wegmann,[1] Benjamin Leutner,[1] Stefan Dech,[1,3] Penelope Vounatsou,[4,5] Eliézer K. N'Goran,[6,7] Giovanna Raso,[4,5] Jürg Utzinger[4,5]

[1]Department of Remote Sensing, Institute for Geography and Geology, University of Würzburg, Würzburg; [2]Institute for Environment and Human Security, United Nations University, Bonn; [3]German Remote Sensing Data Centre, German Aerospace Centre, Oberpfaffenhofen, Germany; [4]Department of Epidemiology and Public Health, Swiss Tropical and Public Health Institute, Basel; [5]University of Basel, Basel, Switzerland; [6]Unité de Formation et de Recherche Biosciences, Université Félix Houphouët-Boigny, Abidjan; [7]Centre Suisse de Recherches Scientifiques en Côte d'Ivoire, Abidjan, Côte d'Ivoire

Correspondence:Yvonne Walz, Department of Remote Sensing, Institute for Geography and Geology, University of Würzburg, Oswald-Külpe-Weg 86, 97074 Würzburg, Germany.
Tel: +49.931.310.
E-mail: yvonne.walz@uni-wuerzburg.de

Key words: Spatial risk profiling; Schistosomiasis; Remote sensing; Ecological relevant model; Côte d'Ivoire.

Acknowledgements: the schistosomiasis prevalence data were derived from the open-access Global Neglected Tropical Disease database (GNTD; http://www.gntd.org). We thank the German Aerospace Centre for providing RapidEye data from the RapidEye science archive. The Moderate Resolution Imaging Spectroradiometer (MODIS) data products are courtesy of the online Data Pool at the National Aeronautics and Space Administration (NASA) Land Processes Distributed Active Archive Center (LP DAAC), USGS/Earth Resources Observation and Science (EROS) Center, Sioux Falls, South Dakota, United States of America (https://lpdaac.usgs.gov/data_access). This publication was funded by the German Research Foundation and the University of Würzburg in the funding programme Open Access Publishing. We thank Michael Hagenlocher for comments on an earlier draft of the manuscript.

Contributions: YW, MW, BL, SD, PV, EKN, GR and JU participated in the design of the study. EKN, GR and JU participated in the collection of the data in Côte d'Ivoire. PV had overall responsibility of the integrity of the GNTD database. YW conducted the analyses and wrote the first draft of the manuscript. MW, BL, GR and JU assisted with the interpretation of results and revised the manuscript. All authors read and approved the manuscript prior to submission.

Conflict of interest: the authors declare no potential conflict of interest.

Funding: this work was financed through a scholarship granted to YW from the *Equal Opportunities for Women in Research and Training* programme at the University of Würzburg, Germany.

Received for publication: 17 July 2015.
Revision received: 12 October 2015.
Accepted for publication: 23 October 2015.

Licensee PAGEPress, Italy
Geospatial Health 2015; 10:398
doi:10.4081/gh.2015.398

Abstract

Schistosomiasis is a widespread water-based disease that puts close to 800 million people at risk of infection with more than 250 million infected, mainly in sub-Saharan Africa. Transmission is governed by the spatial distribution of specific freshwater snails that act as intermediate hosts and the frequency, duration and extent of human bodies exposed to infested water sources during human water contact. Remote sensing data have been utilized for spatially explicit risk profiling of schistosomiasis. Since schistosomiasis risk profiling based on remote sensing data inherits a conceptual drawback if school-based disease prevalence data are directly related to the remote sensing measurements extracted at the location of the school, because the disease transmission usually does not exactly occur at the school, we took the local environment around the schools into account by explicitly linking ecologically relevant environmental information of potential disease transmission sites to survey measurements of disease prevalence. Our models were validated at two sites with different landscapes in Côte d'Ivoire using high- and moderate-resolution remote sensing data based on random forest and partial least squares regression. We found that the ecologically relevant modelling approach explained up to 70% of the variation in *Schistosoma* infection prevalence and performed better compared to a purely pixel-based modelling approach. Furthermore, our study showed that model performance increased as a function of enlarging the school catchment area, confirming the hypothesis that suitable environments for schistosomiasis transmission rarely occur at the location of survey measurements.

Introduction

Schistosomiasis is a neglected tropical disease caused by blood flukes of the genus *Schistosoma*. Schistosomiasis is considered the most important water-based disease. In mid-2003, an estimated 779 million people were at risk of infection and recent estimates suggest that more than 250 million people are infected, mainly in sub-Saharan Africa (Steinmann *et al*., 2006; Hotez *et al*., 2014). Human infection with schistosomes occurs through contact with freshwater bodies where specific snails proliferate that act as intermediate hosts.

School-aged children are at highest risk of infection (Bundy and Blumenthal, 1990), and hence, they are the main target group for epidemiological surveys and control interventions (WHO, 2002).

Remote sensing data have been widely used for spatially explicit risk profiling of schistosomiasis (Seto *et al*., 2002; Stensgaard *et al*., 2005; Brooker, 2007, Simoonga *et al*., 2009; Walz *et al*., 2015). The overarching aim is to support prevention and control, focussing on high-risk areas. In most of the previous applications, *Schistosoma* infection prevalence estimates at the unit of the school were modelled in a pixel-based approach using remotely sensed environmental data for schistosomiasis risk prediction over space (Walz *et al*., 2015). The underlying assumption is that information about the presence and successful completion of the parasite life cycle – features which are implicit in disease prevalence – are linked to proxies of environmental suitability for intermediate host snails and parasites. However, the approach to link school-based schistosomiasis prevalence data with environmental measures obtained from remote sensing has an inherent shortcoming, as shown in Figure 1. Indeed, the location of an epidemiological survey (*i.e*. school) does not spatially superimpose with the location of potential disease transmission. In order to obtain relevant information from remote sensing data to characterise disease-relevant environmental conditions, the information from remote sensing data has to be extracted at the exact location of potential disease transmission sites, which is different from the school location. To our knowledge, this spatial discrepancy between model components as illustrated in Figure 1 has not been addressed in previous schistosomiasis risk profiling studies.

