

Julius-Maximilians-Universität Würzburg  
Institut für Geographie und Geologie  
Lehrstuhl für Fernerkundung

## **Remote sensing for disease risk profiling: a spatial analysis of schistosomiasis in West Africa**

Dissertation zur Erlangung  
des naturwissenschaftlichen Doktorgrades

vorgelegt von

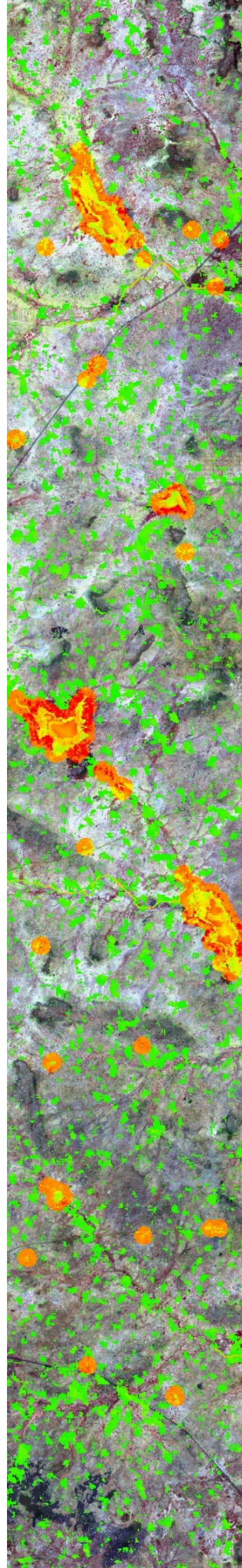
**Yvonne Walz**

aus Lindau am Bodensee

Würzburg, Oktober 2014

Erstgutachter: Prof. Dr. Stefan Dech

Zweitgutachter: Prof. Dr. Jürg Utzinger





Eingereicht am: 13.10.2014

1. Gutachter: Prof. Dr. Stefan Dech
2. Gutachter: Prof. Dr. Jürg Utzinger

Mentor: Dr. Martin Wegmann

Tag der Disputation: 14.01.2015

(Cover picture: Habitat suitability index (HSI) layer on RapidEye data from 18<sup>th</sup> February 2010)



# Abstract

Global environmental change leads to the emergence of new human health risks. As a consequence, transmission opportunities of environment-related diseases are transformed and human infection with new emerging pathogens increase. The main motivation for this study is the considerable demand for disease surveillance and monitoring in relation to dynamic environmental drivers. Remote sensing (RS) data belong to the key data sources for environmental modelling due to their capabilities to deliver spatially continuous information repeatedly for large areas with an ecologically adequate spatial resolution.

A major research gap as identified by this study is the disregard of the spatial mismatch inherent in current modelling approaches of profiling disease risk using remote sensing data. Typically, epidemiological data are aggregated at school or village level. However, these point data do neither represent the spatial distribution of habitats, where disease-related species find their suitable environmental conditions, nor the place, where infection has occurred. As a consequence, the prevalence data and remotely sensed environmental variables, which aim to characterise the habitat of disease-related species, are spatially disjunct.

The main objective of this study is to improve RS-based disease risk models by incorporating the ecological and spatial context of disease transmission. Exemplified by the analysis of the human schistosomiasis disease in West Africa, this objective includes the quantification of the impact of scales and ecological regions on model performance.

In this study, the conditions that modify the transmission of schistosomiasis are reviewed in detail. A conceptual underpinning of the linkages between geographical RS measures, disease transmission ecology, and epidemiological survey data is developed. During a field-based analysis, environmental suitability for schistosomiasis transmission was assessed on the ground, which is then quantified by a habitat suitability index (HSI) and applied to RS data. This conceptual model of environmental suitability is refined by the development of a hierarchical model approach that statistically links school-based disease prevalence with the ecologically relevant measurements of RS data. The statistical models of schistosomiasis risk are derived from two different algorithms; the Random Forest and the partial least squares regression (PLSR). Scale impact is analysed based on different spatial resolutions of RS data. Furthermore, varying buffer extents are analysed around school-based measurements. Three distinctive sites of Burkina Faso and Côte d'Ivoire are specifically modelled to represent a gradient of ecozones from dry savannah to tropical rainforest including flat and mountainous regions.

The model results reveal the applicability of RS data to spatially delineate and quantitatively evaluate environmental suitability for the transmission of schistosomiasis. In specific, the multi-temporal derivation of water bodies and the assessment of their riparian vegetation coverage based on high-resolution RapidEye and Landsat data proved relevant. In contrast, elevation data and water surface temperature are constraint in their ability to characterise habitat conditions for disease-related parasites and freshwater snail species. With increasing buffer extent observed around the school location, the performance of statistical models increases, improving the prediction of transmission risk. The most important RS variables identified to model schistosomiasis risk are the measure of distance to water bodies, topographic variables, and land surface temperature (LST). However, each ecological region requires a different set of RS variables to optimise the modelling of schistosomiasis risk. A key result of the hierarchical model approach is its superior performance to explain the spatial risk of schistosomiasis.

Overall, this study stresses the key importance of considering the ecological and spatial context for disease risk profiling and demonstrates the potential of RS data. The methodological approach of this study contributes substantially to provide more accurate and relevant geoinformation, which supports an efficient planning and decision-making within the public health sector.

# Zusammenfassung

Globale Umweltveränderungen rufen neue Gesundheitsrisiken hervor. Eine Konsequenz sind veränderte Bedingungen für die Übertragung von umweltbezogenen Krankheiten und ansteigende Infektionen mit neu auftauchenden Erregern. Die Motivation für diese Arbeit basiert auf der steigenden Nachfrage, dynamische Veränderungen der Umwelt und deren Beziehung zu Veränderungen von umweltbedingten Krankheiten zu überwachen. Fernerkundungsdaten gehören zu den wichtigsten Datenquellen für die Umweltmodellierung, da diese es ermöglichen, die Landbedeckung flächendeckend, reproduzierbar und in einer adäquaten räumlichen Auflösung zu kartieren.

Ein Forschungsbedarf, der in dieser Studie identifiziert wird, ist die fehlende Berücksichtigung einer räumlichen Diskrepanz innerhalb der bisherigen Vorgehensweise der Modellierung von Krankheitsrisiken mit Fernerkundungsdaten. Typischerweise werden epidemiologische Daten als Prävalenz einer Krankheit aggregiert erhoben, beispielsweise auf Schul- oder Dorfebene. Jedoch repräsentieren diese Punktmessungen weder die räumliche Verteilung von Habitaten, in welchen krankheitsrelevante Arten ihre geeigneten Umweltbedingungen vorfinden, noch den Ort, an dem sich die Menschen infiziert haben. Die Konsequenz ist, dass Messpunkte der Krankheitprävalenz und fernerkundungsbasierte Umweltvariablen, welche das Habitat von krankheitsrelevanten Arten charakterisieren sollen, räumlich nicht übereinstimmen.

Das Hauptziel dieser Studie ist, ein Verfahren für die Anwendung von Fernerkundungsdaten bei der Modellierung von Krankheitsrisiken zu entwickeln, welches sowohl den ökologischen als auch den räumlichen Kontext der Krankheitsübertragung widerspiegelt. Am Beispiel der Krankheit Schistosomiasis werden weitere mögliche Einflussgrößen auf die Modellgüte quantitativ bewertet. Dies sind unter anderem die verschiedenen Skalenniveaus und die Heterogenität von Ökozonen.

In dieser Arbeit werden die Bedingungen, die auf die Übertragung von Schistosomiasis einen Einfluss haben, aus der bestehenden Literatur im Detail ermittelt. Es wird eine konzeptionelle Grundlage entwickelt, die bestehende Zusammenhänge zwischen satellitengestützten Messungen, der Ökologie der Krankheitsübertragung sowie zu den Ergebnissen der epidemiologischen Studien ermittelt. Während eines Aufenthaltes im Untersuchungsgebiet wurde die Eignung der Umwelt für die Übertragung der Schistosomiasis analysiert. Diese Umwelteignung wird durch die Entwicklung eines Habitat-Eignungs-Index (*habitat suitability index*, HSI) quantifiziert und mit relevanten Fernerkundungsvariablen verknüpft. Im nächsten

Schritt werden Inhalte dieses konzeptionellen Modells gezielt für die Entwicklung eines hierarchischen Modellansatzes verwendet, welcher die gemessene Prävalenz in einen statistischen Zusammenhang mit ökologisch relevanten Messungen von Fernerkundungsdaten bringt. Die statistischen Modelle des Risikos, sich mit Schistosomiasis zu infizieren, basieren auf zwei verschiedenen Modellalgorithmen, dem sogenannten Zufalls-Wald Algorithmus (*Random Forest*) und der Regression der partiellen, kleinsten Quadrate (*Partial Least Squares Regression, PLSR*). Der Einfluss von räumlichen Skalen auf die Risikomodellierung wird anhand verschiedener räumlicher Auflösungen der Fernerkundungsdaten ermittelt. Darüber hinaus werden unterschiedlich große Einzugsgebiete mit Hilfe eines Pufferverfahrens (*Buffer*) anhand der Schulen mit Prävalenzmessungen analysiert. Risikomodelle der Schistosomiasis werden für drei ausgewählte Untersuchungsgebiete in Burkina Faso und der Elfenbeinküste erstellt, welche einen ökologischen Gradienten von der Trockensavanne zum tropischen Regenwald sowie von flachen und bergigen Regionen darstellt.

Diese Studie zeigt, dass Fernerkundungsdaten für die räumliche Abgrenzung und eine quantitative Bewertung der Umwelteignung für die Übertragung der Schistosomiasis geeignet sind. Besonders relevante Informationen sind zeitlich dynamische Veränderungen der Wasserbedeckung sowie die Erfassung des Grades der Ufervegetationsbedeckung auf Basis von hochaufgelösten RapidEye und Landsat Daten. Hingegen sind topographische Daten und die satellitengestützten Messungen der Temperatur nur eingeschränkt geeignet um Habitate der Parasiten und Frischwasserschnecken als wesentlichen Bestandteil der Krankheitsübertragung zu charakterisieren. Bei zunehmender Größe des Einzugsgebietes der Schulen verbessern sich die statistischen Modelle und können somit das Übertragungsrisiko besser erfassen. Die wichtigsten Fernerkundungsvariablen für die Modellierung des Schistosomiasis Risikos sind die Distanz zum nächstgelegenen Gewässer, topographische Variablen sowie die Landoberflächentemperatur (*land surface temperature, LST*). Für jede Ökozone muss jedoch eine geeignete Zusammenstellung von Fernerkundungsvariablen getroffen werden. Ein ganz wesentliches Ergebnis der hierarchischen statistischen Modellierung ist eine verbesserte Erklärung des räumlichen Risikos von Schistosomiasis.

Insgesamt unterstreicht diese Studie die Bedeutsamkeit des ökologischen und räumlichen Kontexts für die Abschätzung des Krankheitsrisikos und demonstriert das Potential von Fernerkundungsdaten. Der methodische Ansatz dieser Arbeit kann wesentlich dazu beitragen, genaue und relevante Geoinformationen bereitzustellen. Damit wird eine effizientere Planung und Entscheidungsfindung innerhalb des Gesundheitssektors ermöglicht.



# Résumé

Le changement environnemental global conduit à l'émergence de nouveaux risques pour la santé humaine. En conséquence, les voies de transmission des maladies liées à l'environnement, sont modifiées de même que l'infection humaine avec l'accroissement des nouveaux agents pathogènes émergents. La motivation principale de cette étude est la demande considérable pour la surveillance et le suivi des maladies en relation avec la dynamique des facteurs environnementaux. Les données de la télédétection sont les sources principales utilisées pour la modélisation de l'environnement en raison de leurs capacités à fournir une information de manière spatiale, répétitive et continue pour les grandes surfaces avec une résolution spatiale écologique adéquate.

L'importante lacune de la recherche scientifique identifiée par cette étude est la non considération de la disparité spatiale inhérente dans les approches actuelles de modélisation des risques de la maladie en utilisant des données de la télédétection. Généralement, les données épidémiologiques sont regroupées à l'échelle du village. Toutefois, ces données ne peuvent pas représenter la distribution spatiale des habitats et définir les conditions environnementales favorables à la prolifération des agents pathogènes de la maladie, ni le lieu, où l'infection s'est produite. En conséquence, les données sur la prévalence et les variables environnementales de la télédétection, qui visent à caractériser l'habitat des agents liés à la maladie, sont spatialement disjointes.

L'objectif principal de cette étude est d'améliorer en utilisant la télédétection les modèles de risque de maladie en incorporant l'aspect écologique et spatial de la transmission de la maladie. Illustré par l'étude des personnes infectées de la schistosomiase en Afrique de l'Ouest, cet objectif comprend la quantification du niveau d'impact des régions écologiques sur les performances du modèle.

Dans cette étude, les conditions qui modifient la transmission de la schistosomiase sont examinées en détail. Une approche conceptuelle reliant les données mesurées issues de la télédétection, la transmission de la maladie, l'écologie et des données de l'enquête épidémiologique a été développée. À partir d'une étude sur le terrain, les facteurs environnementaux à la transmission de la schistosomiase ont été évalués, ensuite quantifiés par l'indice de qualité de l'habitat (*habitat suitability index*, HSI) et combiné aux données de la télédétection. Le modèle conceptuel de la pertinence environnementale a été affiné par le

développement d'une approche de modèle hiérarchique qui relie statistiquement la prévalence de la maladie en milieu scolaire avec les mesures écologiques pertinentes de données de la télédétection. Les modèles statistiques de risque de schistosomiase proviennent de deux différents algorithmes; la forêt aléatoire (*Random Forest*) et la régression des moindres carrés partiels (*Partial Least Squares Regression*, PLSR). Le niveau d'impact a été analysé sur la base de différentes résolutions spatiales de données de la télédétection. En outre, des divers degrés carré des bassin de réception ont été analysés autour de mesures en milieu scolaire. Trois sites distinctifs du Burkina Faso et de la Côte d'Ivoire sont spécifiquement modélisés pour représenter un gradient de écozones de savane sèche a forêt tropicale y compris les régions plates et montagneuses.

Les résultats du modèle révèlent l'applicabilité des données de la télédétection pour la délimitation spatiale et l'évaluation quantitative de la pertinence de l'environnement pour la transmission de la schistosomiase. Précisément, la dérivation multi-temporelle des course d'eau et l'évaluation de leur couverture riveraine de végétation a partir des images à haute résolution RapidEye et Landsat jugées adéquate. En revanche, les données d'altitude et de température de la surface de l'eau ont montré certaines limites dans leur capacité à caractériser les conditions de l'habitat des parasites et des escargots en tant que composantes essentielles de la transmission de la maladie. Avec l'augmentation des degrés carrés des bassins de réception observés autour de l'emplacement de l'école, la performance des modèles statistiques augmente, améliorant ainsi la prédiction du risque de transmission. Les plus importantes variables des données de la télédétection identifiées pour modéliser le risque de schistosomiase sont la mesure de la distance des plans d'eau, les variables topographiques, et la température de surface de la terre (*land surface temperature*, LST). Cependant, chaque région écologique nécessite une série différente de variables de données de télédétection afin d'optimiser la modélisation du risque de schistosomiase. Le résultat primordial de l'approche du modèle hiérarchique est sa supérieure performance à expliquer le risque spatiale de la schistosomiase.

Dans l'ensemble, cette étude souligne l'importance cruciale de tenir compte du contexte écologique et spatiale pour le profilage du risque de maladie et démontre le potentiel des données de télédétection. L'approche méthodologique de cette étude contribue de manière substantielle à fournir avec plus de précision et de pertinence l'information géographique, prenant en charge une planification efficace et la prise de décision dans le secteur de la santé publique.

# Table of contents

<b>Abstract</b> .....	<b>i</b>
<b>Zusammenfassung</b> .....	<b>iii</b>
<b>Résumé</b> .....	<b>v</b>
<b>1 Introduction</b> .....	<b>1</b>
1.1 Human health and the environment.....	1
1.2 Remote sensing for modelling disease risk: approaches and challenges.....	4
1.3 Objectives and outline.....	7
<b>2 Geography of Burkina Faso and Côte d’Ivoire</b> .....	<b>11</b>
2.1 The biophysical environment .....	11
2.1.1 Topography.....	11
2.1.2 Geology and soils.....	13
2.1.3 Climate.....	15
2.1.4 Vegetation .....	16
2.1.5 Hydrology .....	18
2.2 The human environment.....	20
2.2.1 Socio-cultural organisation.....	20
2.2.2 Economy and agriculture.....	21
2.3 Schistosomiasis epidemiology.....	23
<b>3 Remote sensing of schistosomiasis risk</b> .....	<b>27</b>
3.1 Schistosomiasis.....	27
3.1.1 Geographical distribution and epidemiology .....	27
3.1.2 Ecology of schistosomiasis transmission .....	28
3.1.3 Schistosomiasis control and elimination .....	38
3.2 Remote sensing .....	39
3.2.1 Fundamentals of satellite remote sensing .....	39
3.2.2 Remote sensing for schistosomiasis risk profiling.....	42
3.3 Modelling schistosomiasis risk .....	50
3.3.1 Social-ecological niche of a disease.....	50

3.3.2	Model approaches .....	53
3.4	Contribution of remote sensing for schistosomiasis risk profiling .....	55
<b>4</b>	<b>Data and pre-processing .....</b>	<b>59</b>
4.1	Epidemiological data .....	59
4.2	Remote sensing data .....	64
4.2.1	Technical description and pre-processing of selected remote sensing data .....	66
4.2.2	Derivation of environmental variables .....	67
4.3	Environmental <i>in situ</i> data .....	72
<b>5</b>	<b>Modelling environmental suitability for schistosomiasis transmission .....</b>	<b>75</b>
5.1	Establishment of mechanistic model .....	77
5.1.1	Model development .....	77
5.1.2	Model composition .....	85
5.1.3	Model validation .....	87
5.2	Results of the mechanistic model approach .....	89
5.2.1	Habitat variable suitability .....	89
5.2.2	Plausibility of modelled environmental suitability .....	92
5.2.3	Model transferability .....	94
5.3	Discussion of remotely sensed environmental suitability .....	96
5.3.1	Remote sensing derived biophysical variables .....	96
5.3.2	Modelling environmental suitability .....	101
5.3.3	Model transferability .....	102
5.4	Summary of environmental suitability model .....	104
<b>6</b>	<b>Modelling schistosomiasis risk .....</b>	<b>105</b>
6.1	Statistical model algorithms .....	107
6.1.1	Random forest .....	108
6.1.2	Partial least squares regression .....	110
6.2	Statistical model procedures .....	112
6.2.1	Multi-scale modelling .....	112
6.2.2	The hierarchical model approach .....	113
6.2.3	The measure of variable importance .....	115
6.2.4	Model validation .....	116
6.3	Results of statistical risk modelling .....	117
6.3.1	How scale matters .....	117
6.3.2	The hierarchical model: a solution to bridge the spatial gap? .....	118
6.3.3	Key remote sensing variables for schistosomiasis risk modelling .....	120
6.3.4	Schistosomiasis risk prediction and validation .....	122
6.4	Discussion of remotely sensed schistosomiasis risk modelling .....	125
6.4.1	Remote sensing data for schistosomiasis risk modelling .....	125
6.4.2	Modelling schistosomiasis risk .....	128
6.5	Summary of schistosomiasis risk modelling .....	129

---

<b>7</b>	<b>Synthesis and outlook .....</b>	<b>131</b>
7.1	Strengths and limitations of remote sensing data for schistosomiasis risk profiling... 131	131
7.2	Transferability of the hierarchical model approach to other environment-related diseases .....	134
7.3	Future research needs.....	135
	<b>References.....</b>	<b>137</b>
	<b>Appendix .....</b>	<b>157</b>
	<b>Abbreviations .....</b>	<b>173</b>
	<b>Glossary.....</b>	<b>177</b>
	<b>List of figures .....</b>	<b>179</b>
	<b>List of tables .....</b>	<b>183</b>
	<b>Curriculum Vitae.....</b>	<b>185</b>
	<b>Acknowledgements .....</b>	<b>189</b>
	<b>Eidesstattliche Erklärung.....</b>	<b>193</b>



# 1 Introduction

## 1.1 Human health and the environment

Population health is central to the three dimensions of sustainable development - society, economy and the environment (UN, 2012a: 27). The connection of human well-being to social capabilities, economic output and environmental resources is both that of a beneficiary and a contributor, which points out the key role of health for sustainable development (Confalonieri et al., 2007: 393; UN, 2012a: 27). Three of the eight Millennium Development Goals (MDGs) focus on health concerning child health (MDG 4), maternal health (MDG 5) and the control of communicable diseases (MDG 6). The remaining goals are key determinants of health, such as poverty reduction, education and environmental sustainability (Dye et al., 2013: IV).

The environment has a fundamental impact on human health (Guernier et al., 2004: 740; Confalonieri and McMichael, 2006: 6) and is estimated to account for 24% of the global disease burden and 23% of all deaths (Prüss-Üstün and Corvalan, 2006: 9). In a medical sense, the environment integrates all factors that are external to human hosts and “can be divided into physical, biological, social, cultural, etc., any or all of which can influence the health status of populations” (IEA, 1995: 53). Thereof, biophysical environments are essential to human health due to the basic need of the human organism for food, water, clean air, shelter and suitable climatic conditions (Corvalan et al., 2005: 12). The relation between the environment and human health is investigated by the scientific discipline of spatial epidemiology, which describes and analyses geographic variations in diseases with respect to environmental, demographic, behavioural, socioeconomic, genetic and infectious risk factors (Elliott and Wartenberg, 2004: 998). The earliest milestone for this spatial linkage between human health and the environment has been recognised already twenty-four centuries ago, when the ancient Greek physician Hippocrates (460-377 BC) articulated the doctrine of “Airs, Waters, Places”, pointing out the relationship between climatic elements, water quality and diseases (Bashford and Tracy, 2012: 513). His observation of regional differences in conditions of living and corresponding differences in prevalent diseases, led him to proclaim the importance of interactions between place and person in determining health and disease (Rosenberg, 2012: 661). In 1849, Snow (1855: 45-48) successfully identified the source of the London cholera epidemic by mapping cholera cases in geographic space. This led to the discovery of a contaminated water pump as the source of the disease. In the mid-19<sup>th</sup> century the Russian parasitologist, Pavlovsky, formulated the concept of landscape epidemiology based on his observations that: (i) some

diseases are limited geographically; (ii) the spatial variation of diseases can be explained by an underlying variation in physical and/or biological conditions that support pathogen, vectors and their reservoirs; and (iii) if abiotic and biotic conditions can be delimited on maps then both contemporaneous risk and future change in risk should be predictable (Pavlovsky, 1966: 155-195; Ostfeld et al., 2005: 328). Nowadays, The acquisition and analysis of the spatial components of disease epidemiology relies on tools such as geographic information systems (GIS), remote sensing (RS) data, and spatial statistics, which enable epidemiological research, disease surveillance and control (Kitron, 1998: 435). Based on these tools, spatial epidemiology offers a variety of ways to identify and map the habitat of disease vectors, to relate it to social-ecological factors and eventually predict the potential risk of disease transmission (Kitron, 1998: 437).

Today, it is clear that population growth and economic development induce high pressure on the global environment and contribute to the human-induced global environmental change (Confalonieri and McMichael, 2006: 6). At the same time, this biophysical environment provides the fundamental elements (environmental media) for the transmission of environment-related diseases (Figure 1-1), which can be categorised as vector-borne diseases (e.g. malaria), water-based diseases (e.g. schistosomiasis), aerosol-borne diseases (e.g. avian influenza - H5N1 virus), soil-borne diseases (e.g. hookworm infection) or food-borne diseases (e.g. salmonellosis) (Bright et al., 2013: 5). Thus, the escalating human pressure on the environment with the consequences of severe changes and degradation of ecosystems have resulted in multiple, mostly negative health impacts (Corvalan et al., 2005: 1; Confalonieri and McMichael, 2006: 8; McMichael, 2013: 1335). Extensive alteration of the natural environment such as large-scale deforestation, expansion of settlements, infrastructure and agricultural land use or human intervention in watersheds, lakes, and river systems triggered widespread changes in the distribution of organisms and biodiversity (Chapin III et al., 2000: 234) and has been accompanied by global increases in morbidity and mortality from a number of environment-related diseases (Patz et al., 2000: 1396). Each environmental change influences the ecological balance and the context within which disease vectors, intermediate hosts or parasites breed, develop, and transmit a disease (Patz et al., 2000: 1395). Degradation of ecosystems may lead to the emergence of new human pathogens, the resurgence of old ones or change the transmission opportunities of established vector-borne pathogens (Kitron, 1998: 442; Patz et al., 2000: 1395; Foley et al., 2005: 571-572). The expected population growth from 7.2 billion in mid-2013 is projected to reach 9.6 billion by 2050 with the largest growth in developing regions, especially in Africa (UN, 2013). Thus, the high pressure on the environment and poor husbanding of natural resources will be further aggravated and might severely influence the habitat conditions and abundance of parasites, vectors, and hosts.

Besides direct anthropogenic degradation of natural environments, climate change is increasingly driving human health impacts. The predicted greenhouse gas concentrations in the atmosphere are expected to increase the global average temperature until the end of this century by between 0.3 and 4.8°C and influence the patterns and amounts of precipitation (IPCC, 2013: 18). Already during the last decades of the 20<sup>th</sup> century, anthropogenic-induced climate change has claimed an estimated 150,000 lives annually resulting for example from increased exposure to thermal extremes, more frequent weather disasters, changing dynamics of disease vectors, seasonality and incidence of food-related and water-borne infections, and crop failures (WHO, 2002b: 72; Patz et al., 2005: 310). Observations of sudden changes in



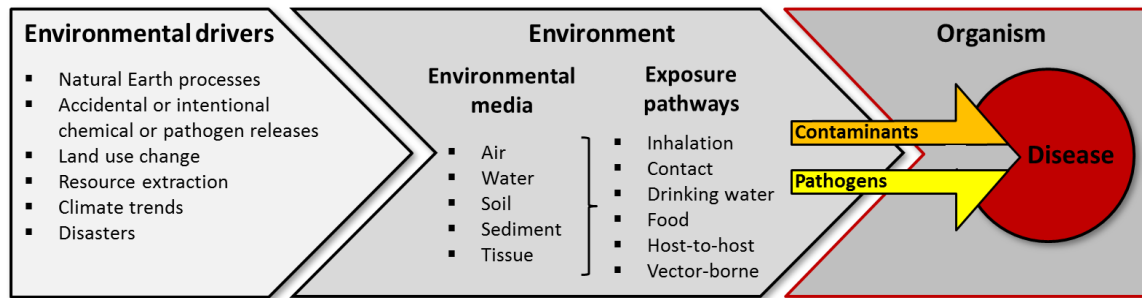


Figure 1-1: Conceptual framework of environment-related diseases. The environment provides disease drivers and enables disease transmission through various media and exposure pathways. Environment-related diseases result of released or transmitted contaminants or pathogens as disease agents (adjusted from Bright et al., 2013: 4)

temperature or rainfall have been related to explosions of vector populations causing epidemics of malaria (Gagnon et al., 2002), Rift Valley fever (Linthicum et al., 1999) or dengue fever (Descloux et al., 2012) in disease-endemic areas (Githeko et al., 2000: 1137). Moreover, climate warming and alterations of rainfall patterns are expected to modify the spatial distribution of climate sensitive diseases resulting either in emerging disease transmission in previously unaffected areas or the disappearance of a disease due to the establishment of unsuitable conditions (Githeko et al., 2000).

Given the dramatic global environmental change, it is a growing concern worldwide that environmental drivers increasingly influence unacquainted exposure to disease agents and pathways of disease transmission with the consequence of new emergence, resurgence, and sudden epidemic outbreaks of environment-related diseases (Gratz, 1999: 51; Weiss and McMichael, 2004: 70; Bright et al., 2013: 4). Thus, there is considerable demand for disease surveillance and monitoring in relation to dynamic environmental drivers (Patz et al., 2000: 1402; Kerr and Ostrovsky, 2003: 299). The key aspects herein lay within the spatially explicit quantification of disease risk, upon which any supplementary step depends. The current challenges in spatial epidemiology must be: (i) to gain a better understanding of the environmental impact on disease ecology; (ii) to identify immediately required action for health authorities and environmental managers; and (iii) to improve spatially explicit predictive models (Patz et al., 2000: 1402).

The crucial basis for predicting disease risk is spatially explicit environmental information. Due to the systematic and consistent view of the Earth at regular time intervals and comparatively low cost, satellite RS data belong to the key data sources for environmental modelling and have proven to be very useful in the assessment of biophysical characteristics of the landscape (Gillespie et al., 2008: 204), the detection of suitable habitat conditions of species (Goetz et al., 2000: 290), and the discovery of environmental changes (Kerr and Ostrovsky, 2003: 299). In combination with GIS, RS data are thought to “revolutionise the discipline of epidemiology and its application in human health” (Hay, 2000: 2). The current approaches and remaining challenges of RS for epidemiological applications are described in the following section.

## 1.2 Remote sensing for modelling disease risk: approaches and challenges

*“Whatever the epidemiological question, if there is an element of environmental input, satellites of one sort or another offer the potential for developing surveillance and early-warning systems to address it on a global scale” (Hay et al., 2000a: xii)*

The prerequisite for RS to contribute to epidemiological research and application is any distinct relation between the physical-natural environment that can be characterised by RS data and the ecology of disease transmission (Hugh-Jones, 1989). Cline (1970) highlighted the opportunity to measure environmental characteristics relevant for disease occurrence by means of RS and record them in a regular fashion. Pavlovsky’s term “landscape epidemiology” (Pavlovsky, 1966: 155) seemed particularly well adapted to highlight the essential benefit of RS that yields relevant information about the disease influencing environment in space and time on a landscape level, which can hardly be acquired with field-based investigations (Cline, 1970: 87). In 1985, the Life Science Division of the National Aeronautics and Space Administration (NASA) initiated the Global Monitoring and Human Health (GMHH) programme and investigated the capability of RS data to predict the spatial and temporal variability in malaria vector population dynamics to assess risk of disease transmission (NASA, 1998). Specific landscape elements such as swamps and unmanaged pasture could significantly explain disease vector abundance, allowing the identification of villages with high human-vector contact (Beck et al., 1994).

Over the past 30 years, the use of RS data and techniques in mapping human and veterinary diseases has increased substantially (Kalluri et al., 2007: 1362). The wealth of scientific literature contains a set of explorative case studies that investigate the informative value of various satellite data and variables with their spatial and temporal properties for selected geographical regions and in relation to specific diseases or disease agents such as vectors, parasites or intermediate hosts (Hay et al., 1997; Beck et al., 2000; Hay et al., 2000b; Yang et al., 2005c; Kalluri et al., 2007; Simoonga et al., 2009; Tran et al., 2010). Since those agents have specific requirements regarding climate, vegetation, soil, and other edaphic factors and are sensitive to changes therein, RS can be used to determine their living conditions and predict potential distributions (Rinaldi et al., 2006: 36). The general idea behind the linkage between RS and disease data is: (i) to identify and map parasite, vector, and host habitats and thereby better understand complex mechanisms of disease transmission; (ii) to monitor changes in those habitats; (iii) to predict changes in vector or host populations based on habitat modification; and (iv) to generate efficient risk maps and early warning systems that can be used to design control programmes (Hugh-Jones, 1989: 244-245; Kitron, 1998: 438; Beck et al., 2000: 225; Myers et al., 2000a). However, the amplex of case studies show that there is not one single best suited methodological approach for RS of diseases (Curran et al., 2000: 44). On the contrary, the choice of methodology is highly specific to a disease, due to the distinct ecological requirements of disease transmission (Beck et al., 2000: 223).

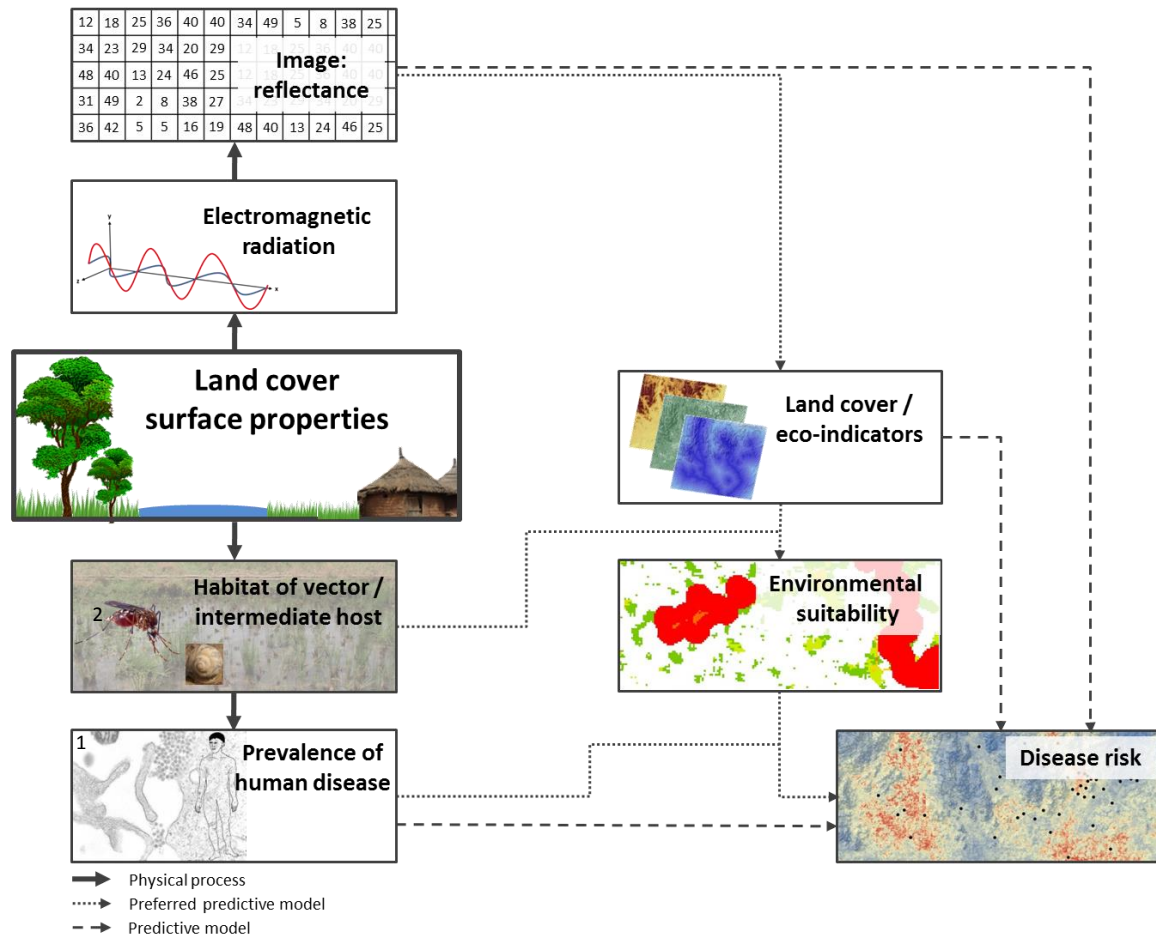


Figure 1-2: Conceptual framework linking remotely sensed images with diseases. The bold solid arrow indicates a physical process between the steps linking remotely sensed image reflectance and human disease, whereas the dotted and dashed arrays reflect the modelling of these processes. The dotted array represents the preferred predictive model approach due to its integration of all relevant steps of information, and the dashed array indicates a predictive model, which is feasible and often applied in practice due to data constraints. Source: modified after Curran et al. (2000: 39); Picture sources: 1. "Dengue virus" by CDC and University of South Carolina - <http://commons.wikimedia.org>; 2. "Gelbfiebermücke (*Stegomyia aegypti*)" by US Department of Agriculture Carolina - <http://commons.wikimedia.org>

RS of human diseases requires interdisciplinary research approaches combining disease-related epidemiological and ecological information with environmental characteristics. The point, where this diverse information is combined, is its geographical location. Thus, an integrative, geographical perspective is required to establish the process chain from integrating specific measurements of each discipline into a spatial database, applying the appropriate methodological approach to combine data, and interpreting added value and the resulting information. A conceptual framework of the linkage between remote sensing and disease-related data (Figure 1-2) illustrates how land cover and surface properties at specific locations bridge the gap between RS measurements and disease occurrence (Curran et al., 2000: 40-41). It varies from physical and well-understood links between surface properties, electromagnetic radiation and image reflectance to mainly empirical links between a vector or intermediate host in the field and a patient in a hospital, the latter being the least understood (Curran et al., 2000: 44). The disease itself, hence the measure of human infection provides a distal relation to specific surface properties, whereas species involved in the disease transmission cycle have a direct relation to the biophysical environment, which is explained by their specific ecological

requirements and environmental habitat (Hirzel et al., 2002: 2028). The preferred model approach to enhance the contextual understanding of disease transmission ecology and predict disease risk based on RS data (Figure 1-2), consists of a preliminary linkage between RS data and disease-related species in its environment and additional information on resulting disease prevalence in humans (Roberts et al., 1991: 273-274). However, due to scarcity of data on disease-related vector or host species, a feasible, but pragmatic predictive model approach (Figure 1-2) is to link RS data directly to human disease prevalence (Hugh-Jones, 1991: 202-203). In this PhD thesis, the linkage between RS data and vector or host specific characteristics results in the information of environmental suitability for disease transmission, whereas the linkage between human disease prevalence and RS data result in the measure of disease risk. To model environmental suitability or disease risk using RS data two approaches may be used, individually or combined. These are: (i) biology-based, mechanistic models; (ii) statistical, correlative models (Malone, 2005; Kearney, 2006). The first approach integrates biological requirements of a species and mechanistic analyses of its fitness known from laboratory or field studies to model environmental preferences, limits of tolerance, and behaviour of the organism (Malone, 2005: 28; Kearney, 2006: 186). The latter approach models a statistical relationship between survey records of observed species or diseases and the corresponding environmental conditions measured at the survey location to predict the established relationship into space and time (Malone, 2005: 27). The aim of both model approaches is to measure the suitability of an environment to establish the niche of a disease as zone, where the pathogen, the vector/intermediate host, and the infected human converge and form a biocenosis in space and time (Malone, 2005: 28).

However, from a geographical point of view, one of the remaining challenges of this defined association between RS data and environment-related diseases is an inherent spatial mismatch between locations of vector or host habitats and human disease prevalence, which again varies in dependence of the disease-specific transmission ecology and the sampling location of respective data. For the case of schistosomiasis, which is the target disease in this thesis to demonstrate this phenomenon, the transmission cycle from human to human requires the parasite to meet specific snails as intermediate host within an aquatic environment. While human infection must happen within aquatic habitats, where parasite and freshwater snails occur, survey measurements of disease prevalence are mainly located at schools where the most vulnerable group is identified. The fact that these disease-related components are not spatially super-imposed has already been mentioned by Curran et al. (2000: 44). At the time of writing this thesis no schistosomiasis risk model based on RS data has recognised this phenomenon, which however, increases uncertainties inherent to the data and modelling approaches – an aspect that must be addressed for improved risk profiling (Brooker, 2007: 1).

Although a large variety of RS data are available and expected to provide useful information for epidemiological studies (Hay et al., 2006), mainly freely available, pre-processed RS products with coarse spatial resolution have been utilised. Most studies use pre-processed, readily available calculations of the normalized difference vegetation Index (NDVI), “while other existing vegetation indices, not directly accessible due to their complex nature, are used rarely but can be helpful for health studies” (Herbreteau et al., 2007: 401). At the same time high spatial resolution RS data have been explored only rarely, whereas these data are expected to be highly advantageous to specifically address the above-mentioned spatial mismatch between disease-

related components, and for the detection of heterogeneous habitat conditions of various disease vectors on a local scale (Goetz et al., 2000: 303; Herbreteau et al., 2007: 402). The current application of RS data results in a limited adaptability of environmental information to address disease-related biological questions at ecologically relevant scales (Herbreteau et al., 2007: 400-401) and points the need to further explore the potential of RS data to target the disease-specific ecology.

Issues of scale and extent are fundamental to spatial statistical analysis, because ecological and epidemiological processes operate differently dependent on scale and area of observation (Robinson, 2000: 92). Disease epidemiology has to deal with an inherent complexity of biological systems, spanning the range of phenomena from those so fine that they operate at the level of the molecule to those so extensive that they can only be studied for large areas (Hay et al., 2000a: xi). With the objective to detect, analyse and explain the spatial heterogeneity of a disease, one has to refer to the variables that describe the data (e.g. vector density, disease cases, micro-habitat conditions), but also to the nature of the spatial units themselves, their size, shape and configuration (Kitron, 1998: 436). In this sense, the application of RS data to profile disease risk needs to consider the appropriate spatial resolution regarding the ecological process under investigation. Environmental suitability and schistosomiasis risk have been modelled from local (Clennon et al., 2004; Raso et al., 2005) to national and continental scales of observation (Clements et al., 2006b; Brooker, 2007; Schur et al., 2013; Stensgaard et al., 2013), most probably resulting in different predictions for the overlapping area. The phenomenon that model results are expected to vary considering different spatial scales (Openshaw, 1984) has not yet been investigated and its impact on model performance never quantified for the case of schistosomiasis risk profiling.

## 1.3 Objectives and outline

As outlined above, previous RS applications for profiling disease risk have not sufficiently addressed the ecological and spatial context of this interdisciplinary research approach. Against this background and based on the specified need for close collaboration between epidemiologists and geographers to fully exploit the potential of RS data and adapt the information to the needs of public health concerns (Mayer, 1983: 1220; Herbreteau et al., 2007: 402-403), this is specifically addressed in this thesis. The aim is to bridge the disciplines of geographical RS, disease ecology, and epidemiology through a conceptual underpinning of their linkages and an explicit spatial analysis of the respective data, illustrated by case studies pertaining to schistosomiasis in West Africa.

**The overall objective of this research is to investigate the potential and optimised application of RS data for modelling environmental suitability and disease risk for schistosomiasis transmission.**

Considering the impact of global environmental change on population health, the establishment of robust RS methods to monitor disease risk under changing environmental conditions is of increasingly high interest. The allocation of up-to-date, accurate, and relevant

geoinformation for planning can essentially contribute to informed decision-making within the public health sector and support the efficient allocation of the sector's limited resources. In contrast to existing schistosomiasis risk predictions that have used RS data, this study focusses on the necessity to challenge the diversity of influencing factors on the accuracy of RS based disease risk models such as the selection of appropriate variables, the impact of different scales, or ecological regions. These research gaps are directly addressed in this thesis by investigating multi-scale RS data from RapidEye, Landsat 5 Thematic Mapper (TM), and the Moderate Resolution Imaging Spectroradiometer (MODIS) on the Terra platform with spatial resolutions ranging from 6.5m to 1km. Its individual pre-processing allows the derivation of a series of RS variables tailored to the specific disease ecology. With the geographical focus of this thesis being West Africa, where both schistosomiasis burden and the need for control remain greatest (Brooker, 2007: 2), the environmental gradient ranging from near desert to tropical rainforest and including flat and mountainous regions provides a useful basis to systematically investigate the impact of different ecological regions. This has most often been neglected by selective case studies but has been shown by Brooker et al. (2001: 1001) to have significant impacts on model accuracy.

To meet the overall objectives of this research, this thesis addresses specifically the following research questions:

- (1) Which RS data and variables are most useful to model environmental suitability and disease risk?
- (2) How can the spatial discrepancy between environmental suitability for schistosomiasis transmission and the measure of disease risk be resolved?
- (3) Which scale is most appropriate for spatial modelling of schistosomiasis risk?
- (4) How do different ecozones impact the performance of schistosomiasis risk models in West Africa?

Figure 1-3 outlines the structure of contents and methods employed in this thesis. The geography and disease epidemiology of the study area in West Africa are described, with focus on the specifically investigated local study sites, in Chapter 2. Given the interdisciplinary character of this thesis, ecological details relevant for disease transmission, its linkage to RS measurements of environmental conditions and a conceptual framework for modelling these data are elaborated for the case study of schistosomiasis in Chapter 3. The procedures of pre-processing epidemiological and RS data and sampling environmental *in situ* data used for the analysis in this thesis are specified in Chapter 4. Chapter 5 investigates the potential of RS data to model and predict environmental suitability for schistosomiasis transmission using the habitat suitability index (HSI). Based on this approach, potential disease transmission sites are spatially delineated, its suitability for schistosomiasis-related snails and parasites quantified, and the transferability of this locally established model to different ecological regions evaluated. In Chapter 6, the potential of RS data to model and predict schistosomiasis risk is elaborated using two different statistical algorithms. Based on the spatial delineation of potential disease transmission sites derived in Chapter 5, a hierarchical model approach is developed and evalu-

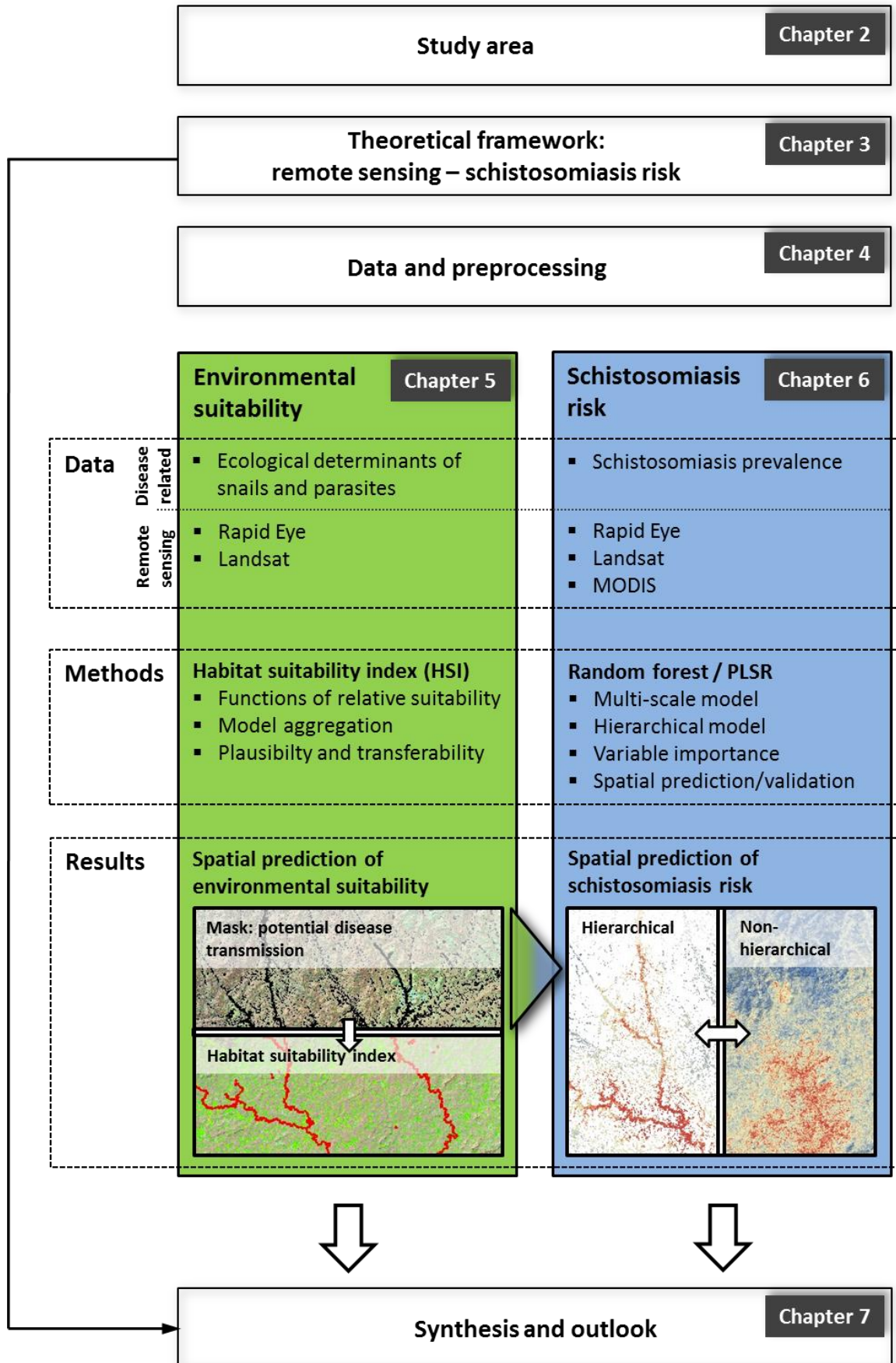


Figure 1-3: Overview of the structure of this thesis

ated to resolve the spatial discrepancy between environmental suitability and schistosomiasis risk. Moreover, this chapter evaluates the importance of the selected RS variables as well as the impact of different scales and ecological regions on model performance. Chapter 7 provides a synthesis of the results of this thesis and gives concise answers to the research questions posed. The transferability of the optimised application of RS data to profile the risk of other environment-related diseases is discussed and an outlook on possible future research questions is provided.



## 2 Geography of Burkina Faso and Côte d'Ivoire

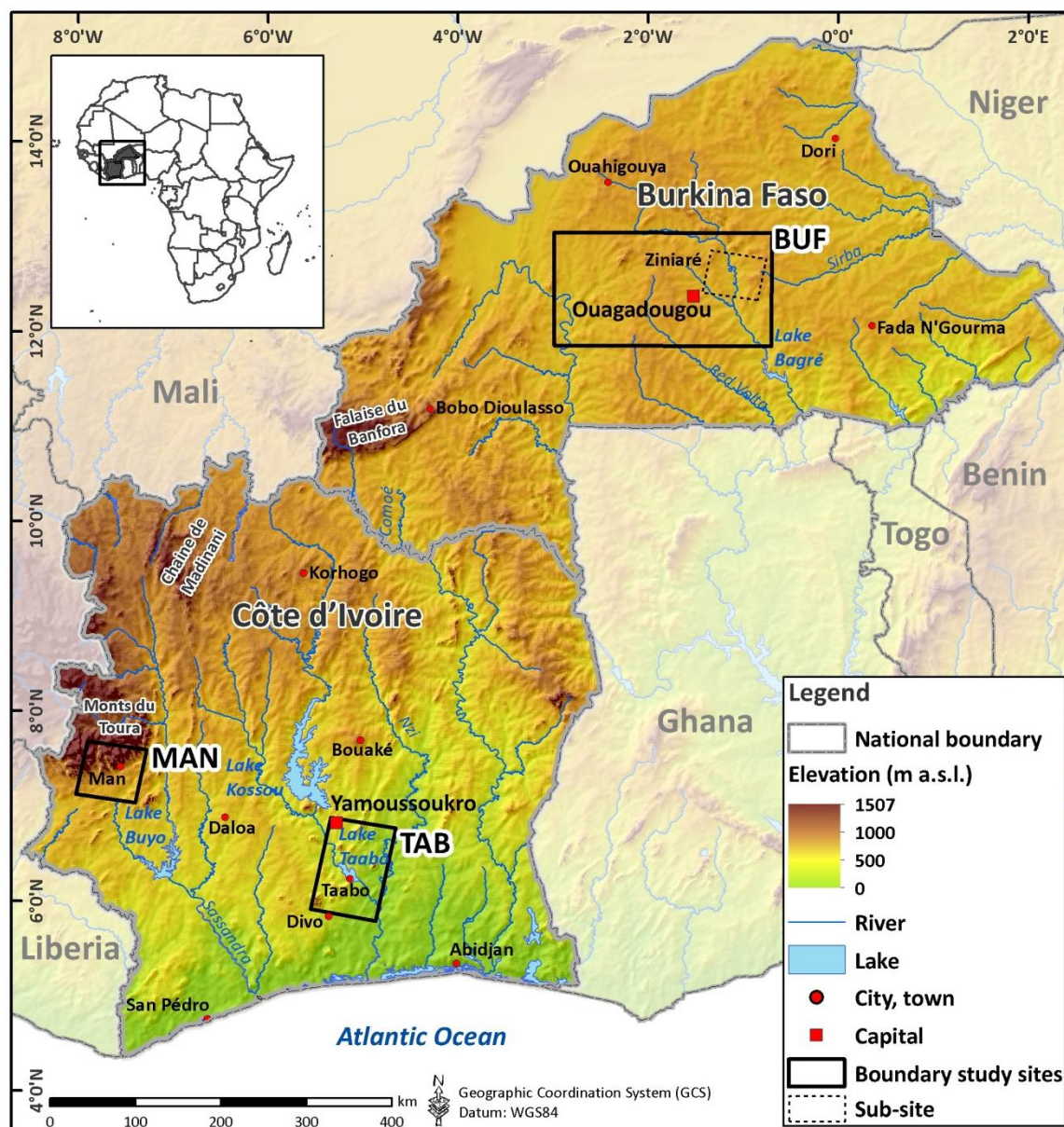
The study area of this thesis is situated in West Africa and comprises the countries Burkina Faso and Côte d'Ivoire (Figure 2-1). In general, sub-Saharan Africa is considered a hotspot of schistosomiasis transmission due to its suitable climate, the existing water management practices together with a high level of poverty and low sanitary standards. Both countries, Burkina Faso and Côte d'Ivoire, are endemic regions of schistosomiasis with prevalence rates estimated higher than 50% for Burkina Faso (Utzing et al., 2011a: 124) and 9.1% for Côte d'Ivoire with a highly focal distribution (Yapi et al., 2014). Three local study sites have been investigated in this study, namely: (i) BUF in central Burkina Faso; (ii) MAN around the city Man in western Côte d'Ivoire; and (iii) TAB around the Lake Taabo in south-central Côte d'Ivoire (see Figure 2-1). They represent a transect of ecozones ranging from dry savannah in the North to tropical rainforest in the South and are characterised by environmental gradients of topography, land cover, and climate. Within each study site, the disease epidemiology is described by highly heterogeneous prevalence rates as shown in more detail in Section 4.1. Thus, these sites well represent the requirements to specifically address the research objectives of this thesis, such as the impact of scale or ecozone on modelling schistosomiasis risk.