Social-ecological processes that govern schistosomiasis transmission – similar to nearly all ecological phenomena – operate across different scales and vary with the scale of observation (Levin, 1992; Schur *et al*., 2012). Thus far, models of schistosomiasis risk have mainly been developed at national, regional and continental scales (Clements *et al*., 2008; Simoonga *et al*., 2009; Schur *et al*., 2013; Chammartin *et al*., 2014; Lai *et al*., 2015) using low spatial resolution remote sensing data (Walz *et al*., 2015). At this scale of observation, it has been concluded that, besides climatic conditions, the impact of humans on habitats of intermediate host snails govern the pattern of schistosomiasis risk (Brooker, 2007; Stensgaard *et al*., 2013). Only few studies modelled schistosomiasis risk at the local scale (Booth *et al*., 2004; Clennon *et al*., 2004; Raso *et al*., 2005; Beck-Wörner *et al*., 2007). These studies showed that topographic variables and the spatial relation between households and water bodies are important in explaining the spatial heterogeneity of disease transmission at the micro-scale (Simoonga *et al*., 2009). Hence, statistical correlation can vary considerably according to the extent of observed area and scale of aggregation (Marshall, 1991), which has been identified as a major constraint of remote sensing-based schistosomiasis risk models (Herbreteau *et al*., 2007).

Two of the most widely used remote sensing variables for schistosomiasis risk profiling are the normalized difference vegetation index (NDVI) and land surface temperature (LST) (Walz *et al*., 2015). An important reason is that both variables are pre-processed, and hence, they are readily available (Herbreteau *et al*., 2007). Additional spectral indices can be calculated from remote sensing data, which might improve models as they are directly linked to the ecology of disease transmission (Herbreteau *et al*., 2007). Important aspects that impact remote-sensing measurements besides the surface conditions *per se* are the composition of the landscape regarding size and heterogeneity of relevant features, such as water bodies and riparian structures. It has been shown that the relevance of remote sensing variables for modelling schistosomiasis transmission varies between different ecozones and has an impact on the model outcome and transferability (Brooker *et al*., 2001). The purpose of this paper is to propose and validate an approach to overcome the spatial discrepancy between model components inherent in pixel-based remote sensing schistosomiasis risk profiling. To this end we propose an approach based on school-catchment with potential refinements by constraining modelling only to potential disease transmission sites. Furthermore, the effect of the scale of observation is investigated through analysis of remote sensing based environmental data within different buffer sizes around school locations. The importance of specific remote sensing-variables is assessed for two different landscapes in Côte d'Ivoire using a suite of vegetation- and water-related indices and other variables relevant to the disease transmission ecology.

Materials and Methods

Ethics statement

We present a secondary analysis with data derived from previously published studies that had received ethical approval by the respective institutional review boards and national ethics committees (see for study site Man (Utzinger *et al*., 2000) and (Raso *et al*., 2005), and for study site Taabo (N'Goran *et al*., 1997, 1998). The studies were conducted according to national and international guidelines. All the data received for the current analyses were anonymised.

Study area

The study area comprises two sites in Côte d'Ivoire: i) Taabo with a surface of 8476 km² situated in south-central Côte d'Ivoire around the man-made Lake Taabo, and ii) Man with a surface of 4381 km² in a mountainous region in western Côte d'Ivoire. Both sites are located in the tropical rainforest zone. The temperature is characterised by a diurnal climate with higher temperature ranges between day and night than between seasons of the year. The mountainous climate zone in Man shows one extensive wet season (March-October), whereas two peaks of precipitation are typical in the study site Taabo (March-July and September-October) (Dipama, 2010). Lake Taabo has

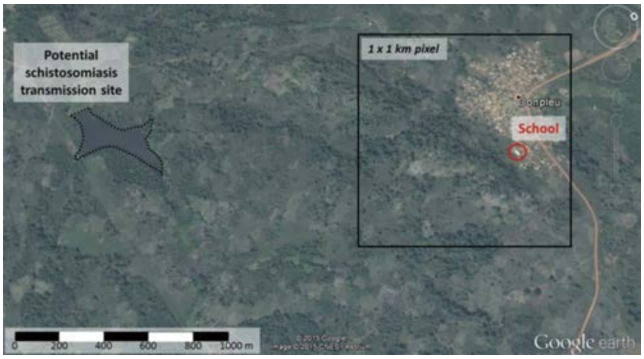

Figure 1. Illustration of the spatial discrepancy among the measurement of *Schistosoma* infection prevalence at the unit of the school (surrounded in red), the location where disease transmission has potentially occurred (freshwater body surrounded by a black dotted line), and the currently used pixel-based measure of remote sensing data given for the example of a 1x1 km pixel in Dompleu, Côte d'Ivoire (located within the study site in the Man region).

been impounded in the late-1970s, mainly for agricultural and hydropower production purposes (Savane, 2010). The study area is an endemic schistosomiasis region, predominantly by *S. haematobium*, where moderate to high transmission rates are observed (N'Goran *et al*., 1997; WHO, 2010). In the Man region, *S. mansoni* is the main schistosome species (Utzinger *et al*., 2000; Raso *et al*., 2005; Chammartin *et al*., 2014; Assaré *et al*., 2015). The spatial distribution of disease prevalence shows a typical focal pattern and ranges from 0 to 100% prevalence in both study sites.

Data

For this study, 6.5 m resolution RapidEye data (Blackbridge, 2015) with cloud coverage below 5%, a 30 m global digital elevation model (GDEM) of the Advanced Spaceborne Thermal Emission and Reflection radiometer (ASTER) (JSS, 2014), and multi-temporal data from the Moderate Resolution Imaging Spectroradiometer (MODIS; 250 m and 1 km) were used. Pre-processing of RapidEye data involved orthorectification, and atmospheric and topographic correction using Atcor (Richter and Schläpfer, 2012) within the Catena pre-processing tool of the German Aerospace Centre (DLR, 2014). The ASTER GDEM data provided the basis for topographic analyses given a vertical root mean square error (RMSE) of 8.68 m (Meyer, 2011). Terra MODIS products are provided as level 3 data and therefore corrected for radiometric, geometric, atmospheric and bi-directional effects (Vermote and Vermeulen, 1999). The Terra MODIS LST product has been validated for multiple validation sites with wide ranges of surface and atmospheric conditions and showed a deviation of ±1 K for temperatures between 263 and 323 K (Wan *et al*., 2004). For reasons of spatial prediction of the models, all datasets were resampled to 30 m resolution, using nearest neighbour resampling for MODIS and bilinear resampling for RapidEye. Table 1 summarises all remote sensing variables used for this study.

Data on human infection with schistosomiasis were derived from the open-access Global Neglected Tropical Disease database (GNTD; http://www.gntd.org) (Hürlimann *et al*., 2011). Overall, 75 survey measurements were extracted for the Man region and 38 for Taabo. The GNTD database includes relevant information such as number of people examined, number of people infected with species-specific schistosomes, date of survey, description of the survey location, method used for sample recruitment and diagnostic techniques (Hürlimann *et al*., 2011). Pre-selecting steps of parasitological data included the removal of surveys before 1980 to consider the maximal life span of the parasites (30 years) (Jordan and Webbe, 1993). Surveys that were not conducted at schools and that used insensitive diagnostic techniques (*e.g.* direct faecal smear for detection of *S. mansoni* eggs) were removed from the database, as explained elsewhere (Schur *et al*., 2011).