### 2.1 The biophysical environment

#### 2.1.1 Topography

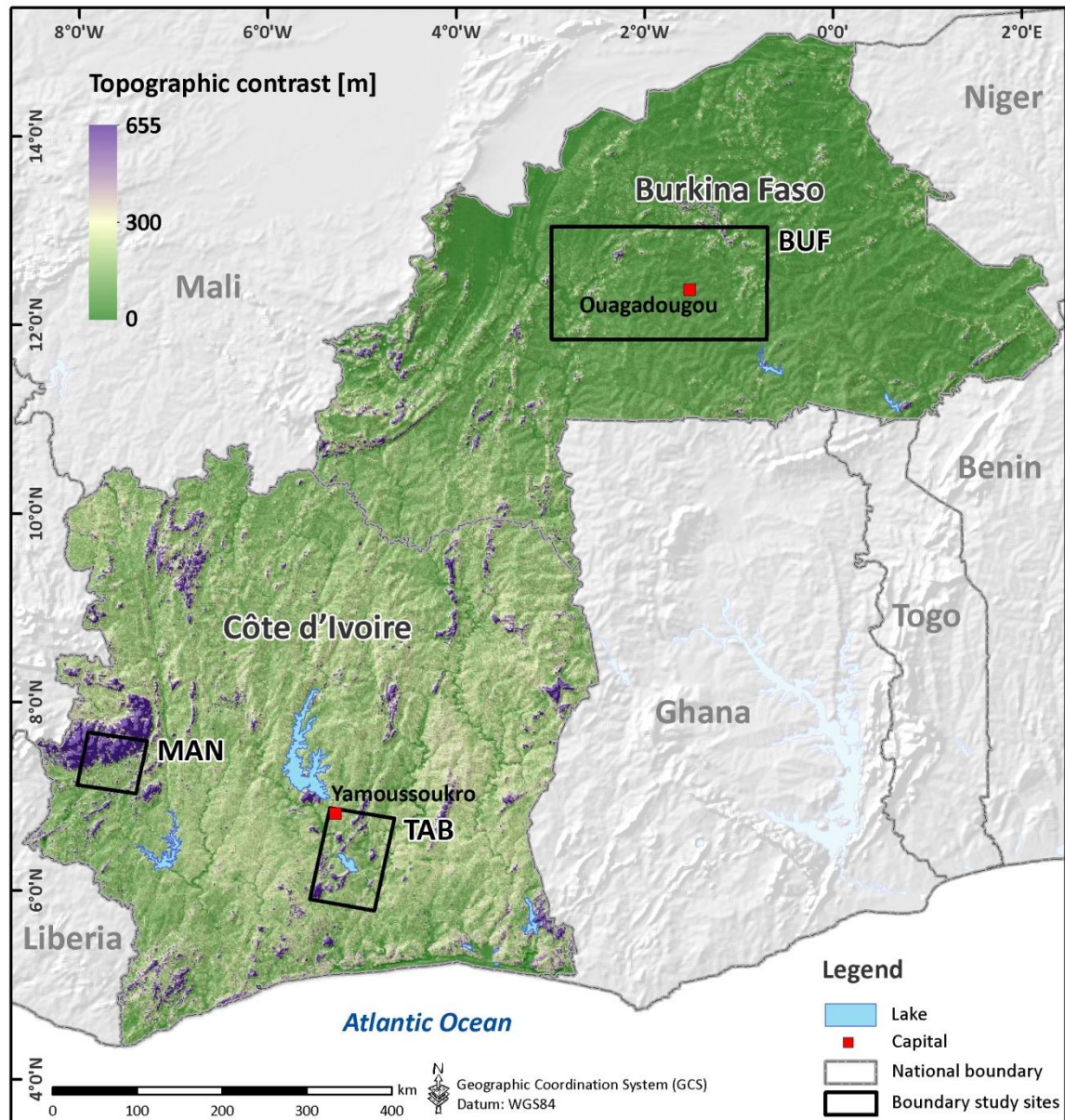
The relief of the study area is generally flat. The only mountains are located in and around the study site MAN in western Côte d'Ivoire with few peaks higher than 1,000m (Savane, 2010b: 122) (Figure 2-1). The southern region of Côte d'Ivoire is characterised by a plane coastal strip that is to the North delimited by a belt of dissected table lands varying from 300 to 600m in altitude (Poorter et al., 2004: 7). An extensive zone of high plateaus spans from the centre of Côte d'Ivoire to the North and covers most of Burkina Faso with only few isolated reliefs, e.g. in the East and North-west of Côte d'Ivoire or the South-west of Burkina Faso (Savane, 2010b: 122-123) (Figure 2-1). The regional contrast between elevations is of high relevance with respect to schistosomiasis transmission risk as topographic contrast essentially shapes the runoff

characteristics of surface water. Figure 2-2 illustrates the contrast (difference between minimum and maximum) of elevation data within 1km grid cells and illustrates the regions with steep topographic brims in dark purple, whereas flat terrains appear in green. The topographic contrast varies strongly between the three selected study sites from high contrast in MAN, few peaks of high contrast in TAB, and a single elevated outcrop in the predominantly flat terrain of the study site BUF.



Data source:  
 Elevation: Shuttle Radar Topography Mission (SRTM) by NASA, JPL, DLR and ASI  
 River and lakes: Savane, 2010b: 127 & Dipama, 2010b: 135; provided by Dr. Joachim Eisenberg (SFB 268), University of Frankfurt, Germany  
 National boundary: GADM ([www.gadm.org](http://www.gadm.org))

Figure 2-1: Overview of the study area Burkina Faso and Côte d'Ivoire in West Africa. The selected study sites BUF, MAN and TAB are indicated by the black rectangles with the solid lines. The sub-site Ziniaré in BUF represents the footprints of available high-resolution RapidEye data. Both study sites in Côte d'Ivoire are fully covered by RapidEye data.



Data source:

Elevation: Shuttle Radar Topography Mission (SRTM) by NASA, JPL, DLR and ASI

Lakes: Savane, 2010b: 127 & Dipama, 2010b: 135 provided by Dr. Joachim Eisenberg (SFB 268), University of Frankfurt, Germany

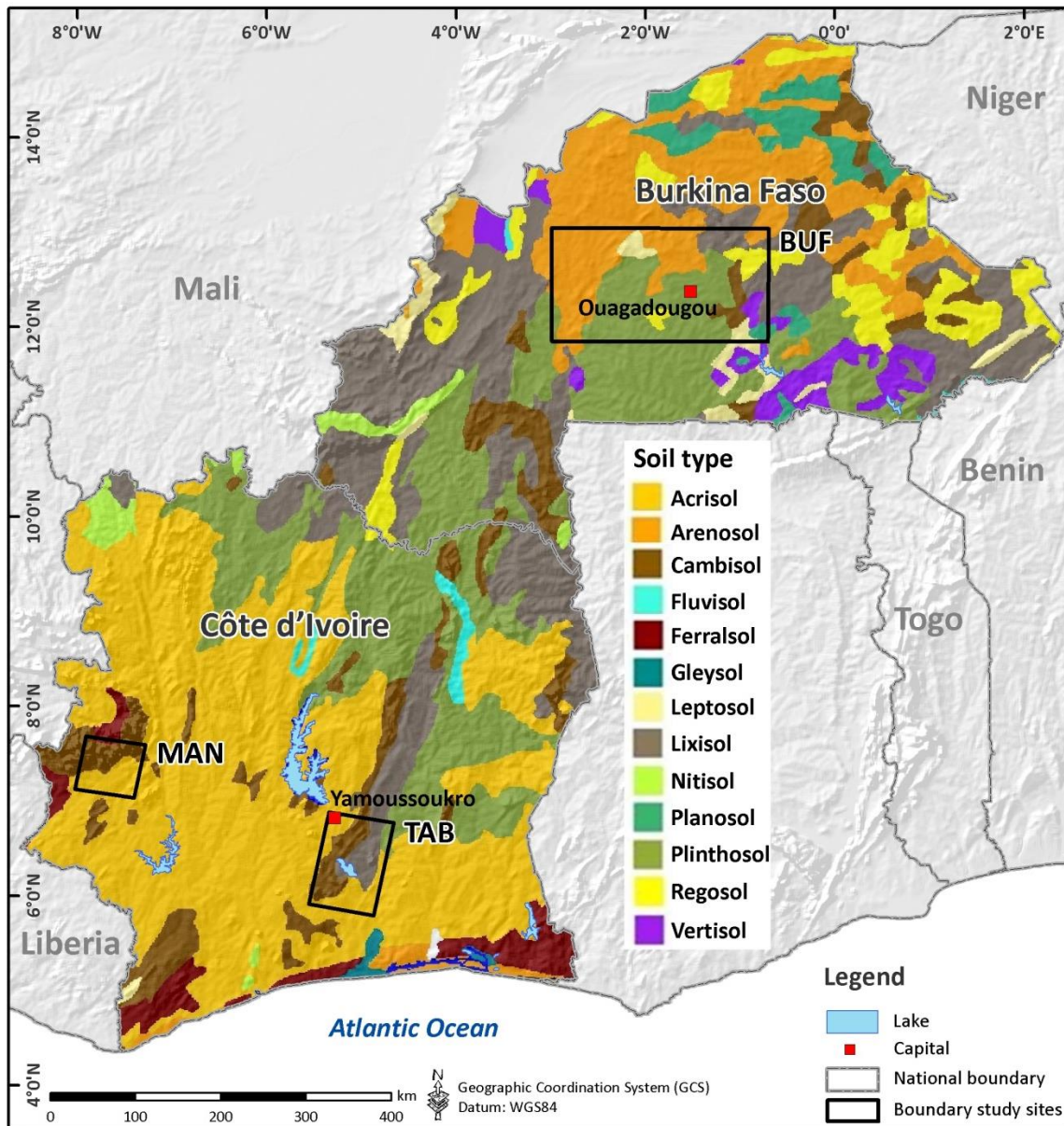
National boundary: GADM ([www.gadm.org](http://www.gadm.org))

Figure 2-2: Topographic contrast in the study area derived from minimum and maximum elevation of SRTM data within 1 km grid cells.

## 2.1.2 Geology and soils

From a geological perspective, the study area is located on the south-eastern margin of the West African Craton, a stable and spacious unit of Precambrian basement. Granites, gneisses, quartz, and schists form the main parental bedrock from which soils develop (Poorter et al., 2004: 8-9; Reichert et al., 2010: 35). Regions that differ from this largely homogeneous crystalline base are a sandstone massif rich in aquifers at the western border of Burkina Faso, rows of fossil dunes in its northern boundary and the sedimentary basin at the southern coastal region of Côte d'Ivoire (Dipama and Anne, 2010: 126-128; Savane, 2010b: 122).

The distribution of soils in the study area is illustrated in Figure 2-3. The zonal soil of the humid region of Côte d'Ivoire represented by the study site MAN is Acrisol with inclusions of Cambisol, whereas the typical soil below tropical rainforest mainly towards the coastal region is Ferralsol. The soil types within the study site TAB in Côte d'Ivoire are a combination of Acrisol, Cambisol, Lixisol and Plinthosol. In Burkina Faso, the Plinthosol and Lixisol are the zonal soil types of the less humid climate. The northern region of Burkina Faso is characterised by Arenosol, which is the typical zonal soil for dry regions. The study site of Burkina Faso is characterised by the combination of Plinthosol and Arenosol with inclusions of Regosol, Leptosol, Cambisol, Lixisol, Gleysol, and Vertisol (FAO et al., 2012).



Data source:  
 Elevation: Shuttle Radar Topography Mission (SRTM) by NASA, JPL, DLR and ASI  
 Lakes: Savane, 2010b: 127 & Dipama, 2010b: 135 provided by Dr. Joachim Eisenberg (SFB 268), University of Frankfurt, Germany  
 National boundary: GADM ([www.gadm.org](http://www.gadm.org))  
 Soil types: Harmonized World Soil Database Version 1.2 (FAO et al., 2012)

Figure 2-3: Soil types of the study area

Soils in the study area are typically low in nutrients and leaching is the predominant pedogenetic process (except for higher elevation areas with younger soils) resulting in low cation availability (0–2cmol/kg) (Poorter et al., 2004: 8-9; Traore and Anne, 2010: 130). The areas around streams are characterised by hydromorphic pedogenesis that evolves under the influence of an excess of temporary water resulting in Gleysols (Traore and Anne, 2010: 130-132). With respect to the potential risk for transmission of schistosomiasis, the water holding capacity of soils is of relevance due to its impact on snail survival. This has been estimated to be low (<25mm water per meter of soil) in the northern dry region of Burkina Faso, moderate (25–100mm) for most of the study area, and high (100–200mm) for the hydromorphic soils along the streams and seasonal flooded areas (USDA-NRCS, 1998).

### 2.1.3 Climate

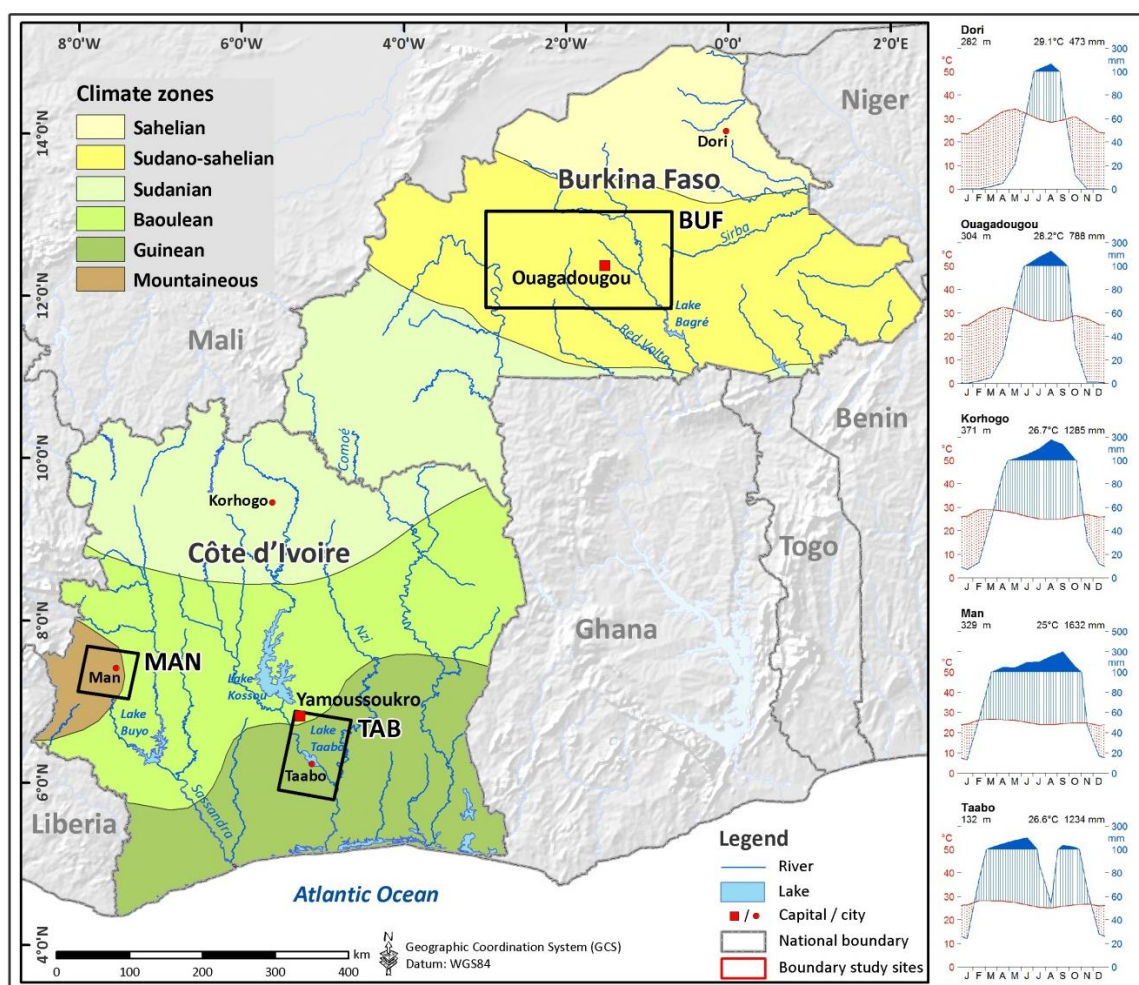
The climate zones of the study area illustrated in Figure 2-4 and described in Table 2-1 show a moisture gradient increasing from North to South. Interannual variations of rainfall are mainly determined by the oscillation of the intertropical convergence zone (ITCZ), where humid maritime and dry continental air masses meet and the convection process close to the Sun zenith creates favourable conditions for rainfall (Poorter et al., 2004: 7-8; Dipama, 2010a: 122-124). The northern end of the study area (Sahelian climate zone) is the zone of lowest rainfall and is characterised by a high rate of potential evapotranspiration due to the high temperatures throughout the year. The Sudano-Sahelian climate zone covers most of Burkina Faso that experiences one rainy season from May to October with generally mid-range temperatures between 20 and 30°C. The border region of Burkina Faso and Côte d'Ivoire is characterised by the Sudanian climate with one prolonged rainy season from April to November (Dipama, 2010a: 122-124). Towards the south, Côte d'Ivoire is exposed to four seasons per year, because the ITCZ transits this region both during its northward and southward movement (Chmielewski et al., 1998: 156-157). The difference between the Baoulean climate in central Côte d'Ivoire and the Guinea climate in the South is that the monthly amounts of rainfall between April and June and August and October are relatively

homogeneous to the Baoulean climate, whereas there is a peak of very heavy rainfalls in the coastal region of Côte d'Ivoire during the first rainy season from May to July (Figure 2-4) compared to the second rainy season between October and November (Savane, 2010a: 125). In contrast, the mountainous climate zone is azonal and results from orographic modification of the atmospheric circulation (Chmielewski et al., 1998: 323-325). A mountainous climate zone with one extensive rainy season is

*Table 2-1: Characteristics of climate types in the study area of Burkina Faso and Côte d'Ivoire (Dipama, 2010a: 124; Savane, 2010a: 124)*

Type of climate	Precipitation (mm/a)	Characteristics of season
Sahelian climate	<600	2 seasons: dry, rainy (rainy: 2-3 months)
Sudano-sahelian climate	600–900	2 seasons: dry, rainy (rainy: 4-5 months)
Sudanian climate	>900-1,700	2 seasons: dry, rainy (rainy: 5-6 months)
Baoulean climate	1,500-2,200	4 seasons: 2 dry, 2 rainy
Guinean climate	1,300-2,400	4 seasons: 2 dry, 2 rainy
Mountain climate	1,500-2,200	2 seasons: dry, rainy (rainy: 6-7 months)

found in the region around the city Man in western Côte d'Ivoire. The temperature in sub-tropical and tropical climate zones is characterised by a diurnal climate with higher temperature ranges between day and night than between seasons of the year (Chmielewski et al., 1998: 71).



Data source:  
 Elevation: Shuttle Radar Topography Mission (SRTM) by NASA, JPL, DLR and ASI  
 Rivers, lakes and climate zones: Savane, 2010b: 125 & Dipama, 2010b: 123 provided by Dr. Joachim Eisenberg (SFB 268), University of Frankfurt, Germany  
 National boundary: GADM (www.gadm.org)  
 Climate data: WorldClim (Hijmans et al., 2005)

Figure 2-4: Climate zones of the study area and Walther Lieth climate diagrams of selected sites in the study area. The climate diagrams have been created based on estimates between 1950 and 2000 (Hijmans et al., 2005)

## 2.1.4 Vegetation

The vegetation zones of the study area are in correspondence with the climate zones described in Section 2.1.2 ranging from the Sahel in the northern end over several savannah biomes to the tropical rainforest in the southern end of the study area (Figure 2-5) (White, 1983). However, in this region, human land use strongly modifies the climate-related, characteristic zonal vegetation. The Sahel zone is characterised by sparse vegetation of thorn bush and savannah scrub, where grasses are short due to extensive grazing (Gornitz, 1985: 290). Towards the South, the transition to the Sudanian savannah is reached with yearly rainfalls exceeding 600mm. The vegetation in this zone covering the study site BUF - originally characterised by dense shrub thickets with scattered trees – is degraded to open tree savannah due to repeated burning practices and permanent land cultivation (Gornitz, 1985: 288-289). Subsequently, the Sudan-

Guinea savannah is reached, where yearly rainfall exceeds 1,000mm. The former closed woodland has been degraded through centuries of repeated grassfires and farming resulting in an open savannah woodland with scattered fire-resistant short trees and tall grass (Gornitz, 1985: 288). The forest savannah mosaic zone spans the transition zone between the Guinea-savannah and the tropical rainforest, which varies with its latitudinal expansion and is highly modified by anthropogenic land use due to logging and extensive farming, which also affects the north-eastern study site TAB. The southern end of the study area is located in the tropical rainforest zone, which is characterised by an evergreen or semi-evergreen rainforest and exhibits one of the world's hotspots of biodiversity (Myers et al., 2000b: 853). It has an average canopy height between 30 and 50m above ground level (Poorter et al., 2004: 10). However, the average deforestation rate (1981-1990) of tropical rainforest in Côte d'Ivoire has been assessed to be 7.6% per year resulting in a remaining forest cover of 11,230km<sup>2</sup> in 1990 from an original forest cover of 150,000km<sup>2</sup> in 1981 (Chatelain et al., 2004: 15). In general, the study area has experienced an enormous rate of desertification in the North and deforestation in the South due to human pressure from unsustainable farming practices and the need for resources (Gornitz, 1985: 287-288; Poorter et al., 2004: 12; Porembski et al., 2010: 67-68).

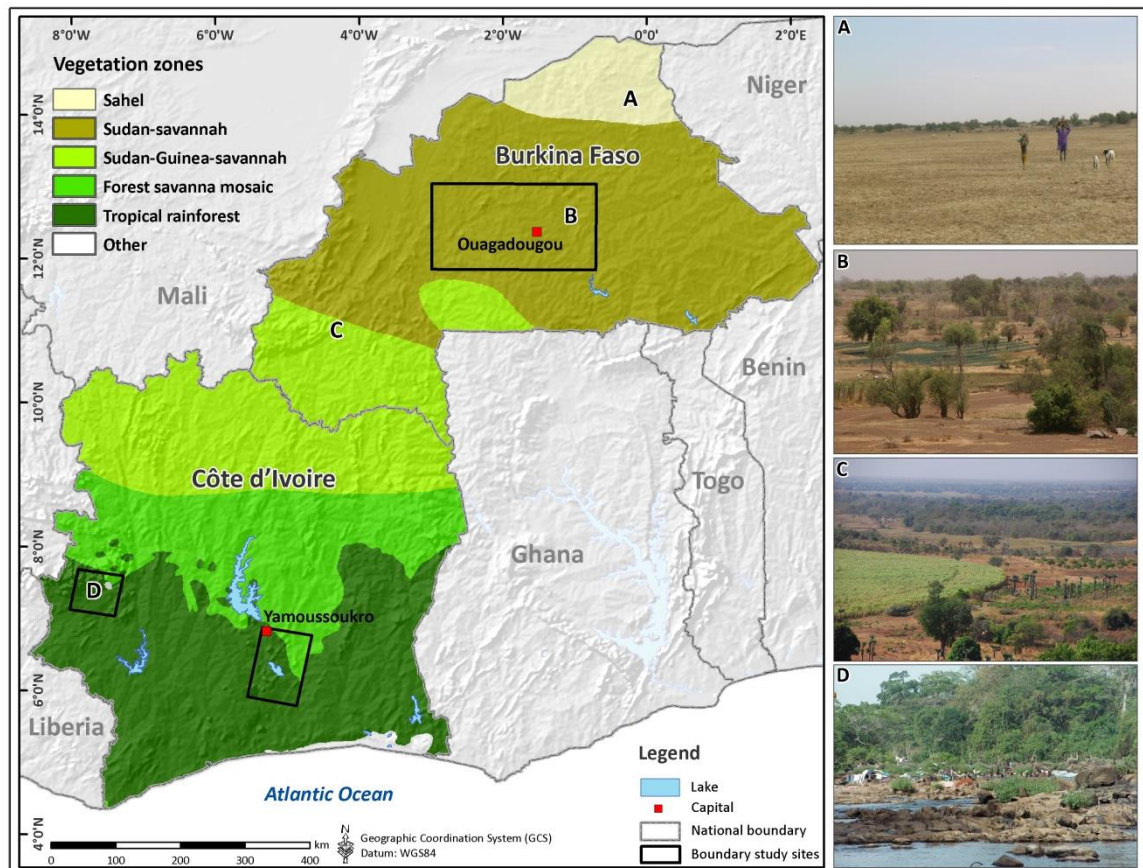


Figure 2-5: Vegetation zones in the study area according to White (1983). Pictures A to D illustrate the vegetation zones from the field perspective for the corresponding sites indicated in the map.

Apart from the zonal vegetation that responds mainly to the climatic gradient, there are azonal vegetation types occurring in areas, where local site conditions dominate. The most relevant examples within the study area are: (i) gallery forests that are found along riverbanks, where the high moisture availability allows them to penetrate deep into the savannah zone, and (ii) rocky outcrops of granitic or gneissic inselbergs and laterite plateaus promoting a multitude of habitat types e.g. for desiccation-tolerant vascular and carnivorous plants especially during the rainy season (Poorter et al., 2004: 10-11; Porembski et al., 2010: 67).

### 2.1.5 Hydrology

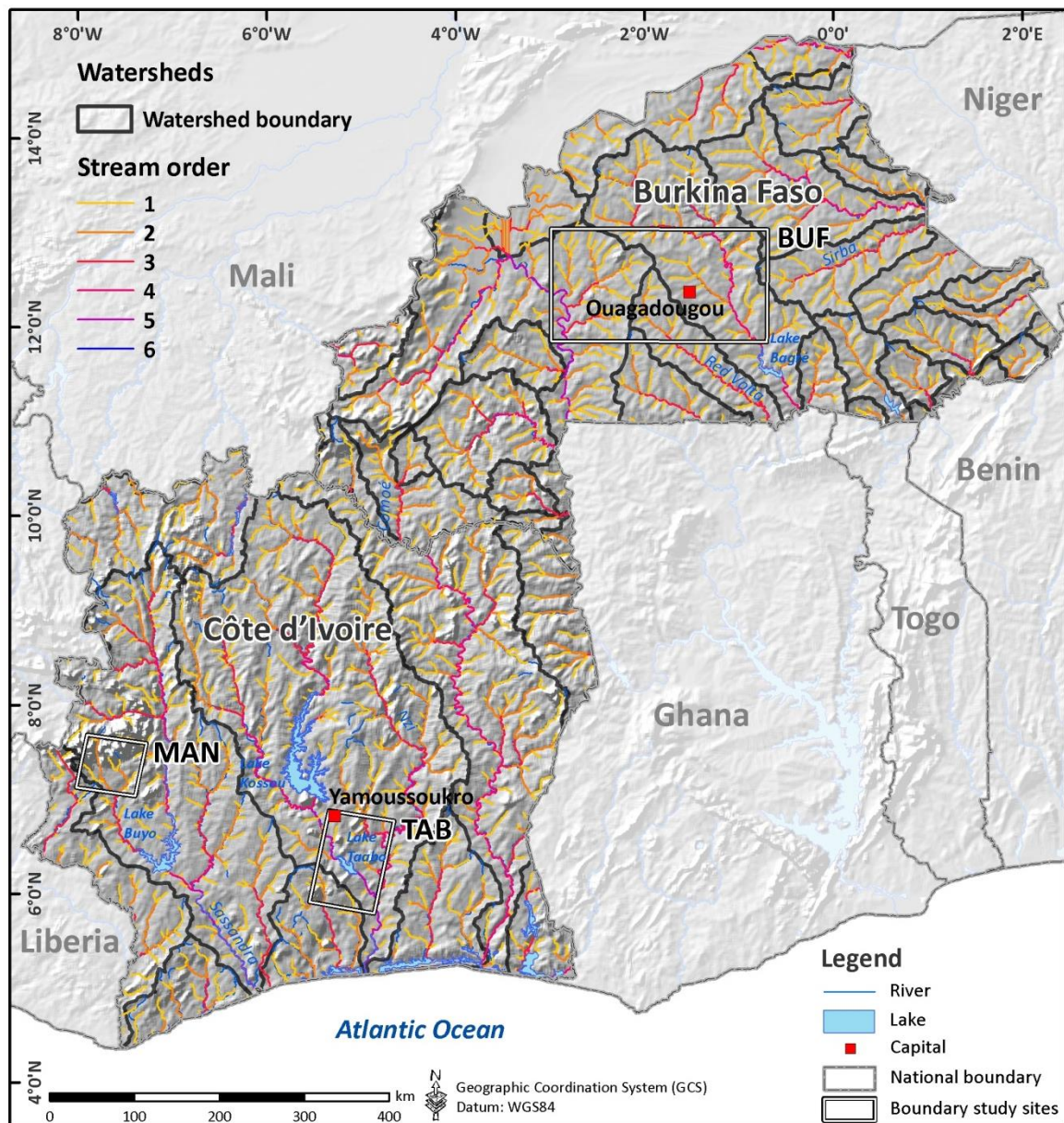
The hydrology corresponds to the most decisive criteria to determine the spatial dynamic of schistosomiasis transmission. In Burkina Faso, rivers are characterised by endorheic drainage, where water coming from the southern and eastern catchments run dry in the Sahelian climate zone. This implies that rivers and specifically their sections towards the North are seasonal. In contrast, rivers in Côte d'Ivoire have an exorheic draining due to their flow from the northern elevation towards the South mounding into the Atlantic Ocean often via estuaries and lagoon systems (Figure 2-6) (Dipama, 2010b: 135; Savane, 2010c: 127).

Despite the dry climatic conditions and the flat relief in Burkina Faso, the river network is relatively dense due to the discharge coming from three major outflow basins: the Niger in the East, the Volta in the South-east, and the Comoé in the South-west (Dipama, 2010b: 134-135). This river network is further subdivided by nearly 2,100 dammed lakes, which provide surface water for manifold usage, especially during the dry season when rivers temporarily dry out. Artificial dammed lakes strongly modify the river hydraulic, which results in e.g reduced water flow velocity with huge impact on the environmental suitability for transmission of schistosomiasis (N'Goran et al., 1997; Dianou et al., 2003). This environmental feature is very well demonstrated based on the satellite images available for the study site BUF. Additionally, one finds ponds where rainwater fills topographic depressions, these play a crucial role in pastoral life especially in the northern part of Burkina Faso, where they constitute the only major water source besides the sparse confined groundwater wells (Dipama, 2010b: 134). Unfortunately, these important hydrologic features cannot be demonstrated on the map of the study area due to the lack of respective data. However, specific sites of hydrological importance such as small-scale reservoirs, temporary ponds, dried out sections of rivers or irrigation systems have been visited in the study site BUF and are described in Section 4.3. These hydrological features are slightly different in Côte d'Ivoire. There are 570 artificial water reservoirs created predominantly for agricultural and hydroelectric power production (Savane, 2010c: 126-128) with a high impact on schistosomiasis transmission (N'Goran et al., 1997; Steinmann et al., 2006). However, due to the more humid climate, the agriculture is less dependent on the storage of water through constructed dams and irrigation practices as witnessed in Burkina Faso. Furthermore, the high amount of rainfall in the study sites MAN and TAB is expected to have an impact on water flow velocity of the presumably perennial rivers.

The hydrological potential is composed of renewable surface and groundwater resources and varies between the North (Burkina Faso) and the South (Côte d'Ivoire) of the study area. In Burkina Faso, renewable surface water contributes with only 27% and groundwater with 73% to the hydrological potential of the country (28.5 billion m<sup>3</sup>) (Dipama, 2010b: 134-136). The total renewable water resources per capita and year are 715m<sup>3</sup> (FAO, 2014a). This is dissimilar to Côte



d'Ivoire, where renewable surface water contributes with 49% and groundwater with 51% to the hydrological potential of the country (76.7 billion m<sup>3</sup>) (Savane, 2010c: 128-129). Here, the total renewable water resources amount to 3,940m<sup>3</sup> per capita and year (FAO, 2014a).



Data source:  
 Elevation: Shuttle Radar Topography Mission (SRTM) by NASA, JPL, DLR and ASI  
 Rivers, lakes and watersheds: Savane, 2010b: 127 & Dipama, 2010b: 135 provided by Dr. Joachim Eisenberg (SFB 268), University of Frankfurt, Germany  
 Stream order derived from DEM of SRTM data  
 National boundary: GADM ([www.gadm.org](http://www.gadm.org))

Figure 2-6: Hydrology of the study area illustrated by the extent of watersheds, rivers, lakes and the stream order modelled from a digital elevation model (DEM). The derivation of stream order is described in Section 4.2.2

## 2.2 The human environment

### 2.2.1 Socio-cultural organisation

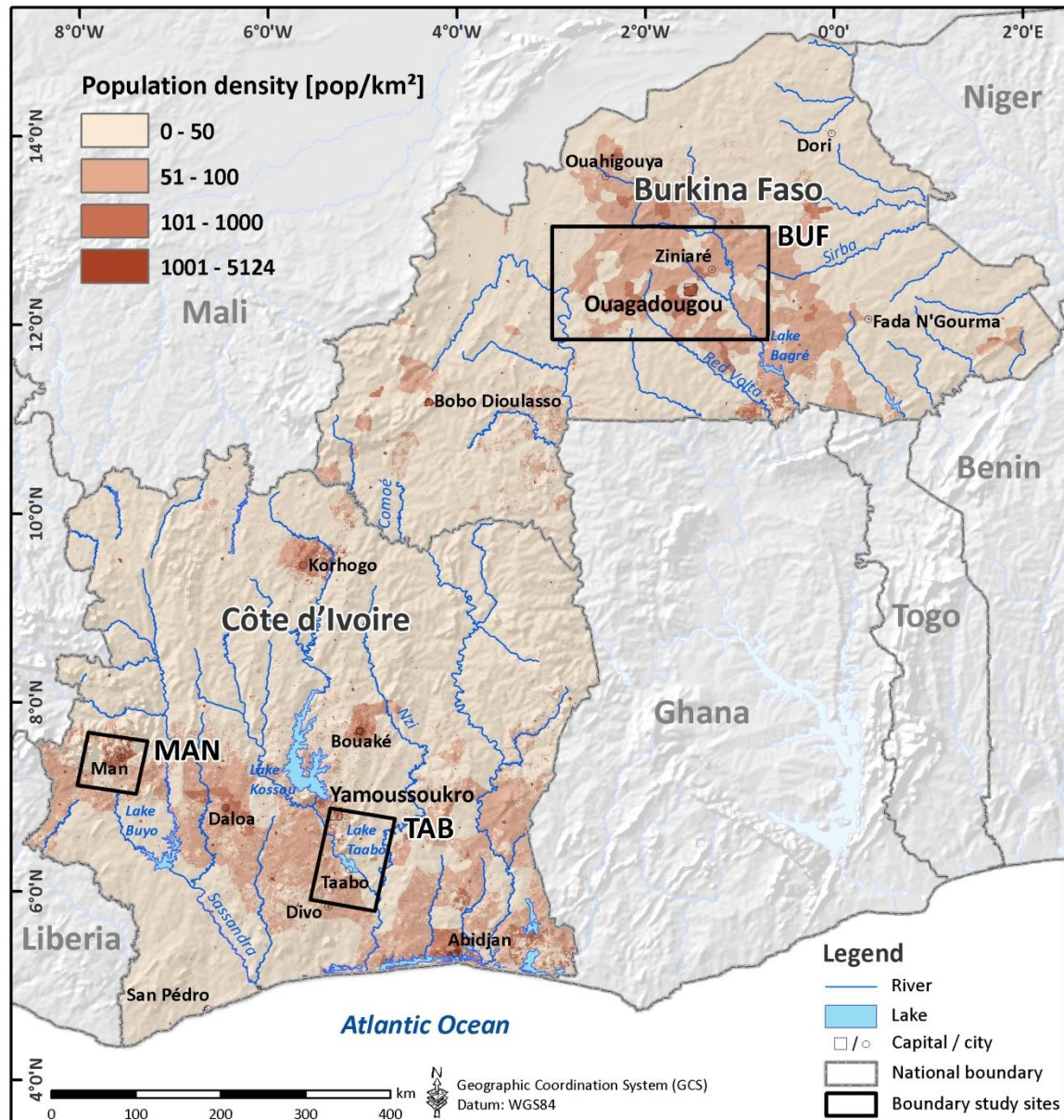
As in almost all developing countries, the population of the study area experienced rapid growth with an average annual growth rate of 3.4% in Burkina Faso and 2% in Côte d'Ivoire between the years 2002 and 2012 (Table 2-2) (FAO, 2014a). The total population in the year 2012 has been estimated to be around 17.5 million inhabitants in Burkina Faso and 20.6 million in Côte d'Ivoire. Thereof, the majority of the population in Burkina Faso (73%) lives in rural areas, whereas the proportion between rural (48%) and urban (52%) population is relatively balanced in Côte d'Ivoire (FAO, 2014a). The capitals Ouagadougou (Burkina Faso) and Yamoussoukro (Côte d'Ivoire) had an estimated population (in the year 2011) of 2.1 and 1.0 million, respectively (UN, 2012b). The comparably low number of inhabitants in Yamoussoukro can be explained by the relatively recent transfer of the Ivorian capital status from Abidjan (4.3 million inhabitants in 2011) to Yamoussoukro in 1983. Still today, Abidjan is the economic capital of the country and seat of the government (UN, 2012b), presumably due to its favourable location close to the sea.

The distribution of the population varies significantly within both countries, which is illustrated in Figure 2-7. In Burkina Faso, the central region around the capital followed by the western region around the economic capital Bobo-Dioulasso are characterised by higher population density in contrast to the very low population density in the eastern parts of the country (Senghor, 2010a: 139-140). In Côte d'Ivoire, the South is more densely populated than the North. The cities with highest population densities are Abidjan, Bouaké, Daloa, and Yamoussoukro.

In Burkina Faso, census results showed significantly more women than men in the years post 1985, i.e. 93 men per 100 women (Senghor, 2010a: 138-139). Overall, there are 60 ethnic groups and 60 languages spread among the Burkinabe population, whereby the principle ethnic group, the Mossi, account for with more than 48% of the population. They are followed by the Fulani with more than 10% and the Bobo, the Gourmantché, and the Gurunsi each with more than 7%. In Burkina Faso, the major religions practiced are Islam and Christianity, whereby both are often practiced in tandem with traditional indigenous beliefs (Senghor, 2010a: 140). Almost a quarter of the population in Côte d'Ivoire is comprised of foreigners who immigrated almost exclusively from the neighbouring countries being members of the Economic Community of West African Nations (ECOWAS). Major ethnic groups in Côte d'Ivoire are: Akan (41%), Mandé North and South (26.5%), Voltaic and Gur peoples (17.6%), and the Krou (12.7%). The religious composition of the Ivorian population is characterised by 38.6% following Islam, 30.4% Christians, 11.9% Animists, and 17.4% that are not a member of any religious group (Kouassi, 2010: 140-141).

The educational level of adolescents between 15 and 24 years of age indicated by the rate of literacy is generally lower in Burkina Faso (39.3%) than in Côte d'Ivoire (67.5%) (Table 2-2). More than a third of the children of primary school age do not go to school in Burkina Faso (35%) and Côte d'Ivoire (38%) (UNESCO, 2011). However, high uncertainty is inherent in these national indicators as demonstrated by another source, where over 70% of the population of 7 years or older has no education referring to class attendance in Burkina Faso (M.E.F., 2008; Senghor, 2010b: 142). The literacy rate is higher among the young population compared to adults and

generally more than 10% higher among the male population compared to the female population (UNESCO, 2011)



Data source:

Elevation: Shuttle Radar Topography Mission (SRTM) by NASA, JPL, DLR and ASI

Rivers and lakes: Savane, 2010b: 127 & Dipama, 2010b: 135 provided by Dr. Joachim Eisenberg (SFB 268), University of Frankfurt, Germany

National boundary: GADM ([www.gadm.org](http://www.gadm.org))

Population density: WorldPop (Linard et al., 2012)

Figure 2-7: Population density in the study area derived from the WorldPop database (Linard et al., 2012)

## 2.2.2 Economy and agriculture

Burkina Faso and Côte d'Ivoire are categorised as low-income and lower-middle-income countries with an estimated gross domestic product (GDP) of US\$ 10.44 billion and 24.68 billion, respectively (Table 2-2) (World Bank, 2014). In Burkina Faso, the agricultural sector has with 52% the largest contribution to the GDP and accounts for 79% of the country's export mainly from livestock, food, and cash crops. Further economic income results from mining (e.g. gold, zinc, copper), embryonic industries predominantly for agro-alimentary and textile production and a

growing tertiary sector (e.g. hotel and restaurant industry, handicrafts) (Senghor, 2010b: 142-148). In Côte d'Ivoire, the tertiary economic sector represented with 43.7% the highest activity (I: 28%, II: 20.3%) resulting a GDP growth rate of above 2%, given the situation in the year 1996. However, since the coup d'état in 1999 and the 2002 rebellion, Côte d'Ivoire experienced repetitive crises with great disruptions to its economy. In 2010 and 2011, another political crisis has massively destabilised the Côte d'Ivoire. These conflicts lead to decreasing economic activities and increasing poverty (Doumbia, 2010: 146-147). Nevertheless, given the latest data from the World Bank (2014) and UNESCO (2011), the GDP of Côte d'Ivoire is still more than double the GDP of Burkina Faso and the poverty rate of 46% is far less than that of Burkina Faso (73%). The most actual information on malnutrition rates of children under 5 years of age in Côte d'Ivoire comes from the year 2007, where it was high at 29.4%. In Burkina Faso, the very high malnutrition rate estimated in the year 2006 (37.6%) decreased significantly to 26.2% in 2010 (World Bank, 2014). An important socio-economic indicator in the context of this research represents the access to improved sanitation facilities. Not even a quarter of the population in Côte d'Ivoire (23.9%) is supplied, whereas the situation is even worse in Burkina Faso (18%) (World Bank, 2014). Especially among the poor, the access to basic services such as education or sanitation as well as to health centres and hospitals is difficult (Doumbia, 2010: 148).

*Table 2-2: Selected socio-cultural and socio-economic indicators for the study area Burkina Faso (BF) and Côte d'Ivoire (RCI). Life expectancy at birth is given in years representing the situation in the year 2011; Infant mortality rate is denoted as number of deaths per 1000 live births for the year 2011; the percentage of population growth rate has been calculated for the time span between the year 2002 and 2012; literacy rate is denoted as percentage of above 15 year old adults (a) or adolescents (ac) between 15 and 24 years who were literate in the year 2007 (BF), respectively in the year 2011 (RCI); school absenteeism is given as percentage of children at primary school age who were out of school for the year 2011 (BF), respectively for the year 2009 (RCI); access to improved sanitation is indicated as percentage of the population with access to improved sanitation facilities for the year 2011; malnutrition prevalence is given as percentage of children under 5 years derived from the weight for age ratio for the year 2010 (BF), respectively for the year 2007 (RCI); the gross domestic product (GDP) is given in billion US\$ and estimated for the year 2012; the level of poverty is given as percentage of the population that live with less than US\$ 2 per day and represents the situation in the year 2009 (BF), respectively in the year 2008 (RCI).*

Indicator	Burkina Faso	Côte d'Ivoire	Source
Life expectancy [years]	55	55	UNESCO (2011)
Infant mortality rate [‰]	82	81	UNESCO (2011)
Population growth [%]	3.4	2	FAO (2014a)
Literacy [%]	28.7 (a), 39.3 (ac)	56.9 (a), 67.5 (ac)	UNESCO (2011)
School absenteeism [%]	35	38	UNESCO (2011)
Access to sanitation [%]	18	23.9	World Bank (2014)
Malnutrition prevalence [%]	26.2	29.4	World Bank (2014)
GDP [billion US\$]	10.44	24.68	World Bank (2014)
Poverty [%]	73	46	UNESCO (2011)

As already mentioned in Section 2.1.4, humans have converted most of the original forest into savannah and agricultural land and only few natural parks and sacred sites remain (Janssen et al., 2010: 90). Agriculture emerges in the northern part of the study area, as soon as river water or sufficient amount of precipitation allows this. In general, West African agriculture is predominantly rain fed and therefore dependent on the seasonal and spatial distribution of

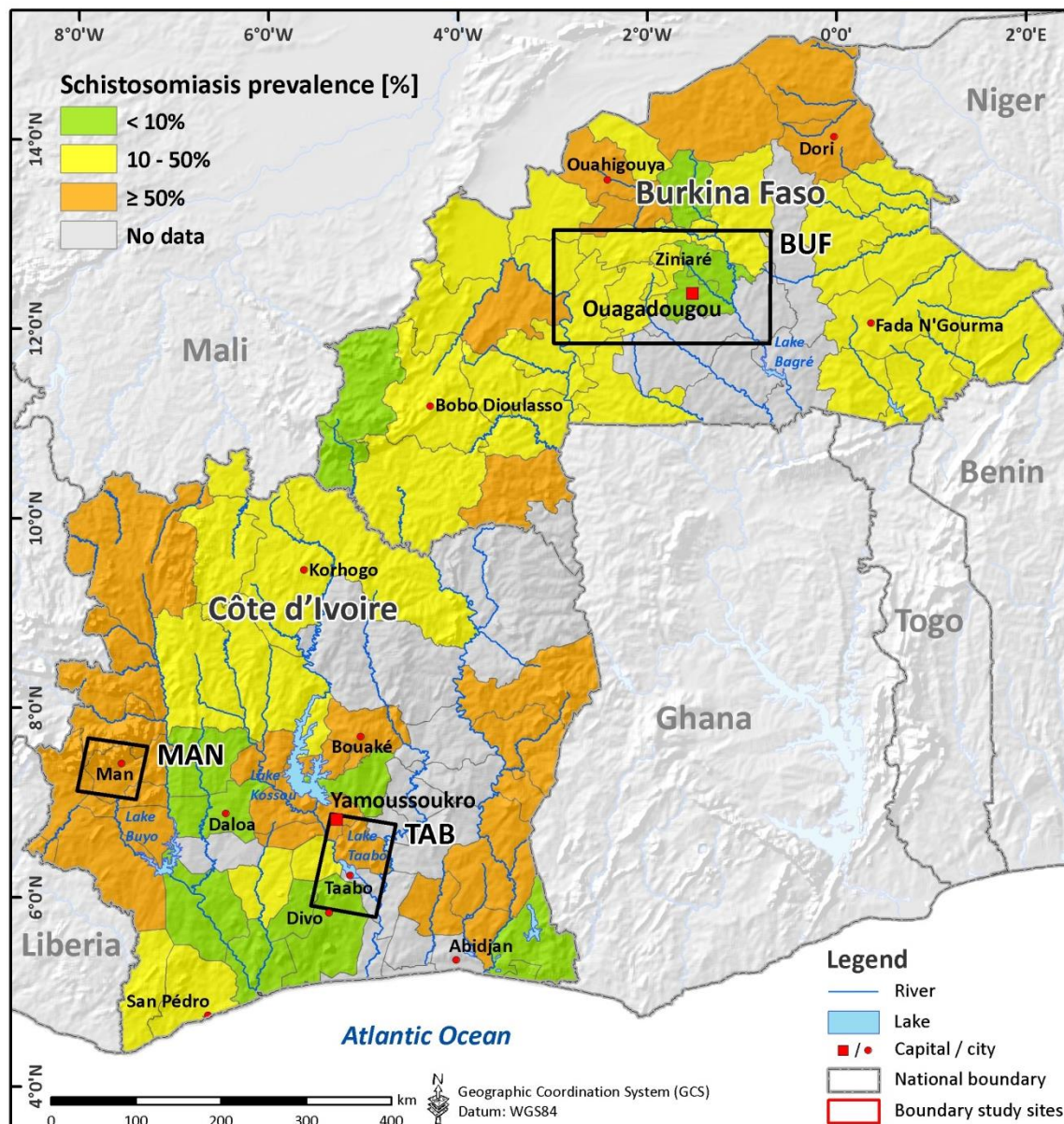
precipitation (Janssen et al., 2010: 88). However, especially in northern and central Burkina Faso the collecting of rain water through a variety of artificial river dams and pools (Section 2.1.5) is an established practice to prolong or even enable agriculture. Agricultural practices are mainly based on traditional techniques for crop production (i.e. millet, maize, sorghum). Only the large-scale and water-intensive production of cotton and rice is driven by politically supported modern agricultural techniques. In the northern area, livestock farming is extensively practiced by traditional herdsman and farmers (Senghor, 2010b: 143-144). In Côte d'Ivoire, the most widely applied agricultural method is the use of fire for clearing in slash-and-burn cultivation, which constitutes the cutting a section of forest or thinning out savannah to prepare or renew land for rain-fed crop farming (Kouassi and Ahoussi, 2010: 150-151). In Burkina Faso, of the more than 8 million economically active inhabitants, 7.4 million are active in agriculture, whereas in Côte d'Ivoire, from 7.8 million economically active inhabitants, only 2.8 million are active in agriculture (FAO, 2014a).

## 2.3 Schistosomiasis epidemiology

The study area is an endemic schistosomiasis region where moderate to high transmission rates have been observed (WHO, 2010b). The general epidemiology and control strategies of schistosomiasis are described in Sections 3.1.1 and 3.1.3, respectively. In the following, specific epidemiological characteristics and national activities of disease control with relevance to this thesis are briefly documented for Burkina Faso and Côte d'Ivoire.

In Burkina Faso, a review by Poda et al. (2004) has confirmed that no districts were free of schistosomiasis transmission and its spatial distribution had a typical focal pattern. The most frequent infections spread all over Burkina Faso result from the *Schistosoma haematobium* parasite, whereas *Schistosoma mansoni* was less frequent and located only in six districts in the southern and western part of the country (Poda et al., 2004). Epidemiological surveys that have been conducted between 2003 and 2007 revealed a prevalence rate of *S. haematobium* between 1.7 and 81.7% with the severest infection rate in the northern region (Figure 2-8) (Dadjoari, 2011). The corresponding snails prevalent in Burkina Faso are *Bulinus truncatus*, *Bu. senegalensis* and *Bu. globosus* for *S. haematobium* and *Biomphalaria pfeifferi* for *S. mansoni*. The heterogeneity in transmission pattern of schistosomiasis is closely linked to the spatial distribution of preferred breeding sites of the respective snail hosts and the principal point of contact between people and the parasites (Boelee et al., 2009: 13). Based on snail surveys in Burkina Faso carried out between 1985 and 1995, Poda (1996: 57-59) has found that 41% of intermediate snail hosts have been found in small reservoirs, 34% in rivers, 20% in temporary ponds, 3% in irrigation channels and 2% in natural lakes. This shows the importance of small reservoirs with respect to the distribution of schistosomiasis in Burkina Faso (Boelee et al., 2009: 16). In general, changes in the natural hydraulic of water systems through dam construction are an amplifying factor for the proliferation of mollusc species and parasite exchange (Poda, 1996; Dianou et al., 2003; Boelee et al., 2009). In northern Burkina Faso, the climate-induced water shortage results in a concentration of domestic activities around reservoirs and temporary ponds that are mostly contaminated with the parasite (Poda et al., 2004). Beyond the 14° northern latitude, parasite transmission takes place only at ponds available during the rainy season, whereas temporary rivers exist typically for very short periods and have a fast drainage,

which does not allow snail development (Boelee et al., 2009: 14). Towards the South between 14° and 12° northern latitudes, the transmission foci of schistosomiasis are numerous small water reservoirs, dam lakes together with horticultural activities and isolated perennial rivers all of which result from an increasing amount of rainfall (Boelee et al., 2009: 14-15). In the Sudanian climate zone (Figure 2-4) between 12° and 10° northern latitudes most rivers and water reservoirs are permanent. However, in this region the overall prevalence rate of schistosomiasis is the lowest of the entire country. Here, the focal points of high prevalence rate (up to 80%) were observed around large irrigated areas (Boelee et al., 2009: 15).