Statistical models

Data were analysed by means of two different statistical algorithms: i) a non-parametric random forest (RF) algorithm (Breiman, 2001), and ii) a parametric partial least squares regression (PLSR) (Wold *et al*., 1984). Both model algorithms are, by design, capable to deal with the prediction of continuous response data, low numbers of training samples and multi-collinear predictor variables. Prevalence of *Schistosoma* infection (continuously scaled between 0 and 100%) was the response variable to be explained by predictor variables consisting of remotely sensed environmental variables (Table 1). As a prerequisite for the PLSR model, all variables were standardised using the z-score transformation, which centres the data in relation to their mean value and scales them to unit variance (Wold *et al*., 2001).

Modelling procedures

In this study, an ecologically relevant approach to model schistosomiasis risk using remote sensing data was developed. The underlying hypothesis of this approach postulates that solely aquatic sites and their vicinity or sites where water may potentially accumulate are ecologically relevant and should contribute to the model. Furthermore, environmental data need to reflect habitat requirements of parasites and intermediate host snails, and consequently to schistosomiasis transmission. We selected a buffer around the school locations to represent the catchment area (Malone *et al*., 2001; Steinmann *et al*., 2006) and masked all areas where water: i) was not detected from remote sensing images during dry and at the end of the wet season, and ii) cannot potentially accumulate due to topography. This mask excludes areas where disease transmission cannot occur (Figure 2). The results of this ecologically relevant model approach were compared to results from the full catchment model approach, where no mask of potential disease transmission sites was applied (Figure 2) in

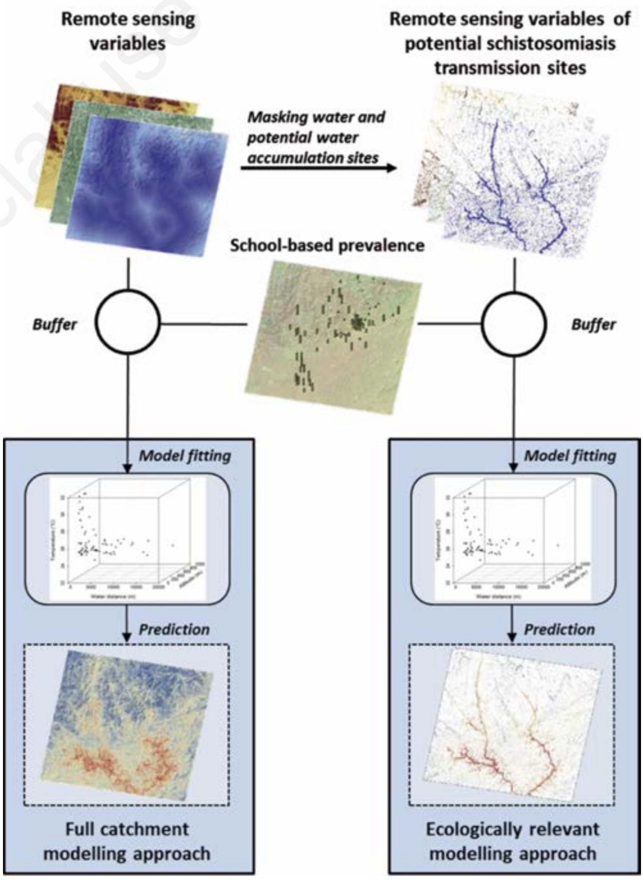

Figure 2. Flowchart of the ecologically relevant and the full catchment modelling approaches for schistosomiasis risk profiling. The ecologically relevant modelling approach considers only potential disease transmission sites for spatially explicit risk profiling of schistosomiasis using remotely sensed environmental variables, whereas the full catchment modelling approach is based on remote sensing measurements for a selected buffer zone around the school location.

order to determine the effect of transmission site selection.

The impact of the selected extent of the school catchment area was investigated by spatial buffer analysis for a defined area around the point measurement of the school location. The models were processed for buffer radii ranging from 0 to 5000 m at steps sequentially increasing by 100 m. The ecologically relevant approach was only feasible for radii of at least 500 m. At lower radii, there were no transmission sites, as explained by too close proximity to the school location. The remote sensing variables were aggregated by mean values for each buffer extent, except for the variables settlements and streams, which were aggregated by their sum (area of settlements) and maximum (stream order), respectively.

In order to determine which remote sensing variables explain the spatial distribution of schistosomiasis prevalence, each predictor's individual variable importance was calculated using the mean decrease in accuracy measure implemented in the RF algorithm. This measure calculates for each predictor variable how much it reduces node impurity, which is calculated as difference between the residual sum of squares before and after the split and summed up over all splits and trees (Hastie *et al*., 2009). Moreover, differences in variable importance between the two study sites were compared.

All RF and PLSR model approaches have been validated through internal cross-validation using the *caret* package (Kuhn, 2008) in R (R Core Team, 2012). The measures used to evaluate model quality were the coefficient of determination (R^2) and the root mean squared error (RMSE). The data splitting for the internal cross-validation was adjusted to the sample size of each study site and resulted in 10-fold cross-validation at the Man study site and 5-fold cross-validation in Taabo. This ensured a comparable number of validation data per group of test data. The cross-validation procedure was repeated 10 times for each model and the best model defined through the minimum RMSE was selected as the final model. Within the cross-validation procedure, the model-algorithm specific parameters *mtry* (=number of predictors tried per split) and *ncomp* (=number of components used for prediction) were iterated to tune the RF (Liaw and Wiener, 2002) and PLSR model (Geladi and Kowalski, 1986), respectively.

Results

Figure 3 depicts the performance of schistosomiasis risk models using the RF and PLSR algorithms at the two study sites in Côte d'Ivoire. The graphics illustrate model fits of the ecologically relevant and the full catchment modelling approaches for all buffer radii from 0 to 5000 m in steps of 100 m. For both PLSR models (Taabo and Man), the selected buffer radius around the school location revealed a considerable effect on model accuracy, as indicated by lowest R^2 values if environmental variables were directly extracted at the school location and higher R^2 values as buffer areas around the school location increased. However, this positive effect of increasing buffer sizes on model performance was not recognisable by the RF algorithm. The ecologically relevant model approach yielded significantly better results on average for the Taabo site compared to the full catchment model approach ($P<0.001$, paired Wilcoxon signed rank test). At the Man study site, this superior performance of the ecologically relevant model approach was not reflected. Instead, both models were not significantly different for the PLSR model ($P=0.78$) and the full catchment approach performed better with the RF model ($P<0.001$). Model performance was generally better in the Taabo site compared to the Man region. The ecologically relevant modelling approach reached a maximum explanatory power for a selected buffer radius between 4 and 5 km in Taabo, where the PLSR model explained up to 70% of the spatial distribution of *Schistosoma* infection prevalence. Meanwhile, the RF algorithm performed slightly worse than PLSR, as indicated by lower R^2 values and explaining a maximum of 67% of the observed distribution. For the case of the full catchment modelling approach, the maximum explanatory power was also achieved for the PLSR model in Taabo, however, with a generally lower R^2 of up to 0.63.

The RMSE of different modelling approaches and algorithms at the study sites reflected the trends described in Figure 3 and are illustrated in Figure 4. It confirmed – similar to the maximum R^2 values – a minimum error of 36.3% of the predicted prevalence for the PLSR modelling approach with a selected buffer radius between 4 and 5 km in the Taabo site. Our study explored a multitude of schistosomiasis-relevant remote sensing variables listed in Table 1 and assessed their individual contribution to explain the spatial distribution of disease risk. Figure 5 shows the importance of the most relevant variables for the ecologically relevant modelling approach in the two study sites and highlights that variable importance differs between the different investigated settings in Côte d'Ivoire. For example, the variable *Euclidean distance from nearest freshwater body* was highly important for all school catchment radii in the Man region and had comparably little importance for the model in the Taabo site. The topographic variables *altitude*, *slope*, *sinks* and *streams* were of high importance to explain schistosomiasis risk in Man, but did not contribute to model disease risk in Taabo. At this site, the spatial prediction of schistosomiasis risk was mainly explained by the mean and median of Terra MODIS LST, which was of considerably lower or no importance at all in the Man region. There are no conspicuous differences between the

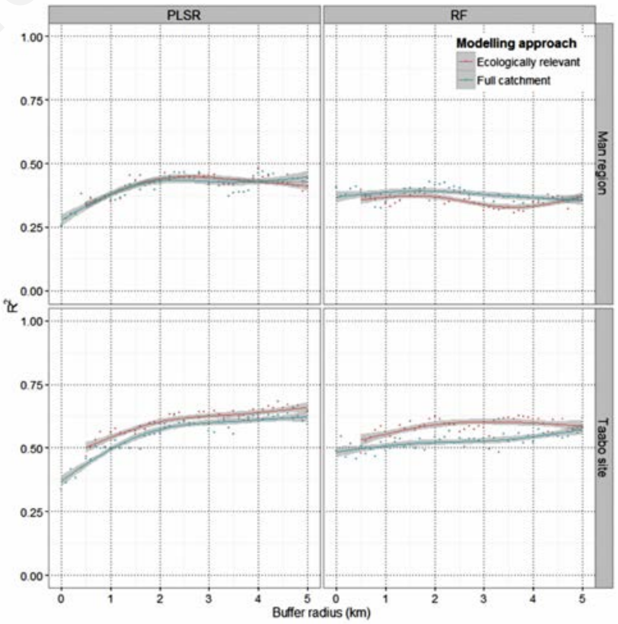

Figure 3. Performance of statistical schistosomiasis risk models using the partial least squares regression (PLSR) and random forest (RF) model algorithms for the two study sites of Man and Taabo in Côte d'Ivoire. Each graph illustrates the model performance indicated by the R^2 for the selected buffer radius around the school location and compares the result of the ecologically relevant (red) and the full catchment (blue) modelling approaches. The lines reflect the results of spline interpolation and the grey bar represents the confidence interval. Note that the extraction of the pixel value with no buffer corresponds to 0 m buffer radius.

ecological and full catchment modelling approaches with regard to the variable importance, except for the marked higher importance of the variable sink depth in the full catchment model in the Man region. This might correspond to the RF algorithm identifying parts of the mask of potential disease transmission sites automatically.

The spatial predictions of the most accurate PLSR models based on the ecologically relevant modelling approach are shown in Figure 6. In general, the spatial predictions depict the risk of infection at potential disease transmission sites. For the Taabo site, the model predicted Lake Taabo in the centre of the study site and several sections of the Bandama River running from north-west to south-east of the study setting crossing Lake Taabo as areas where the risk of schistosomiasis transmission is highest. The spatial prediction of schistosomiasis prevalence in the Man region predicted low disease risk in the mountainous part and high risk along the river valleys running towards the south of the study site.

Discussion

The proposed ecologically relevant modelling approach used in combination with a carefully selected school catchment buffer allows bridging the spatial gap between the measurement of disease prevalence and relevant environmental conditions for schistosomiasis transmission. The issue of anisotropy inherent in the spatial structure of prevalence data (Chammartin *et al*., 2013) is directly addressed by this ecologically relevant modelling approach. The buffer analysis pursued here underscores that the extent of the area observed around a school location has a major effect on the performance of a remote sensing-based schistosomiasis risk model. The improvement of model performance with the observation of an increasing school catchment area confirms the spatial conflict of schistosomiasis risk profiling using an environmental measurement at the unit of the school. The results of the PLSR model (Figures 3 and 4) indicate that starting from a distance of 2 km and up to 5 km, the school catchment area integrates more relevant environmental conditions that represent schistosomiasis transmission conditions corresponding to school-based measures of prevalence. The effect of different scales of aggregation on model performance has previously been shown for aggregating data on different administrative levels (pixel, district and national), where treatment needs were either over- or underestimated depending on the level of spatial aggregation and focality of disease distribution (Schur *et al*., 2012). A general improvement of the previously applied pixel-based approach at the location of schools is provided by this study, both conceptually and from a point of view of model performance. From the 32 remote sensing variables investigated for schistoso-

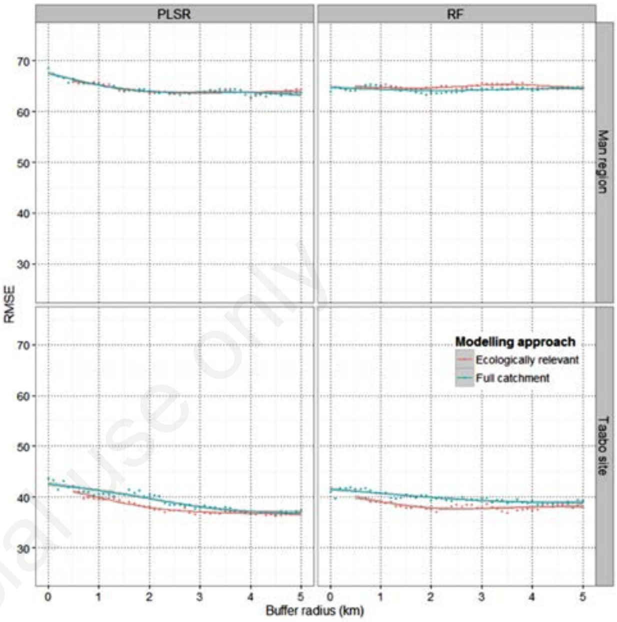

Figure 4. Performance of statistical schistosomiasis risk models using the partial least squares regression (PLSR) and random forest (RF) model algorithms for the two study sites of Man and Taabo in Côte d'Ivoire. Each graph illustrates the model performance indicated by the root mean square error (RMSE) for the selected buffer radius around the school location and compares the result of the ecologically relevant (red) and the full catchment (blue) modelling approaches. The lines reflect the results of spline interpolation and the grey bar represents the confidence interval. Note that the extraction of the pixel value with no buffer corresponds to 0 m buffer radius.