Data source:  
Elevation: Shuttle Radar Topography Mission (SRTM) by NASA, JPL, DLR and ASI  
Lakes: Savane, 2010b: 127 & Dipama, 2010b: 135 provided by Dr. Joachim Eisenberg (SFB 268), University of Frankfurt, Germany  
National boundary: GADM (www.gadm.org)  
Schistosomiasis prevalence: Dadjoari, 2011 (BF), ICL, 2014 (IC)

Figure 2-8: Geographic distribution of schistosomiasis prevalence in the study area on district level based on estimates of the national schistosomiasis control programme of Burkina Faso between 2003 and 2007 (Dadjoari, 2011) and estimates from epidemiological studies between 1998 and 2005 in Côte d'Ivoire (ICL, 2014).

Similar to Burkina Faso, schistosomiasis transmission is endemic nationwide in Côte d'Ivoire with more than 21.5 million inhabitants who require treatment (WHO, 2010a). Both parasites, *S. haematobium* and *S. mansoni* are prevalent in Côte d'Ivoire. Their corresponding snails in Côte d'Ivoire are *Bu. forskalii*, *Bu. globosus*, *Bu. truncatus* and *Bio. pfeifferi*, respectively (Kinanpara et al., 2013: 110). Epidemiological studies that have been conducted between 1998 and 2005 revealed prevalence rates of *S. haematobium* between 5.5 and 38.6% and of *S. mansoni* between 13.6 and 57.4%, whereas the western and eastern districts of the country as well as the central region around Lake Kossou show highest transmission rates with prevalence greater 50% (Figure 2-8). Similar to the situation in Burkina Faso, the construction of large dams such as Lake Kossou or Lake Taabo led to a marked increase of schistosomiasis prevalence predominantly caused by *S. haematobium* (N'Goran et al., 1997: 541). The typical sites for potential disease transmission in Côte d'Ivoire are dam lakes, natural ponds, irrigation canals, and river confluences or bulges, where current of the river is slow and human contact frequent (Kinanpara et al., 2013: 110).

The surveillance and control of schistosomiasis are subject to supervision by national health authorities with the consequence that national borders delineate the respective efforts and applied practices between countries. In Burkina Faso, with assistance from the Schistosomiasis Control Initiative (SCI) (Fenwick et al., 2009) nation-wide control of the disease has been implemented by the following steps: (i) identifying the most heavily infected regions; (ii) training local health staff and teachers; (iii) providing health education to the local population; and (iv) distributing the drug praziquantel to treat against the disease. Thereby, during mass treatment campaigns conducted between 2004 and 2006 more than 6 million children aged between 5 and 15 years have been treated in Burkina Faso (MoH and PNLSc, 2010: 6). More than 97% of the financial resources invested for schistosomiasis control in Burkina Faso stem from funding through the United States Agency for International Development (USAID) and the SCI, and less than 3% result from the national budget (MoH and PNLSc, 2010: 20). Since 2008, Burkina Faso is no longer a target country of the SCI (ICL, 2013).

In Côte d'Ivoire a national control programme has been established in 1998, but due to limited funding and subsequent civil unrest, mass drug administration never happened at this scale (Tchuem Tchuente and N'Goran E, 2009: 1741-1742). Thus, in the year 2010 Côte d'Ivoire has been classified into the first group within the Integrated Control of Schistosomiasis in Sub Saharan Africa (ICOSA) project, coordinated by the SCI, because no treatment had been given previously. The objective was to map schistosomiasis in 66 targeted districts within Côte d'Ivoire by 2014 and distribute community- and school-based treatment dependent on the mapping results. Treatments begun in June 2012 and were scheduled to cover more than one million until May 2014. At the time of writing the current thesis, mapping has been completed (Eliézer K. N'Goran, personal communication). If mapping results in full endemicity of the country as already indicated by historical data, the project covered 12% (1.84 million) of the total population of Côte d'Ivoire with treatment (ICL, 2013).





## 3 Remote sensing of schistosomiasis risk

### 3.1 Schistosomiasis

Schistosomiasis is a parasitic disease in humans caused by blood flukes of the genus *Schistosoma*. The transmission cycle of the disease from human to human requires the parasite to meet specific snails as an intermediate host. These snails release the parasite in a development stage where it can infect humans within an aquatic environment.

The disease has already been known to occur in Egypt and Mesopotamia amongst the earliest agricultural civilizations of the great river valleys (Farooq, 1973: 1-2). As early as the 16<sup>th</sup> century BC, haematuria, a typical sign of urogenital schistosomiasis, has been depicted in hieroglyphs in Egyptian papyri and paintings (Farooq, 1973: 2; Adamson, 1976: 177; Jordan, 2000: 9). Calcified parasite eggs have been found in the kidneys of Egyptian mummies from the 20<sup>th</sup> dynasty between 1184 and 1087 BC (Ruffer, 1910: 16; Farooq, 1973: 2). In 1851, the German physician Theodor Bilharz discovered the parasite *S. haematobium* recovered post-mortem from the mesenteric veins of Egyptians and demonstrated their relationship to haematuria and eggs passed in the urine (Bilharz, 1852: 72-73; Sturrock, 1993b: 1). Still today, the term bilharzia is used as a synonym for schistosomiasis in association with this discovery (Utzinger et al., 2011b: 10)

#### 3.1.1 Geographical distribution and epidemiology

From a global public-health perspective, schistosomiasis is the most significant water-based disease (Steinmann et al., 2006: 411). Global statistics on disease burden suggest that 779 million people are at risk (Steinmann et al., 2006: 414-415) and about 440 million people are currently infected (Colley et al., 2014: 2259) with 97% of all infections occurring in Africa (Figure 3-1). There is a considerable discussion regarding the true burden of the disease (King et al., 2005: 1564-1566; Gryseels et al., 2006: 1113; King and Dangerfield-Cha, 2008: 73; Utzinger et al., 2011a: 124-125; Murray et al., 2012: 2204). It ranges from lowest estimates of 1.7 million disability-adjusted life years (DALYs) (WHO, 1999: 104), an updated considerably higher value of 4.5 million (WHO, 2002a: 2) to a maximum of 70 million DALYs assessed by King and Dangerfield-Cha (2008: 73). The most recent estimate of 3.3 million DALYs due to schistosomiasis is given by Murray et al. (2012: 2204). The difficulty in assessing the true burden of schistosomiasis are the resulting end-organ pathologies, impaired growth and development of children, chronic

inflammation, anaemia, and other health consequences of the disease. Despite a comparatively low mortality rate of 11,700 deaths estimated in the year 2010 (Lozano et al., 2012: 2105), the burden of the disease is estimated to be equivalent to malaria or HIV/AIDS (Hotez and Fenwick, 2009: 1).

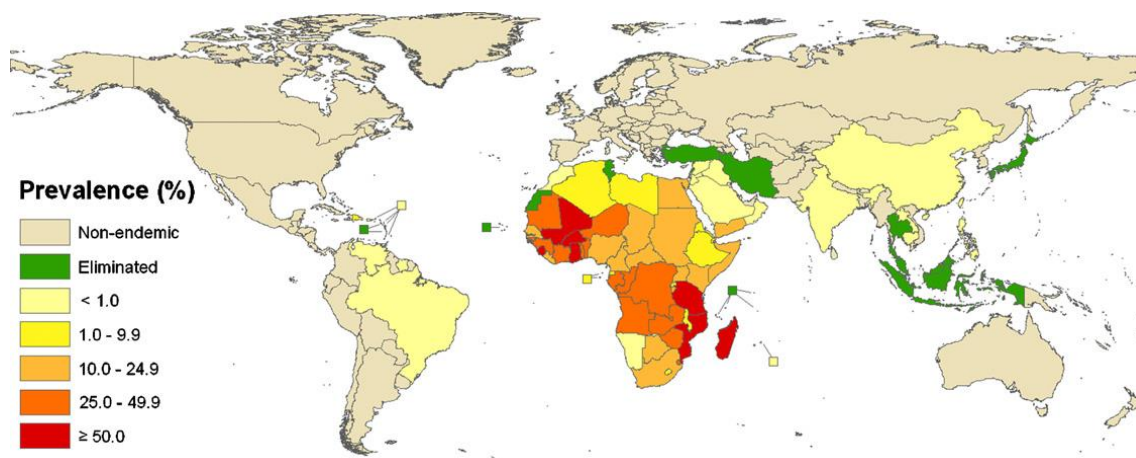


Figure 3-1: Global distribution of schistosomiasis. Source: Utzinger et al. (2011a: 124)

Schistosomiasis is a typical disease of poverty (WHO, 2013) that is widespread where access to clean water and basic sanitation is lacking, hygiene is at a sub-standard level and health infrastructure is weak or non-existent (Bruun and Aagaard-Hansen, 2008: 50; Utzinger et al., 2009: 1863; King, 2010: 102; Utzinger et al., 2011a: 122). In endemic parts of the world, the prevalence of schistosomiasis is intimately linked with water resources development projects and irrigated agriculture (Hunter et al., 1993; Steinmann et al., 2006: 411). The modification of flowing hydrological regimes to stagnant water bodies enabled the spreading of the disease to previously non-endemic areas (Dianou et al., 2003: 107-108; Fenwick, 2006: 1077), a situation that might be further exacerbated by climate change and an increasing pressure of humans on environmental resources as outlined in Section 1.1 (Martens et al., 1997; Yang et al., 2005b: 131; Zhou et al., 2008: 192-193; Utzinger et al., 2011a: 122).

### 3.1.2 Ecology of schistosomiasis transmission

*“Transmission of schistosomiasis is the result not only of interplay between humans, snails and parasites, but also of complex demographic, environmental, biological, technological, political, socio-economic and cultural processes” (Bruun and Aagaard-Hansen, 2008: 1)*

As stated here, it becomes clear that human acquisition of schistosomiasis is multi-faceted and complex, integrating various disciplines to understand and research this process. The parasite life cycle, its intermediate snail host, and its definitive vertebrate hosts provide the fundamental basis that schistosomiasis transmission from human to human can occur (Figure 3-2). There are six schistosome species parasitising humans, namely, *S. haematobium*, *S. mansoni*, *S. japonicum*, *S. intercalatum*, *S. mekongi* and *S. guineensis*. The former three are the most widespread and important from a public health point of view (Utzinger et al., 2011b: 10-11). To complete a successful life cycle, *S. haematobium* is transmitted almost exclusively by snails of the genus

*Bulinus*, of which however, not all species are susceptible (Sturrock, 1993a: 33). All intermediate hosts of the *S. mansoni* parasite belong to the genus *Biomphalaria* (Sturrock, 1993a: 33). Both snail genera are pulmonate snails of the *Planorbidae* family and live in aquatic environments (Sturrock, 1993a: 33). In contrast, *S. japonicum* is transmitted by snails of the genus *Oncomelania*, which results in a different transmission ecology due to the amphibious habitat of these snails (Sturrock, 1993b: 4-5). For the scope of this research only the transmission ecology of *S. haematobium* and *S. mansoni* and its aquatic intermediate hosts are relevant due to their endemic distribution within the study area of Burkina Faso and Côte d'Ivoire.

Cercariae represent the infective development stage of the parasites that enter the human body through penetrating the intact skin within an aquatic environment (Sturrock, 1993b: 12). The whole process of penetration through the skin of humans is completed within a few minutes (Sturrock, 1993b: 15). Each successful cercaria travels through the blood circulation of its host to reach the liver and develop into a single adult worm (Sturrock, 1993b: 15-16). After reaching sexual maturity and the pairing of the worms, female worms release eggs (Sturrock, 1993b: 16). These become either trapped in tissues of the human host causing pathology due to immune reactions and progressive damage to organs, or leave the human body with excrements (urine in the case of *S. haematobium* and stool in the case of *S. mansoni*) with the perspective to continue the parasite life-cycle (Jordan and Webbe, 1993: 125). The time span between penetration of cercariae and the first passage of eggs in excreta varies between 34-35 days for *S. mansoni* (Clegg, 1965: 140) and 70 days for *S. haematobium* (Smith et al., 1976: 104). The mean life span of the adult worm in humans range from 3 to 10 years (Jordan and Webbe, 1993: 110), however, observations indicate a maximal life span of adult schistosome worms for more than 30 years (Chabasse et al., 1985: 643; Jordan and Webbe, 1993: 109-110).

Eggs from the parasite contain fully developed larvae (i.e. miracidium) ready to hatch if they reach freshwater (Sturrock, 1993b: 5-6). Excreted eggs can remain viable for about a week if they are not exposed to excessive heating or desiccation (Upatham, 1972: 274-275; Sturrock, 1993b: 10). Miracidia are chemo-sensitive and swim with a speed of about 2mm per second searching for their appropriate intermediate host snail species (Sturrock, 1993b: 10). They remain active for 8 to 12 hours, but their infectivity starts to drop rapidly within 4 to 6 hours after hatching (Sturrock, 1993b: 11). A successful miracidium penetrates the body of the snail within a few minutes (Sturrock, 1993b: 11).

Within the freshwater snail, the miracidium can further develop to cercariae. Following an asexual reproduction of the parasite within the snail, one single miracidium can give rise to many hundreds or thousands of cercariae for several months although snails can also become self-cured and stop cercarial shedding (Sturrock, 1993b: 12). Snails can fundamentally modify the parasite life cycle and with it the transmission of the disease and its risk, due to specific environmental conditions that influence snail survival, reproduction and the competence of a snail to develop the parasite from a miracidium to cercariae (Table 3-1).

Similar to the miracidium, a cercaria is the second, free living, infective schistosome larva which does not feed and is adapted to live in freshwater (Sturrock, 1993b: 12). After snails have shed cercariae, their life span varies between 48 to 72 hours according to their initial food reserves (glycogen) and their intensity of spending it through active swimming (Sturrock, 1993b: 14). However, infected snails can produce between 250 and 600 cercariae per day life-long and

cercarial output might reach several thousand depending on size of the snail (Sturrock, 1993a: 64).

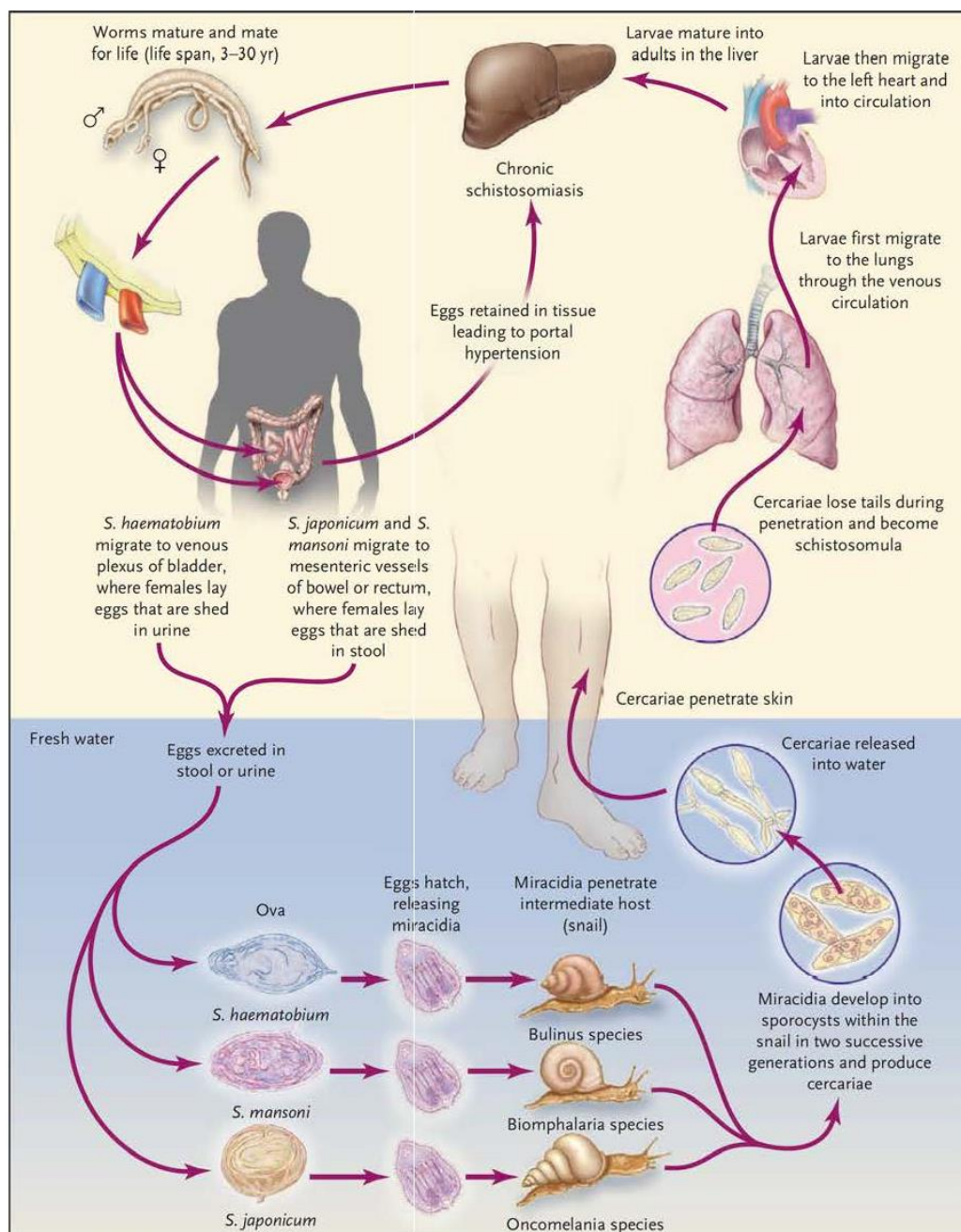


Figure 3-2: Consecutive steps of the parasite life cycle of schistosomiasis transmission: Eggs of the parasite are excreted by infected humans with faeces for *S. mansoni* or urine for *S. haematobium*. When eggs meet water they start hatching and parasite larvae are released as miracidia, search for their appropriate snail species as intermediate host and penetrate its tissue. Following further development in the snail host, the parasite is released by the snail as the infective stage of cercaria that can infect humans through the penetration of the intact skin when humans enter a water body. Within the human body, the parasite migrates through the blood vessel to the portal vein in the liver, where growth and sexual pairing starts. Dependent on the schistosome species, the adult worms embed themselves in the mesenteric venules of the bowel or bladder and continuously produce eggs that are excreted by humans via stool or urine, respectively. If eggs become trapped, morbidity is caused by egg calcification and chronic inflammation of the tissue.

Source: (King, 2009: 107)

The ecology of parasite and snail species as well as human characteristics stimulate, inhibit or modify the dynamics of disease transmission in a specific manner. The environment has thereby an essential impact on the ecology of disease transmission such as successive rates of reproduction, survival, and probability that the next step in the parasite life cycle is reached (Hairston, 1973: 278-279). In the following paragraphs, factors that specifically influence the life cycle of *Schistosoma* parasites, the ecology of intermediate snail hosts, and the vulnerability of humans to become infected with the parasite are reviewed in detail. An overview of the relevant variables and its effects on the ecology of disease transmission is presented in Table 3-1.

### The *Schistosoma* parasites

The **water temperature** has a major influence on the length of the prepatent period of a *Schistosoma* parasite, which is again related to its abundance in the environment. Within the temperature range of 10-30°C, the hatching of the eggs is stimulated by a combination of light and dilution of the shell that further affects the osmotic pressure (Sturrock, 1993b: 10). The length of the prepatent period from penetration of the miracidium to initial shedding of cercariae by the snail varies with temperature between the minimum of 17 days at 30-35°C and several months towards cooler temperatures (Sturrock, 1993b: 12). Pflüger (1980) and Pflüger et al. (1984) found a species-specific length of the prepatent period for *S. mansoni* and *S. haematobium* parasites. The general increasing length of the prepatent period with decreasing water temperature reached the developmental null point of parasite development at 14.2°C for *S. mansoni* (Pflüger, 1980: 162) and 15.3°C for *S. haematobium* (Pflüger et al., 1984: 99). The development time of *S. mansoni* parasites in the snails ( $y$ ) given in days has been approximated by Equation 3-1,

$$y = \frac{268}{x - 14.2} \quad \text{Equation 3-1}$$

where  $x$  represents the measured water temperature and 268 has been calculated as the sum of biological relevant temperatures until cercarial shedding becomes constant (Pflüger, 1980: 162-164). This relation was slightly shifted for the case of *S. haematobium*, where the theoretical developmental null point was reached at 15.3°C and the sum of the constant time-temperature product resulted in 295 (Pflüger et al., 1984: 99). It has further been observed, that the maximum of *S. haematobium* cercarial shedding was reached at water temperatures around 25°C (Pflüger et al., 1984: 100-101).

**Water flow velocity** influences the spatial distribution of the parasite. Stagnant water refers to highest cercarial density, whereas flowing water can transport the parasite passively for considerable distances (Jordan and Webbe, 1993: 97). Very slow moving water with a speed of approximately 0.1m/s is beneficial to allow the widespread dissemination of the parasite and meet its intermediate and definitive host (Upatham, 1973: 296; Sturrock, 1993a: 62). At the same time, active parasite mobility has been observed to be stimulated by various components of sebum (sweat) secreted by humans (Haas and Schmitt, 1982: 304-305; Stirewalt et al., 1983: 366).

*Schistosoma* parasites in the stage of miracidia have natural **predators** such as fish and diverse carnivorous invertebrates that feed on them and thereby reduce their abundance by a certain degree (Gibson and Warren, 1970: 835; Jordan and Webbe, 1993: 122).

**Sunlight** is a particular stimulus for the release of cercariae from infected snails, whereas the number of infected cercariae produced is mainly influenced by the size of the snail and its ambient temperature (Sturrock, 1993b: 13).

Table 3-1: Overview of parasite-, snail- and human-related factors that modify, retain or intensify the cycle of schistosomiasis transmission

Parasite-related factors	Effect on schistosomiasis transmission	Reference(s)
Temperature	Length of prepatent period Activity, survival and infection rate of free-living stages of the parasite	Pflüger (1980) Sturrock (1993b)
Water flow velocity	Passive transport of parasites in flowing water determines cercarial density	Jordan and Webbe (1993)
Predators	Fish and carnivorous invertebrates reduce parasite population as natural predators	Gibson and Warren (1970); Jordan and Webbe (1993)
Sunlight	Stimulation of cercarial shedding	Sturrock (1993b)
Pathogenicity	Different strains of <i>S. mansoni</i> and <i>S. haematobium</i> result in geographical variations of disease severity	Stirewalt (1973)
Species	Different efficiency in identifying and infecting snails	Sturrock (1993b)
Snail-related factors	Effect on schistosomiasis transmission	Reference(s)
Water temperature	Fecundity, mortality and rate of reproduction	Abdel-Malek (1958); Shiff (1964); Shiff and Garnett (1967); Appleton (1978); Pflüger (1980); Pflüger et al. (1984)
Water flow velocity	Flow velocity > 0.3m/s result that snails become dislodged and swept away	Scorza et al. (1961); Appleton (1978); Sturrock (1993a)
Vegetation	Food supply Surface to crawl and deposit egg masses Increase of dissolved oxygen	Abdel-Malek (1958)
Substratum	Nature of substratum is related to snail abundance	Abdel-Malek (1958); Appleton (1978)
Water depth	Snails generally found in shallow water near the margins of their habitats. Below 1.5 to 2 m, snails have little importance for the transmission of schistosomiasis	WHO (1957)
Fluctuations of water level	Permanence of available habitats determines the distribution patterns of snails	Abdel-Malek (1958); Appleton (1978)
Rainfall	Creation of temporary snail habitats Increase of water flow velocity Supports contamination of water Passively transports snails when rainfall is heavy	Appleton (1978); Jordan and Webbe (1993); Sturrock (1993a)
Turbidity	Turbidity can impact the reproduction cycle	Harrison and Farina (1965); Appleton (1978)
Water chemistry/ quality	Low pH, refuse from factories directly harm snails High snail abundance where water is polluted with human excrements	Deschiens (1954); Abdel-Malek (1958); Appleton (1978); Sturrock (1993a)
Sunlight	Completely shaded pools provide unsuitable habitat Activity of snails is high in direct sunlight	Abdel-Malek (1958)
Predators/pathogens	Natural predators, parasites and pathogens may limit the abundance of snails	Abdel-Malek (1958)
Species	Variation of susceptibility to parasite and efficiency to produce cercariae	Mulvey and Vrijenhoek (1982); Sturrock (1993a)

Human-related factors	Effect on schistosomiasis transmission	Reference(s)
Water contact behaviour	Exposure of the skin to parasite infested water is the prerequisite for human infection.	Bundy and Blumenthal (1990)
Hygiene	Contamination of water due to excrements of infected humans in or aside water	Farooq et al. (1966); Huang and Manderson (1992); Jordan and Webbe (1993)
Gender	Relationship between gender and risk of infection is culturally variable and a determinant of water contact activities	Husting (1983); Chandiwana (1987); Huang and Manderson (1992)
Age	Highest risk for children as consequence of degree of exposure and low level of immunity	Bundy and Blumenthal (1990); Huang and Manderson (1992); Butterworth (1993)
Immunity	Resistance to reinfection can be developed by the human body as a consequence of previous infections	Butterworth (1993)
Ethnic origin	Variation in the susceptibility to infection	Jordan and Webbe (1993)
Religion	Religious rules are related to water contact behaviour and disease exposure	Huang and Manderson (1992); Jordan and Webbe (1993)
Socioeconomic status	Relation to hygiene, the availability of protected water supplies and ability to cope with the disease	Lima e Costa et al. (1987)
Migration	Population movements can modify spatial patterns of disease transmission through both introduction of the parasite or the acquisition of infection	Doumenge and Mott (1987); Bundy and Blumenthal (1990); Jordan and Webbe (1993)
Occupation	Work related to water increases the exposure and risk of infection (fishermen, farmer, etc.)	Farooq et al. (1966); Huang and Manderson (1992)
Location of house	Location of house in relation to suitability of closest water source can influence infection status	Mota and Sleigh (1987); Huang and Manderson (1992); Clennon et al. (2006)
Prevention/control measures	Spatial pattern of disease transmission can be highly modified by mass treatment campaigns and successful preventive measures	Webbe and Jordan (1993); Clements et al. (2009b); Zhang et al. (2012)

Internal factors of the parasite may modify the risk of infection such as different strains of *S. haematobium* and *S. mansoni* having differing **pathogenicity**. This could also account for geographical variations in severity of human schistosomiasis (Stirewalt, 1973: 30-31). The efficiency of snail and human infection varies with **species**. Hence, miracidia of *S. haematobium* have shown to need more individuals to infect their intermediate snail host than those of *S. mansoni* (Sturrock, 1993b: 11). Additionally, intermediate host snails of *S. haematobium* are more dispersed than those of *S. mansoni*, which further results in lower field snail infection rates for *S. haematobium* (Sturrock, 1993b: 11).

### **The freshwater snail as intermediate host**

Similar to the parasite, survival, fecundity, and rate of reproduction of freshwater snails are sensitive to **water temperature**. Despite snails having broad tolerance ranges of their ambient temperature between day and night or seasons, the most favourable range lies between 18 and 32°C (Appleton, 1978: 4). Snails of the genus *Bulinus* show a distinct peak of maximal reproduction at 25°C (Shiff, 1964: 103), whereas *Biomphalaria* represent a plateau of high reproduction rates between 20 and 27°C (Shiff and Garnett, 1967: 437-438; Appleton, 1978: 5). The correlation between thermal regimes and snail fecundity allowed deriving a critical level of 120-179 degree hours greater than 27°C per week. When this limit was exceeded, snails were absent from this habitat (Appleton, 1978: 7). In contrast to the parasite, snails are more sensitive to warm conditions and mortality increased to 100% when exposed to 36°C and higher for a few days (Pflüger, 1980: 164). However, in sub-tropical regions of Africa, the impact of temperature on the limitation of snail distribution is only relevant for very small water bodies exposed to continuous high temperatures (Abdel-Malek, 1958: 788).

Regarding water **flow velocity**, freshwater snails have a noticeable narrow tolerance range (Appleton, 1978: 10) and become dislodged when flow velocity exceeds approximately 0.3 m/s (Scorza et al., 1961: 194). A nearly linear, negative correlation between the density of *Biomphalaria spp.* and water flow velocity has been derived until the aforementioned limit (Scorza et al., 1961: 193; Appleton, 1978: 10). It has also been shown, that snails were being dispersed along streams and irrigation schemes (Clennon et al., 2007: 690), which can result in the agglomeration of snails and parasites in downstream areas, especially during and following sufficient rains. Beck-Wörner et al. (2007: 961) showed that habitat suitability for *Bio. pfeifferi* was increasing with higher stream order assuming that higher stream order was linked to decreased flow velocity and low stream order was related to streams more likely to desiccate during the dry season.

**Vegetation** determines the habitat suitability of freshwater snails in several ways. First, the presence of aquatic vegetation is positively linked to an increase of the amount of dissolved oxygen and the consumption of carbon dioxide (CO<sub>2</sub>) and thereby linked to movement and reproduction of pulmonate snails (Abdel-Malek, 1958: 804). Second, snails seek broad-leafed vegetation as surfaces to crawl and deposit their egg masses on. Third, the periphyton, which encrusts the submerged parts of the plant, provides the food supply for snails (Abdel-Malek, 1958: 813).

The nature of **substratum** of a water body is related to snail abundance. Whereas firm mud rich in decaying organic matter provides a favourable habitat for snails, clean sand, semi-liquid mud or bottom loose of organic matter does not provide a suitable snail habitat (Abdel-Malek, 1958: 794-795). The availability of food and firm surfaces for oviposition modify the selection of the substratum by the snail (Appleton, 1978: 15).

**Water depth** is related to the distribution of freshwater snails, which are generally found in shallow water near the margins of their habitats as a relation of food, shelter and light conditions (WHO, 1957: 11). Despite the fact that snails are able to survive at a depth of 10 m, they are rarely found below 1.5-2 m. However, the presence of snails in deeper water has little importance for the transmission of schistosomiasis (WHO, 1957: 12).



Sudden **fluctuations of water levels**, such as irrigation channels with certain pump schemes, provide habitats of low suitability to establish a snail population (Abdel-Malek, 1958: 792). Thus, the permanence of available habitats is one further criterion to determine the population and distribution patterns of host snails. However, snail populations are also able to persist in temporary habitats through their ability to aestivate during periods of drought in sheltered spots, under vegetation, on mud or in mud crevices (Abdel-Malek, 1958: 792; Appleton, 1978: 12). In general, species of *Bulinus* are more successful to withstand periods of prolonged desiccation surviving up to one year through burying themselves beneath the substratum compared to species of *Biomphalaria* (Appleton, 1978: 13). During this time, development of the parasite and cercarial shedding may be suspended temporarily (Pitchford et al., 1969: 370).

**Rainfall** modifies snail habitat conditions in manifold ways. If rainfall is heavy, snail populations will be reduced through being swept away, as flow velocity of water is at the same time increasing (Appleton, 1978: 10; Sturrock, 1993a: 53). Furthermore, temporary habitats are created by enduring rainfall events and snails can establish a population either if they survived desiccation or by being passively transported to the temporary habitat with the discharge of the rain. Another relevant aspect with respect to the probability of disease transmission is that rainfall directly supports the contamination of water through washing human faeces with large amounts of parasite eggs into the potential snail habitats (Jordan and Webbe, 1993: 118).

Water **turbidity** due to a high content of suspended minerals (360 mg/l) can impact the reproduction cycle of freshwater snails through smothering egg masses, preventing development and hatching of eggs, however, adult snails were not affected (Abdel-Malek, 1958: 790; Harrison and Farina, 1965: 329-330). Growth of aquatic plants is limited due to high turbidity and thereby habitat conditions become deteriorated (Abdel-Malek, 1958: 790).

With respect to **water chemistry** and **quality**, a low pH value in general may be directly harmful to snails (Sturrock, 1993a: 51). Their frequency was found to be proportional to water hardness with clear preference for very hard waters (Appleton, 1978: 3). Maximum tolerated concentrations and lethal concentrations of certain ions for snail species have been quantified by Deschiens (1954: 918). In general, *Bulinus spp.* show a greater tolerance to changing chemical conditions than *Biomphalaria spp.*, however, the latter genus was found to have higher tolerance to chloride (Cl<sup>-</sup>) and sodium (Na<sup>+</sup>) concentrations (Deschiens, 1954: 917-918; Abdel-Malek, 1958: 796). Industrially polluted waters were found to be unsuitable for intermediate host snails, whereas abundance was high near human habitations, which pollute water with their excrements potentially containing large amounts of the parasite (Abdel-Malek, 1958: 793).

Similar to the stimulation of cercarial shedding with daylight, snails themselves were observed to be noticeably active in **sunlight** (Abdel-Malek, 1958: 789). Egg masses of snails are often seen in direct sunlight and are apparently unaffected (Abdel-Malek, 1958: 789). Furthermore, sunlight corresponds to the flourishing of aquatic weeds, the abundance of microflora and thereby a high content of dissolved oxygen rendering the water highly suitable for snails (Abdel-Malek, 1958: 789; Appleton, 1978: 14). In contrast, completely shaded pools provide unsuitable habitat conditions and snails remain absent (Abdel-Malek, 1958: 789).

There are several vertebrate (e.g. crabs, fish, amphibians, birds, mammals) and invertebrate (e.g. insects, other snails) **predators** that influence the abundance of snails (Abdel-Malek, 1958: 810-812). Parasites such as leeches or trematodes and **pathogens** such as fungi, virus and

bacteria may be pathogenic to snails (Abdel-Malek, 1958: 806-810). However, both, predators and pathogens would mainly limit the abundance of aquatic snails (Abdel-Malek, 1958: 806).

Finally, the respective snail **species** have differing genetic predispositions that result in an intraspecific variation of the susceptibility to the miracidium of a parasite as well as its efficiency to produce cercariae (Mulvey and Vrijenhoek, 1982: 1199; Sturrock, 1993a: 59).

### **The humans as definitive host**

Humans are the definitive host, where the schistosome parasite grows, pairs and reproduces itself. Due to the life-long reproduction of the parasite within humans and the continuous excretion of eggs through faeces or urine, humans are at the same time potentially circulating the disease, when hygiene is sub-standard (WHO, 2013). Many of the human-related factors that influence transmission of the disease are in consequence of each other, which is described in the following paragraph and illustrated in Table 3-1.

The entry of the infective stage of the parasite by percutaneous penetration is fundamentally dependent on the behaviour of the prospective human host. Hence, **water contact behaviour** is the major, decisive factor related to the risk of infection with schistosomiasis. Even when the environmental setting provides most suitable conditions for the transmission of the disease, infection does not occur, if people do not either enter the water body or protect themselves from direct contact with the water. Major activities leading to infection were identified to be personal hygiene, swimming or bathing in water and washing clothes (Bundy and Blumenthal, 1990: 267). Less critical activities were the washing of objects, fetching water and crossing water bodies most likely due to the shorter duration of water contact (Bundy and Blumenthal, 1990: 267).

The contamination of surface waters and their surroundings with human faeces or urine containing *Schistosoma* eggs is the second major pillar for the transmission of the disease and is preliminary defined by the **hygiene** of the human population (Jordan and Webbe, 1993: 117). The rate of infection has been observed to be significantly higher in persons living in houses without a latrine and access to piped water (Farooq et al., 1966: 293; Huang and Manderson, 1992: 183; Grimes et al., 2014 - under review).

The relationship between **gender** and risk of infection is equivocal and varies with the cultural background of the people (Huang and Manderson, 1992: 180). In some regions, higher prevalence of infection in women could be related to the fact that water-related activities were four times greater for women than men (Husting, 1983: 28-29), whereas this was *vice versa* when men dominated the activities with exposure to water (Chandiwana, 1987: 502). Therefore, the predictive power of infection risk based on gender is poor (Chandiwana, 1987: 504; Huang and Manderson, 1992: 180)

**Age** of humans influences the risk of acquiring the disease in manifold ways. In general, children and adolescents are the highest risk group with a peak at around 10 years (Bundy and Blumenthal, 1990: 278). On the one hand this is due to high exposure when fetching and playing in water (Huang and Manderson, 1992: 179) and on the other hand there is a low level of innate **immunity** to the disease (Bundy and Blumenthal, 1990: 282). It has been shown that humans are able to develop defence mechanisms that modify the effects of exposure with increasing age (Butterworth, 1993: 347-349). Hence, this typical convex shape of age-prevalence and age-

intensity with respect to schistosomiasis is argued to be related to a slow acquisition of immunity to reinfection following a slow death of adult worms from early infections (Bundy and Blumenthal, 1990: 278; Butterworth, 1993: 350).

Furthermore, it has been observed that the acute stage of the disease was rarely found in indigenous populations, yet very often in travellers to endemic areas (Jordan and Webbe, 1993: 124). Differences in the susceptibility to the disease with respect to **ethnic origin** of humans have been attributed to the immunological response influenced by different ancestral experiences with the infection (Jordan and Webbe, 1993: 124).

**Religion** plays a role when respective rules govern practices that may significantly affect patterns of water use (Huang and Manderson, 1992: 183). For example, ritual washing five times a day before prayer as required by male Muslims, significantly affects the prevalence of schistosomiasis in the respective communities (Jordan and Webbe, 1993: 101).

Schistosomiasis is a typical disease of poverty (King, 2010; WHO, 2013), which accentuates the high relevance of the **socioeconomic** status of a population with respect to the transmission of the disease. In African countries, being poor is very often related to poor hygiene and housing, limited access to clean water, subsistence farming and low educational level – all of them are established factors that impact transmission of schistosomiasis (Utzing et al., 2011a: 122-123). A very important aspect of a low socioeconomic status is further related to the lacking ability of a household to cope with the disease through seeking medical care and avoiding further infection through protective measures.

Often, large water resource development projects attract **migrant workers** and their families to the hot spots of potential disease transmission (Jordan and Webbe, 1993: 88). Semi-permanent or seasonal workers migrate to large agricultural projects for harvesting, often related to extensive irrigation schemes (Bella et al., 1980; Cheesmond, 1980; Bundy and Blumenthal, 1990: 88). These people are on the one hand at high exposure to becoming infected with the disease if they enter endemic areas, on the other hand they can also (re-) introduce the parasite into controlled or non-endemic areas (Jordan and Webbe, 1993: 88). It has also been observed that after new rail and highway systems have been opened in West Africa, the spatial distribution of schistosomiasis has followed (Doumenge and Mott, 1987: 11).

A positive relationship between the **occupation** of a person and risk of infection is not surprising, if this person is exposed to water during work (Huang and Manderson, 1992: 182). It has been shown that, especially farmers and farm laborers as well as fishermen and boatmen, had specifically higher prevalence rates compared to factory workers (Farooq et al., 1966: 306-308).

The **location of a household** in relation to the suitability of a water body to transmit the disease has shown to be highly relevant with respect to the level of prevalence (Huang and Manderson, 1992: 183-184). Thus, a local study in Ghana has shown that high infection levels were clustered around ponds known to contain snails that shed cercariae of *S. haematobium*, and prevalence was low in households close to a river where the intermediate host snails were rarely found (Clennon et al., 2006). The study by Mota and Sleight (1987) resulted that the relative location of a house to snail-free or snail-colonised water sources resulted as key influence on infection status in Brazil.

The application of various **prevention** and **control measures** (Section 3.1.3) has a remarkable impact on the transmission of the disease. Following mass treatment campaigns, the level of prevalence and morbidity decreases dramatically (Webbe and Jordan, 1993: 407-408; Koukounari et al., 2007: 659). However, examples (e.g. Clements et al., 2009b; Zhang et al., 2012) have demonstrated that the level of reinfection is often high, when prevention measures have been neglected and water bodies are still infested with the parasite. In the following section, the activities to prevent and control the disease are described in more detail.

### 3.1.3 Schistosomiasis control and elimination

Major strategies to prevent the transmission of the disease are: (i) to reduce human contact with infected water; and (ii) to reduce further contamination of water with eggs from the parasite (Webbe and Jordan, 1993: 405). In May 2001, at the 54<sup>th</sup> World Health Assembly (WHA) held at World Health Organisation (WHO) headquarters in Geneva, resolution WHA 54.19 was endorsed, urging member states to attain “a minimum target of regular administration of chemotherapy to at least 75% and up to 100% of all school-age children at risk of morbidity by 2010” and to “promote access to safe water, sanitation and health education through intersectoral collaboration” (WHA, 2001: 1). This resolution together with the guidelines presented in a subsequent WHO Technical Report Series (WHO, 2002a), were considered as the precursors for integrated and sustainable control of schistosomiasis (Utzinger et al., 2003: 1932; Stothard et al., 2009: 1668). The most significant response to WHA resolution 54.19 was the launch of the SCI, a charitable institution initiated with a start-up grant of US\$ 30 million from the Bill & Melinda Gates Foundation in 2002 (SCI, 2014). The major objectives of SCI are to implement and evaluate control of schistosomiasis and integrate these programmes into the structures of national health ministries to develop sustainable control (Fenwick et al., 2009: 1720-1721).

In practice, the antischistosomal drug praziquantel is administered to at-risk populations and results in significant reductions of morbidity due to schistosomiasis (Doenhoff et al., 2009: 1825). Through its availability at low cost (approximately US\$ 0.1-0.2 per treatment of a school-aged child) (WHO, 2002a: 11), it is distributed at large scale during survey and mass treatment campaigns at school locations. However, examples have shown that this control strategy failed to be sustainable, as re-infection occurred if prevention was neglected (Clements et al., 2009b: 7; Zhang et al., 2012: 5). Therefore, a shift from morbidity control to transmission control with focus on the snails and potential infection sites together with prevention measures tailored to the prevailing social-ecological system is the challenge of actual and future control of schistosomiasis (Stothard et al., 2009: 1672; Utzinger et al., 2011a: 132; Rollinson et al., 2013: 436).

RS data and methods have been investigated predominantly with the aim of supporting the work of national control programmes through prioritising areas of risk to most efficiently allocate available resources for disease prevention and control (Brooker, 2002: 211; Simoonga et al., 2009: 1687), which is reviewed in Section 3.2.2.

## 3.2 Remote sensing

In the PhD thesis presented here, RS refers to the use of Earth observing satellites. Satellite RS systems can be categorised into passive and active systems. The former measure the magnitude of electromagnetic radiation (EMR) reflected and emitted from the Earth's surface. The latter generate and emit their own radiative energy and capture the reflected returns from the Earth surface with a modification dependent on surface structure and condition. In this work, only passive systems have been used and shall therefore be considered in detail.

### 3.2.1 Fundamentals of satellite remote sensing

The fundamental principle of passive RS is based on the interaction of EMR originating from the Sun or the Earth itself with the atmosphere and the Earth surface as measured by a remote sensor. Hence, RS is composed of three main components: EMR carrying the information about the Earth surface, the capture of EMR using a remote sensor and the processing and analysis of the received signals. A comprehensive overview of the fundamentals of RS is given by Jensen (2000).

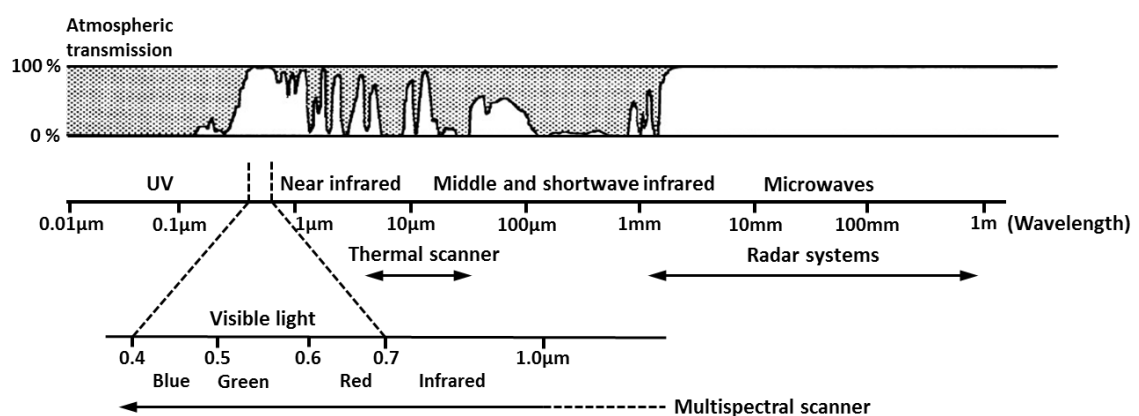


Figure 3-3: Electromagnetic spectrum. Source: modified after Albertz (2001)

Besides the Sun, electromagnetic energy is emitted by any object with a temperature above the absolute zero (0 Kelvin,  $-273^{\circ}\text{C}$ ), including water or vegetation. The propagation of this energy results from an electromagnetic wave that transports energy with the speed of light and a distinct relation between wavelength and frequency through space. The longer the wavelength, the lower is its frequency. The amount of radiant energy is called the radiant flux measured in watts [W]. According to Niels Bohr and Max Planck, the exchange of radiant energy between surface and electromagnetic wave results from the transfer of energy in discrete packets called quanta or photons. The wavelength of EMR is described by the electromagnetic spectrum (Figure 3-3), of which only the spectral range between  $0.3\text{--}14\ \mu\text{m}$  is of relevance for passive RS. This range is typically classified further into the visible light ( $0.4\text{--}0.7\ \mu\text{m}$ ), the near, middle, and shortwave infrared ( $0.7\text{--}3\ \mu\text{m}$ ), and the thermal infrared spectrum ( $3\text{--}14\ \mu\text{m}$ ) (Figure 3-3).

The radiant flux interacts in a characteristic way with the Earth’s surface. EMR is reflected and absorbed specifically in dependence of the physical properties of any surface such as type, structure, texture, moisture content, and chemical composition. Reflectance, the amount of reflected EMR at specific wavelengths, calculated as the ratio of reflected and incident radiation, refers to the spectral properties, which are specific

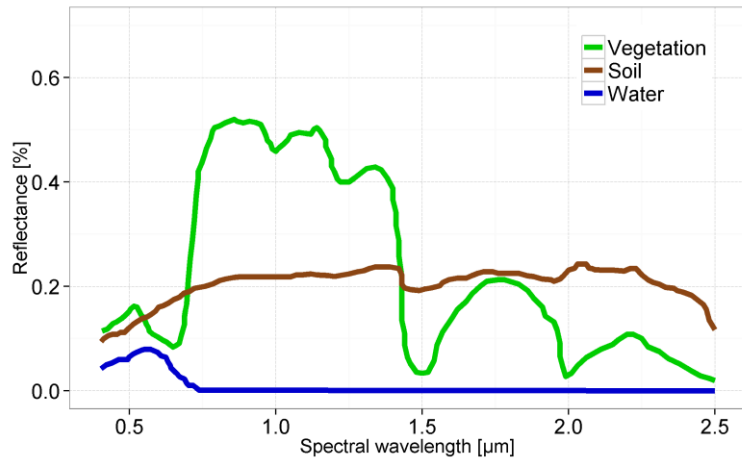


Figure 3-4: Spectral signature of vegetation, soil and water in dependence on the spectral wavelength.

for surface conditions and provide the fundamental focus of RS research. Non-reflected incident radiation is either absorbed or transmitted. Figure 3-4 exemplifies the typical spectral reflectance properties of a surface with green vegetation, bare soil, and water. The spectral properties of vegetation are characterised by a specifically high reflectance in the near infrared compared to very low reflectance in the blue and red parts of the visible spectrum due to the

high chlorophyll absorption at these wavelengths. In contrast, soil reflectance shows a continuous increase from low reflectance at short wavelengths to higher reflectance at longer wavelengths, while the overall reflectance is markedly reduced for moist soils compared to dry soils. Due to the high absorption of EMR by water, this type of surface is characterised by very low reflectance values.

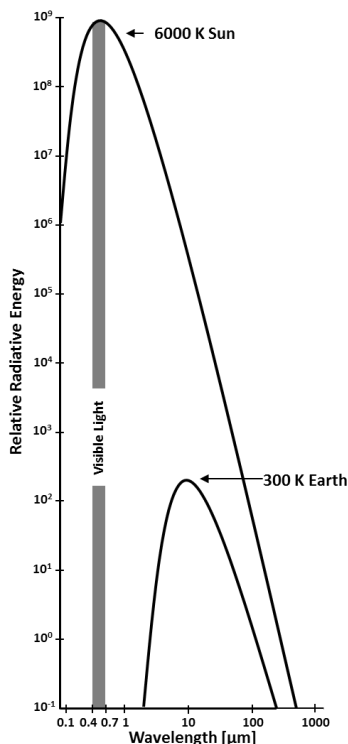


Figure 3-5: Energy emitted from black body radiators for Sun and Earth as a function of wavelength. Source: modified after Jensen (2000: 33)

Apart from reflected EMR, the Earth also emits EMR as a function of surface temperature. The magnitude of EMR emitted from blackbodies in dependence of surface temperature and spectral wavelength is described by the Planck’s law, the Stefan-Boltzmann law, and the Wien’s displacement law. A comprehensive overview of these thermal RS fundamentals is provided by Kuenzer and Dech (2013). Very few terrestrial surfaces act as perfect blackbody radiator and therefore emit radiation in a modified way according to the specific emissivity of a surface, which can be expressed as ratio between the surface radiant flux and the blackbody radiant flux at a given temperature. The dominant spectral wavelength where maximal EMR is emitted by the Earth surface has its peak at around 10 µm (Figure 3-5).

The measurement of reflectance and emittance by a satellite remote sensor makes use of the so-called atmospheric windows (Figure 3-3) that correspond to wavelength regions in

which the atmosphere transmits most of the incident, reflected or emitted electromagnetic energy. Surface reflectance can be captured by the satellite sensor in the spectral wavelengths of visible light and portions of the reflective infrared. The thermal infrared region of the electromagnetic spectrum allows for the thermal measurement of thermal emission from the Earth between 3 and 14  $\mu\text{m}$ .

However, the measurement of surface reflectance and emissivity by a remote sensor is altered by three main factors, namely: (i) the atmosphere; (ii) the variation in sun illumination; and (iii) variation in viewing angles of the sensor. The atmosphere influences EMR in both directions, as it first reduces the amount of solar irradiance illuminating a ground object, secondly it acts as reflector itself adding a scattered, extraneous path irradiance to the signal, and thirdly it attenuates the reflected signal in a wavelength dependent way (Lillesand and Kiefer, 2000: 21-22). The viewing geometry results from the position of the sensor and the variation of its scan angle noted as view zenith and view azimuth angle, whereas the variation in illumination angles is caused by the Earth's orbit. The constellation of Sun illumination, sensor view angle, and surface reflectance properties has an impact on the at-sensor detected brightness. However, all three disturbing factors mentioned here can be corrected for by performing an atmospheric correction for the respective satellite image as well as accounting for directional reflectance effects (bi-directional reflectance distribution function; BRDF in short) (Lillesand and Kiefer, 2000: 31).

Following this image correction process, the received signal has to be analysed with respect to the objective of interest. The characteristic spectral properties of surfaces (Figure 3-4) can be depicted by the signal measured at the remote sensor with its designed band combinations (see Table 4-2). The reflectance at specific spectral wavelengths can be directly investigated in relation to the phenomenon of interest. However, in most cases, indices are calculated from spectral reflectance at different wavelengths to enhance information and normalise images (e.g. atmospheric effects). The most prominent index is the NDVI given in Equation 3-2,

$$NDVI = \frac{nir - red}{nir + red} \quad \text{Equation 3-2}$$

where *nir* corresponds with the measurement at near infrared spectrum of the sensor and *red* with the visible red light. This index responds to the change in the amount of green biomass and chlorophyll content, where the red band records the high absorption of chlorophyll and the near infrared band records the high reflectance when vegetation is dense and growing vigorously. For passive RS, there are multiple vegetation-, water-, and soil-related indices available. An overview of the indices with relevance to the schistosomiasis disease as investigated by this thesis is provided in Section 4.2.2.