Table 1. Overview of remote sensing data and variables used for schistosomiasis risk modelling in this study.

Data source	Variables		Spatial resolution (m)	Time of acquisition	Reference(s)
RapidEye	Spectral reflectance band 1-5		6.5	3 January 2015 (Taabo), 5 January 2015 (Man)	Blackbridge (2015)
	Spectral indices	NDVI		-	
		NDWI			McFeeters (1996)
		EVI			Huete *et al*. (2002)
		SAVI			Huete (1988); Huete and Liu (1994)
		MSAVI			Qi *et al*. (1994)
	Euclidean distance to water, settlement area				-
ASTER GDEM	Altitude, slope, sink depth, stream order		30	July 2009 - September 2011	(JSS, 2014)
MODIS (MOD13Q1)	NDVI, EVI (minimum, maximum, mean, median)		250	January 2010-December 2010	Huete *et al*. (1999)
MODIS (MOD11A2)	LST and ΔLST (day/night) (minimum, maximum, mean, median)		1000	January 2010-December 2010	Wan (1999)

NDVI, normalized difference vegetation index; NDWI, normalized difference water index; EVI, enhanced vegetation index; SAVI, soil-adjusted vegetation index; MSAVI, modified soil-adjusted vegetation index; LST,

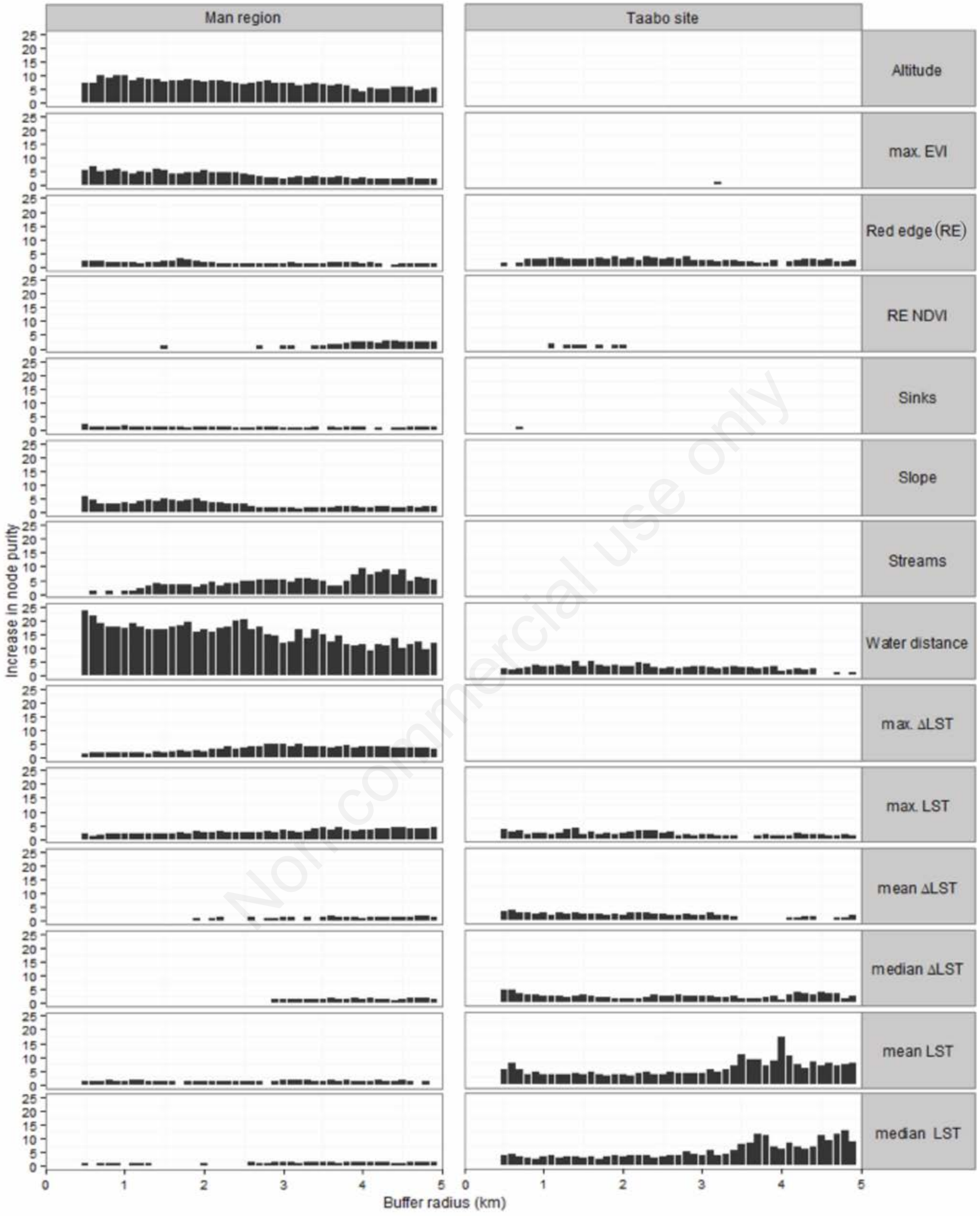

Figure 5. Variable importance of selected remote sensing variables for each buffer radius in the two study sites of Man and Taabo in Côte d'Ivoire, illustrated for the ecologically relevant model. The measure of variable importance is indicated by the increase in node purity as implemented in the random forest algorithm. All variables are listed in Table 1.

miasis risk profiling in this study (Table 1), there were only few that contributed to an explanation of the spatial pattern of schistosomiasis. Furthermore, the importance of remote sensing variables varied strongly from one study site to another (Figure 5). In the mountainous region around the town of Man, schistosomiasis risk has predominantly been explained by the distance to freshwater bodies and topographic variables, which fits to the strong topographic gradient and confirms the positive correlation between schistosomiasis prevalence and stream order of the nearest water body or altitude as derived by Beck-Wörner *et al*. (2007). The variable sink depth was identified as new important variable by the full catchment modelling approach in the Man region. Topographic variables are considered as proxies for relevant ecological conditions for schistosomiasis-related parasite and intermediate host snail species, such as water accumulation, flow velocity and dispersal (Beck-Wörner *et al*., 2007; Clennon *et al*., 2007ì). In the more levelled lake-side region around Taabo, mean and median LST and temperature difference between day- and night-time as well as distance to the closest freshwater bodies were the most important variables to explain the spatial distribution of schistosomiasis risk (Figure 5). As shown before by Malone *et al*. (1994) for the Nile delta in Egypt, remotely sensed temperature is a useful proxy to model hydrological conditions. The differential importance of variables according to study site indicates that valuable information from remote sensing data for specific landscapes exists. Their importance might blur when applied in models across different regions, where different relationships dominate. This confirms the result of Brooker *et al*. (2001) in Tanzania, where only those models that were fitted within the same ecological zone showed reasonable performance. The high spatial resolution of RapidEye (6.5 m) and ASTER GDEM (30 m) were relevant to delineate the potential disease transmission sites for this ecologically relevant modelling approach. However, the statistical model revealed that the variables that best represent the specific landscape characteristics with regard to potential habitats of parasites and intermediate host snails were important explanatory variables. As RapidEye data do not provide information on either topography or temperature, data from this sensor did not contribute relevant variables for the statistical models.

The comparison between the ecologically relevant and the full catchment modelling approaches has shown that the former approach improved schistosomiasis risk models in the study site Taabo only. Analyses demonstrated that close to 70% of the variance could be explained by the ecologically relevant model approach using remote sensing data only. Given the complex social-ecological interaction that underlies schistosomiasis transmission (Utzinger *et al*., 2011), remote sensing can provide a valuable contribution for schistosomiasis risk assessment in this study region. However, the comparably poor performance of schistosomiasis risk models in the Man region could be related to the difficulty in detecting water bodies due to its small-scale heterogeneity and frequent coverage by riparian forests. Additionally, the contribution of socioeconomic and cultural factors seems more relevant than environmental variables to explain the spatial risk of schistosomiasis on the local scale in this part of Côte d'Ivoire (Raso *et al*., 2005).

Conclusions

The superior model performance of the ecologically relevant modelling approach in the Taabo site was well reflected by an appropriate spatial prediction of disease risk according to the distribution of reference data. However, the predicted disease risk of single hotspots not in the vicinity of the well-investigated water sites around Lake Taabo cannot be evaluated due to the lack of validation data on disease preva-

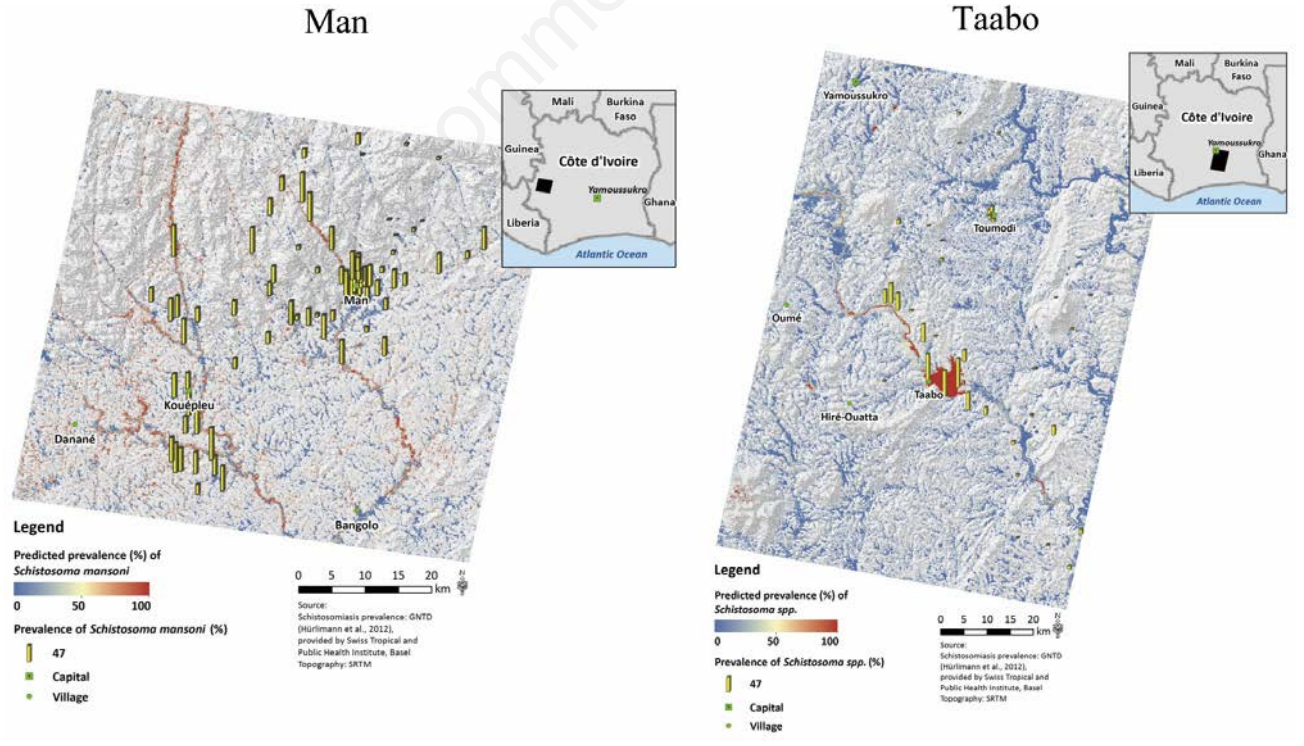

Figure 6. Spatial prediction of schistosomiasis prevalence for the study sites Man and Taabo in Côte d'Ivoire, based on an ecologically relevant model approach. This spatial prediction was derived from the best fitted models with a buffer radius of 4000 m and 4400 m for the two study sites of Man and Taab, respectively. The height of the yellow bar reflects the level of prevalence between 0 and 100%.

lence. The ecologically relevant modelling approach explicitly models the potential disease transmission sites, which generally avoids the prediction of high disease risk at non-suitable areas such as steep topographic elevations, as found isolated in risk predictions of the full catchment model approach. This spatial prediction can support the planning of disease prevention and control measures in a spatially explicit manner and supports the identification of new causal relations of disease transmission in different ecological regions. It will be interesting to further validate our models in other schistosomiasis - endemic parts of sub-Saharan Africa and to monitor model performance in the face of enhanced control programme activities.