The measurement of thermal RS is especially used to measure the surface temperature, which is the temperature of the radiating surface and therefore often referred to as skin temperature (Czajkowska et al., 2002: 254). However, the term land surface temperature (LST) is thereby used ambiguously, as it includes a wide range of surface temperature variables such as kinetic, thermodynamic, radiometric, canopy and even air temperature (Norman and Becker, 1995). The calculation of surface temperature in  $^{\circ}\text{C}$  from the energy retrieved at the sensor includes sensor calibration, atmospheric correction and the approximation of emissivity effects, which can be realised by assigning specific emissivity values to each land cover class. Despite different land surfaces varying only little in their emissivity (e.g. water: 0.98, plant leaves: 0.96,

sand: 0.93), a difference in spectral emissivity from 1 to 0.99 typically results in an increase of 1-2°C (Schmugge et al., 1998: 124).

### 3.2.2 Remote sensing for schistosomiasis risk profiling

Owing to the life cycle of the parasite and the mode of infection with the parasite, the transmission of schistosomiasis is spatially and temporally restricted to water bodies inhabited by snails and parasites. Schistosomiasis, as environmental disease, warrants a relevant contribution of RS technologies with respect to the spatial distribution of the disease (Malone, 2005: 27). The first application of RS to predict the probability of spatial occurrence of human schistosomiasis using Landsat 5 TM data was published in 1984 for the Philippines by Cross et al. (1984). Ten years later, diurnal temperature differences derived from data of the Advanced Very High Resolution Radiometer (AVHRR) from the National Oceanic Atmospheric Administration (NOAA) have been related to survey measurements of schistosomiasis prevalence in Egypt (Malone et al., 1994). As thermal differences between day and night reflect regional hydrologic conditions (Jensen, 2000: 393), the significant inverse relationship showed well the predictive ability of RS data for schistosomiasis transmission risk (Malone et al., 1994: 716-718). Since then, the investigation of RS in relation to schistosomiasis has experienced considerable growth and interest. To demonstrate this, the online library PubMed (<http://www.ncbi.nlm.nih.gov/pubmed/>) has been accessed in January 2014 with the following terms and Boolean operators: “remote sensing” OR “geographic information system” OR “mapping” OR “prediction” AND “schistosomiasis” AND “Africa” (cf Simoonga et al., 2009: 1684) for the time span between 1995 and 2013, which resulted in 93 publications, of which 31 were relevant reviews or case studies using RS data for spatial modelling of schistosomiasis risk. However, within the process of reviewing the literature as well as its cited references, the number of studies increased to 37.

The development and potential of RS and its combination with GIS-based spatial analyses has been reviewed by various groups (Bergquist et al., 2000; Brooker and Michael, 2000; Abdel-Rahman et al., 2001; Brooker, 2002; Malone, 2005; Brooker et al., 2006; Brooker, 2007; Simoonga et al., 2009). It can be summarised that the contribution of RS and GIS for schistosomiasis risk profiling is composed of its ability to: (i) determine the geographical limit of disease distribution due to ecological constraints of disease transmission; (ii) further investigate the context of disease ecology and epidemiology through its spatial relation; (iii) support prevention and control through prioritising areas of disease risk; and (iv) provide early warning for areas where disease transmission could become established. However, most studies that have been reviewed had an integrative focus combining the aforementioned objectives by some means or other.

The first step to analyse spatial disease risk consists of the geographical mapping of empirical survey data on disease prevalence and/or intensity, among which the “Atlas of the global distribution of schistosomiasis” by Doumenge and Mott (1987) is deemed the pioneer work. It has been shown that simple geographical mapping of the disease already provides useful information to highlight endemic areas for which further information is required, to quantify population at risk and to estimate the cost of disease intervention programmes (Brooker et al., 2000a: 1459-1462; Standley et al., 2009: 42-45; Hodges et al., 2011: 3-4; Kabatereine et al., 2011: 4-8). The comparison between surveys of different time steps on the very spot has been



investigated to monitor disease development in space (Tchuem Tchuente et al., 2012: 4-6) and to evaluate the impact of disease intervention programmes (Clements et al., 2009b: 3). However, it is widely discussed that the comparison between different maps of survey data is critical due to a high variability of survey methodologies used in practice (Brooker et al., 2000b: 305-306; Brooker et al., 2009: 4-5). To give some examples, examination methods to detect the parasite in faeces have variable sensitivities, the age of the examined population varies between surveys, or the location of the survey varies between rural or urban areas as well as between schools, households or community health centres (Brooker et al., 2009: 5). Even today, despite the global positioning system (GPS) facilitated ease of geo-locating survey sites, measurements of disease prevalence and intensity are painstaking and costly, which explains well the paucity of reliable epidemiological data especially in the most affected areas of sub-Saharan Africa (Simoonga et al., 2009: 1687). To overcome the problem of data paucity, the well-known influence of the environment on disease distribution has been used for modelling and spatial prediction of disease risk for non-sampled locations. The conceptual background and methodological modelling approaches are detailed in Section 3.3.

Following the seminal work of Malone et al. (1994), a large number of studies have investigated RS data and derived environmental variables and their relation to human infection or snail occurrence with the objective of modelling and predicting schistosomiasis risk for various regions and scales (Table 3-2). This review shows that the most frequently used RS data are from the NOAA-AVHRR and later the MODIS sensors with ground resolutions of 1.1km and 250m, respectively. High resolution data from Landsat 5 TM (30m) have only marginally been analysed and very high-resolution RS data (1m) have solely been investigated for one study site in Kenya (Clennon et al., 2004; Clennon et al., 2006; Clennon et al., 2007). Topographic information from either Shuttle Radar Topography Mission (SRTM) data or the global 30 arc-second elevation (GTOPO30) model has been added in most studies as predictor variable. The environmental variables most commonly used were the NDVI and LST, hypothesised to represent surrogate measures of environmental moisture and temperature, respectively (Malone et al., 2001: 62). The availability of NOAA-AVHRR and MODIS data at no charge and the web-accessed supply of pre-processed imagery boosted studies that investigated these data (Herbreteau et al., 2007: 401). However, there are many other vegetation or moisture-related indices that are not directly accessible due to a more complex nature, which are rarely used for health studies (Herbreteau et al., 2007: 401). Many studies (e.g. Clements et al., 2009a; Koroma et al., 2010; Hodges et al., 2012) have used spatial information of perennial water bodies and river networks from the GeoNetwork platform provided by the Food and Agricultural Organization of the United Nations (FAO, 2014b). However, the acquisition dates back to the 1990s and it lacks information on temporal dynamics. Actualisation would be highly relevant to monitor environmental changes such as construction of dam lakes or irrigation schemes (Steinmann et al., 2006). Satellite RS provides data and methodological procedures to map and monitor water bodies and other disease relevant variables such as water temperature, turbidity or vegetation coverage (Tran et al., 2010).

Reference data for the spatial analysis with RS data were always point data, either of human infection prevalence most frequently geo-located at schools, or snail occurrence located in sampled water bodies (Table 3-2). Overall snail data were very rarely available and most analyses were based on infection of schoolchildren surveyed. In most cases, infection of

schoolchildren has been substantiated based on parasitological examinations, and few case studies were built upon prevalence data sampled based on morbidity questionnaires (e.g. Clements et al., 2008b). Epidemiological data of human infection or snail sampling often had a temporal mismatch of several years between the sampling and the acquisition of RS data (e.g. Brooker and Clements, 2009; Schur et al., 2011b). However, this has been considered a negligible drawback since schistosomiasis is a chronic disease with a life-span of adult worms being typically several years (Jordan and Webbe, 1993: 110). Therefore, spatial variability in long-term synoptic environmental factors is hypothesised to have more influence on transmission success and infection patterns than seasonal variability in a location (Brooker and Clements, 2009: 592). When environmental data such as NDVI or LST are used to predict the risk of schistosomiasis, “in effect, one is predicting the environmental requirements for a particular snail species (infected with a particular parasite species) - and not the human parasitic infection *per se*” (Simoonga et al., 2009: 1687). An analysis of RS data with respect to snail abundance and disease prevalence showed that snail distribution generally corresponded to the prediction model of schistosomiasis prevalence, however, the best model of snail distribution showed different ranges of temperature than found in the schistosomiasis prevalence model (Kristensen et al., 2001; Malone et al., 2001). One challenge to date to further improve RS and GIS-based risk mapping, is to account for the spatial mismatch between the measurement of human infection and the location where disease transmission may occur (Simoonga et al., 2009: 1687). This aspect and a potential solution to overcome this spatial conflict when using RS data is directly investigated in this thesis.

Environmental analyses using RS data provide the opportunity to understand more completely the process underlying broad-scale patterns of schistosomiasis distribution and can help to potentially improve our knowledge of schistosomiasis infection ecology (Brooker, 2002: 210). To give an example, the local study of Raso et al. (2005) found that – besides age, sex, and socioeconomic status – rainfall pattern and elevation significantly explained the geographical variation of *S. mansoni* distribution in the Man region in Western Côte d’Ivoire. For the same region, Beck-Wörner et al. (2007) found a significant correlation for stream order of the closest river, the water catchment and altitude. For sub-continental East Africa, a negative correlation resulted from the distance to water body and elevation with respect to the distribution of *S. mansoni* infection intensity (Clements et al., 2006b). In contrast, a study in sub-continental West Africa reported only a negative correlation with distance to perennial inland water bodies with respect to *S. haematobium* infection intensity and significance for elevation failed (Clements et al., 2009a). Both sub-continental studies investigated a nearly 20-year mean of LST and NDVI from NOAA-AVHRR, however, they provided no significant contribution to the final model (Clements et al., 2006b: 716; Clements et al., 2009a: 924). In contrast, in Tanzania, a positive effect of both LST and NDVI could be derived for prediction of *S. haematobium* (Brooker et al., 2001). In another case study in Tanzania, the NDVI was already rejected in a preparatory variable selection process (Clements et al., 2006a). In Nigeria, LST resulted as the only significant environmental variable to predict urogenital schistosomiasis (Ekpo et al., 2008). This shows that predictor variables and resulting models are tailored to the reference data, the scale of observation and the geography of the study site. A reasonable impact of different ecological zones on predictor performance and model outcome has been established by Brooker et al. (2001: 1001) in Tanzania. This phenomenon is taken up by this thesis and specifically investigated for the ecozones of savannah and tropical rainforest in study sites of West Africa.

This review shows that different scales of RS data have been investigated for the risk profiling of schistosomiasis with spatial resolutions ranging from 1 m to 8 km (Table 3-2). Most of the studies have investigated data from a single sensor with predominantly low spatial resolution, such as NOAA-AVHRR at 1.1 km. A common procedure was to combine the data with remotely sensed topographic information. There are several approaches, where multi-scale RS data such as 1m Ikonos data, 30m Landsat, and 1km MODIS data have been analysed in an integrated way (Abdel-Rahman et al., 2001; Kabatereine et al., 2004; Raso et al., 2005; Raso et al., 2006; Clennon et al., 2007; Vounatsou et al., 2009). However, these studies used multi-scale data predominantly to cover a broad spectrum of potentially relevant information such as land cover classification based on high-resolution data and climatic surrogates based on multi-temporal low resolution data. What is missing in all studies, but directly addressed by this thesis, is a direct comparison of RS metrics derived at different spatial resolutions to evaluate the stability of predictor performance for varying spatial scales. This issue of scale is essential to be considered in order to better understand disease ecology based on the linkage between biological and RS data (Brooker, 2002: 210).

RS and GIS have proven to be useful for planning and implementing disease intervention and control programmes by excluding areas where schistosomiasis is unlikely to be a public health problem and modelling priority areas of increased transmission risk (Brooker et al., 2001: 1004; Brooker, 2002: 211). Clements et al. (2008a) predicted regions with a probability of schistosomiasis transmission greater than 50% to design mass treatment campaigns according to the criterion of the WHO in Burkina Faso, Mali, and Niger. Estimates of the number of schoolchildren at risk of high prevalence have been predicted for West Africa (Schur et al., 2011b) and Tanzania (Brooker et al., 2001), where additionally expectable programme costs were calculated based on the model predictions. A greater demand of treatment resources resulted if data were aggregated on provincial level compared to the national level due to the integration of large spatial heterogeneities of disease risk on the sub-national level, again indicating the high relevance of scale for risk profiling (Schur et al., 2012). Many studies focused on modelling the risk of polyparasitic co-infections of schistosomiasis and soil-transmitted helminth infections with the objective to enhance cost-effectiveness through integrated control measures (e.g. Raso et al., 2006; Schur et al., 2011a; Hodges et al., 2012). Morbidity control through mass treatment campaigns has not proven to be sustainable (Section 3.1.3), because a water site can be converted to a high-risk transmission zone if it is (re-) contaminated by single untreated individuals. This suggests the importance of water-site factors to achieve the shift from morbidity to transmission control or even local elimination (Stothard et al., 2009: 1672; King, 2010: 100).

Table 3-2: Overview of RS data and environmental variable investigated for spatial analyses of schistosomiasis from 1995 onwards. References are listed manifold if the study investigated data from more than one remote sensor.

Satellite sensor	Environmental variable	Sensor combination	Geographic area	Temporal coverage	Reference data		Reference(s)
					Snail	Human infection	
NOAA-AVHRR spatial resolution: 1.1km	Diurnal temperature difference NDVI	Landsat	Egypt	Monthly time series 1990-1991	Snail occurrence, infection rate (water survey 1 km distance to population survey)	Survey prevalence (rural health units)	Abdel-Rahman et al. (2001)
	LST, NDVI	SRTM	Tanzania	Monthly time series 1985-1998	-	Survey prevalence (school)	Brooker et al. (2001)
	LST, NDVI	-	Ethiopia – East Africa	Annual + seasonal composites 1992-1996	-	Survey prevalence (town/village) – 5km buffer	Malone et al. (2001)
	LST, NDVI	-	Ethiopia	Annual + seasonal composites 1992-1996	Snail occurrence	Survey prevalence (town/village) – 5km buffer	Kristensen et al. (2001)
	LST, NDVI	-	Cameroon	1985-1998	-	Survey prevalence (district – stratified at school level)	Brooker et al. (2002)
	LST, NDVI	Landsat GTOPO30	Uganda	Not indicated	-	Survey prevalence, infection intensity (school/village)	Kabateraine et al. (2004)
	LST, NDVI	-	Kenya, Ethiopia, Uganda	Annual + seasonal composites 1992-1995	Snail occurrence	CEGET/WHO atlas – 5km buffer	Malone et al. (2004)
	Land cover	MODIS Landsat METEOSAT GTOPO30	Côte d'Ivoire	1992/1993	-	Survey prevalence (school)	Raso et al. (2005); Raso et al. (2006)
	LST, NDVI	GTOPO30	Tanzania	1982-1998	-	Survey prevalence (school)	Clements et al. (2006a)
	LST, NDVI	GTOPO30	East Africa	1982-2000	-	Survey prevalence (school)	Clements et al. (2006b)
	Land cover	GTOPO30	Côte d'Ivoire	1992/1993	-	Survey prevalence (school)	Raso et al. (2007)
	LST, NDVI	-	Burkina Faso, Mali, Niger	Not indicated	-	Survey prevalence (school)	Clements et al. (2008a)

Satellite sensor	Environmental variable	Sensor combination	Geographic area	Temporal coverage	Reference data		Reference(s)
					Snail	Human infection	
NOAA-AVHRR spatial resolution: 1.1km	LST, NDVI	-	Nigeria	2001-2002	-	Survey prevalence (school)	Ekpo et al. (2008)
	LST, NDVI	-	Zambia	1992-1995	Snail occurrence, cercarial shedding	Survey prevalence, infection intensity (school/village)	Simoonga et al. (2008)
	LST, NDVI	-	East Africa	1982-2000	-	Survey prevalence (school)	Brooker and Clements (2009)
	LST, NDVI	-	Tanzania	Not indicated	-	Survey prevalence (school)	Clements et al. (2008b)
	LST, NDVI	-	Burkina Faso, Mali,	1982-1998	-	Survey prevalence (school)	Clements et al. (2009a)
	NDVI	-	Sudan	Not indicated	-	Survey prevalence (village)	Sturrock et al. (2009)
	LST, NDVI	SRTM	Sierra Leone	Not indicated	-	Survey prevalence (school)	Koroma et al. (2010); Hodges et al. (2012)
	LST, NDVI	-	Ghana	Not indicated	-	Survey prevalence, Infection intensity (school/village)	Soares Magalhães et al. (2011)
	LST, NDVI	NOAA-AVHRR Landsat METEOSAT GTOPO30	Côte d'Ivoire	Monthly time series January + November 2002	-	Survey prevalence (school)	Raso et al. (2005); Raso et al. (2006)
	LST, NDVI	-	Uganda	Annual + seasonal composites 2000-2003	Snail occurrence	Survey prevalence (school)	Stensgaard et al. (2005)
LST, NDVI	-	Uganda	Annual + seasonal composites 2000-2003	Snail occurrence	Survey prevalence (school)	Stensgaard et al. (2006)	
LST, NDVI, land cover	GTOPO30	West Africa	2000-2008	-	Survey prevalence (school)	Schur et al. (2011b)	
LST, NDVI Land cover	GTOPO30	East Africa	2000-2009	-	Survey prevalence (school / community)	Schur et al. (2011c); Schur et al. (2013)	

Satellite sensor	Environmental variable	Sensor combination	Geographic area	Temporal coverage	Reference data		Reference(s)
					Snail	Human infection	
MODIS spatial resolution: 1km, 500m, 250m	LST, NDVI	METEOSAT SRTM	Côte d'Ivoire	Not indicated	-	Survey prevalence (school)	Younatsou et al. (2009)
Landsat TM / ETM+ spatial resolution: 30m, 60m	Spectral bands: (blue(1), red(3), mir(5), thermal(6)) NDVI Tasseled cap: brightness, greenness, wetness	NOAA-AVHRR	Egypt	May 1990	Number, distribution, infection rate	Survey prevalence (rural health units)	Abdel-Rahman et al. (2001)
	Water body map	NOAA-AVHRR GTOPO30	Uganda	March 2000	-	Survey prevalence, infection intensity (school/village)	Kabaterine et al. (2004)
	Settlements, roads, rivers	NOAA-AVHRR MODIS METSAT GTOPO30	Côte d'Ivoire	January + November 2002	-	Survey prevalence (school)	Raso et al. (2005); Raso et al. (2006)
	NDMSI (normalized difference moisture stress index)	Ikonos SRTM	Kenya	June 1986 + January 2003	Snail / shell occurrence	-	Clennon et al. (2007)
Ikonos spatial resolution: 1m	Land cover	Landsat SRTM	Kenya	March 2001	Snail / shell occurrence	-	Clennon et al. (2004); Clennon et al. (2006); Clennon et al. (2007)
METEOSAT spatial resolution: 8km	Rainfall estimates	NOAA-AVHRR MODIS Landsat GTOPO30	Côte d'Ivoire	September 2001 – August 2002	-	Survey prevalence (school)	Raso et al. (2005); Raso et al. (2006)

Satellite sensor	Environmental variable	Sensor combination	Geographic area	Temporal coverage	Reference data		Reference(s)
					Snail	Human infection	
METEOSAT spatial resolution: 8km	Rainfall estimates	MODIS SRTM	Côte d'Ivoire	Not indicated	-	Survey prevalence (school)	Vounatsou et al. (2009)
SRTM spatial resolution: 90m	Altitude	NOAA-AVHRR	Tanzania	February 2000	-	Survey prevalence (school)	Brooker et al. (2001), Beck-Wörner et al. (2007), Vounatsou et al. (2009)
	Altitude, slope, stream order, catchment	-	Côte d'Ivoire	February 2000	-	Survey prevalence (school)	
	Altitude, slope, drainage network	Landsat Ikonos	Kenya	February 2000	Snail / shell occurrence	-	Clennon et al. (2007)
	Altitude	NOAA-AVHRR	Sierra Leone	February 2000	-	Survey prevalence (school)	Koroma et al. (2010)
GTOPO30 spatial resolution: 30sec (ca. 1km)	Altitude	NOAA-AVHRR Landsat	Uganda	1994-1997	-	Survey prevalence, infection intensity (school/village)	(Kabatereine et al., 2004)
	Altitude	NOAA-AVHRR MODIS METEOSAT Landsat	Côte d'Ivoire	1994-1997	-	Survey prevalence (school)	Raso et al. (2005); Raso et al. (2006)
	Altitude	NOAA-AVHRR	Tanzania	1994-1997	-	Survey prevalence (school)	(Clements et al., 2006a)
	Altitude	NOAA-AVHRR	East Africa	1994-1997	-	Survey prevalence (school)	(Clements et al., 2006b)
	Altitude	NOAA-AVHRR	Côte d'Ivoire	1994-1997	-	Survey prevalence (school)	Raso et al. (2007)
	Altitude	MODIS	West Africa	1994-1997	-	Survey prevalence (school)	Schur et al. (2011b)
	Altitude	MODIS	East Africa	1994-1997	-	Survey prevalence (school/community)	Schur et al. (2011c)

### 3.3 Modelling schistosomiasis risk

The question of how diseases are distributed on Earth in space and time has a long history, which has inspired epidemiologists, ecologists and geographers to seek explanations (Section 1.1). To overcome the paucity of epidemiological and intermediate host-related data, spatial model approaches have become the instrument of choice with the potential to discover relationships between disease occurrence and environmental and geographic conditions to predict this established relation for non-sampled locations (Simoonga et al., 2009: 1687). The conceptual principle for modelling disease risk is based on modelling the ecological niche of disease-related species (e.g. parasites or disease vectors) to understand the ecology of disease, characterise its habitat distribution and predict suitable environmental conditions in space (Peterson, 2006: 1822-1823) (Box 1). However, to model the risk of a human disease such as schistosomiasis, the human as the final host needs to be addressed in the model approach. The ecological niche concept to model species distribution must therefore be extended by human-related environmental factors and adjusted to a social-ecological niche of disease (Section 3.3.1). Different approaches of modelling disease risk are presented in Section 3.3.2. The contextual relation between disease ecology and RS variables as well as its overall contribution for the multi-faceted and complex system of schistosomiasis transmission risk are consolidated in the next Section 3.4.

#### Box 1. Terms and definitions

**Risk** is defined as effect of uncertainty on objectives (International Organization of Standardization (ISO) 31000: Risk management) and implies a future event with an uncertainty if and how the entity of interest is affected by a certain phenomenon. For this research, risk is defined as probability of humans to become infected with the parasite.

**Niche** is a subset of those environmental conditions (determined in an n-dimensional model space, Figure 3-6a) which affect a particular organism, where the average absolute fitness of individuals in a population is greater than or equal to one (Kearney, 2006: 187)

**Environment** in a medical sense integrates all factors external to humans but interact with them, e.g. physical, biological, social and cultural environment (IEA, 1995: 53). With respect to the parasite and snail species, environment includes biotic and abiotic phenomena surrounding and potentially interacting with the organisms (Kearney, 2006: 187).

**Habitat** is a description of a physical place (i.e. geographical space, Figure 3-6b), at a particular scale of space and time, where an organism either actually or potentially lives (Kearney, 2006: 187)

#### 3.3.1 Social-ecological niche of a disease

The transmission cycle of a disease is a composite phenomenon that represents multiple interactions between a set of species, such as a pathogen, a human or animal host that become infected, and a vector to enable disease transmission (Peterson, 2006: 1822). It follows that the spatial occurrence of a disease is determined by the combination of complexities of the



occurrence of disease component species as well as effects of chance events (Peterson, 2006: 1822). Within the concept of landscape epidemiology, Pavlovsky explained the nidality (=focality) of diseases through the “source of disease” (=nidus) in the spatial domain (Pavlovsky, 1966: 7) being associated with specific landscapes (Pavlovsky, 1966), whereas the complexity of a nidus depends on the transmission requirements of a pathogen (Reisen, 2010: 469). The complex pattern and influences of environmental variation on disease distribution have been investigated by estimating specific ecological niches of disease-related species (Peterson, 2006: 1822). According to the concept of landscape epidemiology, humans become infected when they travelled into the nidus and come in contact with the pathogen (Pavlovsky, 1966: 9). However, the disjunct distribution of infection and human residence complicates the understanding of disease epidemiology and transmission (Reisen, 2010: 462).

The theoretical concept of the ecological niche of species has been established by Grinnell (1917). It hypothesises that the geographic distribution of a species is determined by an n-dimensional set of ecological conditions under which the species can maintain its population (Grinnell, 1917: 115-118; Hutchinson, 1957: 416; Peterson, 2006: 1822). The distribution of a species in its environment is determined by abiotic conditions that correspond with its physiological limits, biotic factors that may be positive (e.g. mutualists, symbionts) or negative (competitors, predators), the dispersal abilities of a species dependent on landscape configuration and its evolutionary capacity to adapt to new conditions (Soberon and Peterson, 2005: 2). A merely abstract formalisation of the ecological niche is the fundamental niche, which from an ecological point of view is the potential niche of a species driven by its environmental requirements and delineated by its physiological (in-)tolerance in absence of biotic interactions (Hutchinson, 1957: 416). The fundamental niche reflects Liebig’s *law of the minimum* (von Liebig, 1840) revealing that species distribution and abundance are determined by periods when conditions are at a minimum rather than times of the year when conditions are suitable (Odum and Barrett, 2005: 178). In contrast, the realised niche inhabited by a species represents both, the environmental dimensions in which species can survive and reproduce as well as its functional role within its biotic environment (Hutchinson, 1957: 418). The realised niche is thus often depicted as a subset of the fundamental niche (Figure 3-6), however, this is not necessarily true for the case of positive biotic interactions (Franklin, 2010: 37). The differentiation between fundamental and realised niche is important when modelling species distribution based on environmental data as it determines whether the distribution is predicted from theoretical physical constraints or field observations (Guisan and Zimmermann, 2000: 153). Both concepts are applied in this thesis by: (i) modelling the fundamental niche of schistosomiasis-related parasites and snails based on a mechanistic approach; and (ii) by modelling the realised disease niche (Figure 3-6) based on a statistical approach. These mentioned model approaches are further explained in Section 3.3.2.

Spatial modeling of human diseases have been fundamentally based on the ecological niche concept to model the distribution and habitat conditions of disease-related species (Peterson et al., 2002; Peterson, 2006; Ayala et al., 2009; Mak et al., 2010). However, from a conceptual point of view, the ecological niche concept can straightforwardly be expanded by human dimensions (Ojiem et al., 2006: 81), factoring in all relevant demographic, behavioural and socioeconomic drivers that play a role in the disease transmission cycle (Figure 3-6). The integration of human, socioeconomic, institutional, and cultural conditions into the ecological niche of plant species

has been conceptualised as social-ecological niche and applied in agricultural sciences (Ojiem et al., 2006; Guto et al., 2012; Mtei et al., 2013).

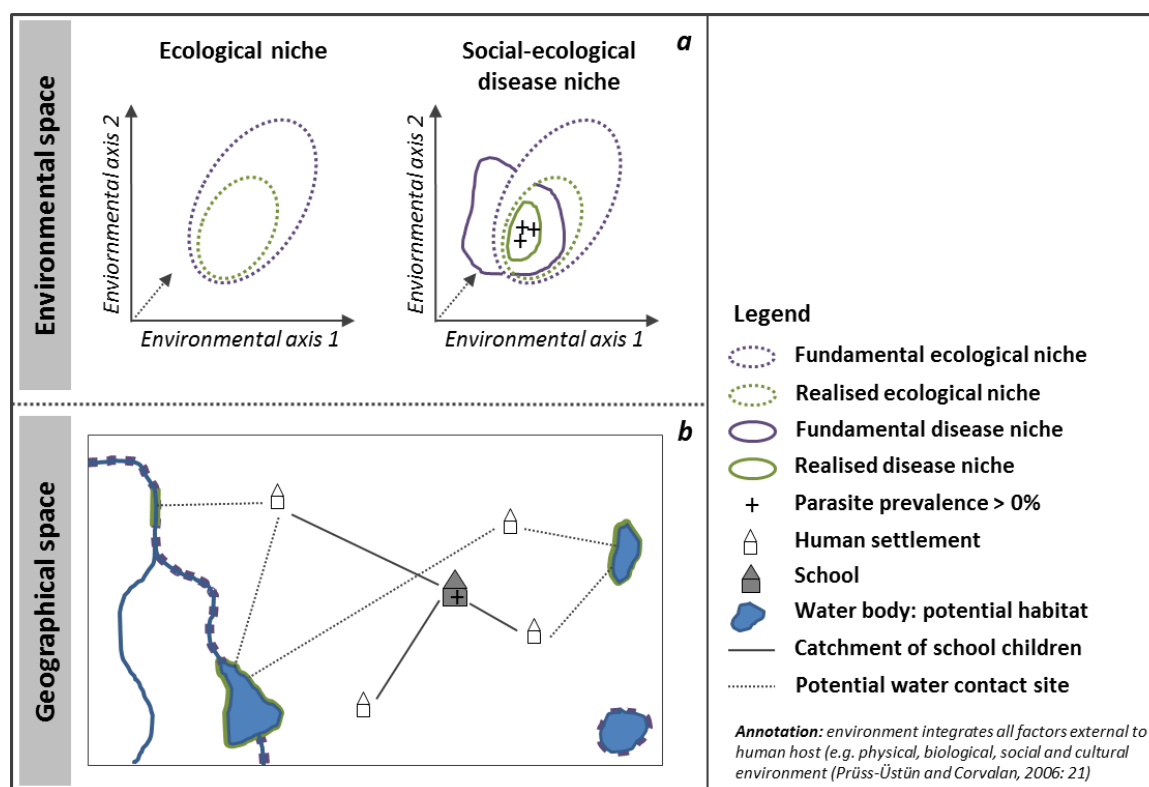


Figure 3-6: Theoretical concept of ecological and social-ecological disease niche in environmental space (a) and the spatial relation of schistosomiasis transmission in geographical space (b).

(a) The set of environmental conditions of the snail and parasite species modelled in an  $n$ -dimensional space determine the ecological niche (modified after Franklin, 2010: 36). The integration of the human environmental factors into the model results in the social-ecological niche. The specific niche conditions (fundamental, realised) can be predicted into geographical space.

(b) The school location corresponds to the geo-located measure of disease prevalence in humans (illustrated as +) and the water body as potential parasite and snail habitat represents the location, where human infection potentially occurs, if environmental conditions where suitable and humans susceptible. It is shown, that there are complex interactions between the measure of disease prevalence, infection at specific water contact site and human behaviour to seek water site. The water habitats, where disease transmission has occurred refer to the realised disease niche (bordered with green solid lines). The water sites bordered with purple dotted lines provide suitable habitats for parasites and snails and refer to the fundamental ecological niche.

With regard to schistosomiasis, Malone (2005: 28) declared that the aim of modelling disease risk is to measure the relative suitability of the environment to establish a “disease niche”, which describes conditions of a zone where parasite, intermediate, and final host coincide. Considering the ecology of schistosomiasis transmission (Section 3.1.2), such a biocenosis can become established, when environmental conditions are suitable for the respective parasites, its intermediate snail host and when susceptible humans enter this habitat. Figure 3-6a illustrates the theoretical concept of the social-ecological disease niche. To establish this niche (Box 1) of schistosomiasis (Malone, 2005: 28), environmental conditions derived from observations, where parasite, snail and humans converge and infection has occurred are modelled within an  $n$ -dimensional space. Regarding the social-ecological niche of schistosomiasis, the fundamental disease niche corresponds with habitats, where parasites and their corresponding snails could

potentially establish a population and susceptible humans potentially access this habitat. The realised disease niche reveals environmental conditions, where parasite, snail, and human infection essentially converge, which partly result of observations from inhabited disease niches. Figure 3-6b illustrates this concept as diagram of potential parasite and snail habitats (Box 1) and the dispersal of humans in geographical space, i.e. represented by two-dimensional (x,y) map coordinates. It shows that the integration of human water contact inherits complex spatial relations between human habitations and the localisation of infection, which again is measured at schools. This spatial complexity is further aggravated by varying disease susceptibility and behaviour of individuals (Table 3-1) that highly modify the spatial probability of water contact and disease infection. However, this conceptual underpinning makes clear, that spatial modelling of schistosomiasis risk comprises a complex interplay of social and environmental determinants of risk, which becomes further intricated through the disjunct appearance of disease agents in space.

### 3.3.2 Model approaches

Spatial modelling of disease distribution has been established as an important tool for disease epidemiology and control (Brooker, 2007; Pullan et al., 2011; Hodges et al., 2012). The general idea of modelling habitat relations either aims to understand the relationship between observed species/disease agents and its abiotic and biotic environment, or to test ecological or biogeographical hypotheses about its distribution and ranges (Franklin, 2010: 11-12). These models are widely used to predict the established relationship for locations, where survey data are lacking (Franklin, 2010: 12). Similar to species distribution modelling (Kearney, 2006: 188), there are two approaches for modelling a disease niche and predicting disease risk in space (Malone, 2005: 27-28), the so-called mechanistic (biological or process-based) models that aim to be general and realistic (Franklin, 2010: 105) and statistical (empirical) models designed to correlate empirical facts (Guisan and Zimmermann, 2000: 158). In this thesis, both model approaches have been used, as illustrated in Figure 3-7.

The **mechanistic model** approach aims at modelling a species' niche based on a set of physiological and behavioural traits to make inference on its potential environmental range (Kearney and Porter, 2009: 336). To give an example, the thermal niche of a species represents a fitness component such as survival, growth, development rate or reproduction as a function of body temperature (Kearney and Porter, 2009). For diseases, key features of such physiological requirements can be obtained from laboratory or field-based studies (Malone, 2005: 28), as reviewed in Section 3.1.2 with an emphasis on schistosomiasis. Environmental preferences, limits of tolerance, fitness values or behaviour of disease agents such as the parasite or the intermediate host snail can be modelled based on either direct observations or based on models of an individuals' response to physical variables (Malone, 2005: 28). The objective of this model approach is to link information about species fitness to environmental conditions and predict habitat suitability of this species in geographical space (Kliskey et al., 1999; Kearney and Porter, 2009: 336). As the mechanistic approach is based on direct measurements of physiological variables it models the fundamental niche, does explicitly not consider biotic interactions and has not the potential to take this into account (Soberon and Peterson, 2005: 2). As depicted in Figure 3-7, a mechanistic model approach has been used to model environmental suitability of schistosomiasis-related parasites and snails in this study.

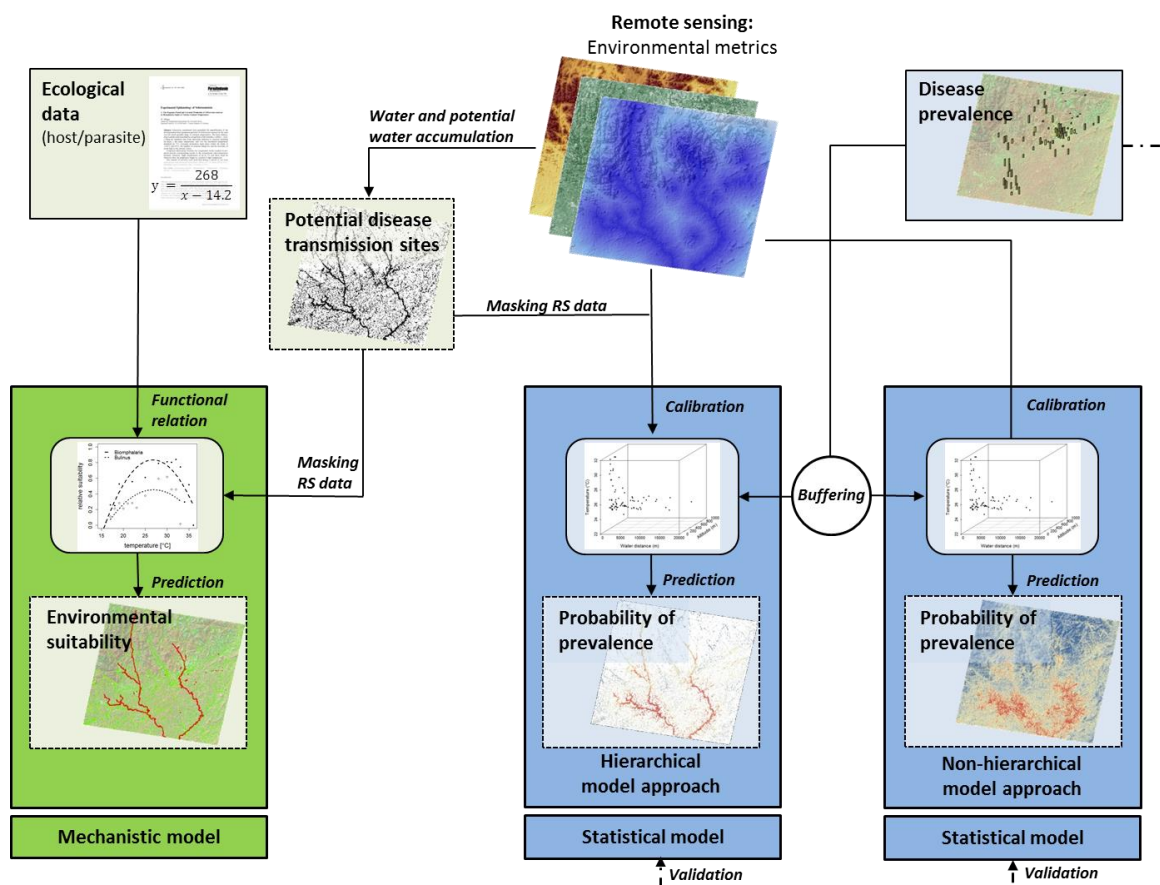


Figure 3-7: Conceptual framework of the model approaches used in this study. A mechanistic model approach has been used to model environmental suitability for disease transmission and a statistical model approach has been used to link school-based disease prevalence data with RS environmental metrics. This study further explored the hierarchical combination of both model approaches with a statistical model that links school-based prevalence directly to the potential disease transmission sites modelled by the mechanistic approach.

The **statistical** approach investigates the correlation between survey records (e.g. human prevalence, infection intensity, snail occurrence) and environmental variables (Malone, 2005: 27). This inductive approach has the objective to identify key environmental features that best describe the range of conditions of reference data and predict the distribution and abundance of a given species or disease by projecting the established relation back into geographical space (Malone, 2005: 27). The assumption of this approach is that habitat suitability corresponds to the intensity of habitat usage by a species, which selects areas that are most satisfying to its life requisites (Schamberger and O'Neil, 1986: 6). In contrast to the mechanistic model approach, a statistical approach is biased by biotic interaction effects since it is based on *in situ* observations, which include those interactions. One can partially account for this bias by combining the geography of other species using single-species models (Soberon and Peterson, 2005: 2). Furthermore, evolutionary effects modify niche characteristics and spatial distribution of species, which are inherent in observational data but cannot be considered by the mechanistic model approach (Peterson and Holt, 2003: 776; Soberon and Peterson, 2005: 8). It has been shown that the mechanistic and statistical model approach are complementary with respect to the information they provide and should be interpreted carefully before being used interchangeably in applications (Soberon and Peterson, 2005: 8). In this thesis, the spatial

delineation of the potential disease transmission sites derived for the quantitative mechanistic model of environmental suitability provided the basis to develop a hierarchical *versus* a non-hierarchical statistical approach to model schistosomiasis risk as illustrated in Figure 3-7.

With respect to schistosomiasis, the mechanistic model approach has been applied by modelling the climatic envelope to establish parasite development based on minimum and maximum temperature thresholds and the number of growing degree days (GDD) required to complete a parasite life cycle (Stensgaard et al., 2013: 381). De Vlas et al. (1996) simulated human-, worm-, and infection-related aspects of schistosomiasis transmission to evaluate and predict the effects of different control strategies based on a stochastic process model. The inferential steps to statistically model the disease niche based on human prevalence or the ecological niche of snail and its parasites have been achieved using diverse algorithms, such as binomial logistic regression (e.g. Brooker et al., 2001; Stensgaard et al., 2005; Koroma et al., 2010), Bayesian inference (e.g. Raso et al., 2005; Vounatsou et al., 2009; Schur et al., 2011b) or genetic algorithms (Stensgaard et al., 2006). Overall, the literature shows that the standard to investigate the spatial relation between the disease and its environment is a descriptive, statistical approach. This could be explained by the straightforward approach of a statistical model, where information is inherent in *in situ* observations compared to the mechanistic model, where all relevant criteria for species occurrence would need to be extracted from theoretical and experimental data and comprehensive validation is difficult.

## 3.4 Contribution of remote sensing for schistosomiasis risk profiling

The review of risk factors that influence the schistosomiasis transmission ecology (Section 3.1.2) and the previous applications of RS data for modelling disease risk (Section 3.2.2) are synthesised in this section. Table 3-3 provides an overview of the linkage between disease-related risk factors and RS variables structured according to the steps of the parasite life cycle. The distribution of the filled rows, which correspond to risk factors that can potentially be measured by RS data, shows that specifically parasite- and snail-related risk factors are covered by RS measurements. However, it shows also very clear that RS has its natural limitations and cannot detect all relevant risk factors such as intrinsic factors related to parasites, snails, and humans as well as some chemical and biological aspects. Thus, remotely sensed environmental measurements can contribute substantially to characterise the habitat conditions of parasites and snails and can moreover detect human settlements and their spatial relation to suitable habitat conditions. However, against this background it has to be kept in mind that RS-based approaches for disease risk profiling always have gaps of information to be filled by other disciplines.

Table 3-3: Overview of the contribution of RS for schistosomiasis risk assessment. The grey highlighted cells correspond with relevant factors of disease transmission, that can be to some extent measured by RS data

	Impact factor on disease transmission	RS (proxy) variable	Ecological impact	
<b>Parasite</b>	Water body	Near and middle infrared reflectance	Hatching of eggs; infection of snail and human host	
	Water temperature	Thermal infrared (emissivity)	Length of prepatent period; activity, survival and infection rate	
	Water flow velocity	Topography: slope angle, curvature	Determination of maximal cercarial density; passive transport of parasite	
	Predators	NA	Reduction of parasite population	
	Sunlight	Shaded habitats (tree coverage)	Stimulation of cercarial shedding	
	Pathogenicity	NA	Severity of disease in humans	
	Species	NA	Different efficiency of snail infection	
	<b>Snail</b>	Water body	Near and middle infrared reflectance	Fundamental habitat of snail to maintain a population
		Water temperature	Thermal infrared (emissivity)	Fecundity, mortality and rate of reproduction
		Water flow velocity	Topography: slope angle, curvature	Determination of snail density; passive transport of snail; food availability
Vegetation		Visible and near infrared reflectance	Food supply; surface for oviposition; increase of dissolved oxygen	
Substratum		NA	Abundance of snails in water body	
Water depth		Visible and near infrared reflectance	Abundance of snails in water body	
Stability of water level		Temporal dynamic of water body	Abundance of snails in water body	
Rainfall		Cloud thickness and temperature	Creation of temporary snail habitats; modification of water flow velocity; supports contamination of water	
Turbidity		Visible and near infrared reflectance	Reproduction cycle of snails	
Water chemistry/quality		NA	Abundance of snails in water body	
Sunlight		Shaded habitats (tree coverage)	Abundance and activity of snails	

	Impact factor on disease transmission	RS (proxy) variable	Ecological impact
<b>Snails</b>	Predators, parasites and pathogens	NA	Reduction of snail population
	Species	NA	Susceptibility to parasite and cercarial productivity
<b>Humans</b>	Water contact behaviour	NA	Exposure to parasite infested water; contamination of surface waters
	Gender	NA	Determinant of water contact activities (culturally variable)
	Age	NA	Related to degree of exposure and level of immunity
	Immunity	NA	Resistance to reinfection can be developed following previous infections
	Ethnic origin	NA	Susceptibility to infection
	Religion	NA	Religious practices may affect patterns of water use
	Socioeconomic status	NA	Standard of hygiene; access to protected water supply; ability to cope with disease
	Migration	NA	Modification of spatial distribution of disease
	Occupation	NA	Work related to water increases exposure
	Location of house	Settlement mapping	Exposure of population to potential disease transmission sites
	Prevention/control measures	NA	Modification of spatial pattern of disease transmission

In the context of modelling schistosomiasis risk, RS variables can provide either direct measurements of the feature of interest, which are, for example, the measure of land surface temperature, water or vegetation. These variables are derived from respective RS metrics such as land surface emissivity or surface reflectance at appropriate wavelengths with respect to their spectral signature (see Figure 3-4). They can directly reflect habitat conditions and provide information about the potential impact on the ecology of disease transmission. On the other hand, RS data provide proxy variables, where the remotely sensed measurement is not representing the respective measure influencing disease transmission but being indirectly linked to the requested information. To give an example based on RS data, the slope of a land surface can be measured from topographic modelling, however, the relevant information to be drawn from this proxy measure would be water flow velocity as a decisive criterium for profiling risk of disease transmission. Due to an additional step of modelling information to ecological indicators, the potential sources of errors that affect the data may increase.

Based on this theoretical background it was the aim to evaluate the potential contribution of RS data and variables for profiling schistosomiasis risk. Unfortunately, information on rainfall was not available at a useful spatial resolution to gain relevant information for the scale of observation used in this study. Furthermore, there were constraints regarding appropriate *in situ* data to model water depth and turbidity from RS measurements. However, the former variable has been approximated from field-based estimations (Section 4.3). Apart from these few exceptions, all RS variables described in Table 3-3 were investigated regarding their potential contribution for schistosomiasis risk profiling in this thesis.



## 4 Data and pre-processing

This chapter presents the acquisition, pre-processing, and quality of epidemiological and RS datasets. Furthermore, environmental *in situ* data with focus on potential disease transmission sites visited in Burkina Faso are described.

### 4.1 Epidemiological data

Epidemiological data on schistosomiasis prevalence given as ratio between infected people and surveyed people have been accessed for the countries Burkina Faso and Côte d'Ivoire from the Global Neglected Tropical Disease database (GNTD) freely available under [www.gntd.org](http://www.gntd.org) (Hürlimann et al., 2011). Additional information of relevance for this study were included in this database, such as prevalence of parasite species, number of examined people, date of survey, description of the survey location (e.g. school, community, hospital), and the method used for sample recruitment and diagnostic techniques (Hürlimann et al., 2011: 4). The data of the GNTD were obtained from a systematic review of peer-reviewed journals and grey literature, ministries of health records in schistosomiasis-endemic countries and data from surveys conducted by research institutions (Hürlimann et al., 2011: 2-3). Data points retrieved from publications and reports were geo-referenced retrospectively using the GEOnet Names server ([earth-info.nga.mil/gns/html](http://earth-info.nga.mil/gns/html)), topographic or sketch maps or Google web search (Hürlimann et al., 2011: 4). An additional database of school- and community-based surveys has been provided by the national schistosomiasis control programme in Burkina Faso following a personal communication with Dr. Moussa Dadoari (Dadoari, 2011). These data represent 86 survey locations of *S. haematobium* species in Burkina Faso, each with 120 examined schoolchildren between 2003 and 2007. Epidemiological data used for this research refer to the species *S. haematobium* and *S. mansoni*, which are the most prevalent species in the study region.

Pre-selecting steps of epidemiological data are illustrated in Figure 4-1 and are based on selection criteria proposed by Schur et al. (2011b: 4). Surveys before 1980 were removed, because the parasite can reach a maximal life span of only 30 years within the human host. To establish a comparable database, only surveys of schoolchildren have been selected for the database to avoid age effects. The sensitivity of diagnostic techniques to establish parasite prevalence is an indicator for data reliability. Surveys based on low diagnostic sensitivity have therefore been removed from the database as proposed by Schur et al. (2011b: 4). If schools were surveyed more than once, the most recent survey has been selected for the database of

this study. Assuming that school locations closer than 500m have a quasi-identical catchment area, those sample locations were combined by summing up the number of examined and number of positive schoolchildren resulting in an updated prevalence rate for this specific location. The geo-location error following the combination of points at this distance is negligible in comparison to the retrospective geo-location procedure of surveys inherent in the database.

The spatial distribution of schistosomiasis prevalence data for the study area is illustrated in Figure 4-2. It can be seen that sample locations in Burkina Faso are distributed evenly in contrast to the more focal distribution in Côte d'Ivoire. For the selected study sites BUF, MAN, and TAB, the availability of epidemiological data was comparable (Figure 4-2). In the following, the properties of the data available for each of the three selected study sites are described in detail (Table 4-1). Data points of school prevalence are count data (Elliott et al., 2006: 5) that represent the sample of schoolchildren. Overall sample size is generally low and varies between the three study sites being highest in the MAN site (75), followed by the BUF site (74) and lowest in the TAB site (38). In contrast, the areal extent is lowest in the MAN (4,381 km<sup>2</sup>) and highest in the BUF site (32,826 km<sup>2</sup>). The number of individuals screened per school (=sample resolution) is an indicator of data reliability. The majority of school locations in Côte d'Ivoire have a sample resolution between 80 and 100 individuals per school and around 50 children per school in BUF (Table 4-1). The observed cases of disease prevalence range from 0 to 100%, allowing to model prevalence and discriminate between low (<10%), moderate (10 - <50%) and high prevalence (>50%) according to the recommendations of the WHO (2002a: 34). Table 4-1 shows that this classification scheme is covered by the epidemiological

data for all three study sites. The year of survey ranges from 1980 to 2007. The older surveys are predominantly in BUF and the most recent ones in MAN. Minimum distance between school locations varies from around 500 m in MAN to 3.26 km in BUF and maximum distance from 62.2 km in MAN to 252.6 km in BUF, whereas the study site TAB represents intermediate ranges for both distance metrics. The average nearest neighbour ratio investigates spatial dispersal of point data (Clark and Evans, 1954). A value less than one exhibits a clustered pattern and a value greater one indicates a trend toward dispersion. The p-value indicates whether a trend is statistically significant. All three study sites resulted a random distribution of the sample points, as no trends have been statistically significant.