References

Assaré RK, Lai YS, Yapi A, Tian-Bi YNT, Ouattara M, Yao PK, Knopp S, Vounatsou P, Utzinger J, N'Goran EK, 2015. The spatial distribution of *Schistosoma mansoni* infection in four regions of western Côte d'Ivoire. Geospat Health 10:69-79.

Beck-Wörner C, Raso G, Vounatsou P, N'Goran EK, Rigo G, Parlow E, Utzinger J, 2007. Bayesian spatial risk prediction of *Schistosoma mansoni* infection in western Côte d'Ivoire using a remotely-sensed digital elevation model. Am J Trop Med Hyg 76:956-63.

Blackbridge. 2015. Satellite imagery product specifications. Available from: www.blackbridge.com/rapideye/upload/RE_Product_Specifications_ENG.pdf

Booth M, Vennervald BJ, Kenty L, Butterworth AE, Kariuki HC, Kadzo H, Ireri E, Amaganga C, Kimani G, Mwatha JK, Otedo A, Ouma JH, Muchiri E, Dunne DW, 2004. Micro-geographical variation in exposure to *Schistosoma mansoni* and malaria, and exacerbation of splenomegaly in Kenyan school-aged children. BMC Infect Dis 4:13.

Breiman L, 2001. Random forests. Mach Learn 45:5-32.

Brooker S, 2007. Spatial epidemiology of human schistosomiasis in Africa: risk models, transmission dynamics and control. T Roy Soc Trop Med H 101:1-8.

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

Bundy DAP, Blumenthal U, 1990. Human behaviour and the epidemiology of helminth infection. In: C Barnard, JM Behnke (eds.) Parasitism and host behaviour. Taylor and Francis, London, UK, pp 264-89.

Chammartin F, Houngbedji CA, Hürlimann E, Yapi RB, Silué KD, Soro G, Kouamé FN, Utzinger J, Raso G, Vounatsou P, 2014. Bayesian risk mapping and model-based estimation of *Schistosoma haematobium-Schistosoma mansoni* co-distribution in Côte d'Ivoire. PLoS Neglect Trop D 8:e3407.

Chammartin F, Hürlimann E, Raso G, N'Goran EK, Utzinger J, Vounatsou P, 2013. Statistical methodological issues in mapping historical schistosomiasis survey data. Acta Trop 128:345-52.

Clements ACA, Garba A, Sacko M, Touré S, Dembelé R, Landouré A, Bosqué-Oliva E, Gabrielli AF, Fenwick A, 2008. Mapping the probability of schistosomiasis and associated uncertainty, West Africa. Emerg Infect Dis 14:1629-32.

Clennon JA, King CH, Muchiri EM, Kariuki HC, Ouma JH, Mungai P, Kitron U, 2004. Spatial patterns of urinary schistosomiasis infection in a highly endemic area of coastal Kenya. Am J Trop Med Hyg 70:443-8.

Clennon JA, King CH, Muchiri EM, Kitron U, 2007. Hydrological modelling of snail dispersal patterns in Msambweni, Kenya and potential resurgence of *Schistosoma haematobium* transmission. Parasitology 134:683-93.

Dipama JM, 2010. The biophysical environment: climate. In: A Thiombiano, D Kampmann (eds.) Biodiversity atlas of West Africa. Vol. 2. Brice Sinsin & Dorothea Kampmann, Cotonou & Frankfurt/Main, Burkina Faso.

DLR, 2014. CATENA Available from: http://www.dlr.de/eoc/desktopdefault.aspx/tabid-5444/9113_read-17838/

Geladi P, Kowalski BR, 1986. Partial least-squares regression: a tutorial. Anal Chim Acta 185:1-17.

Hastie T, Tibshirani R, Friedman J, 2009. The elements of statistical learning: data mining, inference, and prediction. Springer, New York, NY, USA.

Herbreteau V, Salem G, Souris M, Hugot JP, Gonzalez JP, 2007. Thirty years of use and improvement of remote sensing, applied to epidemiology: from early promises to lasting frustration. Health Place 13:400-3.

Hotez PJ, Alvarado M, Basáñez MG, Bolliger I, Bourne R, Boussinesq M, Brooker SJ, Brown AS, Buckle G, Budke CM, Carabin, H, Coffeng LE, Fèvre EM, Fürst T, Halasa YA, Jasrasaria, R, Johns NE, Keiser J, King CH, Lozano R, Murdoch ME, O'Hanlon S, Pion SDS, Pullan RL, Ramaiah KD, Roberts T, Shepard DS, Smith JL, Stolk WA, Undurraga EA, Utzinger J, Wang M, Murray CJL, Naghavi M, 2014. The Global Burden of Disease study 2010: interpretation and implications for the neglected tropical diseases. PLoS Neglect Trop D 8:e2865.

Huete A, Didan K, Miura T, Rodriguez EP, Gao X, Ferreira LG, 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. Remote Sens Environ 83:195-213.

Huete A, Justice C, van Leeuwen WJD, 1999. MODIS vegetation index (MOD 13): algorithm theoretical basis document. Version 3. NASA. Available from: http://modis.gsfc.nasa.gov/data/atbd/atbd_mod13.pdf

Huete AR, 1988. A soil-adjusted vegetation index (SAVI). Remote Sens Environ 25:295-309.

Huete AR, Liu HQ, 1994. An error and sensitivity analysis of the atmospheric- and soil-correcting variants of the NDVI for the MODIS-EOS. IEEE T Geosci Remote 32:897-905.

Hürlimann E, Schur N, Boutsika K, Stensgaard AS, Laserna de Himpsl M, Ziegelbauer K, Laizer N, Camenzind L, Di Pasquale A, Ekpo UF, Simoonga C, Mushinge G, Saarnak CFL, Utzinger J, Kristensen TK, Vounatsou P, 2011. Toward an open-access global database for mapping, control, and surveillance of neglected tropical diseases. PLoS Neglect Trop D 5:e1404.

Jordan P, Webbe G, 1993. Epidemiology. In: P Jordan, G Webbe, RF Sturrock (eds.) Human schistosomiasis. CAB International, Wallingford, UK, pp 98-106.

JSS, 2014. ASTER GDEM Japan Space Systems (J-spacesystems). Available from: http://gdem.ersdac.jspacesystems.or.jp/search.jsp

Kuhn M, 2008. Building predictive models in R using the caret package. J Stat Softw 28:1-26.

Lai YS, Biedermann P, Ekpo UF, Garba A, Mathieu E, Midzi N, Mwinzi P, N'Goran EK, Raso G, Assaré RK, Sacko M, Schur N, Talla I, TchuemTchuenté LA, Touré S, Winkler MS, Utzinger J, Vounatsou P, 2015. Spatial distribution of schistosomiasis and treatment needs in sub-Saharan Africa: a systematic review and geostatistical analysis. Lancet Infect Dis 15:927-40.