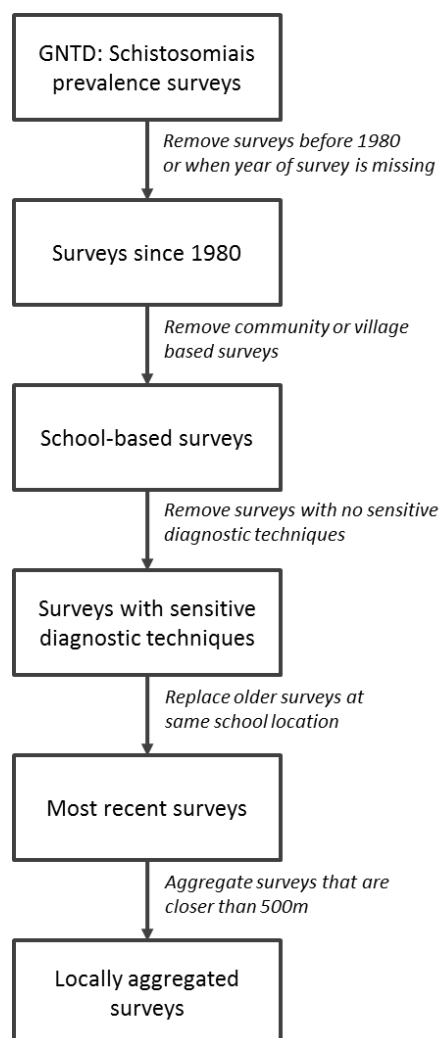


Figure 4-1: Steps in pre-selection of epidemiological data

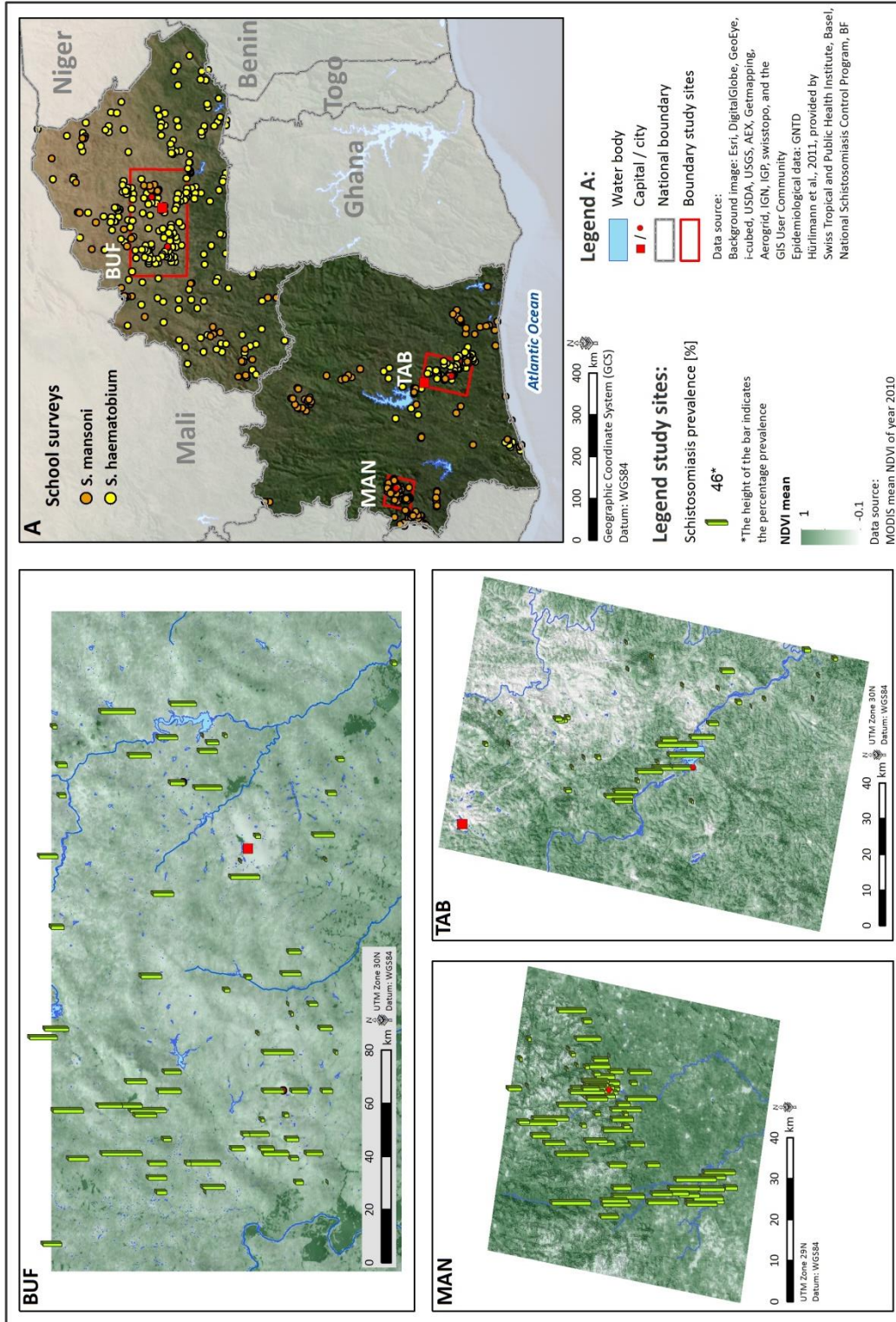
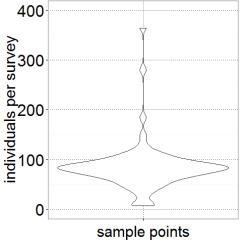
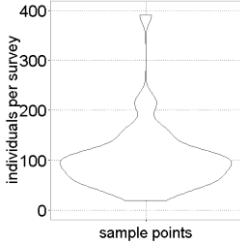
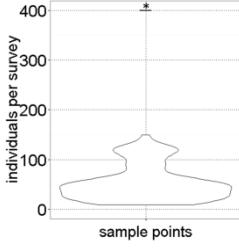
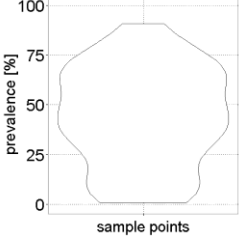
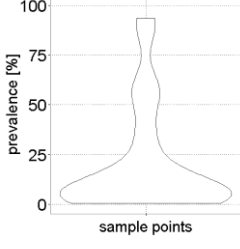
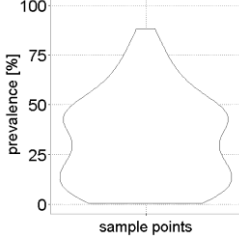
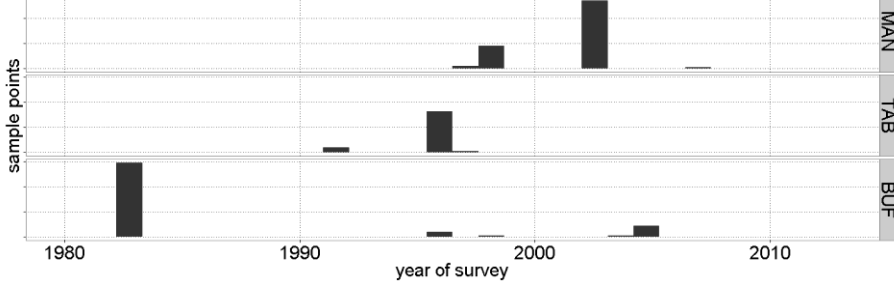
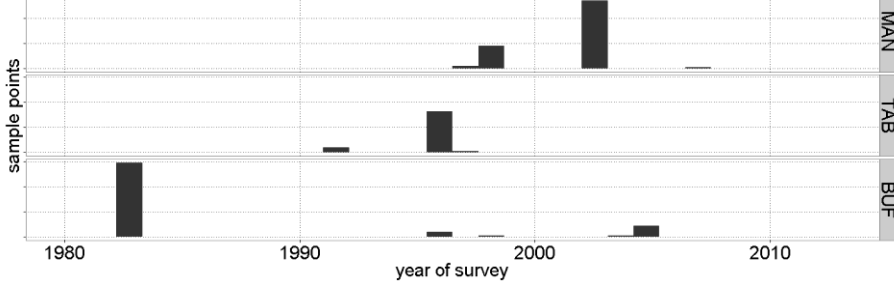
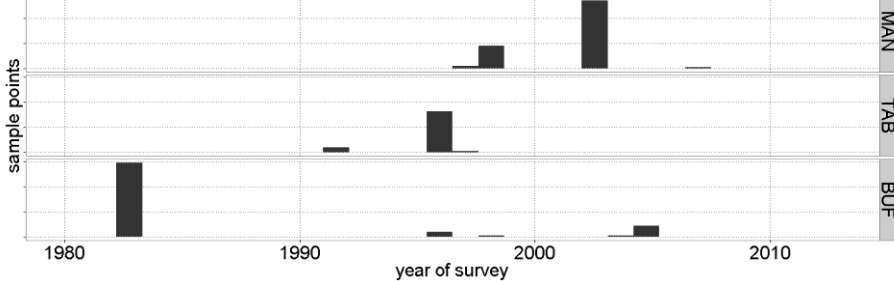


Figure 4-2: Spatial distribution of epidemiological data from the GNTD for Burkina Faso and Côte d'Ivoire (Hürlimann et al., 2011) and the data from the National Schistosomiasis Control Programme in Burkina Faso (Dadjoari, 2011) at school locations.

Table 4-1: Description of the epidemiological database for each of the three study sites in Côte d'Ivoire (MAN, TAB) and Burkina Faso (BUF). Sample size describes the number of school locations surveyed in each study site. The area extent indicates the area of each study site. Sample resolution gives the number of individuals surveyed per school location illustrated by a violin plot. The wider the plot at the x-axis the more school locations have been sampled with the number of individuals as indicated on the y-axis. The prevalence distribution is illustrated by a similar violin plot showing the prevalence between 0 and 100% on the y-axis. The year of survey is shown in histogram plots for all three study sites. The minimum and maximum distance describes the distance in km between the two closest and farthest school locations, respectively. The point pattern of sampled school locations for the study sites has been analysed using the nearest neighbour ratio (NNR).

\* outliers in BUF with 3,153 and 1,537 individuals/school removed from display

	MAN	TAB	BUF
Sample size (number of schools)	75	38	74
Area extent (km <sup>2</sup> )	4,381	8,476	32,826
Sample resolution* (individuals/school)			
Prevalence distribution			
Year of survey			
Minimum distance between schools (km)	0.52	0.74	3.26
Maximum distance between schools (km)	62.2	101.1	252.6
Point pattern index (NNR)	1.03 (p=0.66)	0.89 (p=0.18)	1.07 (p=0.22)

Data on schistosomiasis prevalence are spatially correlated as common exposures to environmental conditions influence transmission similarly at neighbouring locations (Vounatsou et al., 2009: 1695). The phenomenon of spatial autocorrelation is a typical phenomenon in ecology (Legendre, 1993: 1659) and stipulated by the ‘first law of geography’ (Tobler, 1970) stating that “near things are more related than distant things” (Tobler, 1970: 236). Hence, the values of neighbouring measurements are either “more similar (positive autocorrelation) or less similar (negative autocorrelation) than expected for randomly associated pairs of observations” (Legendre, 1993: 1659). The autocorrelation inherent in spatial data has consequences on statistical models as many models are not explicitly spatial and make the standard assumption that observations are independent (Franklin, 2010: 138). However, it has at the same time been observed that prevalence of *Schistosoma spp.* has a typically focal distribution and therefore low spatial autocorrelation (Abdel-Rahman et al., 2001: 50).

To understand spatial autocorrelation inherent in the data used in this study, in a first step the overall relationship between distance and prevalence similarity between school locations has been tested using the Mantel test (Mantel, 1967: 213). This investigates regression coefficients between the observed pattern and a randomized pattern with one of the matrices being shuffled (Koenig, 1999: 22-23). This test has been processed using the “ade4” package in R (Dray and Dufour, 2007). A Monte Carlo simulation based on 9,999 replications showed positive correlation for the study sites MAN and BUF and negative correlation for the study site TAB. The Null-Hypothesis of unrelated matrices has been rejected only for the study site MAN ( $p < 0.001$ ). Based on this result, spatial autocorrelation changes significantly with distance for the study site MAN, and no significant change between autocorrelation and distance has been established in TAB and BUF. This reflects the high sample density given in sample size per area extent in MAN compared to much lower sample density in other study sites.

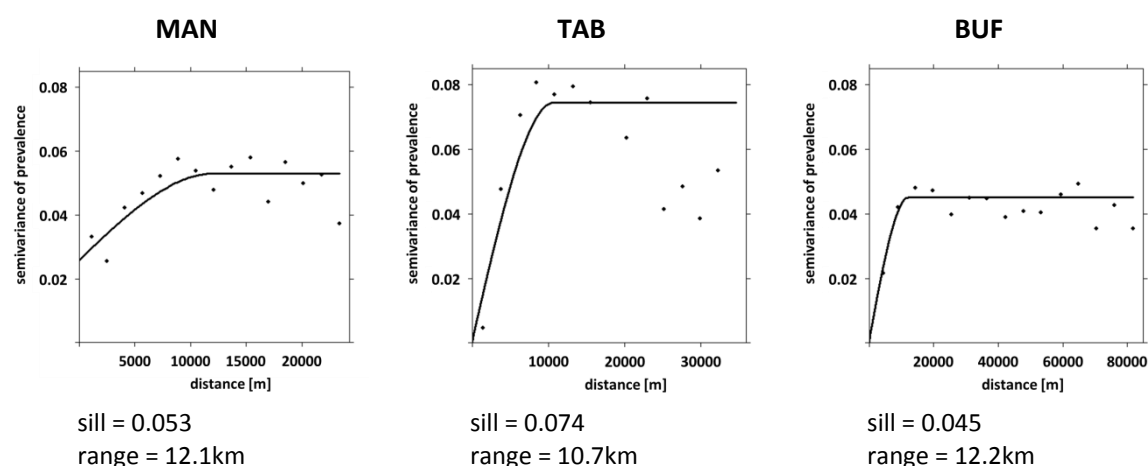


Figure 4-3: Patterns of the spatial structure of schistosomiasis prevalence in the three study sites MAN, TAB and BUF. The dots represent the empirical semivariogram and the solid line shows the best-fitted omnidirectional semivariogram based on a spherical model for de-trended prevalence data. The sill indicates the maximum of the modelled semivariance and the range indicates the distance, for which spatial independence of two school locations is given. Thus, spatial autocorrelation is inherent in the data until the distance of 12.1 km in MAN, 10.7 km in TAB, and 12.2 km in BUF.

In a second step, the spatial dependence of schistosomiasis prevalence has been quantitatively assessed based on the estimation of spatial semi-variance between all pairs of observations according to Fortin and Dale (2005: 132-138). The empirical semivariogram of the

sample data provided the basis to fit an omni-directional, spherical and de-trended semivariogram model using the ordinary least-squares (OLS) regression method provided by the “gstat” package (Pebesma, 2004). The most relevant characteristics of the fitted semivariogram model are the total observed variation of the variable (= sill), the modelled variability at distances smaller than the shortest empirical distances (= nugget) and the distance at which two observations could be considered independent (= range) (Karl and Maurer, 2010: 197).

Figure 4-3 presents semivariograms for the schistosomiasis prevalence in the three selected study sites of Burkina Faso and Côte d’Ivoire. In each setting, the semivariogram exhibits a spatial structure for distances between 10 and 12 km and after this distance there was no spatial autocorrelation in the database. However, if spatial autocorrelation is inherent in the data, either data that are closer to each other than the indicated distance need to be removed or this spatial structure needs to be considered by the model algorithm.

## 4.2 Remote sensing data

Remotely sensed land surface characteristics of different spatial resolutions were derived from RapidEye (6.5 m), Landsat 5 TM (30 m), and Terra MODIS (250 m/500 m/1 km) data for the purpose of multi-scale analysis. The spectral bands that correspond between the sensors are the blue, green, and red bands in the visible spectrum and the near infrared bands (Figure 4-4). RapidEye provides an additional red edge band (4) and both Landsat 5 TM and Terra MODIS data provide further spectral bands in the middle infrared and shortwave infrared wavelength, which are also suitable for the generation of relevant spectral indices (Figure 4-4). The technical details of these three sensors are listed in Table 4-2.

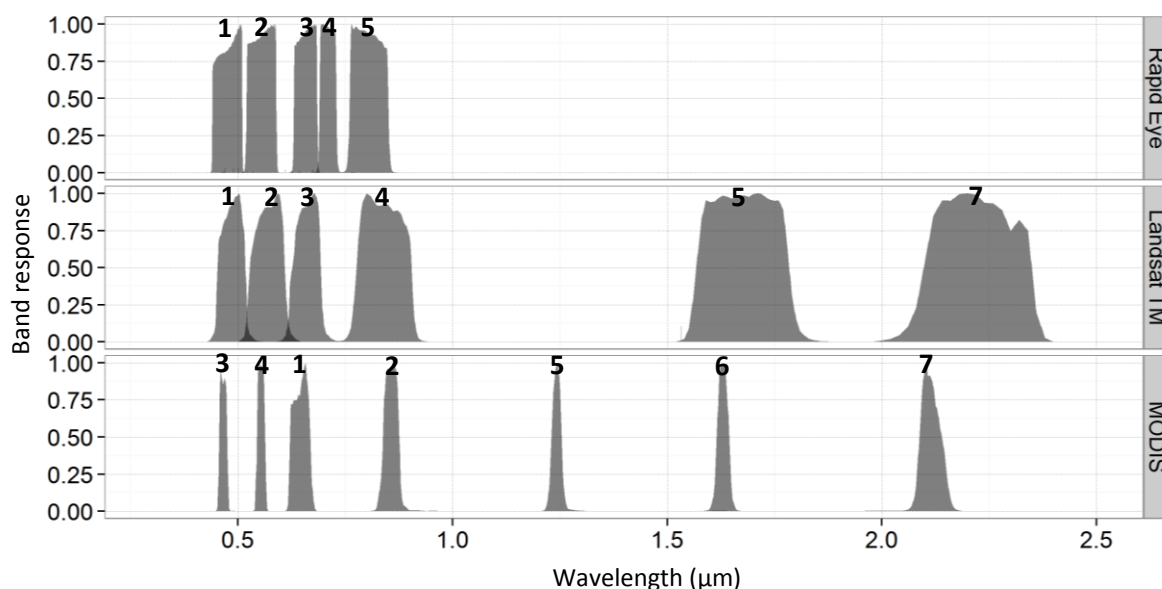


Figure 4-4: Spectral bands of multi-scale RS data from the sensors RapidEye, Landsat 5 TM, and Terra MODIS used in this study. The band numbers according to the sensor configuration are noted above each band. Source of spectral response function: RapidEye (Blackbridge, 2013), Landsat 5 TM (USGS, 2014), and Terra MODIS (NASA, 2014).

For the multi-scale analysis, data from the year 2010 and 2011 have been selected due to the availability of high-resolution RapidEye data. The temporal correspondence of RS data with

epidemiological data is considered acceptable due to the 30 year life span of adult worms and the chronic characteristic of the disease. The specific scale-related analysis of RS data with different spatial resolutions has focused on the study site MAN, where coverage of multi-scale RS data was most appropriate. In contrast, the study site TAB was lacking coverage of corresponding Landsat data with acceptable level of cloud coverage and BUF was not sufficiently covered by RapidEye data. The available scenes used for the multi-scale analysis in this study are noted in Table 4-3.

*Table 4-2: Technical details of the multi-scale RS data used in this study. Numbers in round brackets refer to the spectral band number of the respective sensor. Nir = near infrared, mir = middle infrared, swir = shortwave infrared, tir = thermal infrared.*

	<b>RapidEye</b>	<b>Landsat 5 TM</b>	<b>MODIS</b>
<b>Start</b>	29 <sup>th</sup> August 2008	1 <sup>st</sup> March 1984	18 <sup>th</sup> December 1999 (Terra) 4 <sup>th</sup> May 2002 (Aqua)
<b>Repetition rate</b>	Daily (off-nadir) 5.5 days (at nadir)	16 days	Daily (> 30°N/S), else every two days
<b>Swath width</b>	77 km	185 km	2,330 km
<b>Spatial resolution</b>	6.5 m (nadir)	30 m (120 m tir band 6)	250 m (1-2) 500 m (3-7) 1 km (tir 20-23 and 31-32)
<b>Spectral resolution</b>	Blue (1): 0.44 – 0.51 µm Green (2): 0.52 – 0.59 µm Red (3): 0.63 – 0.685 µm Red edge (4): 0.69 – 0.73 µm Nir (5): 0.76 – 0.85 µm	Blue (1): 0.45 – 0.52 µm Green (2): 0.52 – 0.60 µm Red (3): 0.63 – 0.69 µm Nir (4): 0.76 – 0.90 µm Mir (5): 1.55 – 1.75 µm Tir (6): 10.40 – 12.50 µm Swir (7): 2.08 – 2.35 µm	Red (1): 0.62 – 0.67 µm Nir (2): 0.841 – 0.876 µm Blue (3): 0.459 – 0.479 µm Green (4): 0.545 – 0.565 µm Mir (5): 1.23 – 1.25 µm Swir1 (6): 1.628 – 1.652 µm Swir2 (7): 2.105 – 2.155 µm

Besides the multi-scale analysis of specifically derived RS variables, several other data products have been analysed with respect to the spatial risk of schistosomiasis transmission. The statistical metrics of Terra MODIS vegetation indices (MOD13Q1 product), daytime LST (MOD11A2) and the calculated difference between day and night-time LST have been investigated in this study, since these metrics respond to ecological requirements of the parasite and snail as documented in Section 3.2.2. Relevant topographic characteristics have been derived from the 30m Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) global digital elevation model (GDEM), which offers higher spatial resolution of topography in contrast to other topographic data (90m and more) so far used for schistosomiasis risk assessment. Other remotely sensed measurements relevant for schistosomiasis transmission used in this study were precipitation derived from Tropical Rainfall Measuring Mission (TRMM) data and information on the spatial distribution of human settlements provided as an established product from TerraSAR-X data.

## 4.2.1 Technical description and pre-processing of selected remote sensing data

**RapidEye**, launched in 2008, is a constellation of five Earth imaging satellites with identical and equally calibrated sensors able to acquire multispectral images on a near-daily basis (Blackbridge, 2013). This multi-temporal high-resolution RS data provides a very useful database for the analysis of snail and parasite habitat dynamics. For this study, archive data from the years 2010 and 2011 have been accessed through the RapidEye Science Archive (RESA) at the German Aerospace Centre (DLR, 2014b) as level 1B (L1B) data. Basic pre-processing applied to L1B data include radiometric and geometric sensor corrections (Blackbridge, 2013: 7). For Burkina Faso, images free of clouds or acceptable cloud coverage below 5% could be received for the dry season in January/February and for the end of the rainy season in October/November (Table 4-3) covering the sub-site within the study site BUF (see Figure 2-1). In the tropical region of Côte d'Ivoire, only dry season images between January and February were sufficiently free of cloud contamination. Further pre-processing steps involved the transformation into the Universal Transverse Mercator (UTM) projection with the World Geodetic System 1984 (WGS84) using bilinear interpolation, orthorectification using a 90 m SRTM digital elevation model (DEM), as well as atmospheric and topographic correction using ATCOR (Richter and Schläpfer, 2012) within the CATENA pre-processing tool of the DLR (2014a). Geometric position of RapidEye data has been further adjusted based on sampled ground control points (GCPs) in Burkina Faso.

**Landsat 5 TM** has been launched on 1<sup>st</sup> of March 1984 and was the longest lasting satellite (decommissioned on 5<sup>th</sup> June 2013) of the Landsat family established in 1972. Data from the Landsat 5 TM can be accessed at no cost from the Earth Resources Observation and Science (EROS) Centre at the United States Geological Survey (USGS) pre-processed using the Level 1 Product Generation System (LPGS). For this study, L1B data have been acquired for the study sites MAN in western Côte d'Ivoire and BUF in central Burkina Faso with time steps in correspondence with available RapidEye data (Table 4-3). However, the study site TAB in southern Côte d'Ivoire has not been covered by Landsat 5 TM for this time period. As this study site is located in the border area of a Landsat tile, data from the Landsat 7 Enhanced Thematic Mapper Plus (ETM+) sensor with large gaps due to the failed Scan Line Corrector (SLC) and at the same time largely contaminated with clouds have not provided an alternative. L1B data have been processed further using ATCOR 2/3, version 8.2 (Richter, 1996; Richter and Schläpfer, 2012). Clouds have been detected and masked based on thresholds in the visible blue band and thermal infrared band six of Landsat 5 TM data (Martinuzzi et al., 2007). The geolocation accuracy of Landsat 5 TM data has been inspected based on sampled GCPs in Burkina Faso and RapidEye imagery. Position accuracy of the pre-processed data has been in the sub-pixel level of a 30m Landsat 5 TM pixel and therefore no additionally correction has been performed.

**MODIS** is a multi-spectral (36 bands) and multi-temporal (daily to every day repeat coverage) remote sensor and part of the Earth Observing System (EOS), which was designed by the NASA to provide observations of terrestrial, atmospheric and oceanic phenomena and processes on a global scale (Justice et al., 2002: 3). The first MODIS was integrated on the Terra (EOS AM-1) spacecraft together with four other sensors and successfully launched on 18<sup>th</sup> December 1999. A second mission of MODIS was started on 4<sup>th</sup> May 2002 on-board the Aqua (EOS PM-1) spacecraft. MODIS data from Terra and Aqua have operational reception, are available at no cost



and still today acquire images on a near-daily basis (Justice et al., 2002: 3). MODIS provides seven spectral bands (bands 1-7) useful to derive land surface characteristics (Table 4-2) and six thermal infrared bands (bands 20-23 ranging from 3.66-4.08  $\mu\text{m}$  and bands 31-32 ranging from 10.78-12.27  $\mu\text{m}$ ) relevant to derive land surface temperature. Data products from the MODIS sensors can be directly downloaded from the Land Processes Distributed Active Archive Centre (LP DAAC) from the USGS (2013) including further information such as quality assessment layers, viewing angles or observation times, if images were composited. For this study, surface reflectance of Terra MODIS data have been accessed in correspondence to the available RapidEye and Landsat 5 TM data through the MOD09GQ and MOD09GA products, which provide daily acquisition of spectral bands 1-2 at 250 m and 3-7 at 500 m resolution, respectively (Table 4-3). These Terra MODIS products are pre-processed as level 3 data and therefore corrected for radiometric, geometric, atmospheric, and bi-directional effects (Vermeulen and Vermote, 1999). Additionally, the MOD13Q1 product, a 16-day composite of vegetation indices with spatial resolution of 250 m, and the MOD11A2 product, an eight-day composite of LST and emissivity with spatial resolution of 1 km, have been investigated in this study. Data composition is done using a per-pixel Terra MODIS-specific compositing method that is dependent on the number of cloud-free observations available (Huete et al., 2002: 198). The Terra MOD13Q1 data include the NDVI and the enhanced vegetation index (EVI) from level 3 processed reflectance data (Huete et al., 1999). The Terra MOD11A2 comprises day-time and night-time LSTs retrieved under clear-sky conditions and emissivities estimated in bands 31 and 32 from land cover types (MOD12 product) per pixel at day and night-time (Wan, 1999). Both products have been downloaded for the year 2010 to match the high-resolution data. 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  $\pm 1$  K for temperatures between 263 K and 323 K (Wan et al., 2004: 272).

The **ASTER** is an imaging instrument, which is on-board the Terra satellite together with MODIS. In this study, the ASTER GDEM has been used as basis for topographic analyses. Altogether 82 tiles of the ASTER GDEM2 (Version 2) have been downloaded from Japan Space Systems (JSS, 2014) and mosaicked to cover the complete study area of Burkina Faso and Côte d'Ivoire with 30m spatial resolution, given a vertical root mean square error (RMSE) of 8.68 m (Meyer, 2011: 6).

**TerraSAR-X** is a German radar satellite from a joint venture of the DLR and the European Aeronautic Defence and Space Company (EADS) Astrium designed with a spatial resolution ranging from 1-16 m. Further technical details of this satellite can be found at DLR (2014c). In this study, a processed settlement mask derived from TerraSAR-X data has been used as described in the following Section 4.2.2.

### 4.2.2 Derivation of environmental variables

Environmental variables that can be related to the ecology of disease transmission (Section 3.1.2) were derived from spectral reflectance captured in the RS data and from the DEM (Section 4.2.1). Thereby indices were calculated to enhance the signal of specifically addressed features (e.g. water, vegetation, etc.) taking advantage of the surface specific differences in reflectance (see Figure 3-4). The NDVI, the EVI, the soil-adjusted vegetation index (SAVI), the modified soil-adjusted vegetation index (MSAVI), the normalized difference water index (NDWI), and the

modified normalized difference water index (MNDWI) were the vegetation and water indices used in this study and will be discussed in the following section.

The NDVI (Equation 3-2), which has already been described in Section 3.2.1, is the most prominent spectral index to enhance the vegetation signal detected by a remote sensor. Vegetation indices are considered as proxy for water availability through vegetation monitoring and can also delineate water, if index values result negatively. The EVI is one further development of the NDVI with the aim of optimising the signal with improved sensitivity in high biomass regions, a decoupling of the canopy background signal and a reduction of atmospheric influences (Huete et al., 2002: 196). The EVI (Equation 4-1) is expressed as

$$EVI = G \frac{p_{NIR} - p_{red}}{p_{NIR} + C_1 p_{red} - C_2 p_{blue} + L} \quad \text{Equation 4-1}$$

where  $p$  are atmospherically corrected surface reflectances,  $L$  is the canopy background adjustment that addresses nonlinear differential near infrared (*NIR*) and *red* radiant transfer through a canopy, and  $C_1$ ,  $C_2$  are coefficients of the aerosol resistance term, which uses the blue band to correct for aerosol influences in the red band. The coefficients adopted in the EVI algorithm based on MODIS data are  $G$  (gain factor) = 2.5,  $C_1 = 6$ ,  $C_2 = 7.5$ ,  $L = 1$  (Huete et al., 1994; Huete et al., 2002: 196).

The SAVI introduces a soil calibration factor,  $L$ , to the NDVI equation to minimise soil background conditions that exert considerable influence on partial canopy spectra and the calculated vegetation index (Huete, 1988: 296-299). The SAVI (Equation 4-2) is expressed as

$$SAVI = \frac{nir - red}{nir + red + L} (1 + L) \quad \text{Equation 4-2}$$

where an  $L$  value of 0.5 in reflectance space has been identified to minimise soil brightness variations and eliminates the need for additional calibration for different soils (Huete, 1988: 306; Huete and Liu, 1994: 897).

The MSAVI is based on the SAVI and uses an iterative, continuous  $L$  function to optimise soil-adjustment that varies with the amount of vegetation present. A large  $L$  value would best describe soil-vegetation interactions for low vegetation amounts, while  $L$  should become smaller with increasing vegetation amounts (Qi et al., 1994: 123). The MSAVI (Equation 4-3) includes visible red ( $p_{red}$ ) and near infrared ( $p_{NIR}$ ) reflectance and is expressed as follows (Qi et al., 1994: 124)

$$MSAVI = \frac{2p_{NIR} + 1 - \sqrt{(2p_{NIR} + 1)^2 - 8(p_{NIR} - p_{red})}}{2} \quad \text{Equation 4-3}$$

The NDWI was developed by McFeeters (1996) with the aim to delineate open water. The NDWI (Equation 4-4) is expressed as follows

$$NDWI = \frac{(green - nir)}{(green + nir)} \quad \text{Equation 4-4}$$

where spectral reflectance at visible *green* wavelength corresponds with the maximum reflectance of water and reflected near infrared radiation (*nir*) indicates the spectral wavelength, where water absorption is very high and the contrast between water and terrestrial vegetation or soil greatest (McFeeters, 1996: 1429).

The MNDWI aims to enhance the delineation of open water features as proposed by McFeeters (1996) by using the middle infrared (*mir*) signal instead of the near infrared (*nir*) band and thereby reduce noise experienced from built-up land (Xu, 2006: 3026-3027). In this study, due to unavailable *mir* bands provided by RapidEye data, the MNDWI (Equation 4-5) could be calculated for Landsat 5 TM and Terra MODIS data only and is expressed as follows

$$MNDWI = \frac{(green - mir)}{(green + mir)} \quad \text{Equation 4-5}$$

At this point, it should be mentioned that Gao (1996) has also named a NDWI, however, this index uses band combinations of the near infrared and middle infrared wavelength and aims to detect vegetation water liquid and thus is different from McFeeters' NDWI.

*Table 4-3: Overview of RS environmental variables used in this study and its availability for the three selected study sites. RE=RapidEye, LS=Landsat 5 TM, MOD=Terra MODIS, GUF=Global Urban Footprint, MOD13Q1 and MOD11A2 describe value added products of the Terra MODIS sensor, ASTER GDEM=Global Digital Elevation Model of the ASTER sensor. If (RE) is noted in brackets, only the sub-site of BUF (Figure 2-1) is covered by RapidEye data. The acquisition dates of RapidEye, Landsat 5 TM and Terra MODIS for multi-scale analysis of the study site MAN are 5<sup>th</sup> January 2011, 12<sup>th</sup> January 2011 and mean from 5<sup>th</sup> to 12<sup>th</sup> January 2011, respectively. The Rapid Eye data used for the study site TAB were acquired on 3<sup>rd</sup> January 2011. The Rapid Eye data that cover the sub-site of the study site BUF were acquired on 18<sup>th</sup> February 2010 (dry season image) and 27<sup>th</sup> October 2010 (wet season image). The full study site of BUF was covered by Landsat TM 5 data acquired between 29<sup>th</sup> January and 21<sup>st</sup> February 2010 for the dry season and between 4<sup>th</sup> November and 22<sup>nd</sup> December 2010 for the wet season. An extensive list of single RS environmental variables used for schistosomiasis risk modelling in each study site is given in the Appendix (Table A 1).*

Environmental variables	MAN	TAB	BUF
Spectral reflectance	RE, LS, MOD	RE	(RE), LS (dry/wet)
NDVI	RE, LS, MOD	RE	(RE), LS (dry/wet)
EVI	RE, LS, MOD	RE	(RE), LS (dry/wet)
SAVI	RE, LS, MOD	RE	(RE), LS (dry/wet)
MSAVI	RE, LS, MOD	RE	(RE), LS (dry/wet)
NDWI	RE, LS, MOD	RE	(RE), LS (dry/wet)
MNDWI	LS, MOD	-	LS
Water body	RE/LS	RE	(RE), LS
Euclidean distance from water body	RE/LS	RE	(RE), LS
Area of human settlements	RE (mapping)	RE (mapping)	GUF
- statistical metrics -			
NDVI (min, max, mean, median)	MOD13Q1	MOD13Q1	MOD13Q1
EVI (min, max, mean, median)	MOD13Q1	MOD13Q1	MOD13Q1
LST (min, max, mean, median)	MOD11A2	MOD11A2	MOD11A2
$\Delta$ Temp (min, max, mean, median)	MOD11A2	MOD11A2	MOD11A2
- topographic characteristics -			
Altitude	ASTER GDEM	ASTER GDEM	ASTER GDEM
Slope	ASTER GDEM	ASTER GDEM	ASTER GDEM
Sink depth	ASTER GDEM	ASTER GDEM	ASTER GDEM
Stream order	ASTER GDEM	ASTER GDEM	ASTER GDEM

The most important class of land cover with respect to transmission of schistosomiasis is **water**. The spectral signature of water is illustrated in Figure 3-4 and represents its maximum of reflectance in the visible wavelengths and shows nearly full absorption of EMR towards the near

infrared and middle infrared. The best mode of discriminating water from land using optical RS provide data at near infrared and middle infrared wavelengths between 0.74 and 2.5  $\mu\text{m}$ , because in this region, water bodies appear very dark in contrast to land surfaces, where vegetation or soil appears bright due to reflection of significant amounts of EMR. However, most natural water bodies contain a variety of organic (e.g. phytoplankton) and inorganic (e.g. suspended minerals) constituents that highly modify the spectral properties of water (Jensen, 2000: 385). This allows on the one hand estimating water quality and turbidity, on the other hand it proves to be difficult to distinguish land and water surfaces. For this study, a water mask has been derived from RapidEye and Landsat 5 TM imagery by thresholding the NDWI with a cut-off value set to zero, above which the spectral index delineates water (McFeeters, 1996: 1429-1430). However, in the study site MAN in western Côte d'Ivoire, the outcome of threshold-based water delineation was poor due to the very small ponds and river lines, which are further partly covered by dense riparian vegetation. The exposure of dark rocks in the mountainous region has further contributed to misclassification of water in this region. Hence, in this area, the water bodies have been mapped using a hierarchical procedure: First, a supervised classification of water/non-water based on RapidEye and Landsat 5 TM data as well as the NDWI has been applied using the random forest classification algorithm (Breiman, 2001). Secondly, topographic landscape elements improbable for the establishment of water bodies were excluded by thresholding curvature and slope calculated from the ASTER GDEM. The resulting water mask has been refined manually by visible inspection of RapidEye data and very high-resolution data available in Google Earth© (GoogleInc., 2010). The derivation of a water mask with spatial resolution of Terra MODIS data did not seem appropriate for the small-scale structure of water bodies as present in the selected study sites. Additionally, the Euclidean distance from water bodies has been calculated for the study sites.

**Human settlements** provide a highly relevant piece of information for the analysis of schistosomiasis transmission risk in space, because only where humans are present, is there a risk of infection. Currently, there are several RS based products available, where human settlement is classified or modelled on global scale. An overview is given by Schneider et al. (2009: 2). However, even the products with highest spatial resolution such as GlobCover with 300m (Arino et al., 2011) or the MCD12Q1 MODIS land cover product with 500 m (Strahler et al., 1999) failed to represent the settlements reasonably. Currently, a settlement mapping approach based on high-resolution RS data from the TerraSAR-X and TanDEM-X missions is being explored for its potential to map urban settlements with 12 m spatial resolution on a global scale (Esch et al., 2010; Esch et al., 2012; Esch et al., 2013). An unpublished beta version of the global urban footprint (GUF) product was provided for this study covering strip wise most areas of Burkina Faso and Côte d'Ivoire. Additionally, information about human settlements has been derived from RapidEye data for the study sites in Côte d'Ivoire by visual delineation of settlements. However, this was not possible for the study site BUF, where rural settlements predominate and consist of dispersed, small agglomerations of loam houses. Due to the extensive area of this study site and the absence of very high-resolution data covering it, the provided beta version of the GUF was used here as it was the only available information. A simple comparison between the GUF and small areas of manually mapped settlements in Google Earth (GoogleInc., 2010) for testing the GUF correspondence resulted that information was useful for urban agglomerations but not appropriate in the rural regions.

Based on Terra MODIS time series data from the year 2010, **statistical metrics** of vegetation indices provided by the MOD13Q1 product (NDVI and EVI) and LST provided by the MOD11A2 product have been calculated. Therefore, Terra MODIS time series data have been masked for clouds based on the provided cloud mask and then summarised by its minimum, maximum, mean, and median value, both on a pixel level. For Terra MODIS LST, statistical metrics have been derived for the daytime temperature measure and the difference between day- and night-time LST.

**Topographic information** about the environment plays a crucial role for schistosomiasis risk profiling, due to the fact that topography can determine to some extent whether water accumulates at a certain location and how fast it flows at the surface. Elevation above sea level provided by the ASTER GDEM was used for this research. In a first step, the DEM was smoothed using a majority filter for a 15x15 moving window. This procedure minimises noise in the data and errors of single pixels. Subsequently, the topographic variables were derived by separate procedures (Figure 4-5). Topographic sinks were described by pixel values, where all neighbouring values were equal or greater than the centre pixel value. The landscape of Burkina Faso has demonstrated an undulating terrain with many natural and man-made sinks on a small scale. To map these sinks and measure its depth, a sink mask has been derived from an image-differencing analysis between the smoothed image and the filled smoothed image. However, it is common that sinks in elevation data are due to errors in the data caused by rounding of elevations to integer numbers. Therefore, all other topographic variables such as elevation, slope and stream order have been calculated from a smoothed image with filled sinks. Slope of the terrain has been derived from the DEM as proxy-indicator for potential flow velocity of surface runoff with inclination calculated in degrees. As water runoff travels from higher to lower altitudes and usually becomes organised in a branched network of stream channels, the flow direction and flow accumulation of water result a stream network of topographic water drainage. The resulting stream network has been ordered according to Strahler (1957) based on an accumulation threshold of 18000 cells, which was the equivalent for the 30m ASTER GDEM to the given SRTM-based threshold given by Beck-Wörner et al. (2007).

**Water surface temperature** of inland water bodies may be obtained using thermal RS techniques. However, unlike land surfaces, water bodies transfer energy primarily through convection between water surface and bottom of the water body. This mixing is responsible for the relatively uniform surface temperature of a water body between day and night in contrast to the high temperature deviations between day and night of land surfaces. Due to the thermal inertia of water bodies, they appear cooler than land during the day and warmer during the night. The emissivity of water is with 0.98-0.99 very close to 1, which allows obtaining relatively accurate water surface temperature measurements by the remote sensor, when effects of the intervening atmosphere are accounted for. However, the RS measurement of water temperature is only a surface measurement of the water body and does not detect any significant change of water temperature that might occur a few meters below the surface (e.g., a thermocline). In this study, the Landsat 5 TM thermal band six has been used to detect water surface temperature as ecological indicator for schistosomiasis-related snail and parasite habitat conditions. The atmospheric correction (Section 4.2.1) has been calculated with a standard emissivity of 0.98, which is appropriate to detect water surface temperature. However, the

120 m spatial resolution of the thermal Landsat 5 TM band 6 is a limitation for the measure of water temperature for water bodies of smaller extent.

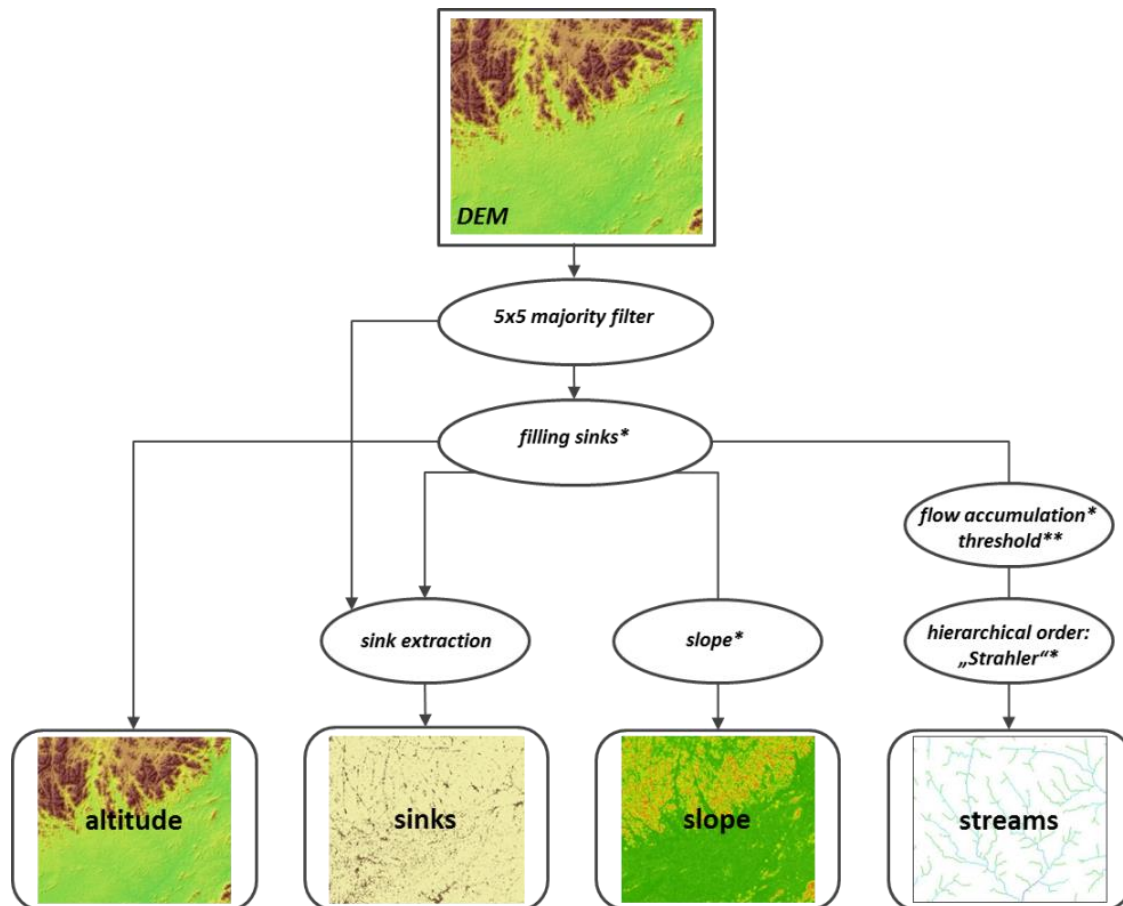


Figure 4-5: Processing chain to derive topographic variables from the ASTER GDEM

\* Function provided in the ArcGIS® spatial analyst toolbox

\*\* Number of cells accumulating water: 18000

## 4.3 Environmental *in situ* data

In the dry season of March 2011, a field trip to central Burkina Faso was conducted for the scope of this research. The objectives of this field trip were: (i) to investigate the landscape structure to discriminate surface characteristics using RS data; (ii) to sample GCPs for RS image geo-rectification and verification; and (iii) to establish locally adjusted criteria that discriminate and evaluate suitability of potential snail and parasite habitats in the field. The planned field trip to the study sites in Côte d'Ivoire had to be cancelled due to political unrests. However, a rough guess of the landscape characteristics could be gained from very high-resolution imagery provided by Google Earth© (GoogleInc., 2010) and field photographs provided by courtesy of Dr. Giovanna Raso, Swiss Tropical and Public Health Institute.

In total, 82 GCPs were sampled during this field trip in the rural landscapes within the sub-site of BUF around Ziniaré (see Figure 2-1) using a non-probability, purposive sampling scheme. The aim was to explore the variety of potential disease transmission sites in this region based on ecological suitability criteria (see Table 3-1) with the drawback to receive a sample that is not representative in a statistical sense. Furthermore, no parasite and snail data were sampled

during this field trip as the sampling of a representative database on snail presence, absence, infection and cercarial density within water bodies was beyond the scope of this research. Instead, at each GCP, a field verification form (Figure A 1 in the appendix), which has been adjusted for the purpose of this study from the standard FAO Land Cover Classification System (LCCS) (Di Gregorio and Jansen, 2005) guided the local analysis of this area around potential disease transmission sites. The following data were recorded: general land cover information such as land cover type, homogeneity, land form and seasonality aspects, type of water body (e.g. pond, channel or river), seasonality (visit during the dry season), type of substratum, water flow velocity, water colour and temperature, water usage, access, and visible human contact. Water body temperature was measured using a handheld digital thermometer at the outer boundary of the water body and resulted between 26.2°C (minimum) and 36°C (maximum). The measurement at the outer boundary of the water surface varied strongly with water depth at the shoreline and is not representative for the complete water body, but indicates a temperature range of water surface temperature in this region. Measurements of water temperature in central and northern Côte d'Ivoire published by Kinanpara et al. (2013: 114) varied between 27.4°C and 35.2°C.

During this field visit, local criteria to evaluate potential disease transmission sites have been derived. Thereby, the following observations were found to be highly relevant with respect to this study:

(1) Several different types of potential disease transmission sites have been discriminated during the field trip for this region (Figure 4-6):

(1.1) Man-made **dam lakes** have been observed to be used for fishing, irrigation of agricultural crops, to water animals, for recreational activities, washing clothes, and personal hygiene.

(1.2) Agricultural **irrigation** practices were observed close to permanent water bodies, where water was pumped by local generators and/or led through natural or concrete channels to the crop field. Thus, irrigated agricultural sites have been established almost exclusively at the shore side of these water bodies and unprotected direct human contact with these water bodies has been observed for almost all locations visited.

(1.3) **Rivers** have been observed to be either stagnant or very slow moving during the dry season. Often, rivers were dried out either completely or at certain passages. Artificial pools were often observed within river beds, where profound holes have been dug to collect water available during the dry season.

(1.4) **Seasonal pools** have been observed either as natural or artificial topographic depressions that have been either filled or dried out.

(2) At all visited water sites human access was not restricted either through natural or man-made barriers.

(3) The rural settlement pattern consists of small-scale agglomerations of several loam houses where one family lives with an average of 30 individuals (personal communication with field assistant Herman Ouoba – graduate student (MSc) from the University of Ouagadougou).

(4) The mode of transportation of the rural population has been observed to be mainly by foot or motor bike, whereas there were individual paths and gravel roads used all over the

sparse vegetated landscape. This information provides insight into the complex spatial relation between school location and human settlements within the catchment areas.

(5) School buildings were often located separately from dispersed rural settlements. The catchment area of a rural school has been estimated to be maximum 5 km walking distance for the pupils (personal communication with field assistant Herman Ouoba).



**Man-made dam lakes**



**Irrigated agriculture**



**Rivers**



**Seasonal pools**

*Figure 4-6: Main types of potential disease transmission sites within the study site BUF*



## 5 Modelling environmental suitability for schistosomiasis transmission

An infection with schistosomes depends largely on the spatial distribution of suitable snails, that act as intermediate hosts and are the prerequisite that a *Schistosoma* parasite reaches the stage to infect humans (see Section 3.1.2). These disease relevant species have specific habitat requirements, which are determined by environmental factors. For example, it has been shown in the laboratory and through field-based investigations that water temperature affects the metabolism of parasites and snails with consequences on parasite activity, survival and infection or snail fecundity, mortality, and rate of reproduction (Table 3-1). Specific temperature thresholds govern the presence or absence of a species and impact its fitness. If temperature conditions were not suitable for a disease-related species, the proliferation of the disease was not successful. The extensive literature on abiotic factors in relation to snail and parasite fitness, reveal conditions that are likely to determine a habitat's suitability or the unsuitability as a disease transmission site (Moodley et al., 2003: 618). Satellite RS data have proven to be useful to assess and monitor biophysical characteristics and detect such habitat conditions (Malone et al., 1994; Malone et al., 2001; Stensgaard et al., 2006).

Biological requirements of a species can be derived from laboratory or field studies (Malone, 2005: 28; Kearney, 2006: 186) and are useful to establish a relationship between a species and its abiotic environmental niche, which can be predicted for locations where survey data are lacking (Franklin, 2010: 12). The modelled environmental preferences, limits of tolerance, and behaviours of organism are summarised by the **HSI**, which was originally developed by the United States Fish and Wildlife Service (USFWS) to estimate the capacity of a habitat to support a species and quantify effects of land management alternatives on species habitats (USFWS, 1981: 10). Thus far, USFWS has applied the HSI within more than 150 species-specific models with the objective to support informed decision making with respect to land management and species conservation (USFWS, 2014). Additionally, the HSI has been rigorously investigated and published in the peer-reviewed literature for marine species (Brown et al., 2000; Vinagre et al., 2006), wildlife (Thomasma et al., 1991; Mitchell et al., 2002; Dussault et al., 2006) and plant species (Store and Kangas, 2001; R ger et al., 2005; Williams et al., 2008). It was aggregated from spatially superimposed species-specific variables of environmental suitability. Although there are numerous HSI models, few have been validated (Duncan et al., 1995: 1361), probably

due to the unavailability of adequate data to support validation (Schamberger and O’Neil, 1986: 7). However, HSI models belong to the most influential management tools in use and provide a transparent basis to further explore species-specific relations with the environment (Brooks, 1997: 163-164). The key contribution of HSI models lies in quantifying both the quality and quantity of available habitats for selected species as well as providing a repeatable assessment procedure based on current environmental data (Ortigosa et al., 2000: 3).

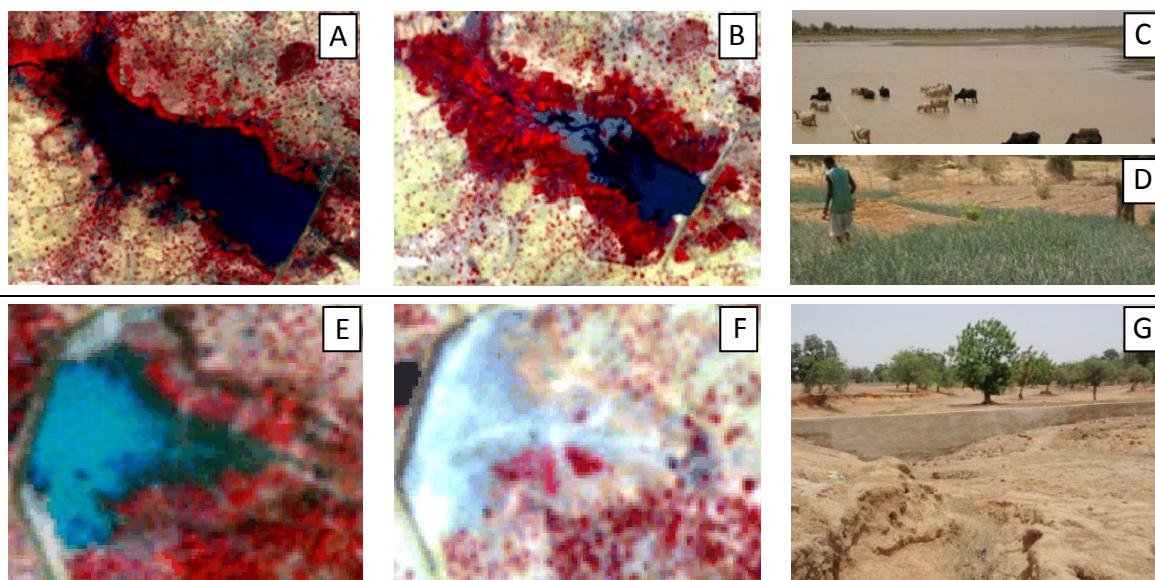


Figure 5-1: Suitability of potential transmission sites for schistosomiasis in Burkina Faso. In the upper row, the left RapidEye image (A) represents a dam lake during the wet season and the RapidEye image in the centre (B) represents the same site during the dry season. The upper field photographs (C and D) illustrate the situation of the dry season image on the ground. This site shows high suitability for potential transmission due to the stability of water persistence (C) and the practice of irrigation agriculture at its waterside (D). In the lower row, the left (E) and central (F) RapidEye images represent another dam lake during the wet and dry season, respectively, and the field photograph (G) illustrates this spot on the ground during the dry season. It shows that this dam lake is a less suitable area for disease transmission due to its temporary drying out.

RS data from the RapidEye sensor were acquired on November 16, 2009 (end of wet season) and February 18, 2010 (dry season). Spectral bands 5-3-1 were displayed as R-G-B colour composite. The field photographs were taken on March 21, 2011 (C and D) and March 22, 2011 (G).

With respect to schistosomiasis, such a mechanistic approach to derive environmental suitability for disease transmission has been investigated by Moodley et al. (2003), Stensgaard et al. (2006), and Stensgaard et al. (2013). All three studies investigated remotely sensed temperature suitability for snails and parasites only. However, RS data provide additional information with respect to environmental suitability for schistosomiasis transmission other than only temperature (see Table 3-3). According to the ecology of schistosomiasis as described in Section 3.1.2, an example, where different environmental suitability for schistosomiasis transmission can be directly derived from RS data, is stability of water level and riparian vegetation coverage, as illustrated in Figure 5-1. Landscape configuration and its temporal dynamic, influence the environmental suitability for potential schistosomiasis transmission and can be captured by RS data.

The **main objective** of this chapter is to assess the potential contribution of RS data to derive environmental suitability for transmission of schistosomiasis by means of a deductive mechanistic HSI modelling approach. Hence, strengths and weaknesses of RS data will be

discussed placing emphasis on: (i) the spatial delineation of environmental conditions, where disease transmission can potentially occur; (ii) quantitative prediction of environmental suitability within potential transmission sites; and (iii) the evaluation of the regional transferability of an established model to different eco-geographic regions.

## 5.1 Establishment of mechanistic model

A mechanistic modelling approach was used to investigate the potential of RS data to assess environmental suitability for the transmission of schistosomiasis. This model is parameterised based on published information on fitness, limits of tolerance and behaviour of disease-related snail and parasite species, and on field data sampled at the eastern sub-site of BUF around Ziniaré (Figure 2-1). Model outcomes are quantitative estimates of environmental suitability for transmission of schistosomiasis. According to Brooks (1997), the general HSI modelling procedure can be divided into three steps, as illustrated in Figure 5-2: First, species-related habitat variables are selected and species-specific requirements are attributed to these variables. Second, the model is composed from single habitat variables, which are scaled between poor and excellent habitat quality (from zero to one) and calibrated based on a sensitivity analysis of different weighting schemes. Third, the model is verified based on the linkage between modelled environmental suitability and field-based estimates of species suitability. In this study, a sensitivity analysis and comprehensive model validation was not feasible as reliable estimates of snail and parasite species population density and fitness were not available. However, the plausibility of modelled environmental suitability in reference to field-based estimates and the transferability of the model to different ecological regions in reference to schistosomiasis prevalence data within a 5 km buffer zone around the survey location were evaluated in this study.

### 5.1.1 Model development

RS data provided the input data to model environmental conditions and ecological limitations relevant for schistosomiasis-related snails and parasites. The model was developed based on the provided information as reviewed in Section 3.1.2. Water temperature, water flow velocity, and habitat stability have been referred to as the most important factors conditioning habitats of freshwater snail species (Appleton, 1978: 1-2). Conditions that influence the longevity and infectivity of free living larval stages of *Schistosoma* parasites are temperature, water flow velocity, turbidity, UV radiation, and exposure to chemical stimuli (Sturrock, 1993b: 11). In general, environmental suitability for schistosomiasis transmission has been considered highest when most favourable conditions for the coexistence of snail and parasite were given and lowest *vice versa*. As soon as conditions for either snail or parasite species become unsuitable, the environment was considered hostile for disease transmission.

The theoretical relations between species fitness and environmental conditions described in Section 3.1.2, were translated into mathematical expressions as proposed by an HSI modelling approach (USFWS, 1981: 42), namely, the functions of relative suitability. The estimation of relative suitability for each habitat variable by utilising continuous functions avoids the loss of information and the increased uncertainty, which would result from a categorical classification

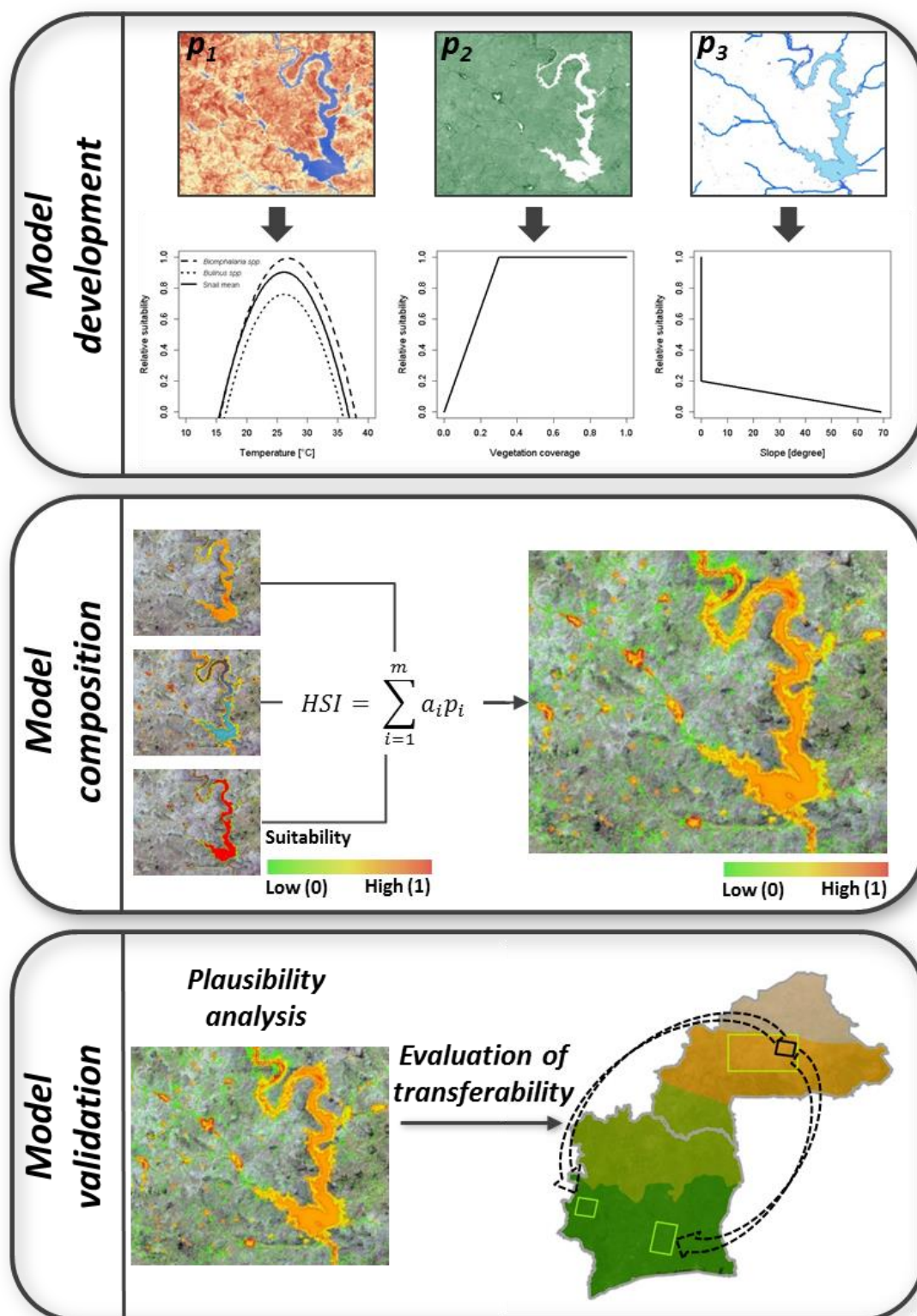


Figure 5-2: Overview of a mechanistic model procedure to derive environmental suitability for schistosomiasis transmission using RS data. The functions of relative suitability developed for the model can represent species-specific information if provided (see dashed and dotted lines that refer to specific snail species suitability with respect to temperature). If the line is solid, the relative suitability addresses snail- and parasite-related suitability in general (see Figure 5-4 for detailed view).

approach (Store and Kangas, 2001). Three methods were applied for the derivation of relative suitability functions of habitat variables (USFWS, 1981: 45), depending on the source of information and data provided. In the event that species response curves were provided, this mathematical relation could be directly transferred to the corresponding environmental variable derived from RS data. An example is given by the function derived for the length of the prepatent period of a parasite within the snail in relation to measured water temperature (Pflüger, 1980; Pflüger et al., 1984). If measurement or thresholds expressed the relationship between species fitness and environmental conditions, functions of relative suitability for specific habitat variables were derived by piece-wise linear or polynomial interpolation between provided values or logical assumptions. This allowed that non-linear relationships between habitat suitability and an environmental variable could be depicted (Store and Kangas, 2001). A third approach uses expert opinions to directly define the relative suitability function of a certain habitat variable. Relative suitability was scaled between 0 and 1 to make the environmental suitability comparable based on the single habitat variables.

The fundamental habitat variable, which is the prerequisite that schistosomiasis transmission can occur, is the availability of water. With RS data, this deterministic factor was mapped (Section 4.2.2) in a direct way by deriving a water mask for the area of interest based on high-resolution RapidEye data. Additionally, the presence of potential water was mapped in an indirect way from topographic information of sinks and drainage lines, where water can potentially accumulate following rainfall or flooding events. The direct measurement of water during dry and rainy season has been further extended with a buffer zone of 200m. This distance was measured from RS data during the dry season to capture the irrigated agricultural sites, which are directly connected to permanent or seasonal water bodies in this study site (Section 4.3). Water and potential water as deterministic habitat variables provide the basis for the HSI model and result in a mask of potential environmental suitability. Regions of potential environmental suitability have been further refined with quantitative information of snail- and parasite-specific habitat conditions, based on non-deterministic habitat factors described in the following paragraphs. Moreover, the binary classification provides the basis for analysis within the hierarchical model approach described in Chapter 6. Based on field expertise (Section 4.3) and the review of disease-related variables that can potentially be measured by means of RS data (Table 3-3), the following snail- and parasite-related habitat variables have been selected to establish an HSI model.

**Habitat stability** is defined as length of water persistence in weeks and was derived for the study area based on RapidEye data from the year 2010. Water bodies persisting less than 4 weeks for the case of *S. mansoni* and less than 5 weeks for *S. haematobium* were considered unsuitable, respectively (Sturrock, 1993a: 65). If water in a snail habitat remains for longer than 6 or 7 weeks, the habitat stability was considered suitable for transmission of *S. mansoni* or *S. haematobium*, respectively (Sturrock, 1993a: 65). The relative suitability function that connects the unsuitable and suitable state of habitat stability results from a linear interpolation between the measured values. The mathematical expressions of relative suitability of habitat stability with respect to proliferation of *S. mansoni* and *S. haematobium* (Equation 5-1 and Equation 5-2) are illustrated in Figure 5-3.

$$S. \textit{mansoni}: \quad f(w) = \begin{cases} 0 & \text{for } 0 < w \leq 4 \\ 0.5w - 2.5 & \text{for } 5 \leq w \leq 7 \\ 1 & \text{for } 8 \leq w \leq 52 \end{cases} \quad \text{Equation 5-1}$$

$$S. \textit{haematobium}: \quad f(w) = \begin{cases} 0 & \text{for } 0 < w \leq 3 \\ 0.5w - 2 & \text{for } 4 \leq w \leq 6 \\ 1 & \text{for } 7 \leq w \leq 52 \end{cases} \quad \text{Equation 5-2}$$

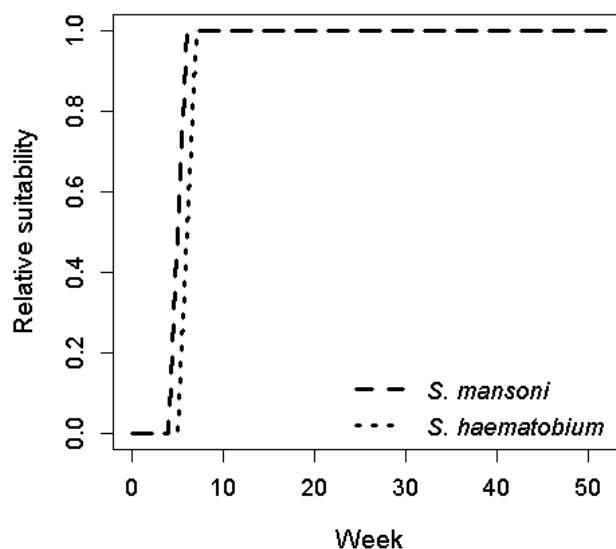


Figure 5-3: Sub-priority functions of relative suitability of habitat stability for *S. mansoni* and *S. haematobium*.

The relation between **water temperature** ( $T$ ) and the length of the prepatent period of parasites was defined as response function for *S. mansoni* by Pflüger (1980) and for *S. haematobium* by Pflüger et al. (1984). These species response functions were directly scaled to functions of relative suitability of temperature, in which the shortest prepatent period corresponded with highest suitability for parasite development (Equation 5-3 and Equation 5-4). Additionally, the relation between water temperature and snail mortality was investigated for snails collected in the field and viewed under laboratory conditions. *Bio. glabrata*, the intermediate host snails of *S. mansoni* showed a mortality rate close to 100% at experimental water temperatures below 16°C and above 36°C (Pflüger, 1980: 163). The mortality thresholds of *Bu. truncatus* snails, the intermediate host snails of *S. haematobium* were at temperatures below 17°C or above 33°C (Pflüger et al., 1984). The laboratory-based measurements of snail mortality (Pflüger, 1980; Pflüger et al., 1984) were interpolated based on a second order polynomial function and resulted respective functional relations for *Bio. glabrata* snails (Equation 5-5) and *Bu. truncatus* snails (Equation 5-6) as illustrated in Figure 5-4.

$$S. \text{ mansonii: } f(T) = \begin{cases} 0 & \text{for } T < 16 \\ -0.003 * \left( \frac{268}{T - 14.2} - 335 \right) & \text{for } 16 \leq T \leq 35 \\ 0 & \text{for } 35 < T \end{cases} \quad \text{Equation 5-3}$$

$$S. \text{ haematobium: } f(T) = \begin{cases} 0 & \text{for } T < 17 \\ -0.006 * \left( \frac{295}{T - 15.3} - 174 \right) & \text{for } 17 \leq T \leq 33 \\ 0 & \text{for } 33 < T \end{cases} \quad \text{Equation 5-4}$$

$$Bio. \text{ glabrata: } f(T) = \begin{cases} 0 & \text{for } T < 16 \\ -4.095 + 0.368 T - 0.007 T^2 & \text{for } 16 \leq T \leq 35 \\ 0 & \text{for } 35 < T \end{cases} \quad \text{Equation 5-5}$$

$$Bu. \text{ truncatus: } f(T) = \begin{cases} 0 & \text{for } T < 17 \\ -2.350 + 0.208 T - 0.004 T^2 & \text{for } 17 \leq T \leq 33 \\ 0 & \text{for } 33 < T \end{cases} \quad \text{Equation 5-6}$$

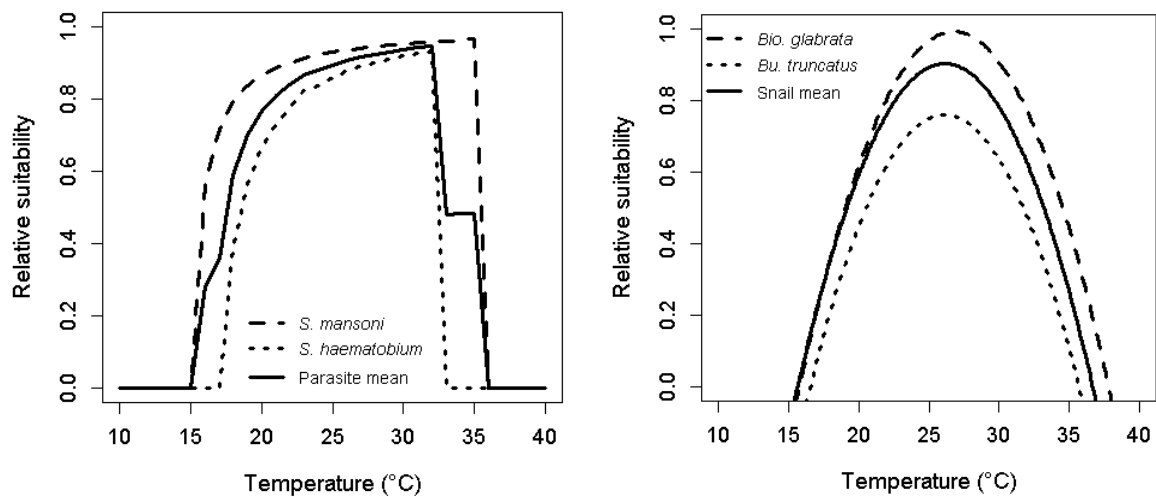


Figure 5-4: Sub-priority functions of relative suitability of water temperature for *S. mansonii* and *S. haematobium* parasites (left) and *Bio. glabrata* and *Bu. truncatus* (right). The mean suitability of parasites and snails are indicated by the solid lines.

A critical value of **water flow velocity** was established at 0.3 m/s (Scorza et al., 1961: 194). In faster flowing water snails become dislodged and the suitability of the habitat drops. The flow velocity of water  $V$  was compared with slopes derived from topographic RS data using the Manning's velocity Equation 5-7 (Albertson and Simons, 1964: 7-24)

$$V = \frac{1.49 * R^{0.66} * S^{0.5}}{n} \quad \text{Equation 5-7}$$

where  $R$  is the hydraulic radius,  $S$  is the line slope and  $n$  is the Manning's roughness coefficient. Within the visited test site around Ziniaré in Burkina Faso, the threshold of 0.3 m/s could be approximated to a slope of 0.00014 degrees. This calculation is based on the assumption that the majority of river beds are gravelled earth channels with some vegetation growth, which are represented by 0.025 for  $n$  (Albertson and Simons, 1964: 7-25).  $R$  is equal to the cross-sectional area of flow divided by the wetted perimeter (Arcement, 1989: 10) and was approximated for this calculation to be 0.7 m. This parameter could not be adjusted to the heterogeneous river

beds in the study site, because the 30 m spatial resolution of topographic data from the ASTER GDEM could not depict this small-scale information at all. Therefore, a mean width and depth of river beds was assumed to be 5 m and 2 m, respectively. However, this approach was applied by Kiel et al. (2006: 318), who suggested that water surface elevations were reasonable to estimate flow velocities from slopes derived from SRTM data for large rivers. The relative suitability function of water flow velocity (Equation 5-8) was derived from linear interpolation between the minimum, the suitability threshold, and the maximum. It was assumed that the relative suitability decreases strongly towards the derived threshold slope of 0.00014 degrees corresponding to a relative suitability of 0.2 and levels out with this general low suitability towards 0 (Figure 5-5).

Flow suitability  
calculated from  
slope:

$$f(S) = \begin{cases} -5714.3 S + 1 & \text{for } 0 \leq S \leq 0.00014 \\ -0.0029 S + 0.2 & \text{for } S > 0.00014 \end{cases} \quad \text{Equation 5-8}$$

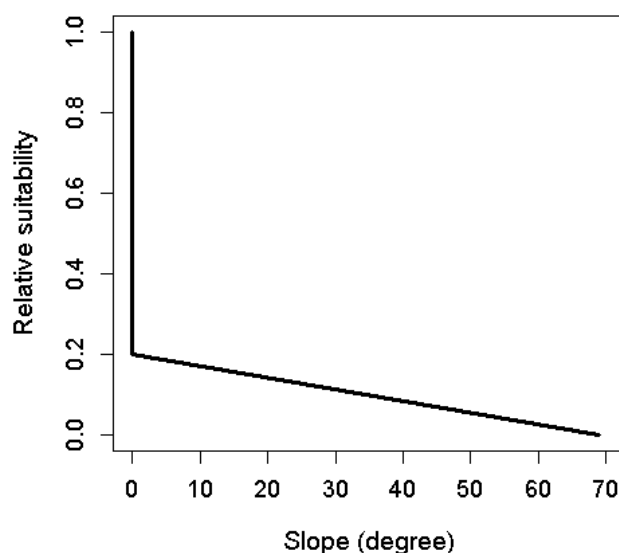


Figure 5-5: Sub-priority function of relative suitability of water flow velocity measured by the proxy of slope.

The relative suitability of **water depth** is expressed by the proxy measurement of Euclidean distance from the shoreline, which was calculated from the polyline boundary of the water masks derived from RS images during the dry and wet season (Section 4.2.2). Intermediate host snails of schistosomiasis are primarily distributed in shallow water at the margins of their habitat (WHO, 1957: 11-12). This information was translated into a decreasing suitability derived from linear interpolation between shoreline and a distance of 200 m inside the water. The threshold was estimated based on the slope to distance ratio between the maximum water level and the current one during the dry season around the visited dam lakes. For this calculation, the threshold was adjusted to 210 m to multiply the 30 m pixel resolution of the satellite data employed in the current study. At distances greater than 2 km from the shoreline to the center of the water body, no suitability for snails transmitting schistosomiasis was assumed. This



threshold was extracted from the extent of the greatest water body in the study site (Equation 5-9 and Figure 5-6).

Water depth  
calculated from  
distance from  
shore:

$$f(x) = \begin{cases} -0.0043x + 1 & \text{for } 0 \leq x \leq 210 \\ -0.000056x + 0.088 & \text{for } 210 > x \leq 2000 \\ 0 & \text{for } x > 2000 \end{cases} \quad \text{Equation 5-9}$$

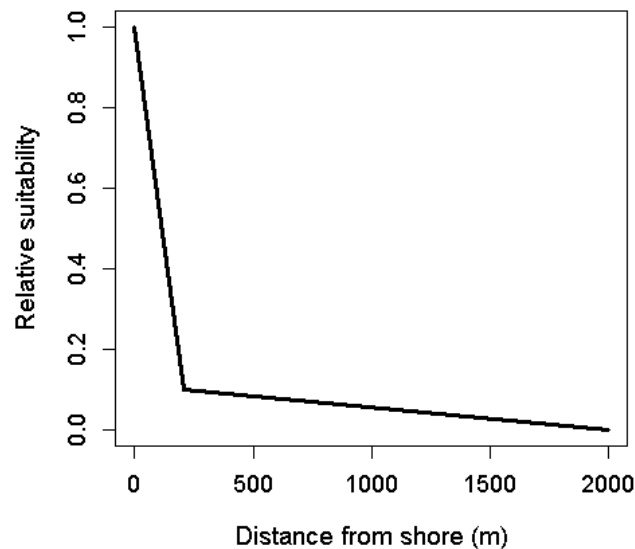


Figure 5-6: Sub-priority function of relative suitability of water depth measured by the proxy of distance from shore in meter.

The relative suitability based on **vegetation coverage** was measure within a 200 m buffer area along detected water sites. According to the theoretical suitability conditions described in Section 3.1.2, higher availability of vegetation positively conditions the habitat for freshwater snails in terms of food supply, surfaces to crawl and deposit egg masses or with respect to the content of dissolved oxygen in water (Table 3-1). The RS approach used in this study does not measure submerged vegetation, however, the buffer zone considers the potential vegetation input into the water body. In this model, the theoretical function of relative suitability with respect to vegetation coverage was derived to detect irrigated agricultural sites visited during the field trip, which corresponded to an average NDVI threshold around 0.3 in dry season RapidEye imagery. Based on this threshold, the sub-priority function of vegetation coverage (Equation 5-10) was a result of linear interpolation between unsuitable conditions for NDVI values smaller or equal 0 and suitable conditions at values of 0.3 and above (Figure 5-7).

**Stream order** was derived for the study area of Burkina Faso and Côte d'Ivoire as described in Section 4.2.2 and resulted in hierarchical levels ranging from order 1 to 7. According to the study of Beck-Wörner et al. (2007), the relative suitability function for stream order (Equation 5-11) was estimated from a linear interpolation between the maximum stream order of highest suitability and the minimum stream order of lowest suitability (Figure 5-8).

Vegetation coverage:

$$f(V) = \begin{cases} 0 & \text{for } V < 0 \\ 3.33 V & \text{for } 0 \geq V \leq 0.3 \\ 1 & \text{for } V > 0.3 \end{cases}$$

Equation 5-10

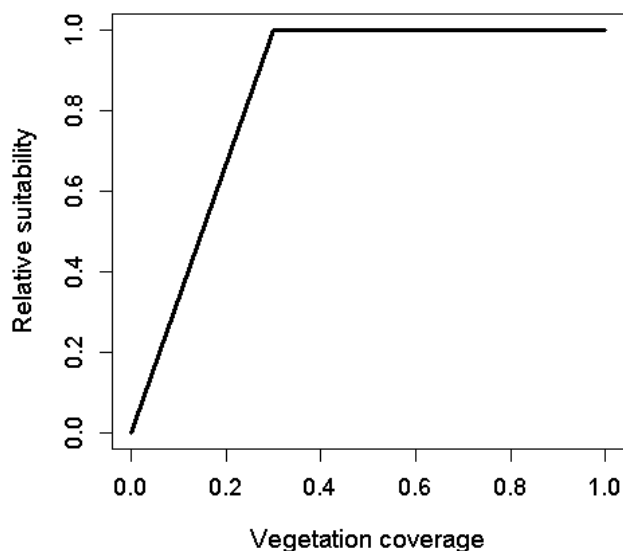


Figure 5-7: Sub-priority function of relative suitability of vegetation coverage derived from the NDVI value.

Stream order:

$$f(\text{stream}) = 0.143 \text{ stream for } 0 \leq \text{stream} \leq 7$$

Equation 5-11

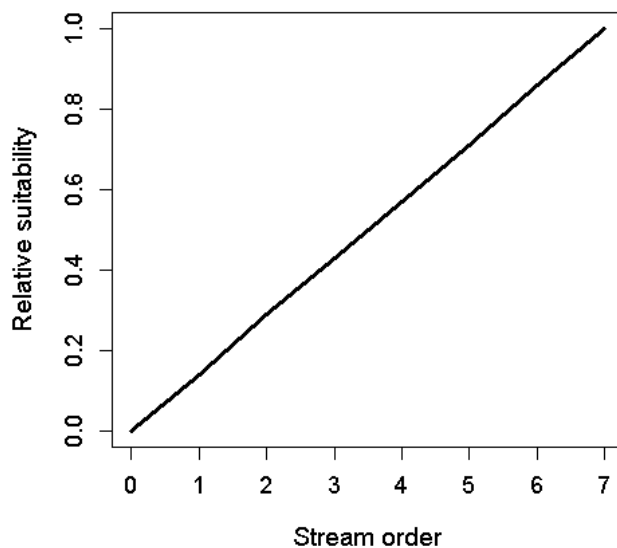


Figure 5-8: Sub-priority function of relative suitability of stream order.

**Sink depth** was derived for the study area of Burkina Faso and Côte d'Ivoire as described in Section 4.2.2 and resulted in a maximum depth of 222 m. The sub-priority function of sink depth

(Equation 5-12) was calculated from linear interpolation between the lowest sink depth of 1 m and the aforementioned maximum sink depth of 222 m (Figure 5-9).

Sink depth:  $f(z) = 0.005z + 0.11$  for  $1 \leq z \leq 222$  Equation 5-12

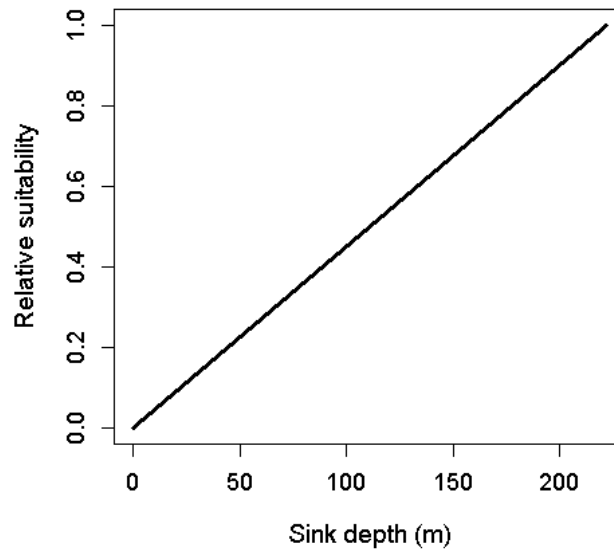


Figure 5-9: Sub-priority function of relative suitability of sink depth.

### 5.1.2 Model composition

Following the assignment of specific suitability characteristics to relevant environmental variables with respect to schistosomiasis transmission (Section 5.1.1), single attributes were combined to an overall HSI for the study site Ziniaré using the multi-criteria decision analysis (MCDA). MCDA is defined as “an umbrella term to describe a collection of formal approaches which seek to take explicit account of multiple criteria in helping individuals or groups explore decisions that matter” (Belton and Stewart, 2002: 2). This definition outlines the three dimensions of MCDA, namely: (i) the formal approach; (ii) the presence of multiple criteria; and (iii) decisions made by individuals or groups of individuals (Mendoza and Martins, 2006: 1). MCDA has been widely used within the GIS community for spatial decision support (Malczewski, 2006) and was investigated with respect to species habitat suitability modelling (Store and Kangas, 2001). In this research, model composition is formally described by the framework provided in Figure 5-10 and integrates multiple criteria relevant for schistosomiasis transmission that are captured by RS data as well as decisions connected to each criterium (Section 5.1.1). This structured approach can provide a basis for evaluating a number of alternative choices on the basis of the selected criteria (Store and Kangas, 2001: 80; DCLG, 2009: 10).

Composition of the mechanistic model of environmental suitability for schistosomiasis transmission discriminated in a first step between directly measured water and areas of potential water accumulation based on topographic information (Figure 5-10). In this study, the HSI (Equation 5-13) was calculated using an additive priority function as adjusted from Store and Kangas (2001: 82), and Walz et al. (2012: 7225).

$$HSI = \sum_{i=1}^m a_i p_i \quad \text{Equation 5-13}$$

*HSI* stands for the habitat suitability index and refers to the global priority of environmental suitability,  $m$  indicates the number of habitat suitability variables,  $a_i$  describes the relative importance of factor  $i$ , and  $p_i$  gives the relative suitability of the factor  $i$ . Due to the lack of appropriate reference data to calibrate the model, the relative importance of factor  $a_i$  was weighted to  $1/m$ .

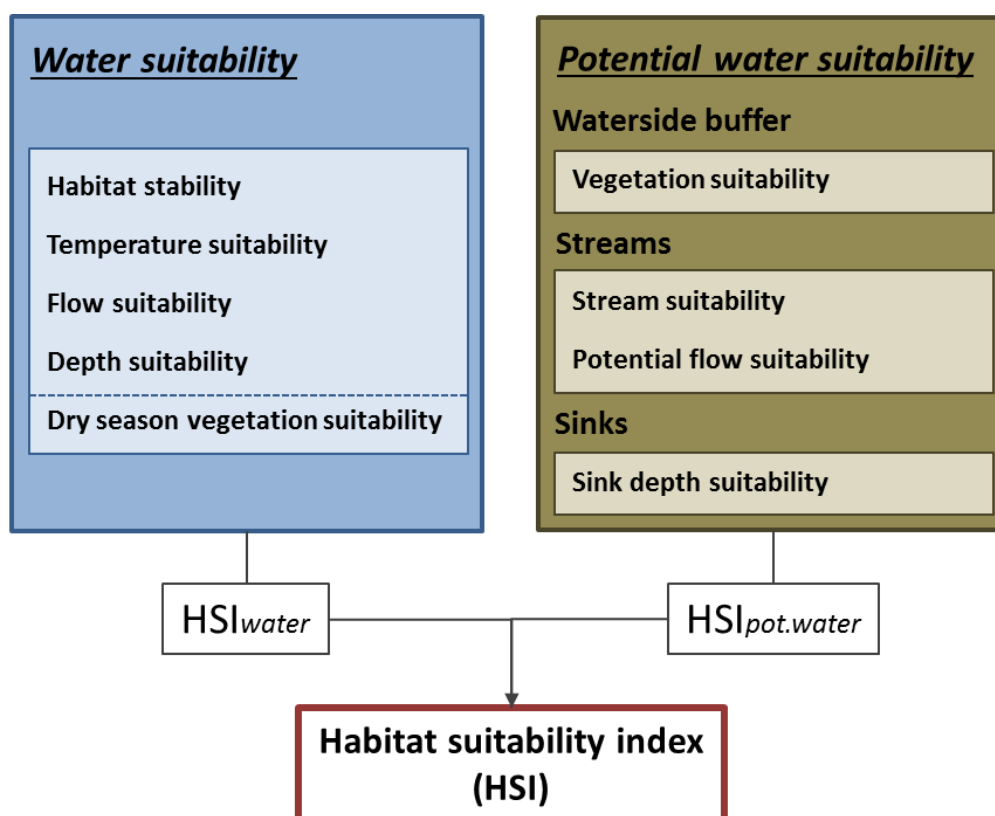


Figure 5-10: Composition of habitat suitability variables to model environmental suitability for schistosomiasis transmission. Water and potential water represent the major decision criteria to spatially delineate potential disease transmission, whereas water and potential water do not superimpose spatially in the model. The  $HSI_{water}$  integrates the upper four water-related suitabilities for the permanent water sites and all five variables for the seasonal water sites as indicated by the blue boxes. Additionally, the  $HSI_{pot.water}$  has been separately calculated for the three groups that do not spatially superimpose, namely the waterside buffer, streams and sinks as illustrated by the boxes.

There is one essential differentiation between the model composition approach illustrated in Figure 5-10 and other GIS-based HSI models presented in the literature introduced above. This is that single habitat variables are not all spatially superimposing. This phenomenon is a consequence of analysing water surfaces, topography, and vegetation coverage, which can predominantly be measured in a spatially exclusive way by means of RS. Vegetation coverage and topography cannot be measured in submerged areas and water characteristics can only be measured at sites where water is detected. Thus, the HSI has been derived separately for spatially superimposing components to capture the relevant suitability at the respective location, which is demonstrated by the hierarchical structure of the model composition in Figure

5-10. Water-related HSI values ( $HSI_{water}$ ) were derived for distinct spatial units of water, namely permanent water and seasonal water, whereas potential water-related HSI values ( $HSI_{pot.water}$ ) were derived separately for the waterside buffer zone, topographic drainage lines (i.e. streams), and sinks. These distinct spatial units were separated during the process of model composition with the aim to highlight the relevant information per site. To give an example, the disaggregated calculation of the  $HSI_{water}$  for zones of permanent and seasonal water allowed for the recognition of seasonally flooded and vegetated zones of high environmental suitability. If spatial composition of the HSI were not be separated in this case, the overall HSI of water would disproportionately decrease due to vegetation suitability around 0 within water bodies measured by RS data. The habitat variable of flow suitability contributes both to the water and the potential water section of the model as the meaning of the measurement is slightly different. The suitability of water flow velocity was directly derived from the slope measured at the water surface, and hence corresponds to the calculation of the flow velocity based on Manning's velocity equation (Equation 5-7), whereas potential flow velocity represents the slope measurement of the terrain measured within topographic drainage lines and is assumed to correspond with the course of (dried) river beds. The overall HSI value for a study site was calculated from the juxtaposition of the  $HSI_{water}$  and  $HSI_{pot.water}$  derived for the defined spatial units.

### 5.1.3 Model validation

The mechanistic model of environmental suitability and RS habitat variables were analysed for plausibility with respect to field reference data of potential schistosomiasis transmission sites visited in Burkina Faso in March 2011. According to the biotope classification of intermediate host snails in Burkina Faso (Poda, 1996: 34-37), seven habitat types (Table 5-1) were selected and identified in the field for this study. These are reservoirs with permanent water or seasonal water, irrigated rice fields, other irrigated crops, rivers with permanent water, seasonally dry river beds as well as topographic sinks. Site specific characteristics were measured or observed in the field (Section 4.3) and most relevant characteristics of selected test sites are listed in Table 5-1. The site specific estimate of environmental suitability is based on knowledge of habitat preferences of *S. haematobium* and *S. mansoni* parasites and *Bulinus* and *Biomphalaria* snail species (Section 3.1.2) together with at-site field measurements as well as scientific findings from Prof. Poda (Poda et al., 1994; Poda, 1996; Poda et al., 1996; Poda et al., 2001; Poda et al., 2004).

Table 5-1: Typical habitat types of schistosomiasis transmission in Burkina Faso. Overview of site-specific results measured or estimated from observations in the study site Ziniaré in Burkina Faso (see Figure 2-1). The RS category refers to the deterministic variable of water or potential water as measured by means of RS data. NA indicates that no measure could be provided.

	Permanent water body	Rice field	Irrigated crops	River water	Seasonal water body	Dry river bed	Dry sink
<b>RS category</b>	Water (permanent)	Pot water	Pot water	Water (permanent)	Water (seasonal)	Pot water	Pot water
<b>Vegetation coverage</b>	NA	75%	60%	NA	30%	20%	20%
<b>Flow velocity</b>	Stagnant	Stagnant	Slowly flowing	Stagnant	NA	(Fluent: erosive river bed)	NA
<b>Water temperature</b>	26.2°C	28.5°C	32°C	31°C	NA	NA	NA
<b>Suitability</b>	High	High	Moderate	Moderate	Moderate	Low	Low

High environmental suitability for schistosomiasis transmission was attributed to dam lakes with permanent water and rice fields. Both sites showed stagnant water within a suitable temperature range. Irrigated rice fields contain a large amount of vegetation, which additionally favours a water site for freshwater snails. Both site types were observed to be accessed by the local population without skin protection measures. Moderate environmental suitability was attributed to irrigated crops, permanent river water, and seasonal dry dam lakes. Lower environmental suitability of irrigated crops in contrast to rice fields was explained by the nature of irrigation practice of crops other than rice. Whilst most rice crops need to be flooded by persistent water, most other crops were observed to be irrigated through regular flooding of man-made dike systems within the field. Hence, environmental suitability is lower in habitats that dry out or where water is only temporarily flowing, which is in line with malacological studies provided by Abdel-Malek (1958: 792). Additionally, the measured water temperature of 32°C (see Table 5-1) is less favourable for intermediate host snail species of the genus *Bulinus* and *Biomphalaria*. Despite the fact that the observed river water appeared to be permanent and stagnant during the dry season due to cut-offs from dried-out sections of the river bed, this site was categorised as moderately suitable. Warm temperatures of the water as well as expected high flow velocities during the rainy season, demonstrated by strong erosive environments closely ahead of this river section, reduced the suitability of this permanent water site. Dam lakes that dried out during the dry season were categorised as moderately suitable as the rate of parasite development and potential human contact were reduced through aestivation. Dried river beds and dried topographic sinks were classified with low environmental suitability for schistosomiasis transmission. Both sites were only temporarily covered by water and the erosive dried river bed indicated high water flow velocity during the rainy season, which would be unsuitable for snail and parasite proliferation. However, natural and man-made cavities were observed within the river beds, where water becomes stagnant and useable during the dry period. These seven field sites were mapped on the dry season RapidEye image from the study sub-site around Ziniaré. The field-based judgement of high, moderate, and low suitability of selected test sites were classified based on the following classification scheme: HSI values greater than 0.66 correspond to high environmental suitability, HSI values between 0.33 and 0.66 represent moderate environmental suitability, and HSI values below 0.33 refer to low

environmental suitability. Areas of no environmental suitability for schistosomiasis transmission correspond to regions that are not covered by the deterministic habitat variable of water or potential water as explained in Section 5.1.1.

Additionally, the modelled environmental suitability was related to schistosomiasis prevalence based on the assumption that high environmental suitability within a catchment area of a school location, would be plausible if the prevalence measured at the school location was high, and *vice versa*. It was the aim to elaborate whether this hypothesis can be confirmed and whether this locally developed model can be transferred to the study site BUF and into different ecozones of MAN and TAB. In order to link modelled environmental suitability to the measured prevalence of schistosomiasis infection, the information of available epidemiological data on school prevalence (Section 4.1) was spatially extended by a circular buffer with a radius of 5km (Kabateraine et al., 2004: 377; Steinmann et al., 2006: 413). Within this buffer region, mean values of habitat variable suitability and the composite HSI were extracted for each school catchment area. The resulting mean suitability was then related to the measured prevalence at the respective school location using the Spearman rank correlation coefficient (Boslaugh and Watters, 2008: 183-184). Despite field data of parasite- and snail-related fitness not being available for this study to directly validate the HSI, disease prevalence provided a highly useful reference to validate environmental suitability as prevalence documents the outcome of the disease transmission process in the environment.

## 5.2 Results of the mechanistic model approach

The mechanistic model of environmental suitability is composed of the suitability of selected habitat variables derived from RS data, which are presented in Section 5.2.1. The results of a plausibility analysis to evaluate the quantitative prediction of environmental suitability within potential schistosomiasis transmission sites are given in Section 5.2.2. Finally, the regional transferability of the model approach established in the study sub-site Ziniaré to different eco-geographic regions in Côte d'Ivoire is presented in Section 5.2.3.

### 5.2.1 Habitat variable suitability

This section describes the resulting suitability of each habitat variable for the water sites, potential water sites, and the derived composite HSI for the model development site of Ziniaré. These results are illustrated in Figure 5-11 according to the model framework shown in Figure 5-10.

With respect to the water sites, the environmental variable of habitat stability resulted in highest suitability, where water was detected during the wet and the dry season representing the permanent water bodies. In contrast, temporary water bodies, which dried out during the dry season, resulted in moderate suitability. The study site showed multiple permanent and seasonal water sites, which fully dried out during the dry season. Water temperature suitability represented the mean temperature suitability for *S. haematobium* and *S. mansoni* parasites and *Bulinus* and *Biomphalaria* snail species calculated from the dry and wet season RS measurements. Most water bodies showed generally moderate to high temperature suitability for parasites and snails, whereas the riparian regions along the water bodies resulted in moderate to low temperature suitability. The relative suitability of water flow velocity showed

that within several small-scale water sites, the suitability of flow velocity appeared to be heterogeneous ranging from very low to very high suitability. However, one would expect stagnant or very slow moving water due to its topographic constitution as dam lake. Water depth suitability showed highest suitability at the boundary of permanent watersides and relative suitability of dry season vegetation coverage highlights areas of very high suitability, where dense vegetation covers a ground that was flooded during the rainy season. The very high suitability of dry season vegetation coverage corresponds mainly to irrigated agriculture, as can be seen on the dry season RapidEye image and was visible during the field survey at visited spots.

Within the potential water sites, the relative suitability of riparian vegetation coverage represented the mean vegetation coverage measured at the dry season and wet season within a 200 m buffer zone around water bodies. The resulting suitability highlighted areas of irrigated agriculture and dense vegetation coverage as highly suitable, whereas highest suitability was reached when dry season and wet season vegetation coverage were high and irrigation was possible throughout the year. Streams resulted in high suitability for the inflow and outflow of the great dam lake in the centre of the study site and low suitability for tributary waters, which was modified by the overlaying potential flow velocity as visible in the the respective image in Figure 5-11. The course of topographically derived streams often did not superimpose with the actual course of river beds, which could be seen in the overlay of derived streams and the RapidEye image. The sink habitat variable resulted preliminary in very low suitability due to its flat character and deeper sinks were masked due to its coverage with water.

Based on the aforementioned habitat variables, the HSI for potential schistosomiasis transmission was calculated and is illustrated in Figure 5-11. It showed a general discrimination between moderate to high suitability in and around water bodies and low suitability at topographic sinks. Individual water sites performed with variable suitability depending on the location in or around the water site. The zoomed-in part of the image in Figure 5-11 shows that high suitability was given in areas at densely vegetated sites in the buffer zone of the water site and in certain areas of the water body, where especially water flow velocity was low. Moderate suitability referred to the littoral zone of permanent water levels, moderately vegetated buffer zones, and certain sectors of the topographic drainage line, where potential flow velocity resulted in high suitability. Low suitability resulted predominantly from topographic sinks as well as the low vegetated region of seasonal flooded land as well as sparsely vegetated regions in the buffer zone around the water site. Especially densely vegetated zones around permanent water sites resulted in high environmental suitability.



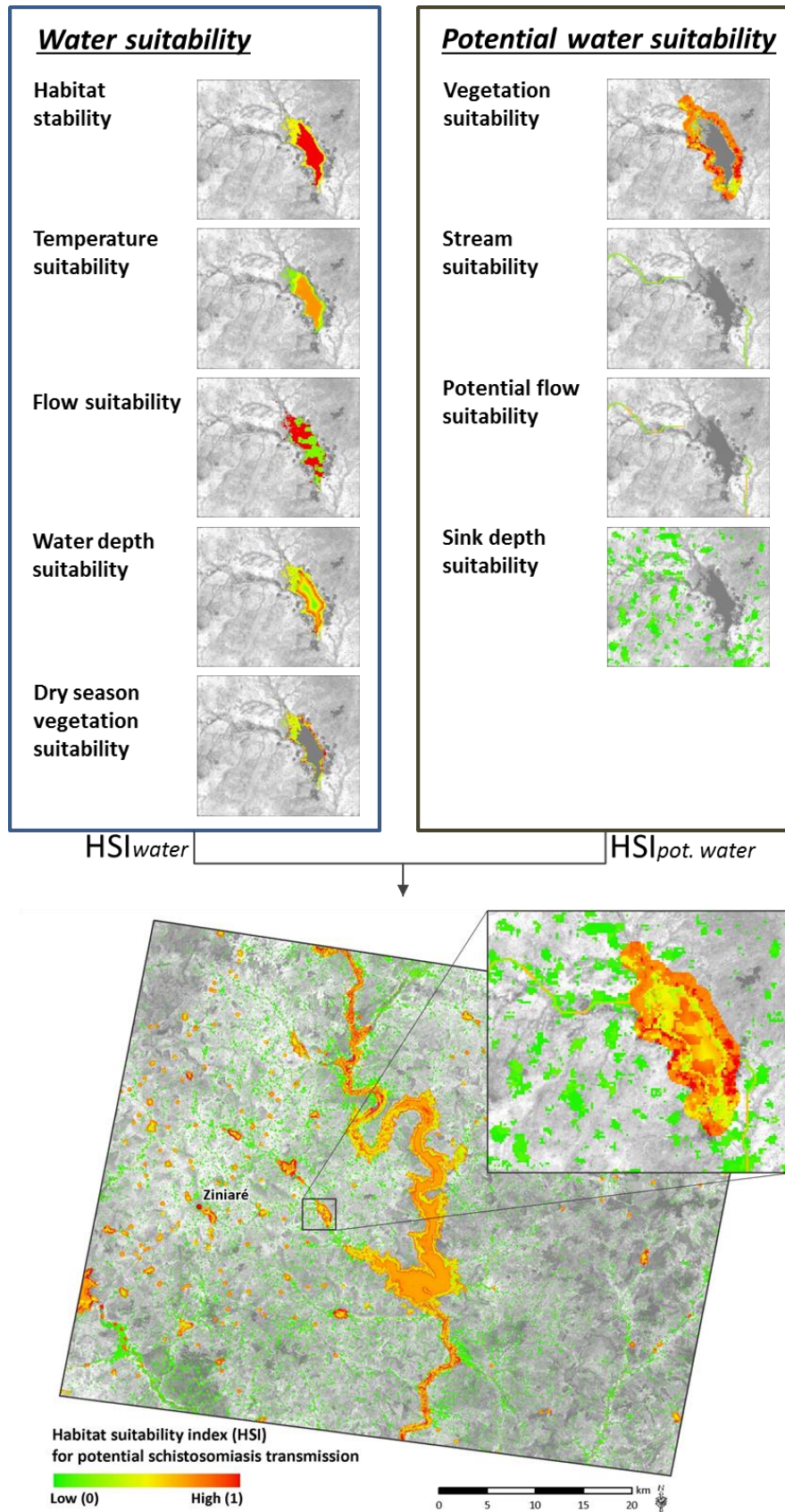





Figure 5-11: Overview of single habitat variable suitability and result of a mechanistic model of environmental suitability for schistosomiasis transmission at the sub-site of Ziniaré in Burkina Faso for the year 2010, based on which the model was developed. A map showing the location of the Ziniaré sub-site in West Africa is given in Figure 2-1.

## 5.2.2 Plausibility of modelled environmental suitability

The modelled environmental suitability was validated with reference to field sites that were visited in March 2011. Table 5-2 and Table 5-3 summarise the resulting suitability of habitat variables and the HSI for water and potential water sites, respectively. The site-specific expert judgement (Section 5.1.3) provided the basis to evaluate modelled environmental suitability within the test site of Ziniaré. The permanent water at the dam lake and the dried river bed sites (Table 5-1) covered more than one pixel of the 30 m resolution suitability image and therefore represented mean values of environmental suitability.

*Table 5-2: Modelled suitability of single habitat variables and the HSI in comparison to field-based expert judgement for water sites.*

*HS = suitability of habitat stability; TS = water temperature suitability; FS = water flow suitability; DS = water depth suitability; dVS = dry season riparian vegetation suitability*





	HS	TS	FS	DS	dVS	HSI	Expert judgement
<b>Permanent dam lake</b> 	1	0.41	0.2	0.81	NA	0.6 (Mod)	High
<b>Permanent river water</b> 	1	0	0.19	1	NA	0.55 (Mod)	Mod
<b>Seasonal dam lake</b> 	0.5	0	0.20	0.5	0.42	0.32 (Low)	Mod

With respect to the water sites, environmental suitability measured at the permanent waterside of a dam lake had a HSI of 0.6. This was composited by the suitability based on habitat stability (i.e. 1), temperature suitability (i.e. 0.41) that resulted from the mean temperature measured by RS data in the dry season (i.e. 30°C) and wet seasons (i.e. 35°C), flow suitability (i.e. 0.2), and water depth suitability (i.e. 0.81). The expert judgement of environmental suitability at the permanent dam lake site was high suitability, which corresponded to a HSI ranging between 0.67 and 1 and did not meet the modelled HSI at this site. The permanent river water test site was correctly detected by the water mask and resulted in a habitat stability of 1 corresponding to permanent water coverage. The temperature suitability provided by the RS

measurement was above the 38°C threshold resulting in suitability of 0. Flow suitability was 0.19, however, the field estimation of flow velocity at this site was stagnant during the dry season, whereas the erosive environment in the field indicated high flow velocity during the wet season. The water depth suitability of 1 reflected well the sub-pixel extent of this river section. A composite HSI of 0.55 resulted in conformity with the estimated moderate environmental suitability at this test site. The dam lake, which dried out during the dry season and was sparsely vegetated was correctly captured by the water mask with a habitat stability of 0.5. The overall composite HSI at this test site has been calculated from all five water-related habitat variables and resulted in an HSI of 0.32, which lies below the expert judgement derived at the field site.

*Table 5-3: Modelled suitability of single habitat variables and the HSI in comparison to field-based expert judgement for potential water sites.*

*bVS = mean vegetation suitability within 200m buffer zone of water; StS: stream suitability; SiS = sink suitability; pFS = potential water flow suitability in streams*

	<b>bVS</b>	<b>StS</b>	<b>SiS</b>	<b>pFS</b>	<b>HSI</b>	<b>Expert judgement</b>
<b>Rice field</b> 	0.97	NA	NA	NA	0.97 (High)	High
<b>Irrigated crops</b> 	0.64	NA	NA	NA	0.64 (Mod)	Mod
<b>Dry river bed</b> 	0.55	NA	NA	NA	0.55 (Mod)	Low
<b>Dry topographic sink</b> 	NA	NA	0.08	NA	0.08 (Low)	Low

The HSI of potential water sites resulted mainly from the suitability of mean vegetation coverage within a 200 m buffer zone of water except for the detection of the dried topographic sink test site. The rice field and irrigated crop test sites were both located within a 200 m buffer zone around detected water sites. The mean vegetation suitability of the rice field was 0.97, which was retrieved from NDVI values between 0.3 (dry season) and 0.6 (wet season), and 0.64 at the irrigated crop site that resulted from NDVI values between 0.2 (dry season) and 0.3 (wet season). Based on the suitability of vegetation coverage, both HSI values well reflected the high and moderate suitability that resulted from the expert judgement for rice and irrigated crop sites, respectively. However, the estimated environmental suitability of rice and irrigated crop sites resulted mainly from the irrigation practice, which varies between a permanently flooded rice field and crops that were irrigated by temporary flooding. This measure was not captured by RS data in this case. The low environmental suitability estimated for the dried topographic sink corresponded well to the measured HSI of 0.08. However, the dried river bed was only captured as a potential water site due to its position within a 200 m buffer zone of water and resulted a HSI value of 0.55 as a consequence of vegetation coverage at the side of the river bed. The HSI measure at this site would have been expected to result from stream and potential flow suitability. However, the topographic drainage lines did not cover this dried river bed.

### 5.2.3 Model transferability

The established model of environmental suitability was transferred to the study sites BUF (Figure 5-12), MAN (Figure 5-13) and TAB (Figure 5-14) with the objective to investigate whether model composition as illustrated in Figure 5-10 would remain reasonable in different regions and ecological settings based on visual inspection in reference to high-resolution RapidEye data. At the same time, the linkage between modelled environmental suitability and school based measures of schistosomiasis prevalence was statistically tested for each study site.