Levin SA, 1992. The problem of pattern and scale in ecology. Ecology

73:1943-67.

Liaw A, Wiener M, 2002. Classification and regression by random forest. R News 2:18-22.

Malone JB, Bergquist NR, Huh OK, Bavia ME, Bernardi M, El Bahy MM, Fuentes MV, Kristensen TK, McCarroll JC, Yilma JM, Zhou XN, 2001. A global network for the control of snail-borne disease using satellite surveillance and geographic information systems. Acta Trop 79:7-12.

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. Am J Trop Med Hyg 50:714-22.

Marshall RJ, 1991. A review of methods for the statistical analysis of spatial patterns of disease. J Roy Stat Soc A Sta 154:421-41.

McFeeters SK, 1996. The use of the normalized difference water index (NDWI) in the delineation of open water features. Int J Remote Sens 17:1425-32.

Meyer D, 2011. ASTER global digital elevation model version 2–summary of validation result. NASA land processes distributed active archive center and the joint Japan-US ASTER science team. Available from: http://www.jspacesystems.or.jp/ersdac/GDEM/ver2Validation/Summary_GDEM2_validation_report_final.pdf

N'Goran EK, Diabaté S, Utzinger J, Sellin B, 1997. Changes in human schistosomiasis levels after the construction of two large hydroelectric dams in central Côte d'Ivoire. B World Health Organ 75:541-5.

N'Goran EK, Utzinger J, Traoré M, Lengeler C, Tanner M, 1998. Use of a questionnaire for quick identification of the principal foci of urinary bilharziasis in central Ivory Coast. Med Trop 58:253-60 [in French].

Qi J, Chehbouni A, Huete AR, Kerr YH, Sorooshian S, 1994. A modified soil adjusted vegetation index. Remote Sens Environ 48:119-26.

R Core Team, 2012. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

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

Richter R, Schläpfer D. 2012. Atmospheric / topographic correction for airborne imagery. ATCOR-4 user guide, version 6.2 BETA. German Aerospace Center (DLR). Available from: http://www.dlr.de/eoc/Portaldata/60/Resources/dokumente/5_tech_mod/atcor4_manual_2012.pdf

Savane I, 2010. The biophysical environment: water resources. In: S Konaté, D Kampmann (eds.) Biodiversity of West Africa. Vol III. Brice Sinsin & Dorothea Kampmann, Cotonou & Frankfurt/Main, Burkina Faso.

Schur N, Hürlimann E, Garba A, Traoré MS, Ndir O, Ratard RC, Tchuem Tchuenté LA, Kristensen TK, Utzinger J, Vounatsou P, 2011. Geostatistical model-based estimates of schistosomiasis prevalence among individuals aged ≤20 years in West Africa. PLoS Neglect Trop D 5:e1194.

Schur N, Hürlimann E, Stensgaard AS, Chimfwembe K, Mushinge G, Simoonga C, Kabatereine NB, Kristensen TK, Utzinger J, Vounatsou P, 2013. Spatially explicit *Schistosoma* infection risk in eastern Africa using Bayesian geostatistical modelling. Acta Trop 128:365-77.

Schur N, Vounatsou P, Utzinger J, 2012. Determining treatment needs at different spatial scales using geostatistical model-based risk estimates of schistosomiasis. PLoS Neglect Trop D 6:e1773.

Seto E, Xu B, Liang S, Gong P, Wu W, Davis G, Qiu D, Gu X, Spear R, 2002. The use of remote sensing for predictive modeling of schistosomiasis in China. Photogramm Eng Rem S 68:167-74.

Simoonga C, Utzinger J, Brooker S, Vounatsou P, Appleton CC, Stensgaard AS, Olsen A, Kristensen TK, 2009. Remote sensing, geographical information system and spatial analysis for schistosomiasis epidemiology and ecology in Africa. Parasitology 136:1683-93.

Steinmann P, Keiser J, Bos R, Tanner M, Utzinger J, 2006. Schistosomiasis and water resources development: systematic review, meta-analysis, and estimates of people at risk. Lancet Infect Dis 6:411-25.

Stensgaard AS, Jorgensen A, Kabatereine NB, Malone JB, Kristensen TK, 2005. Modeling the distribution of *Schistosoma mansoni* and host snails in Uganda using satellite sensor data and geographical information systems. Parassitologia 47:115-25.

Stensgaard AS, Utzinger J, Vounatsou P, Hürlimann E, Schur N, Saarnak CFL, Simoonga C, Mubita P, Kabatereine NB, Tchuem Tchuenté LA, Rahbek C, Kristensen TK, 2013. Large-scale determinants of intestinal schistosomiasis and intermediate host snail distribution across Africa: does climate matter? Acta Trop 128:378-90.

Utzinger J, N'Goran EK, Caffrey CR, Keiser J, 2011. From innovation to application: social-ecological context, diagnostics, drugs and integrated control of schistosomiasis. Acta Trop 120(Suppl.1):121-37.

Utzinger J, N'Goran EK, Ossey YA, Booth M, Traore M, Lohourignon KL, Allangba A, Ahiba LA, Tanner M, Lengeler C, 2000. Rapid screening for *Schistosoma mansoni* in western Côte d'Ivoire using a simple school questionnaire. B World Health Organ 78:389-98.

Vermote EF, Vermeulen A, 1999. MODIS algorithm technical background document: atmospheric correction algorithm: spectral reflectances (MOD09). Version 4.0. NASA. Available from: http://modis.gsfc.nasa.gov/data/atbd/atbd_mod08.pdf

Walz Y, Wegmann M, Dech S, Raso G, Utzinger J, 2015. Risk profiling of schistosomiasis using remote sensing: approaches, challenges and outlook. Parasit Vector 8:163.

Wan Z, 1999. MODIS land-surface temperature algorithm theoretical basis document (LST ATBD). Version 3.3. NASA. Available from: http://modis.gsfc.nasa.gov/data/atbd/atbd_mod11.pdf

Wan Z, Zhang Y, Zhang Q, Li ZL, 2004. Quality assessment and validation of the MODIS global land surface temperature. Int J Remote Sens 25:261-74.

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

WHO, 2010. Schistosomiasis control at a glance. World Health Organization, Geneva, Switzerland.

Wold S, Ruhe A, Wold H, Dunn IW, 1984. The collinearity problem in linear regression. the partial least squares (PLS) approach to generalized inverses. SIAM J Sci Stat Comp 5:735-43.

Wold S, Sjöström M, Eriksson L, 2001. PLS-regression: a basic tool of chemometrics. Chemometr Intell Lab 58:109-30.