The modelled environmental suitability for the study site BUF is illustrated in Figure 5-12. This study site covers the training site around Ziniaré and based on visual inspection results for this wider region were comparable. As already seen in the HSI of the Ziniaré sub-site (Figure 5-11), seasonal and permanent water sites appeared with moderate to high environmental suitability, whereas potential water accumulation due to topographic features resulted in moderate to low suitability. Vegetation coverage within a 200 m buffer zone around water sites reflected well the distribution of irrigated agriculture with respect to dam lakes, which are widespread in this area. Water in rivers was not captured very well by RS data, when river beds appeared narrow or vegetation covered part of them. However, the course of the drainage system was very well reflected by topographic sinks in this study site. Although environmental suitability appeared to be reasonable as evaluated for the test site Ziniaré, the rank correlation analysis resulted in no significant correlation between the level of *Schistosoma* prevalence and environmental suitability in the respective school catchment (Table 5-4). It is demonstrated by the zoomed area in Figure 5-12 that both very low and very high prevalence measured at neighbouring school locations were located in close proximity to potential transmission sites of high environmental suitability in its catchment area. Hence, already based on visual inspection, it can be seen that other factors besides environmental suitability as reviewed in Section 3.1.2 would explain the spatial heterogeneous distribution of disease prevalence in the study site BUF.

Table 5-4: Spearman rank correlation coefficients for human *Schistosoma* prevalence and modelled environmental suitability and its corresponding confidence intervals given in brackets. Therefore, single variable suitability and composite HSI mean values were extracted from a 5 km buffer zone around the measured prevalence.

	BUF	MAN	TAB
Suitability of habitat stability (HS)	-0.14 (-0.36, 0.09)	<b>0.50* (0.30, 0.65)</b>	0.32 (-0.00, 0.58)
Water temperature suitability (TS)	-0.08 (-0.30, 0.15)	<b>0.47* (0.27, 0.63)</b>	-
Water flow suitability (FS)	0.04 (-0.19, 0.27)	<b>0.42* (0.21, 0.59)</b>	0.30 (-0.17, 0.57)
Water depth suitability (DS)	-0.23 (-0.44, -0.01)	<b>0.54* (0.36, 0.68)</b>	-0.15 (-0.45, 0.18)
Dry season riparian vegetation suitability (dVS)	0.02 (-0.21, 0.25)	-	-
Mean vegetation suitability within 200 m buffer zone of water (bVS)	0.22 (0.00, 0.43)	<b>0.47* (0.27, 0.63)</b>	0.05 (-0.27, 0.36)
Stream suitability (StS)	0.09 (-0.14, 0.32)	<b>0.50* (0.30, 0.65)</b>	0.25 (-0.08, 0.53)
Sink suitability (SiS)	-0.12 (-0.34, 0.11)	0.20 (-0.03, 0.41)	-0.10. (-0.41, 0.23)
Potential water flow suitability in streams (pFS)	-0.09 (-0.31, 0.14)	<b>0.42* (0.21, 0.59)</b>	0.30 (-0.02, 0.56)
<b>HSI</b>	-0.12 (-0.34, 0.11)	<b>0.45* (0.25, 0.62)</b>	<b>0.57* (0.31, 0.75)</b>

\*p < 0.01

The modelled environmental suitability of the study site MAN is illustrated in Figure 5-13. Based on visual inspection, the derived environmental suitability did perform reasonable discrimination between low suitability derived in the mountainous region in the northern part of the study site and the high suitability in the lowland of the southern part. Areas of high suitability for disease transmission were predominantly represented by the course of the river, whereas the waterside buffer zone was largely covered by forest, which does not correspond to high environmental suitability for potential schistosomiasis transmission due to irrigated agricultural areas. However, a significant correlation to prevalence of schistosomiasis was found for all habitat variables except for sink suitability and for the HSI.

The resulting environmental suitability for schistosomiasis transmission in TAB (Figure 5-14) showed moderate to high environmental suitability for the major river crossing this study site, the waterside of its intersecting Lake Taabo as well as river tributaries and very small water sites detected predominantly in the northern part of the study site. However, no seasonal water could be detected for this site because RS data from the wet season were not available. Furthermore, high vegetation coverage within a 200 m buffer zone of water sites did not correspond to irrigated agriculture in this region, but reflected high tree coverage, which might have misled the classification into high environmental suitability for schistosomiasis

transmission. However, the result of the statistical evaluation at this site resulted in a significant Spearman rank correlation coefficient of 0.57 between the composite HSI value and the school based prevalence (Table 5-4). This positive relation is confirmed by the spatial distribution of high prevalence rates in close proximity to the Bandama River and Lake Taabo, whereas low prevalence rates are predominantly distributed further away from these hotspots of environmental suitability with few exceptions close to the river outflow of Lake Taabo.

## 5.3 Discussion of remotely sensed environmental suitability

This research has shown that RS data can provide spatial information of environmental suitability for potential schistosomiasis transmission. The validation of derived suitability of a habitat variable in relation to measurements and observations in the field has shown that the agreement between measured suitability and the respective field derived expert judgment varied between remotely sensed variables and test sites. This validation procedure aimed to identify strengths and weaknesses of RS variables to assess environmental suitability for potential schistosomiasis transmission. However, a comprehensive approach to validate the environmental variables and suitability would need a representative set of field data on parasite and snail prevalence and the corresponding environmental metrics at respective sites.

### 5.3.1 Remote sensing derived biophysical variables

The RapidEye and Landsat 5 TM sensor provided useful data sources to detect the small-scale heterogeneity of **water bodies** in the Ziniaré, TAB, and BUF study sites, respectively. For the study site MAN, the spatial resolution of 6.5 m was still too coarse as water bodies predominantly consisted of small reservoirs and rivers, which were furthermore covered by large trees (see Section 4.2.2). Despite the fact that there were only two useful points in time of RapidEye data available for this study, the RapidEye sensor was well designed to capture data on a near-daily basis (see Section 4.2.1), which would allow for the monitoring of the critical limit of water persistence between 4 to 6 weeks for *S. mansoni* and 5 to 7 weeks for *S. haematobium* (Sturrock, 1993a: 65). Despite the theoretical capacity to detect this seasonal dynamic of water bodies and rivers, cloud coverage strongly impacted optical RS data especially in the southern region of Côte d'Ivoire. In this case, active RS based on radar data, which are not limited by cloud coverage, could provide a useful alternative.

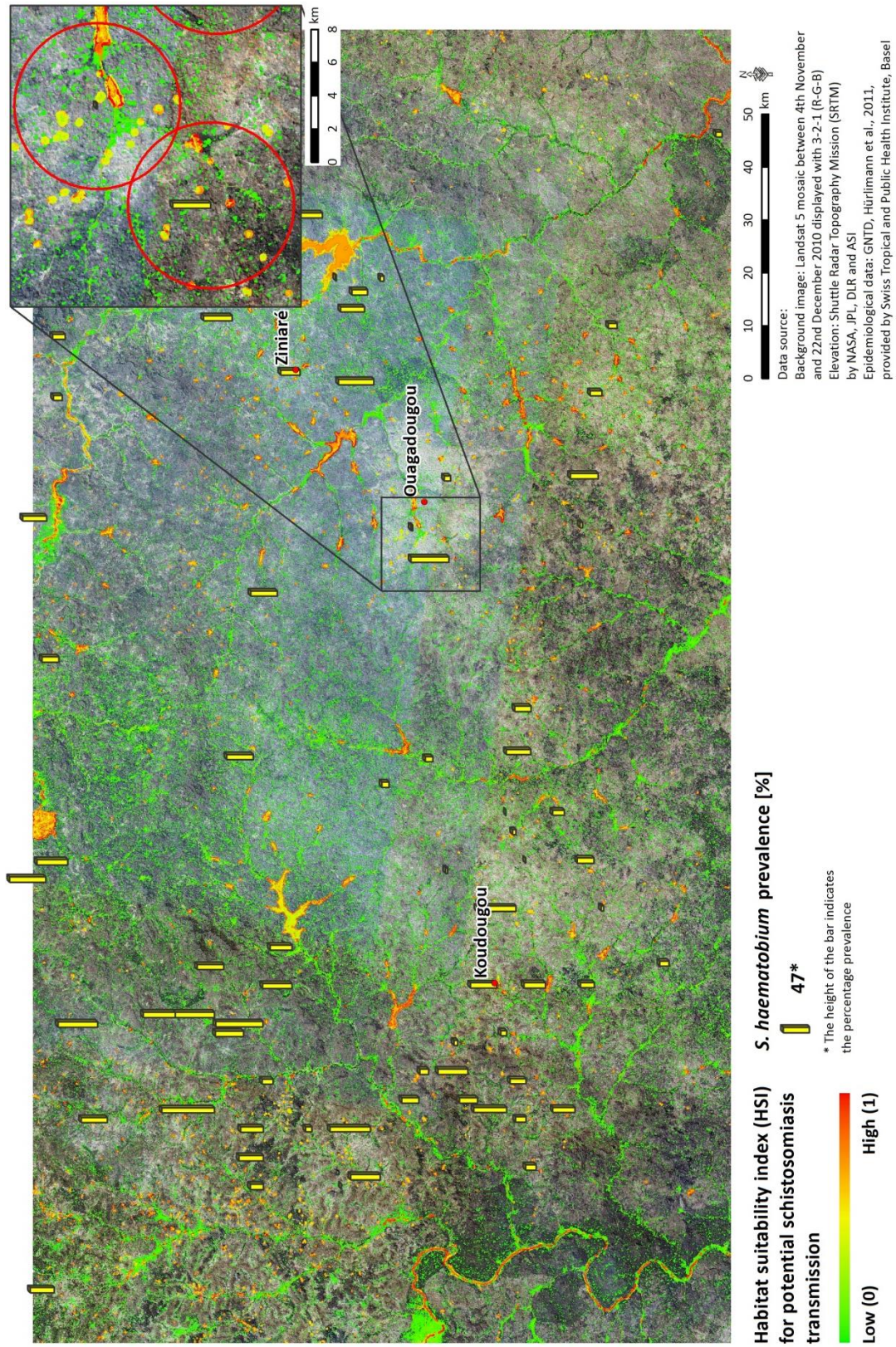


Figure 5-12: Environmental suitability for potential schistosomiasis transmission in the study site BUF, Burkina Faso. A map showing the location of the study site BUF in West Africa is given in Figure 2-1.

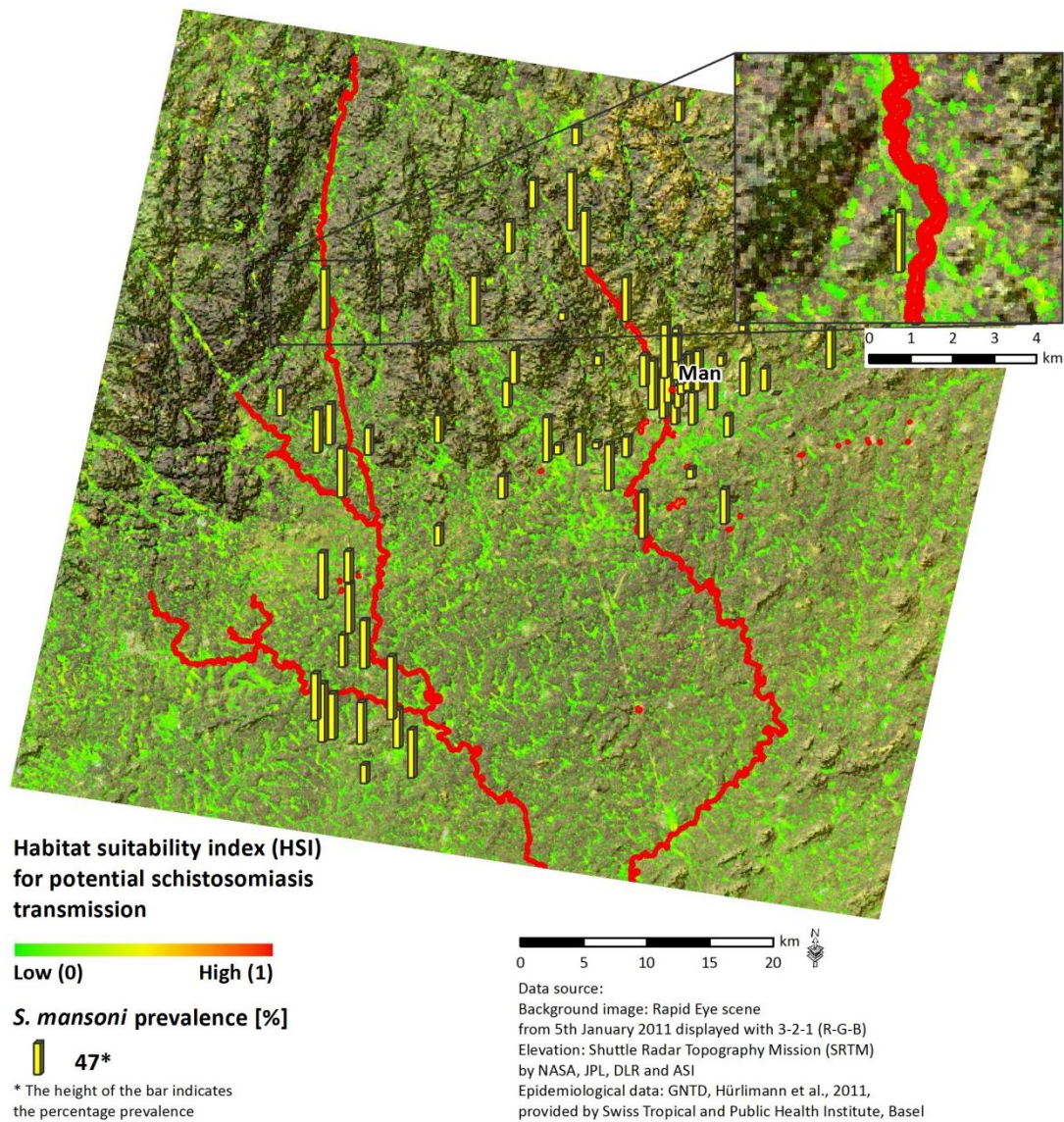


Figure 5-13: Environmental suitability for potential schistosomiasis transmission in the study site MAN, Côte d'Ivoire. A map showing the location of the study site MAN in West Africa is given in Figure 2-1.



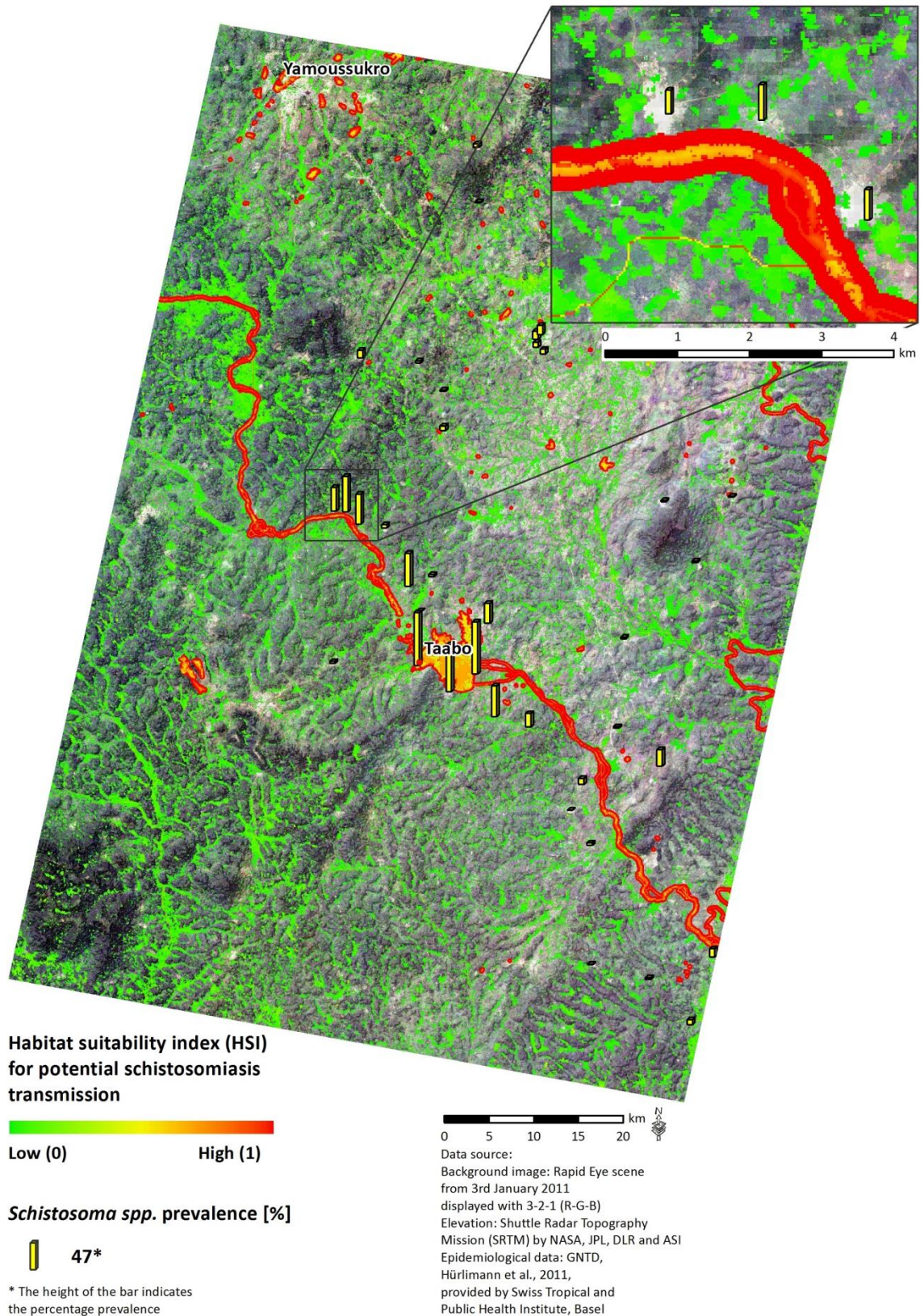


Figure 5-14: Environmental suitability for potential schistosomiasis transmission in the study site TAB, Côte d'Ivoire. A map showing the location of the study site TAB in West Africa is given in Figure 2-1.

Water **temperature** is considered an important habitat variable from an ecological point of view as temperature is critical for both snail and parasite development as reviewed in Section 3.1.2. However, the measure of water surface temperature provided only estimates of water temperature that directly impacts snail and parasite development and reproduction. Especially within isolated and shallow water bodies, extreme temperature variations were detected by Fisher and Mustard (2004: 293). Field-based measurements at the waterside were taken in March 2011, whereas the RS data employed for this study were recorded in February 2010. This might be a further cause of deviating temperatures between the field and RS measurements. However, the measurement deviation in the range of 10°C at the permanent river water supports the assumed impact of the inappropriate spatial resolution of the remotely sensed temperature measurement at this site. Here, the 120 m spatial resolution of the thermal Landsat 5 TM did not capture the linear structure of the river with its transect not exceeding 10m and resulted in mixed pixel information of water and land surface temperature. This results in incorrect surface temperature measurements due to the calibration of surface emissivity for pure water. At this scale, airborne thermal RS data would be more useful to derive water temperature of rivers and streams as shown by Torgersen et al. (2001). Nevertheless, the thermal band of the Landsat 5 TM sensor was evaluated, whereupon an offset error of 0.092 W/m<sup>2</sup> (approximately 0.68 K) was corrected by modifying a calibration coefficient in the processing system post 2007 (Barsi et al., 2007: 552), which was relevant for the pure water pixel at large dam lakes.

Most optical RS data capture visible and near infrared spectral reflectance and are therefore well designed to measure **vegetation coverage**. In this study, vegetation was reasonably detected by high-resolution RapidEye data in comparison to field-based estimates of vegetation coverage for the rice field and the irrigated crops. However, vegetation coverage also plays a crucial role for characterising a potential schistosomiasis transmission site in the subaquatic area, as submerged vegetation conditions dissolved oxygen content of the water body and thereby influences snail activity and reproduction (Section 3.1.2). Hyperspectral RS data have demonstrated the ability to detect submerged aquatic vegetation (William et al., 2003; Vahtmäe et al., 2006; Marcus and Fonstad, 2008). In this study, the measurement of submerged vegetation coverage was not feasible due to the lack of hyperspectral RS data and appropriate field data, which was also beyond the scope of this work.

The **topography** derived from the ASTER GDEM provided information on the topographic structure below the detected water level in all water bodies except for parts of the great dam lake in the study site Ziniaré. Therefore, the measurement of slope for water sites did not always correspond to the water surface as intended for this model. This was well documented by strong heterogeneities of the slope measure within a dam lake, which was evident to have a flat surface of stagnant or very slow moving water. Additionally, the temporal dynamic of flow velocity between dry season and wet season as seen for permanent river water sites could not be captured by the single acquisition of a DEM. However, suitability of flow velocity resulted reasonable for the large dam lake in the Ziniaré study site and was considered a useful proxy to highlight very flat zones within topographic streams. For large rivers, Kiel et al. (2006: 317-318), successfully derived water flow velocity using SRTM data. Topographically derived streams were very often not superimposing with the course of the actual river bed in high-resolution RS data, which was further documented at the dry river bed test site. However, topographic sinks

documented very well the course of river beds and could successfully detect a field measured sink.

### 5.3.2 Modelling environmental suitability

Environmental suitability for schistosomiasis transmission was modelled based on single habitat variables, which were parameterised by **theoretical functions of relative suitability**. These functions were derived from different background information, which is demonstrated by the following examples: habitat suitability related to water temperature was derived from laboratory-based measurements (Pflüger, 1980; Pflüger et al., 1984), whereas stream order suitability was characterised from field-based surveys (Clennon et al., 2007) or spatial analysis (Beck-Wörner et al., 2007). The parasite-related water temperature suitability function was directly provided in the literature, whereas snail-related water temperature suitability was interpolated from provided measurements of snail mortality at given temperatures. The thresholds that characterise the range and course of functions given in this study were either cited in the literature (e.g. habitat stability-related suitability), measured within high-resolution RS data (e.g. vegetation coverage) or directly estimated in the field (e.g. water depth). These mixed sources of information can impact the validity of resulting habitat variable suitability as it is not clear whether this function fits for the modelled location. Due to the lack of appropriate field data on prevalence or fitness of parasites and snails, the parameterisation of these functions could not be verified within this thesis. However, these functions reflect the state-of-the-art of available knowledge with respect to environmental suitability for schistosomiasis transmission and provide the basis for location specific evaluation. In the following paragraph, strengths and weaknesses of selected functions of relative suitability used in this study will be discussed.

The **water flow velocity based suitability** has been parameterised based on the given suitability threshold in relation to snail prevalence and the defined relation (Equation 5-8) between the RS measurement of slope and water flow velocity. However, this distinct relation is still exposed to error sources that need to be considered with respect to input data, namely the inconsistent RS measurement of slope at water sites as discussed in Section 5.3.1 as well as the rough estimation of several variables necessary to compute the Manning's equation (Equation 5-7) at only few locations in the field. To receive more accurate information on flow velocity, it would be necessary to derive these input parameters specifically at location, which is not feasible with the 30m resolution ASTER GDEM within the small scale terrain heterogeneity experienced in this study site. Nevertheless, the water flow velocity based suitability showed reasonable values for large stagnant water bodies such as the large dam lake in the Ziniaré study site and was considered a useful proxy to highlight very flat zones within topographic streams. The theoretical function to parameterise the **water depth-related suitability** was based on subjective thresholds, which were not validated with respect to its relation to water depth but were solely estimated during the field visit at the seasonally dried areas of a dam lake boundary. The thresholds of 210 m and 2 km are not considered appropriate for rivers in the study area, which have a steep gradient towards the main current line and are often narrow. Nevertheless, the distance from water level was considered a useful proxy with respect to dam lakes and provides reasonable information that suitability of schistosomiasis transmission is highest in the very flat littoral zones of a water body as indicated by the WHO (1957). Other studies found that

if water depth measurements were provided, spectral reflectance of airborne multispectral imagery (Gilvear et al., 1995; Winterbottom and Gilvear, 1997), Landsat 5 TM data (Bierwirth et al., 1993) or data from the IRS-LISS sensor (Kumar et al., 1997) were successfully used to estimate water depth based on regression models. The **theoretical function of topographic sink based suitability** represents the logical assumption that the deeper sinks are the longer water persists, which resulted in higher suitability for schistosomiasis transmission. This function derived from linear interpolation between minimum and maximum sink depth is certainly modified by the given sink depth thresholds within the area of interest. A validation of sink depth and its relation to water holding capacities would require additional field data considering precipitation and soil drainage. With respect to precipitation, rainfall measurements derived from the TRMM data were not considered useful for this study site due to the large scale mismatch between a 30 m resolution suitability image and the TRMM resolution of 28 km. This scale mismatch was similar to the soil characteristics provided by the FAO et al. (2012).

The single remotely sensed variables of environmental suitability were aggregated to a composite **HSI** and evaluated in relation to reference test sites scored into the classes low, moderate, and high environmental suitability. Despite the model composition was adapted to the prerequisites of schistosomiasis transmission specifically to the study site Ziniaré, it provides a transparent basis to reproduce and adjust the model. The implausible HSI at the seasonal water in dam lake and the dried river bed could be explained by the RS measurements of water temperature and topography with inappropriate spatial resolution as discussed in Section 5.3.1. However, a precise identification of key factors determining the suitability of any particular habitat is difficult especially from field-based analyses due to complex interactions (Sturrock, 1993a: 50). These interactions are considered to some extent through the aggregation of single environmental variables to an overall HSI. It has to be stated here that the modelled environmental suitability represents the fundamental ecological niche (Section 3.3.1) of *S. haematobium* and *S. mansoni* parasites and *Bulinus* and *Biomphalaria* snail species, which does not necessarily imply that species are abundant at an appropriate site. Parasites and snails can still be absent from apparently suitable habitats, because isolation of individual habitats and (re-) invasion are dictated by chance combinations of factors that permit snail dissemination (Sturrock, 1993a: 50).

### 5.3.3 Model transferability

The mechanistic model of environmental suitability related to schistosomiasis transmission was developed and composed specifically for the study site Ziniaré in Burkina Faso and was then transferred to the study sites BUF, TAB, and MAN to analyse the impact of different geographical settings. The topographic variables performed comparable results in all three study sites, where sinks represented the flat areas of potential water accumulation as potential disease transmission sites with low suitability. Streams traced the water drainage based on topography, however, in all three study sites the streams did not fully correspond with the course of the current river beds visible in high-resolution RS images. The specific landscape configuration of dam lakes and irrigated agriculture was highly relevant for potential schistosomiasis transmission in the BUF region, however, this was not given in the study sites TAB and MAN in Côte d'Ivoire. In the study site TAB there were dam lakes, but not with irrigated agricultural sites at its waterside and in the study site MAN there were hardly any dam lakes visible, but rivers

with dense forests at its watersides. Therefore, the 200 m buffer zone around watersides in Côte d'Ivoire was not considered to represent high environmental suitability given by the index in most areas, which hampered the direct transferability of the model between different geographical settings. However, the deterministic variable of water and potential water sites was well represented in all three study sites and provides a reasonable basis for locally specific analysis of environmental suitability with respect to schistosomiasis prevalence as provided in Chapter 6.

Evaluation of modelled environmental suitability in relation to *S. haematobium* and *S. mansoni* prevalence was based on the , that high environmental suitability within a catchment area of a school location would be plausible if the prevalence measured at the school location was high, and *vice versa*. It is clear that suitable environmental conditions provide the prerequisite that transmission of the disease may occur (Section 3.1.2). However, at the same time there are other factors that modify this relation between environmental suitability and schistosomiasis prevalence, such as local disease intervention measures (Clements et al., 2009b; Zhang et al., 2012), economic development (King, 2010), individual disease susceptibility (Butterworth, 1993; Jordan and Webbe, 1993) or human behaviour (Bundy and Blumenthal, 1990; Schmidlin et al., 2013), whether an infective habitat is entered or protective measures are applied or not. The lack of correlation between the HSI and schistosomiasis prevalence at the BUF site could additionally be related to the temporal gap between environmental suitability assessed for the year 2010 and school-based surveys that were conducted 20 to 30 years ago (see Section 4.1). The composite HSI was plausible in the study site TAB, which could be explained by the clear demarcation of highly suitable disease transmission sites in this region (Figure 5-14) and the non-existence of intervention measures prior to the surveys of the provided data in this region (Section 2.3). This study site exemplifies that aggregation of relevant information is superior to single variables per se. In contrast, in MAN both the single habitat variables and the composite HSI provided a significant plausible relation to the school-based schistosomiasis prevalence measures. Especially the topographic habitat variables performed reasonably in this mountainous study site and in consideration of the ecological requirements for disease transmission. While the plausibility of stream suitability confirmed the results derived by Beck-Wörner et al. (2007), water flow suitability, sinks, and water depth suitability appear to be new variables that could be useful to model environmental suitability and disease transmission risk.

A predictive map of environmental suitability mainly supports prevention and control measures in line with the shift from morbidity to transmission control (Utzinger et al., 2011a: 132). These maps could be useful to target specific schools for initial epidemiological surveillance or to focus on areas of high reinfection potential for more regular re-treatment (Stothard et al., 2002: 474; Brooker, 2007). Despite the numerous constraints and limitations underlying **mechanistic models of environmental suitability**, simple and even untested HSI models were continuously used for different kinds of decision making procedures (Brooks, 1997: 165). The fact that only few HSI models were tested and validated against field data (Duncan et al., 1995) could be explained by the difficulty of sampling appropriate species occurrence and behaviour data. However, despite available information coming from various sources such as field surveys and laboratory tests, the deductive approach applied in this study was considered the only suitable method to derive environmental suitability (Ottaviani et al., 2004). If species presence-

absence data had been available, they would not necessarily imply whether a habitat was suitable or not to establish a population. There is no standard approach to test HSI models as each model defines habitat in a slightly different manner and test data that are consistent with model content and purpose are hardly accessible (Schamberger and O'Neil, 1986: 7). HSI models were particularly affected by subjective judgement and model uncertainty, which occurs as a result of interpretation of data, especially when data are scarce and error-prone. This is particularly the case when expert knowledge is the primary means of informing an HSI and when experts estimate facts or classifications as provided in this research. Uncertainty occurs as a result of the simplification of real processes by models (Ray and Burgman, 2006). However, the model presented in this study can be easily reproduced and adjusted by other experts and in reference to new data and information. A comprehensive validation of this mechanistic model approach would consist of a direct linkage between derived environmental suitability and field data of parasite- and snail-related fitness, which were not available for this study.

## 5.4 Summary of environmental suitability model

In summary, RS data displayed the potential to spatially delineate and evaluate environmental suitability for transmission of schistosomiasis. The prerequisite that disease transmission may occur in the environment could be derived through water bodies and sites of potential water accumulation by means of RS. This delineated environment of potential schistosomiasis transmission could be further evaluated based on the HSI resulting in areas of divergent priorities. The mechanistic model has large components of subjective estimates, however it is transparent, easily reproducible, and can be well adjusted to new findings and data. Remotely sensed temperature and topographic variables did not perform appropriately due to their coarse spatial and temporal resolution. However, water and vegetation well reflected the environmental suitability and could be regularly updated to monitor changing conditions and newly emerging habitats based on this model. This could support an immediate reaction of public health authorities to target prevention measures. The composition of the model needs adjustments to regional landscape structures relevant for schistosomiasis transmission.

## 6 Modelling schistosomiasis risk

RS data have been widely used for spatial modelling of schistosomiasis risk in different geographical settings (Seto et al., 2002; Stensgaard et al., 2005; Yang et al., 2005a; Beck-Wörner et al., 2007; Brooker, 2007; Clennon et al., 2007; Vounatsou et al., 2009; Guimaraes et al., 2010; Schur et al., 2013). Thereby, survey measurements of disease prevalence provided the reference to model environmental data and predict schistosomiasis risk in space. Risk modelling provides a more comprehensive analysis than modelling environmental suitability alone, because the presence and successful completion of the parasite life cycle is implicit in disease prevalence data. However, as demonstrated by the conceptual framework of the social-ecological disease niche (Figure 3-6) and the RS based model of environmental suitability for schistosomiasis transmission (Chapter 5), the location of survey measurement at schools do not spatially superimpose with the RS measurement of schistosomiasis relevant environmental conditions (Figure 6-1). This **spatial divergence** has not been addressed in any of the existing spatial schistosomiasis risk models that statistically link disease prevalence with RS data.

Social-ecological processes of schistosomiasis transmission – similar to nearly all ecological phenomena – operate across different **scales** and vary with the scale of observation (Schur et al., 2012). Thus far, models of schistosomiasis risk have mainly been developed at regional and national scales (Simoonga et al., 2009: 1686) using low spatial resolution RS data (Table 3-2). On this scale of observation, it has been concluded that predominantly climatic conditions determine the spatial risk of schistosomiasis transmission (Brooker, 2007: 3). There are very few studies that modelled schistosomiasis risk on the local scale (Clennon et al., 2004; Raso et al., 2005; Beck-Wörner et al., 2007). These studies showed that besides the environment, socio-economic and demographic predictors were highly relevant in explaining the spatial heterogeneity of disease transmission on the micro-scale (Simoonga et al., 2009: 1686-1687). This underlies the fact that statistical correlation can vary dramatically according to the extent of observed area and scale of aggregation (Marshall, 1991: 431), which has been identified as a major constraint of RS based schistosomiasis risk models. Herbreteau et al. (2007: 401) emphasise the need to address diseases at different relevant scales and investigate the impact of RS image resolution (Herbreteau et al., 2007: 401). Especially high spatial resolution RS data are expected to be highly beneficial for the detection of heterogeneous habitat conditions of disease-related species on a local scale (Goetz et al., 2000: 303; Herbreteau et al., 2007: 402). Simoonga et al. (2009: 1689) concluded that the neglected scale issue “affected usefulness of

developed models and maps for reducing micro-scale transmission through improved resource targeting” (Herbreteau et al., 2007: 1689).

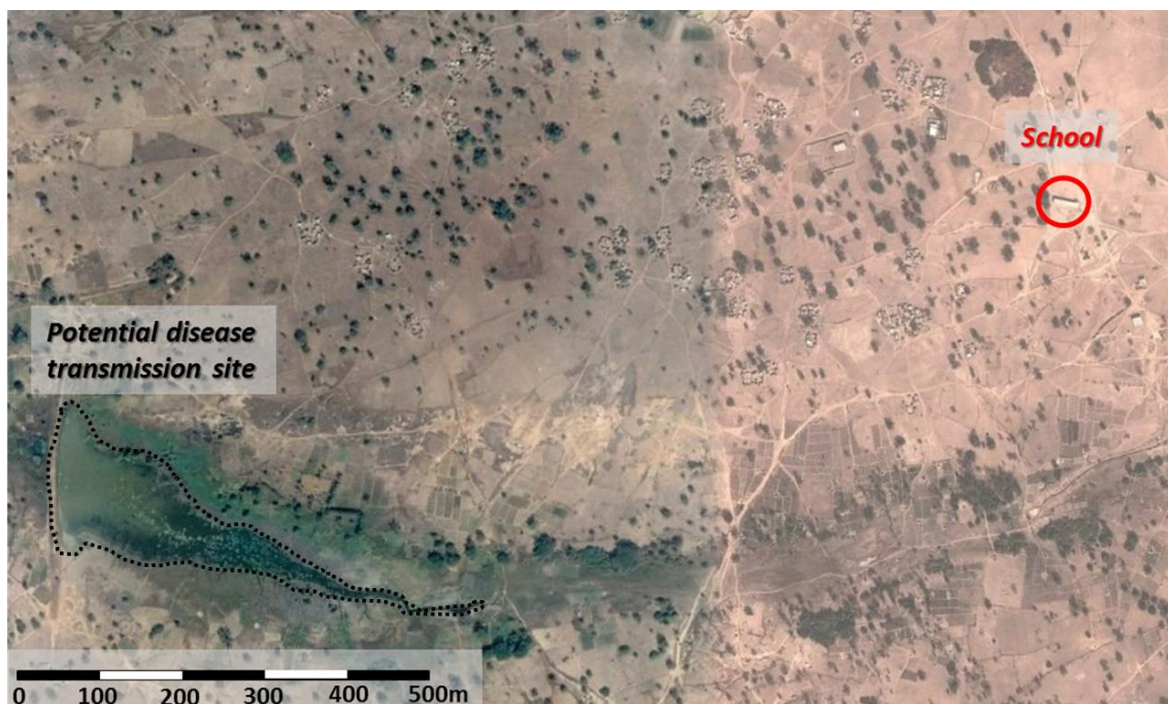


Figure 6-1: Illustration of the spatial discrepancy between the measurement of schistosomiasis prevalence at a school (surrounded in red) and the location where disease transmission has potentially occurred (water body surrounded by the black dotted line). Source: Google Earth (Image ©DigitalGlobe 2011)

**RS variables** used for schistosomiasis risk modelling so far were predominantly NDVI and LST (Table 3-2), mainly because these variables are pre-processed and readily available (Herbreteau et al., 2007: 401). However, further vegetation indices can be calculated from RS data, which are expected to be helpful for health studies (Herbreteau et al., 2007: 401). Furthermore, reviewing the ecology of disease transmission reveals that additional RS variables may improve models of schistosomiasis transmission (Table 3-3). The relevance of an RS variable for modelling schistosomiasis transmission risk is expected to vary between different landscapes and ecological regions, which could have an impact on outcome of the model and its **transferability** between different regions. Important aspects that impact RS 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. Figure 6-2 illustrates the difference in landscape structure covered by the study area, which ranges from savannah in the North to tropical rainforest in the South, all of which are endemic for schistosomiasis transmission. A study by Brooker et al. (2001: 1004) has shown that RS based schistosomiasis risk models performed only reasonably if modelled within the same ecological zone in the Republic of Tanzania.



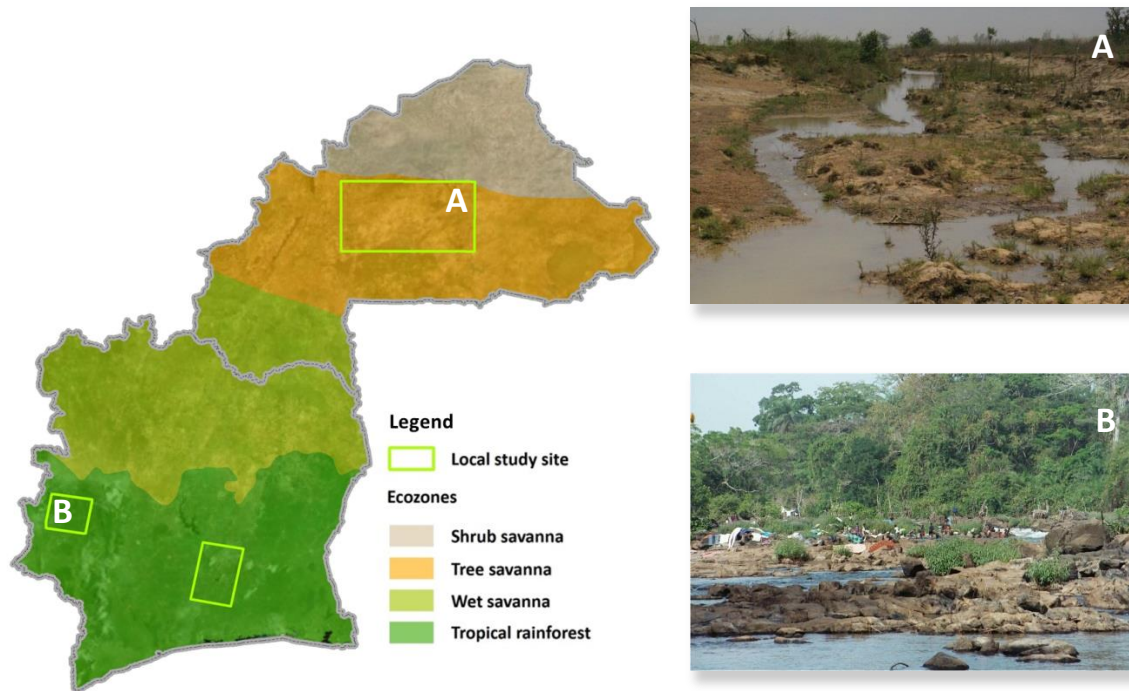


Figure 6-2: The impact of different ecological regions on RS based schistosomiasis risk models. The study area covers different ecological regions ranging from savannah to tropical rainforest. Picture B provided with courtesy of Dr. Giovanna Raso.

The **main objective** of this chapter is to investigate and quantify the potential of RS data for schistosomiasis risk modelling and spatial risk prediction. Therefore, the research gaps introduced above regarding: (i) the issue of scale; (ii) the spatial discrepancy of model components; (iii) the importance of specific RS variables; and (iv) the impact of ecological regions on model performance and transferability will be addressed in detail. The issue of scale is investigated through analysis of multi-scale RS data from RapidEye (6.5 m resolution), Landsat 5 TM (30 m resolution), and Terra MODIS (250 and 500 m resolution). Furthermore, an appropriate scale of observational units around school locations for RS based environmental analysis is investigated. To account for the spatial discrepancy between school-based disease prevalence and relevant environmental conditions, a hierarchical model approach has been developed. Relevant RS predictors were identified from a series of vegetation and water-related indices and other variables tailored to the disease transmission ecology. The impact of ecozones on variable importance and model performance is addressed by comparison of model results from study sites in savannah, tropical, and mountainous regions.

## 6.1 Statistical model algorithms

The data were analysed by means of two different statistical algorithms in order to achieve broader generalizability: the non-parametric Random Forest machine learning algorithm (Breiman, 2001) and the parametric partial least squares regression (PLSR) (Wold et al., 1984). Schistosomiasis prevalence (continuously scaled between 0 and 100%) was the response variable to be explained by predictor variables consisting of a multitude of RS reflectances, indices and other variables (Table 4-3). Both model approaches are by design able to deal with

the substantial collinearity between predictor variables and allow to identify which RS data and variables perform best for the application of schistosomiasis risk modelling. Due to the small extent of the study sites limited by the spatial coverage of available high-resolution data, the sample size was generally low. For the study site TAB it was even smaller than the number of predictor variables. These prerequisites required model algorithms capable of handling and predicting continuous response data, dealing with low numbers of training samples, and multicollinear predictor variables. The following Sections 6.1.1 and 6.1.2 explain the model algorithms of Random Forest and PLSR, respectively, and illustrate how these algorithms can deal with the above mentioned prerequisites.

### 6.1.1 Random forest

The statistical machine learning approach of random forests is an ensemble learning method that operates by constructing a multitude of decision trees. In contrast to conventional statistical procedures, machine learning algorithms learn the relation between predictor and response variables from the data and do not assume a specific statistical model, e.g. normality. While other non-parametric machine learning approaches such as neural networks or support vector machines are rather complex, tree-structured models are based on simple functions of the input variables (Sutton, 2005: 303). Decision trees have the capability to consider non-linear relations between the response and predictor variables and do not require reducing the feature space of predictor variables to a non-correlated data-set, as would be the case for conventional multinomial regression models. In addition, decision tree ensembles are highly robust with respect to outliers in the training data (Breiman et al., 1984: 55-58).

There are different approaches to construct decision trees, namely the Automatic Interaction Detection (AID) developed by Morgan and Sonquist (1963), the Theta-AID (THAID) developed by Morgan and Messenger (1973), and the Classification and Regression Trees (CART) developed by Breiman et al. (1984). CART is the most commonly used approach and represents the decision tree method implemented in Random Forests used in this study.

A single decision tree grows by partitioning the feature space with respect to the response variable, which results in increasingly homogeneous sub-spaces. Thereby, at each node of a tree branch the best binary split, meaning highest informative value with respect to the response variable, is selected (Figure 6-3). The main idea is to select a subset of the response variable training set that is “purer than the data in the parent subset” (Breiman et al., 1984: 23) with the aim of constructing a tree that estimates the response variable noted in the end leaves of the tree with a best set of predictor variables as given by the respective tree branch. In this logic, purity is indicated by the difference between the residual sum of squares (RSS) before and after the split (Hastie et al., 2009: 593). To find the optimal size of a tree, criteria such as the minimum amount of data per leaf can be defined to stop further splitting of a node. A single decision tree is unstable due to the hierarchical structure of the tree, where errors or changes in lower level branches propagate at all subsequent splits of the tree (Hastie et al., 2009: 588). It is thus the strength of ensemble regression trees such as random forests to combine constructed trees and thereby reduce this high variance between single trees to derive a stable model result.

The methodical family of random forests has in common that random processes are implemented to grow single trees and combine them to a forest model. Methodical approaches used in different implementations of random forests are for example the random subspace

method (Ho, 1998), the random split selection (Dietterich, 1999) or bagging and random input selection (Breiman, 1996, 2001). The latter approach from Breiman (2001), is most widespread and mainly termed Random Forest method in the literature (from now on referred to as Random Forest in capitals). In the Random Forest algorithm, the growth of a tree is based on a random subset selection of training data, called bagging (i.e. in bag, see Figure 6-3). This prevents the model from overfitting and provides an independent test data set (i.e. out of bag) for internal model evaluation or the measure of variable importance. Secondly, only a subset of predictor variables ( $=m_{try}$ ) is randomly selected at each node (see Section 6.2.4). Finally, the Random Forest ensemble combines the predictions of all single trees by calculating their mean prediction for regression or by majority voting for classification. The ensemble approach leads to reduced model variance.

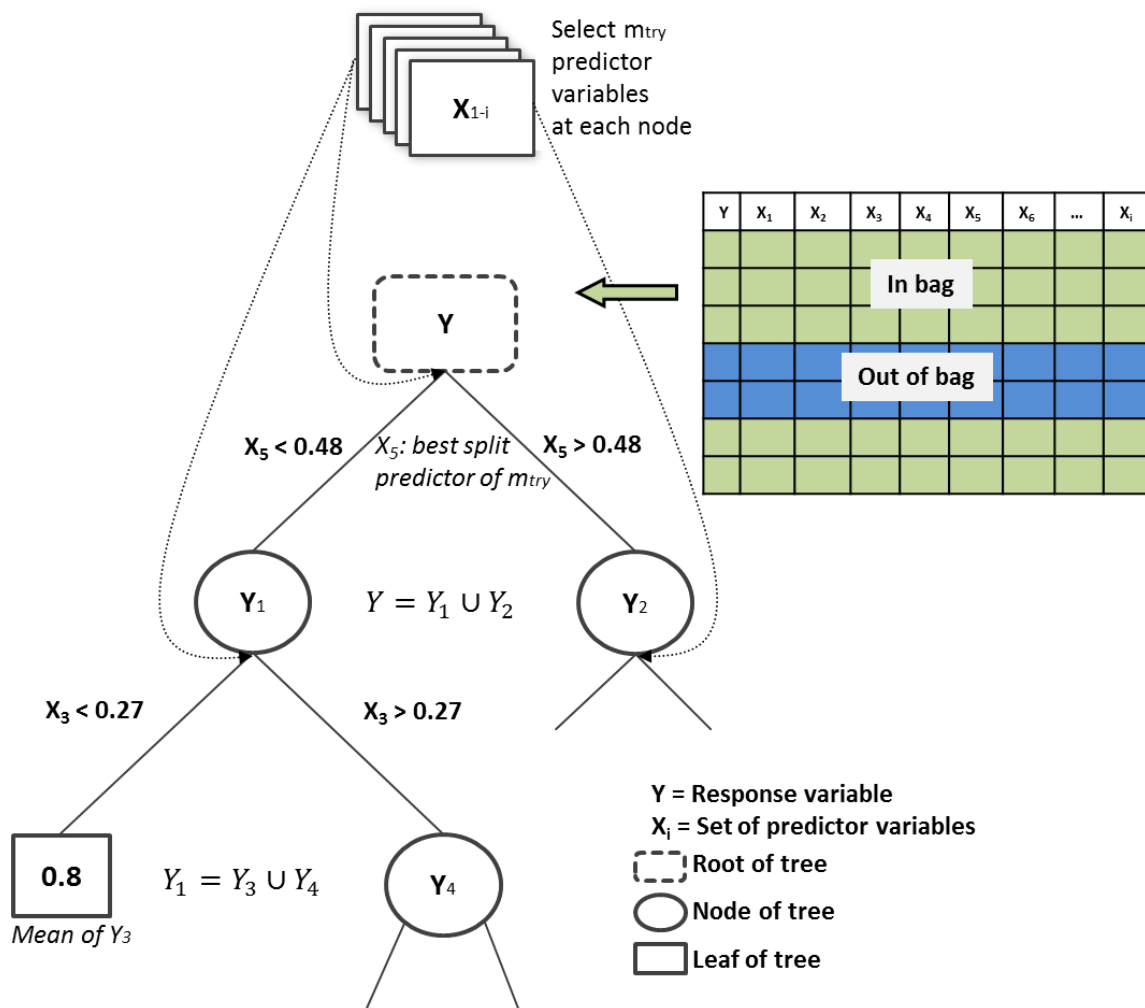


Figure 6-3: Schematic representation of a Random Forest regression tree.  $Y$  corresponds to the response variable, which is the schistosomiasis prevalence in this study.  $X$  refers to the set of predictor variables, which consists of the RS variables listed in Table 4-3. At each node, the response variable will be split into more homogeneous sub-spaces, while a split is explained by a threshold derived from the variable with highest informative value out of the number ( $m_{try}$ ) of randomly selected predictor variables. A leaf represents the mean value of a most pure subset of the response variable, which is best explained by the predictor variables and threshold indicated by its tree branch.

Besides the aim to explore the relationship between response and predictor variables, the Random Forest algorithm is used to predict this relationship for non-sampled locations in the case of spatial modelling. For spatial prediction, the fitted Random Forest is applied to new data and the predictions of all trees are again combined by averaging or majority voting.

## 6.1.2 Partial least squares regression

The PLSR developed by Wold et al. (1984), is a special approach of conventional multivariate regression methods. For PLSR, the original data of both the response (if multivariate) and predictor variables, are transformed to principal components with the aim of capturing most of the information in the predictor variables (X) that is useful for predicting the response variable(s) (Y). A linear regression model is then derived from the scores of the principal components (Figure 6-4). The model thereby overcomes the restrictions of standard regression methods as the high dimensionality of original predictor variables is reduced. The orthogonality of the principle components eliminates the multicollinearity issue. Therefore, PLSR can, unlike multiple linear regression (MLR), analyse data with strongly collinear, noisy, and numerous X variables (Wold et al., 2001: 109). The main goal of PLSR is to derive the relation between X and Y. In contrast to the principal component analysis (PCA), which is mainly designed to decompose the original predictor data to best represent X, the PLSR selects components with highest explanatory power to predict Y. This is achieved by the simultaneous decomposition of X and Y with the constraint of maximizing the covariance between X and Y (Garthwaite, 1994: 122).

In the context of PLSR, components are called latent variables (LVs) and are obtained iteratively (Mevik and Wehrens, 2007: 3). The general procedure to iteratively compute the LVs is: (i) to mean-center and scale the original data matrix; (ii) to compute the scores, loadings and weights of the LVs (Figure 6-4); (iii) to subtract the LVs from the original data (= residual matrix E, see Figure 6-4); and (iv) to iteratively repeat this procedure based on the deflated data matrix (Dayal and MacGregor, 1997). It is continued as long as there are significant LVs (Höskuldsson, 1988: 217) and the residuals can no longer be minimised. To ensure that a best set of LVs from X to predict Y is derived, the computations of LVs consider both the outer relation of the scores and loadings with respect to X and the inner relation between the scores of X and Y with the aim of minimising their mixed relation. In order to obtain orthogonal scores from the original data, it is necessary to obtain weights (Geladi and Kowalski, 1986: 12). Figure 6-4 illustrates the transformation and procedure of the PLSR for the case of one response variable (Y) and a predictor matrix (X) of m variables and n samples, which reflects the general data structure provided for this thesis. Thus, one LV results in: (i) one score for each sample n (matrix T), which provide a new X variable as input for the subsequent MLR analysis; (ii) one loading per variable m (matrix P'), which reflect the regression coefficient obtained for each original variable from the PLSR algorithm; and (iii) one weight pe variable m (matrix w'), which reflects the covariance structure between predictor and response variable.

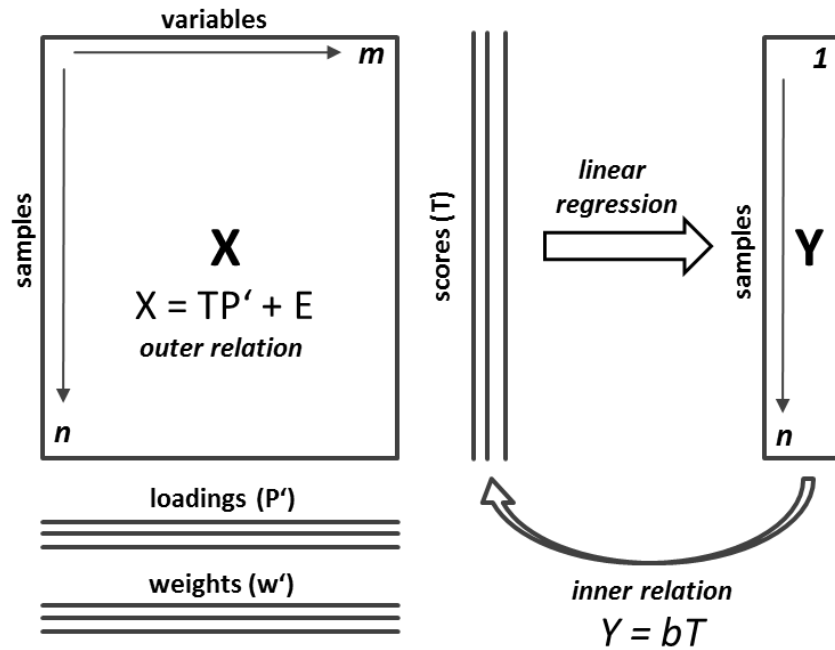


Figure 6-4: Schematic representation of the PLSR approach. This graph shows the PLSR model procedure for the case of a single response variable  $Y$  (corresponds to schistosomiasis prevalence) to be predicted by a multitude of predictor variables  $X$  (correspond to the RS environmental variables).

Source: modified after Wold et al. (2001: 113)

There are different methods of computing LVs for PLSR models: the classical PLS algorithm is the nonlinear iterative partial least squares (NIPALS) procedure developed by Wold (1975). In 1993, Lindgren et al. (1993) developed a new kernel partial least squares (PLS) algorithm, which has been proven to be faster and provide identical results to those obtained from the standard NIPALS procedure (De Jong and Ter Braak, 1994: 169). Whereas the NIPALS deflates both the  $X$  and  $Y$  matrices, the kernel PLS deflates only the  $X$  matrix given the fact that deflation of the  $Y$  matrix was considered optional (Höskuldsson, 1988; Dayal and MacGregor, 1997: 73-74). Due to the data structure of this thesis with only one response ( $Y$ ) variable, the faster kernel PLS was assumed to provide the most suitable algorithm. Furthermore, less prominent and used approaches to compute LVs for PLSR are discussed by Höskuldsson (1995).

Before the analysis, the  $X$  and  $Y$  variables need to be transformed to make their distributions fairly symmetrical (Wold et al., 2001: 113). Thus, in this thesis, the input data have been standardised using the z-score given in Equation 6-1,

$$z = \frac{x - x_{mean}}{s} \quad \text{Equation 6-1}$$

which centers the data  $x$  in relation to their mean value  $x_{mean}$  and scales to unit variance  $s$ . The initial standardisation of the data allows to directly compare the scores and loadings of different PLSR implementations (Mevik and Wehrens, 2007: 3). As in multiple linear regression, the overall purpose of PLSR is to build a linear model between the response variable (or matrix)  $Y$  and the predictor matrix  $X$ . Therefore, the problem of choosing an optimum set of predictors still remains. The number of LVs ( $n_{comp}$ ) used for the PLSR model needs to be specified for the respective PLSR model and is therefore a parameter for tuning the model (see Section 6.2.4).

## 6.2 Statistical model procedures

In this study, specific methodological procedures of modelling schistosomiasis risk were applied to meet the respective objectives introduced above. Due to the prerequisite of scaled data as input for the PLSR model (Section 6.1.2), the full database of epidemiological and RS variables (Chapter 4) was scaled to z-scores in a preparatory step. This scaled database was used as input data for all model approaches used in this study.

### 6.2.1 Multi-scale modelling

When modelling the risk of schistosomiasis with RS data, the issue of scale is inherent in different aspects of the modelled phenomenon. A very prominent scale issue in this regard is the spatial mismatch between sampling units of disease prevalence and remotely sensed environmental information, which is prominent in high spatial resolution RS data (Figure 6-1). This scale issue is influenced by the spatial resolution of RS data and the school catchment area included in the spatial modelling. In the multi-scale modelling approach of this study (Figure 6-5), the impact of the spatial resolution of RS data (scale 1) and the extent of the school catchment area (scale 2) on the model accuracy was investigated based on the study site MAN. This study site was covered by RapidEye, Landsat 5 TM, and Terra MODIS data and provided at the same time the highest sample size of school prevalence. To discriminate between effects due to spatial resolution and due to specific sensor characteristics, the sensor characteristics were analysed separately at a given spatial resolution.

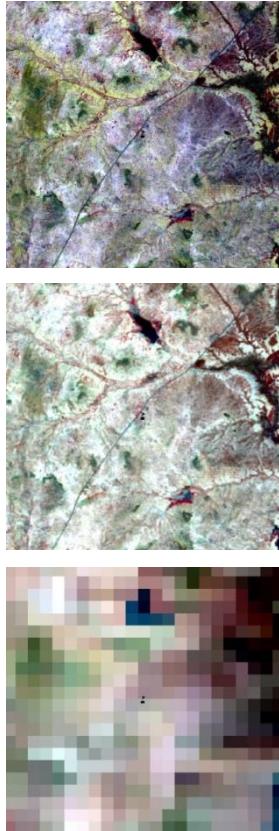
The impact of the spatial resolution of RS data (scale 1) was investigated based on high-resolution RapidEye reflectance and spectral indices. This dataset was thenceforth aggregated to the relevant spatial resolution of 30 m (corresponding to Landsat 5 TM) and 250 m and 500 m (corresponding to Terra MODIS). In a separate analysis, the spatial resolution was represented by data from the respective sensor with the intention to discriminate between the resolution and sensor impact on the model. The sensor impact was furthermore evaluated by comparing model results from associated spectral reflectance and indices available for all three sensors with all available reflectance and indices provided by the respective sensor (Table 6-1).

*Table 6-1: Overview of bands and indices used for multi-scale and multi-sensor analysis. The spectral wavelength of bands is given in Table 4-2. nir = near infrared, mir = middle infrared, swir = shortwave infrared, tir = thermal infrared*

	RapidEye	Landsat 5 TM	Terra MODIS
<b>Cut-set of bands and indices</b>	Blue, green, red, nir, NDVI, EVI, SAVI, MSAVI, NDWI	Blue, green, red, near infrared, NDVI, EVI, SAVI, MSAVI, NDWI	Blue, green, red, nir, NDVI, EVI, SAVI, MSAVI, NDWI
<b>Full set of bands and indices</b>	Blue, green, red, redegde, nir, NDVI, EVI, SAVI, MSAVI, NDWI	Blue, green, red, nir, mir, swir, tir, NDVI, EVI, SAVI, MSAVI, NDWI, MNDWI	Blue, green, red, nir, mir, swir1, swir2, NDVI, EVI, SAVI, MSAVI, NDWI, MNDWI

The impact of the selected extent of the school catchment area (scale 2) was investigated by spatial buffer analysis for a defined area around the point measurement of the school location. The buffer radius has been defined in steps of 100m around points and analysis ceased at a buffer radius of 5km. This distance was estimated to represent the maximum estimated school catchment area (Malone et al., 2001: 62). The RS variables were aggregated by their means for each buffer extent.

**Scale 1:**  
spatial resolution of  
RS data



**Scale 2:**  
catchment area of school

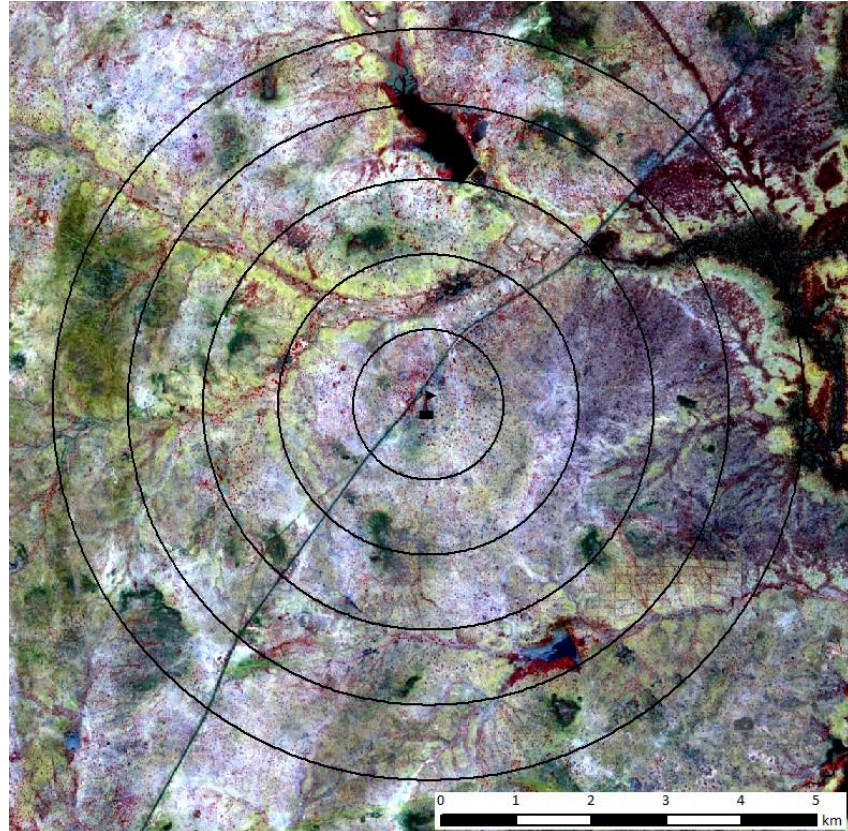


Figure 6-5: Multi-scale analysis of RS environmental factors. The first scale issue addresses the spatial resolution of RS data (left column) for modelling schistosomiasis risk, which is investigated for RapidEye with 6.5 m (top), Landsat 5 TM with 30 m (centre), and Terra MODIS data with 500 m (bottom). The second scale issue considers the catchment area of the school to be considered in the spatial model procedure. The buffer radius for multi-scale analysis with respect to the catchment area ranges from 100 to 5 km from school location.

## 6.2.2 The hierarchical model approach

A hierarchical model was developed to address the spatial mismatch between the school measurement of schistosomiasis prevalence and relevant environmental conditions for disease transmission measured from RS data. For the hierarchical model, the binary mask of potential disease transmission sites derived in Chapter 5 were used to specifically select only relevant regions for modelling schistosomiasis risk (Figure 6-6). The underlying hypothesis of this hierarchical model approach was that only sites where water was detected or water potentially may accumulate, represent relevant environmental information with respect to schistosomiasis transmission and the prevalence measured for the school catchment area.

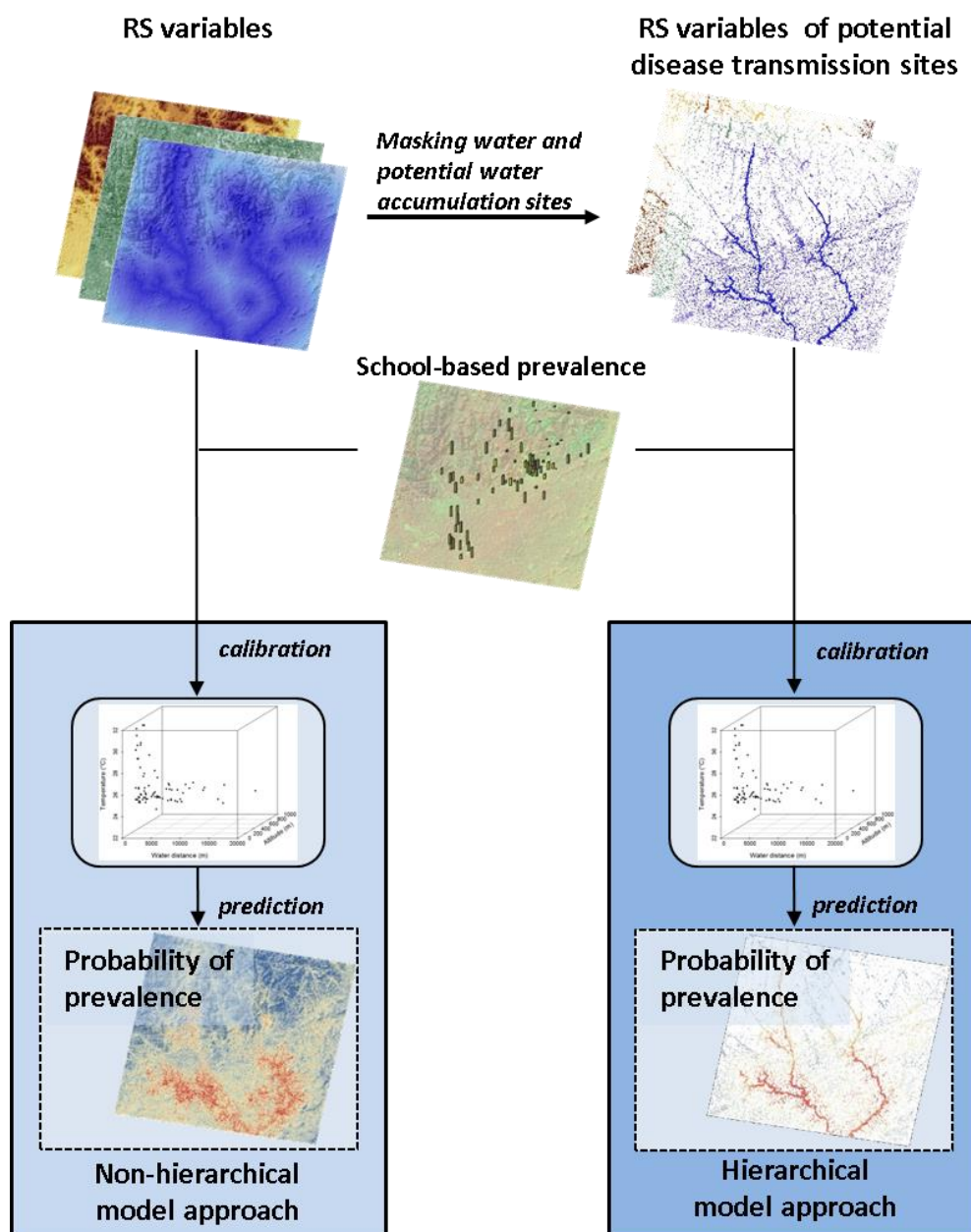


Figure 6-6: Flowchart of the hierarchical model approach considers only potential disease transmission sites derived in Chapter 5 for modelling schistosomiasis risk based on remotely sensed environmental variables.

The impact of this hierarchical model approach was evaluated in reference to the corresponding non-hierarchical procedure, where the respective buffer area around the school location was completely considered by the model. Both approaches rely on the full set of RS variables derived for this thesis (Table 4-3), whereas each study site was modelled based on the best available data set (see Table A 1 in the Appendix). The models were processed for different extents of the school catchment areas ranging from a buffer radius of 500 m to 5 km in steps of 100 m. RS variables were aggregated by their means within the spatial buffer extent, except for the variables settlement area and streams, which were aggregated by their sum and maximum, respectively. The extraction of the data within given buffer zones was based on RS variables resampled to a spatial resolution of 30m using the nearest neighbour method. Thus, the spatial



resolution matches with the provided mask of potential disease transmission sites and makes extraction of small patches feasible as required by the hierarchical model approach.

Additionally, the impact of different ecological regions on model accuracy was further evaluated by comparing the resulting models from each study site with the overall region. Thus, schistosomiasis risk models were separately derived for each study site based on the common cut-set of variables, which were available for all three study sites (see Table A 1 in Appendix) and compared with the cross-ecozonal model. This was based on data from all three sites, where savannah, tropical lakeside and tropical mountainous regions were combined in one cross-ecozonal model approach.

Both models, the non-hierarchical and hierarchical model, were used to predict schistosomiasis risk for non-sampled locations in space. The predicted z-scores were then calculated in reverse to schistosomiasis prevalence between 0 and 100% and further inspected by their spatial mean, standard deviation, and coefficient of variation through all spatial extents modelled.

### 6.2.3 The measure of variable importance

In order to evaluate which of the multitude of RS variables provide valuable information to explain the spatial distribution of the disease, each predictor's individual variable importance was calculated. Moreover, changes in variable importance with ecozone were evaluated.

Random Forest regression provides two measures of variable importance: the increase of the mean squared error (IncMSE) and the cumulative increase in node purity (IncNodePurity). The IncMSE is derived for each predictor variable from the MSE difference between the predictive measure based on the original dataset and based on a permuted dataset, where the predictor in question was randomized. The IncNodePurity calculates for each predictor variable how much it reduces node impurity, which is the difference between RSS before and after the split and sums this up over all splits and trees (Hastie et al., 2009: 593). For the Random Forest regression analysis used in this study, the IncNodePurity measure was used to evaluate the importance of RS predictor variables. Only variables with an increase in node purity greater than one were considered important in this study.

A variable importance measure of the PLSR model is the VIP (Equation 6-2), which reflects how well a predictor variable describes the response variable and how important this information is for the composition of the set of predictor variables. This is due to the inclusion of the weights, which reflect the covariance between the predictor and response variables (Andersen and Bro, 2010: 732). The VIP value for a variable  $j$  is expressed as

$$VIP_j = \sqrt{\frac{\sum_{f=1}^F w_{jf}^2 SSY_f J}{SSY_{total} F}} \quad \text{Equation 6-2}$$

where  $w_{jf}$  is the weight value for variable  $j$  and component  $f$ ,  $SSY_f$  is the sum of the squares of the explained variance for the  $f$ th component and  $J$  is the number of variables.  $SSY_{total}$  is the total sum of squares explained of the response variable, and  $F$  is the total number of components. If the VIP exceeds the threshold value of one it indicates an important variable (Andersen and Bro, 2010: 732)

## 6.2.4 Model validation

All Random Forest and PLSR model approaches have been validated through internal cross validation using the “caret” package in R (Kuhn, 2008). The principle of cross-validation is to split the database into a larger training and a smaller test data set. Based on training data, the model is established and applied to test data, which are at the same time used to evaluate model performance. Typical measures to evaluate model quality in a regression context are the coefficient of determination ( $R^2$ ), which indicates how well data fit a statistical model ranging between 0 and 1, and the RMSE, which measures the difference between predicted and observed values in units of the predicted measure. A third measure used for model evaluation is the Nash-Sutcliffe efficiency (NSE) index, which determines the relative magnitude of residual variance or noise compared to measured data variance, hence the real information provided by the model (Nash and Sutcliffe, 1970). This index can range between one and minus infinity. An index value of one corresponds to the optimal match between simulated and observed data and positive scores indicate that model simulation is better compared to a model based on the mean value of observations. The NSE (Equation 6-3) is commonly used to assess predictive power of hydrological models (Krause et al., 2005), however, it was successfully applied for other model applications such as the evaluation of wind erosion prediction systems (Hagen, 2004) or energy flux analysis (Wong et al., 2010). It is expressed as

$$NSE = 1 - \left( \frac{\sum_{i=1}^n (Y_i^{obs} - Y_i^{sim})^2}{\sum_{i=1}^n (Y_i^{obs} - Y^{mean})^2} \right) \quad \text{Equation 6-3}$$

where  $Y_i^{obs}$  is the  $i^{\text{th}}$  observation for the total number of observations  $n$ ,  $Y_i^{sim}$  is the  $i^{\text{th}}$  simulated value for the corresponding observation and  $Y^{mean}$  is the mean value of observed data.

K-fold cross-validation splits the database into a defined number of parts, of which one part is used to test the model and all other parts are used to fit the model. To keep a comparable amount of validation data per group of test data, the data splitting of internal cross-validation was adjusted according to the sample size of each study site. Models of the study site MAN (sample size = 75; fold size = 7-8) and BUF (sample size = 74; fold size = 7-8) were validated based on a 10-fold cross-validation and models of the study site TAB (sample size = 38; fold size = 7-8) were validated with 5-fold cross-validation. The cross-validation procedure was repeated ten 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  $m_{\text{try}}$  (number of variables selected per split) and  $n_{\text{comp}}$  (number of components used for model), were iterated to tune the Random Forest (Liaw and Wiener, 2002) and PLSR model (Geladi and Kowalski, 1986), respectively.

Due to the limited amount of data points available for this study, only the model outcome of study site MAN could be verified by external validation. At this site, 33 additional school surveys were derived from the pre-selection steps of epidemiological data (Figure 4-1). This independent test set corresponded to schools surveyed a few years earlier than the latest surveys used in the training data base of the model. The predicted prevalence was evaluated by the test data set based on the resulting  $R^2$  of a linear model. The common approach to split the data base into 80% training and 20% test data was used for the internal cross-validation procedure described

above and not considered useful for external validation of the model due to the general shortage of reference data.

## 6.3 Results of statistical risk modelling

This section presents the results according to the objectives introduced above. The impact of scales regarding RS data and area of observation are described in Section 6.3.1. The outcome of the developed hierarchical model to bridge the spatial discrepancy between the prevalence measure and remotely sensed environmental conditions is presented in Section 6.3.2. The most important RS variables for modelling schistosomiasis risk are given in Section 6.3.3. Finally, Section 0 presents the results of spatial schistosomiasis risk predictions and its validation for the study site MAN. The comparison between different ecological regions and used model algorithms are reported in the respective model results of all sections.

### 6.3.1 How scale matters

The impact analysis of scale on schistosomiasis risk modelling referred to the scale of RS data, hence its spatial resolution, and the scale of observation regarding the environment within the catchment area of school locations. Figure 6-7 shows the predictive power of Random Forest and PLSR models based on RapidEye, Landsat 5 TM, and Terra MODIS data for different catchment buffers around the surveyed schools. Although the spatial resolution of RS data *per se*, as tested based on aggregated RapidEye data, showed no impact on the model outcome (see Figure A 2 in the Appendix), the model performance varied strongly between the different sensors used for schistosomiasis risk modelling. Terra MODIS data resulted generally in poor model performance with  $R^2$  values around 0.25, RapidEye data reached higher predictive power and Landsat 5 TM provided superior results.

The scale of observation around the school location shows a marked impact concerning the predictive power of models based on RapidEye and Landsat 5 TM data. In all cases, the  $R^2$  results are lowest, if RS variables were directly extracted at the school location. The models derived from RapidEye and Landsat 5 TM data performed better with an increasing area of observation around school locations. In most cases, the  $R^2$  levelled out at its highest score for analysis within a buffer extent between 2 and 4 km. However, this was different for the Random Forest model of Landsat 5 TM data, which resulted in a peak of explanatory power for the analysis within 3 km around the school locations and deteriorated for larger observational units.

The comparison between the cut-set of bands and indices available for all three sensors and the sensor specific spectral properties showed that Landsat 5 TM data resulted in the highest overall model performance, if the full potential of the Landsat 5 TM spectrum was used. In contrast, RapidEye and Terra MODIS data showed no impact between the cut-set and full set of bands and indices used for modelling. Both model algorithms result in similar trends considering the scale of observation, however, Random Forest reached higher  $R^2$  scores for the Landsat 5 TM model and PLSR reached higher scores for the RapidEye model.

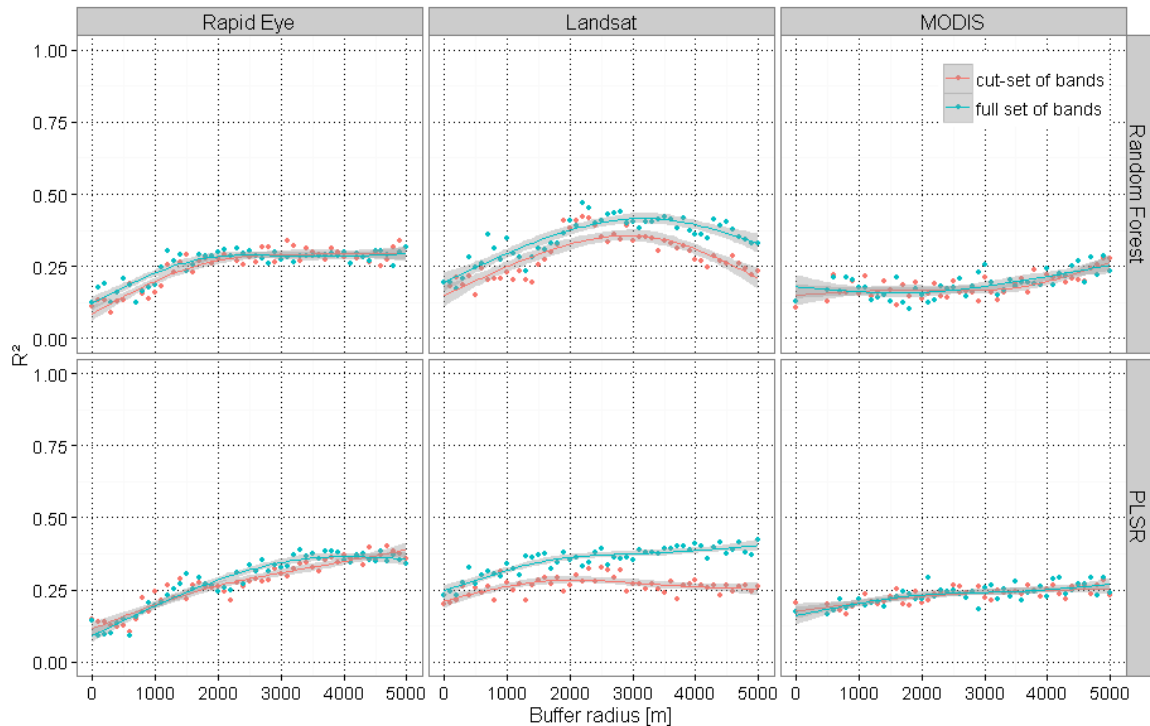


Figure 6-7: Comparison between model results from RapidEye, Landsat 5 TM, and Terra MODIS data for the study site MAN. The red dots represent model results ( $R^2$ ) derived from the cut-set of bands and indices available for all three sensors and the blue dots refer to the model results derived from the full set of bands and indices available for the respective sensor as given in Table 6-1. The lines show the spline interpolation and the grey bars represent the confidence interval. The upper row represents model results from the Random Forest algorithm and the lower row from the PLSR analysis.

### 6.3.2 The hierarchical model: a solution to bridge the spatial gap?

The hierarchical model approach was developed with the aim of overcoming the spatial mismatch between the school-based prevalence and remotely sensed environmental information regarding the ecological process of disease transmission.

Figure 6-8 illustrates the model performance in comparison between the non-hierarchical and the hierarchical model approach for the three study sites and based on the two model algorithms used in this study. It is apparent that the outcomes of both approaches, the non-hierarchical and hierarchical model, vary strongly between the study sites. In BUF, none of the established models performed satisfactorily with  $R^2$  values below 0.3. The performance of the model was better in the study site MAN and highest in the study site TAB. In accordance to this, performance of the hierarchical model approach improved substantially in TAB and improved partially in MAN but did not make a difference in BUF. In TAB, both model algorithms showed a marked improvement of explanatory power through the hierarchical model approach with maximum  $R^2$  scores between 0.6 and 0.7 within buffer zones of 3 km radius and larger. At this site, the PLSR model performed slightly better than the Random Forest algorithm. In MAN, only the PLSR model indicated an improvement for the analysis within the first kilometre around the school locations.

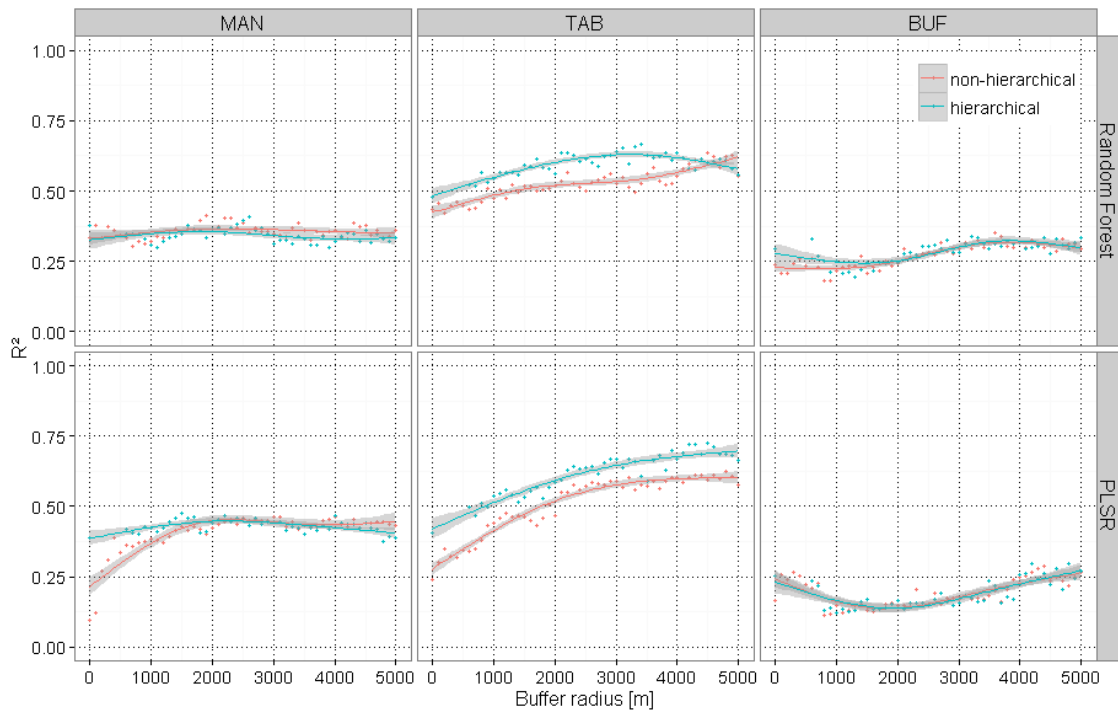


Figure 6-8:  $R^2$  values per selected buffer radius of the non-hierarchical (red) and the hierarchical (blue) models for the three study sites MAN, TAB and BUF based on the Random Forest (upper row) and PLSR model (lower row).

Models were additionally evaluated based on the RMSE (Figure A 3) and the NSE (Figure A 4). Lowest RMSEs around 0.6 were reached in accordance with highest  $R^2$  values around 0.7 and highest NSE scores close to 0.75 for the hierarchical PLSR model in the study site TAB. Similar to the  $R^2$  values in TAB, the NSE resulted slightly lower from the Random Forest model compared to the PLSR model, however, with lower RMSE for the buffer zones between 0 and 2 km. For larger buffer zones, the PLSR model performed superiorly due to a lower RMSE. The three model evaluation criteria  $R^2$ , RMSE, and NSE consistently confirmed that the hierarchical model approach performed superiorly in comparison to the non-hierarchical model approach in the study site TAB with high model efficiency confirmed by the NSE index. In MAN, the NSE of the PLSR model performed superiorly to the  $R^2$  values with scores up to 0.6 for the hierarchical approach, whereas for the Random Forest model this index resulted minor scores below 0.3. However, this was different for the RMSE, whose scores corresponded well to the results of the  $R^2$  values derived in MAN. In BUF, the poor model performance indicated by the  $R^2$  values was confirmed by the consistently high RMSE and low NSE, except for some outliers of NSE scores greater than 0.4 for the hierarchical PLSR model within buffer extents of 4 to 5 km radius.

The strong impact of the geographical region on modelling schistosomiasis risk and the outcome of modelling across ecozones is illustrated in Figure 6-9. The deviation in explanatory power between the poorest models in BUF and best models in TAB reaches up to 40%. The combination of training data from different ecological regions resulted in poor model performance with  $R^2$  scores below 0.25. This is further confirmed by the RMSE (Figure A 5) and NSE (Figure A 6).

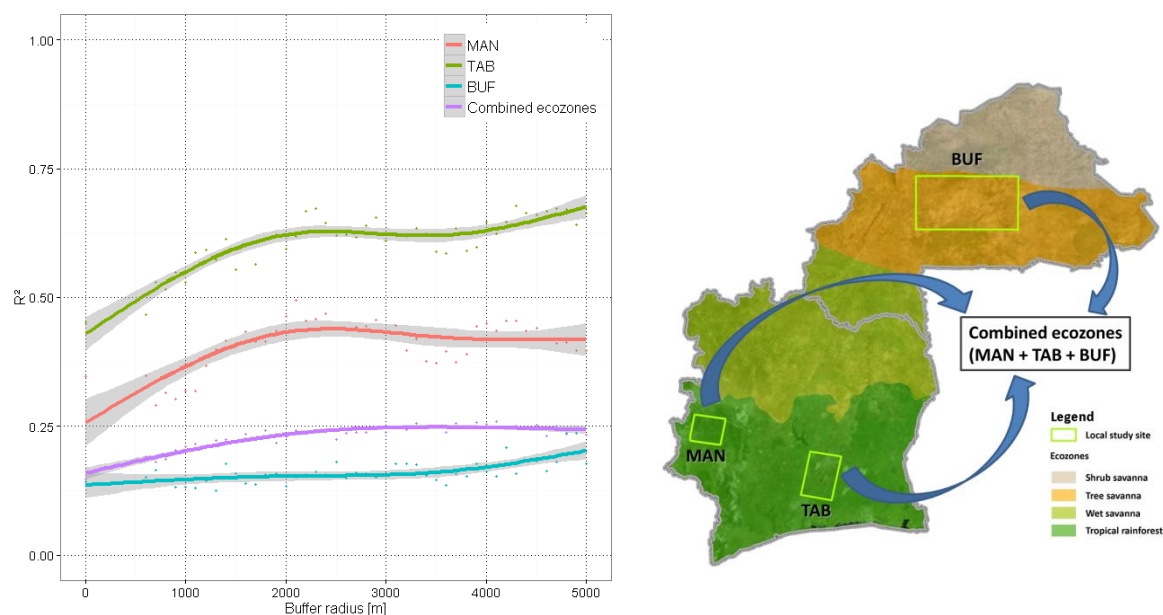


Figure 6-9: Hierarchical model performance in different ecological regions and for the cross-ecozonal model using PLSR

### 6.3.3 Key remote sensing variables for schistosomiasis risk modelling

Until now, the NDVI and LST were the most commonly used RS variables for schistosomiasis risk modelling. The objective of this thesis was to explore the multitude of disease relevant RS variables listed in Table 4-3 and assess their individual contribution to modelling the spatial distribution of disease risk. The importance of the most relevant variables in the hierarchical Random Forest model is illustrated in Figure 6-10 for each of the three study sites. An overview of all variables with importance greater than one is given in Figure A 7 for the non-hierarchical Random Forest model, in Figure A 8 for the hierarchical Random Forest model and in Figure A 9 for both PLSR model approaches in the Appendix.

Variable importance resulted differently depending on the study site investigated (Figure 6-10). For example, the variable “euclidean distance from water body” was highly important for all school catchment radii in MAN and had comparably little importance regarding the study sites TAB and BUF. Furthermore, the topographic variables “altitude”, “slope”, and “streams” were of high importance in the hierarchical Random Forest model for MAN, whereas the non-hierarchical approach additionally considered the variable “sinks” important at this site (Figure A 7). In contrast, the spatial prediction of schistosomiasis risk in the study site TAB was mainly explained by the mean and median of Terra MODIS LST, which was not considered important in the study sites MAN and BUF. However, maximum LST contributed more in the MAN site and minimum LST contributed more in BUF. Based on the Random Forest model, in BUF no variables are highlighted specifically important with nearly all variables contributing marginally to the model (Figure A 7 and Figure A 8).

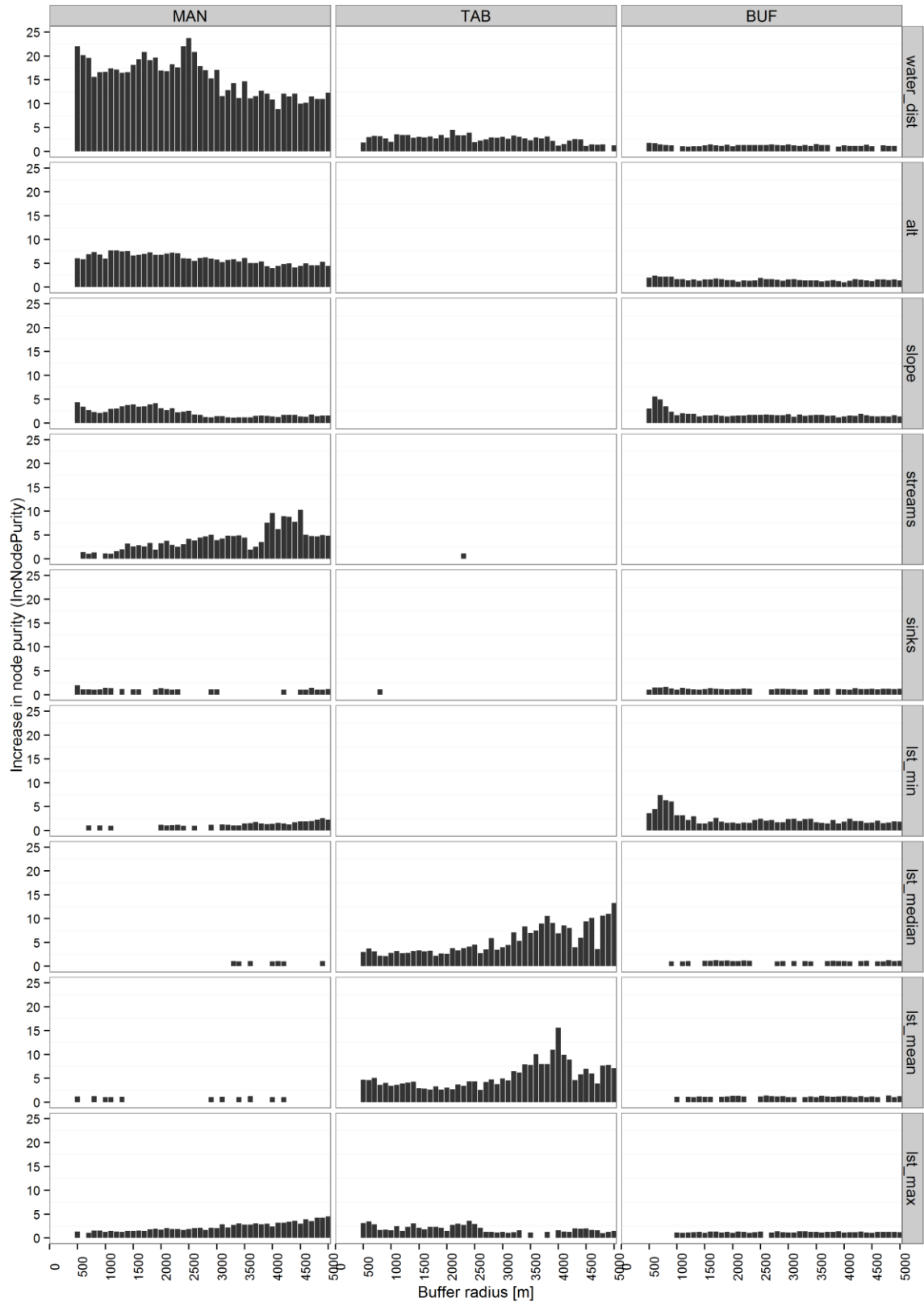


Figure 6-10: Variable importance measured for selected variables in comparison between the three study sites given by the increase of node purity from the Random Forest hierarchical model. The nomenclature of selected RS variables is explained in Table A 1.

The PLSR models resulted in generally different variable importance (Figure A 9) compared to the Random Forest models. In MAN, “streams” and “settlement area” resulted in VIP greater than one for selective buffer extents around school locations, whereas “euclidean distance from water body” was not considered important by this model. In TAB, the PLSR model considered “streams” as the major variable of high importance for both the non-hierarchical and hierarchical model, whereas Terra MODIS LST reached the threshold only for two buffer extents of the hierarchical model approach. In the study site BUF, the PLSR model considered “settlement area” as the important variable to model schistosomiasis risk.

### 6.3.4 Schistosomiasis risk prediction and validation

The overall objective to model schistosomiasis risk based on RS data, is to generate spatial information of disease transmission risk to support planning, intervention, and control initiatives of public health authorities. The RS based models used in this study were therefore used to predict the established relation between disease prevalence and environmental conditions for the non-hierarchical and the hierarchical model approach. Figure 6-11 and Figure 6-12 present examples of spatial predictions based on the Random Forest and PLSR model for the study sites MAN and TAB, respectively. Spatial predictions in the study site BUF (see Figure A 10 in the Appendix) were not analysed further due to its overall poor performance in this study.

Figure 6-11 illustrates the spatial predictions from the Random Forest and PLSR models for the non-hierarchical and hierarchical approach in the study site MAN. The validation of this analysis based on an external test data set is expressed by the  $R^2$  for each model. The Random Forest model predicted low risk in the mountainous part and high risk along the river valleys running towards the South of the study site. Despite the resulting  $R^2$  value of 0.54 being slightly higher compared to the one of the hierarchical model ( $R^2 = 0.51$ ), the north-south running river in the western part of the study site and specifically the small-scale structures of high disease risk along the river valley north of the city Man, were better predicted according to the visual indication of measured prevalence. In contrast, the PLSR model resulted in a more fuzzy distribution of low and high risk regions and showed similar results of spatial predictions between the non-hierarchical and hierarchical approach. However, the external validation indicated that the non-hierarchical approach of the PLSR model performed better ( $R^2 = 0.51$ ) than the hierarchical one ( $R^2 = 0.44$ ). This slightly better performance of the non-hierarchical PLSR model compared to the hierarchical PLSR model was found for almost all scales of observation around schools, whereas no difference between hierarchical and non-hierarchical approach and scales of observations was indicated by the Random Forest model (Figure A 11).



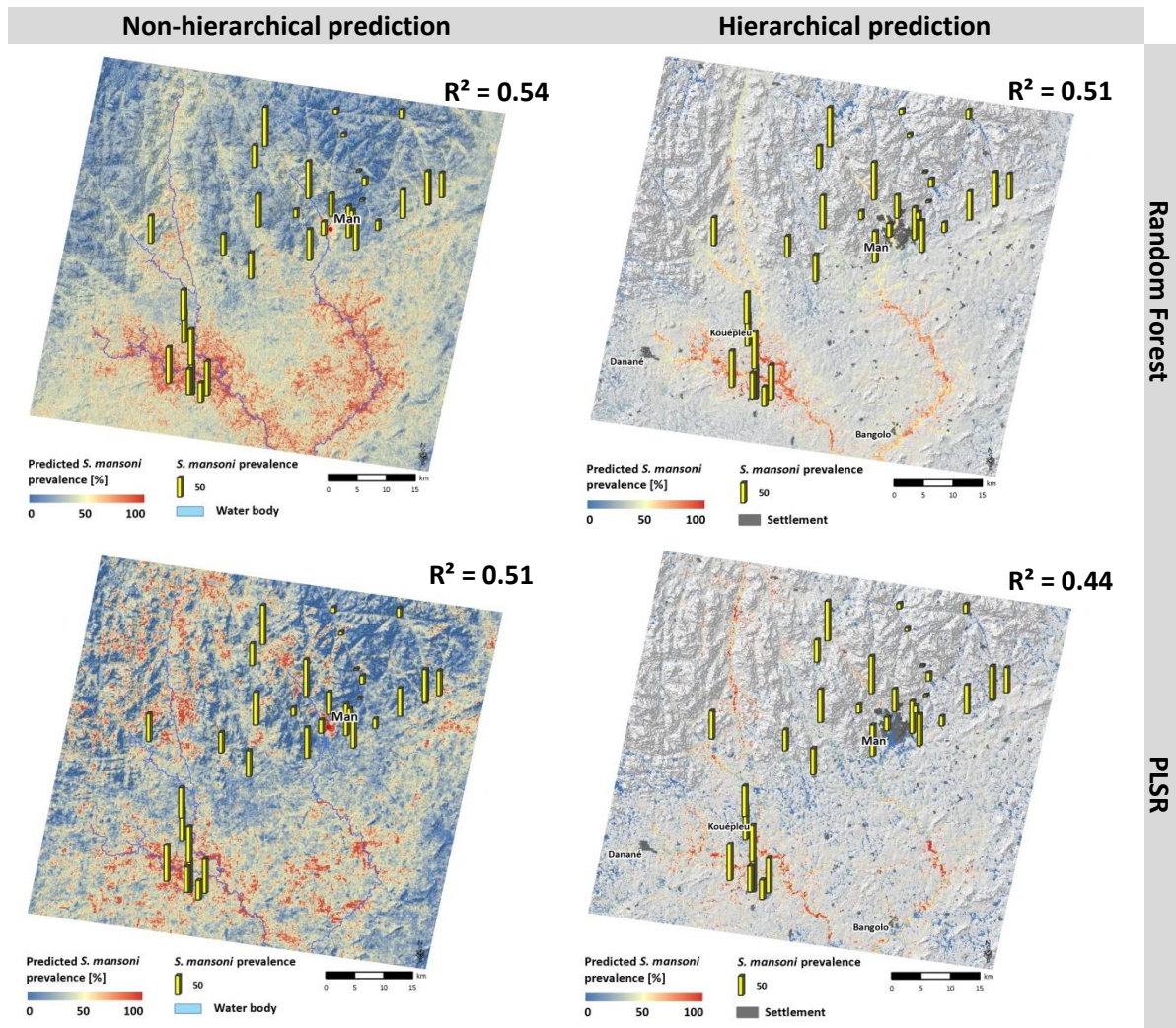


Figure 6-11: Comparison of spatial risk prediction between non-hierarchical and hierarchical model approach at the study site MAN based on Random Forest (upper row) and PLSR model (lower row). This example is based on spatial modelling within a buffer extent of 3 km around school location. The  $R^2$  value presents the result of the linear model between predicted and test data. Yellow bars indicate the school locations of the test data: the higher the bar the higher the measured prevalence.

Figure 6-12 illustrates the spatial predictions from the Random Forest and PLSR models for the non-hierarchical and hierarchical approach in the study site TAB. For this study site, no additional data for external validation were available. The four models predicted the Lake Taabo in the centre of the study site as area of high risk for disease transmission. The Bandama river running from north-west to south-east crossing Lake Taabo was identified as a high risk area by the hierarchical models of both algorithms and the non-hierarchical approach of the PLSR model. However, the course of the river in the north-eastern corner of the study site was only predicted as a high risk area by the hierarchical Random Forest model. Additionally, several hotspots of high disease risk were predicted by Random Forest models that correspond to settlements and topographic elevations in TAB. Those were predicted with less risk by the PLSR model.

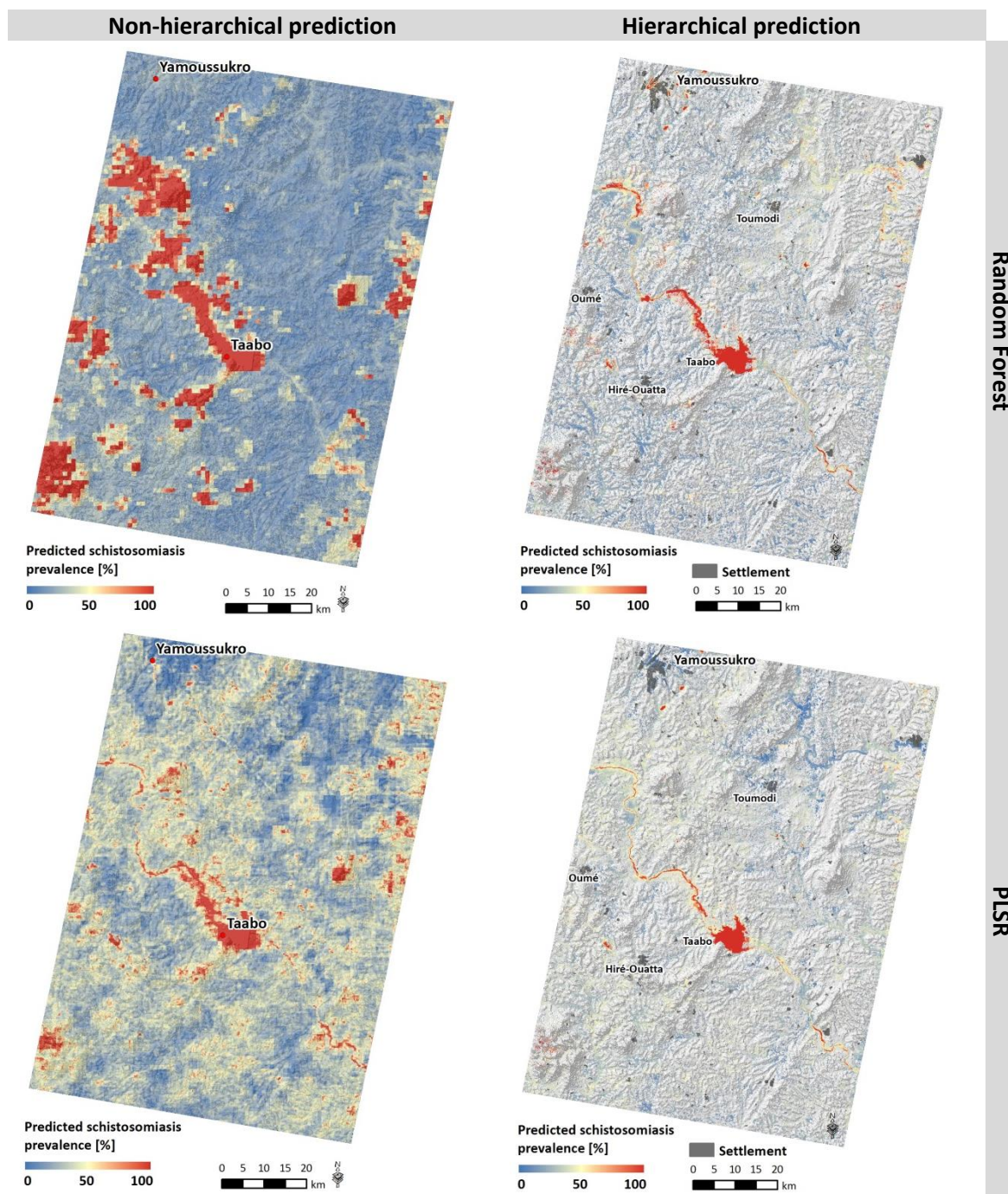


Figure 6-12: Comparison of spatial risk prediction between non-hierarchical and hierarchical model approach at the study site TAB based on Random Forest (upper row) and PLSR model (lower row). This example is based on spatial modelling within a buffer extent of 3km around school location.

The mean predicted prevalence, its standard deviation and the coefficient of variation derived from buffer extents between 0 and 5 km, are illustrated in Figure 6-13 for the non-hierarchical and hierarchical Random Forest model approach in TAB. Both the non-hierarchical and hierarchical model approaches, predicted a very high mean value of disease prevalence for Lake Taabo from all buffer extents. This predicted mean value indicates that parasite prevalence and human infection due to water contact is estimated very high in Lake Taabo. The course of the Bandama river shows a mean predicted prevalence around 50% with a moderate standard deviation and a low coefficient of variation. The coefficient of variation results in higher scores

for the non-hierarchical approach in comparison to the hierarchical model and shows large areas in the southern part of the study site with highest scores of variation between different buffer extents analysed. Similar results were provided by the PLSR model as illustrated in Figure A 12, however, with a generally lower coefficient of variation and standard deviation between different buffer extents. The results of mean predicted prevalence, standard deviation, and coefficient of variation from buffer extents between 0 and 5 km for the study site MAN are illustrated in Figure A 13 for the Random Forest model and Figure A 14 from the PLSR model. Moderate to high mean prevalence was predicted from both model algorithms and approaches close to the course of the river, however, the Random Forest model predicting the high impact of distance to the river appearing as buffer of mean moderate to high prevalence. All models of MAN resulted a very low mean prevalence for the mountaineous region in the north-east of the study site. Similar to TAB, the standard deviation and coefficient of variation resulted generally lower from Random Forest models compared to PLSR models.

## 6.4 Discussion of remotely sensed schistosomiasis risk modelling

RS data have shown a high potential for spatial risk modelling and prediction of schistosomiasis risk. This study has shown that with the presented approach not the spatial resolution of RS data *per se*, but the extent of the area observed around a school location has a major impact on the model performance. From the 60 RS variables investigated for schistosomiasis risk profiling, there were very few that explained most of the disease variation. The Random Forest and PLSR algorithms were used to investigate a non-hierarchical and hierarchical approach for modelling schistosomiasis risk with the hierarchical approach performing considerably superiorly. In the following, the RS data and sensors used for schistosomiasis risk modelling (Section 6.4.1) and the performance of spatial modelling and risk prediction (Section 6.4.2) are discussed.

### 6.4.1 Remote sensing data for schistosomiasis risk modelling

RS data provide information on environmental conditions for *Schistosoma* parasites, corresponding snails, and schistosomiasis transmission sites. Statistical disease risk models link these environmental conditions to school-based measures of disease prevalence, which represents the school catchment area and aim at indicating the environmental suitability for diseases transmission in this area. Due to this discrepancy between point measurement and school catchment area, the impact of scale on schistosomiasis risk models based on RS data was investigated by this research.

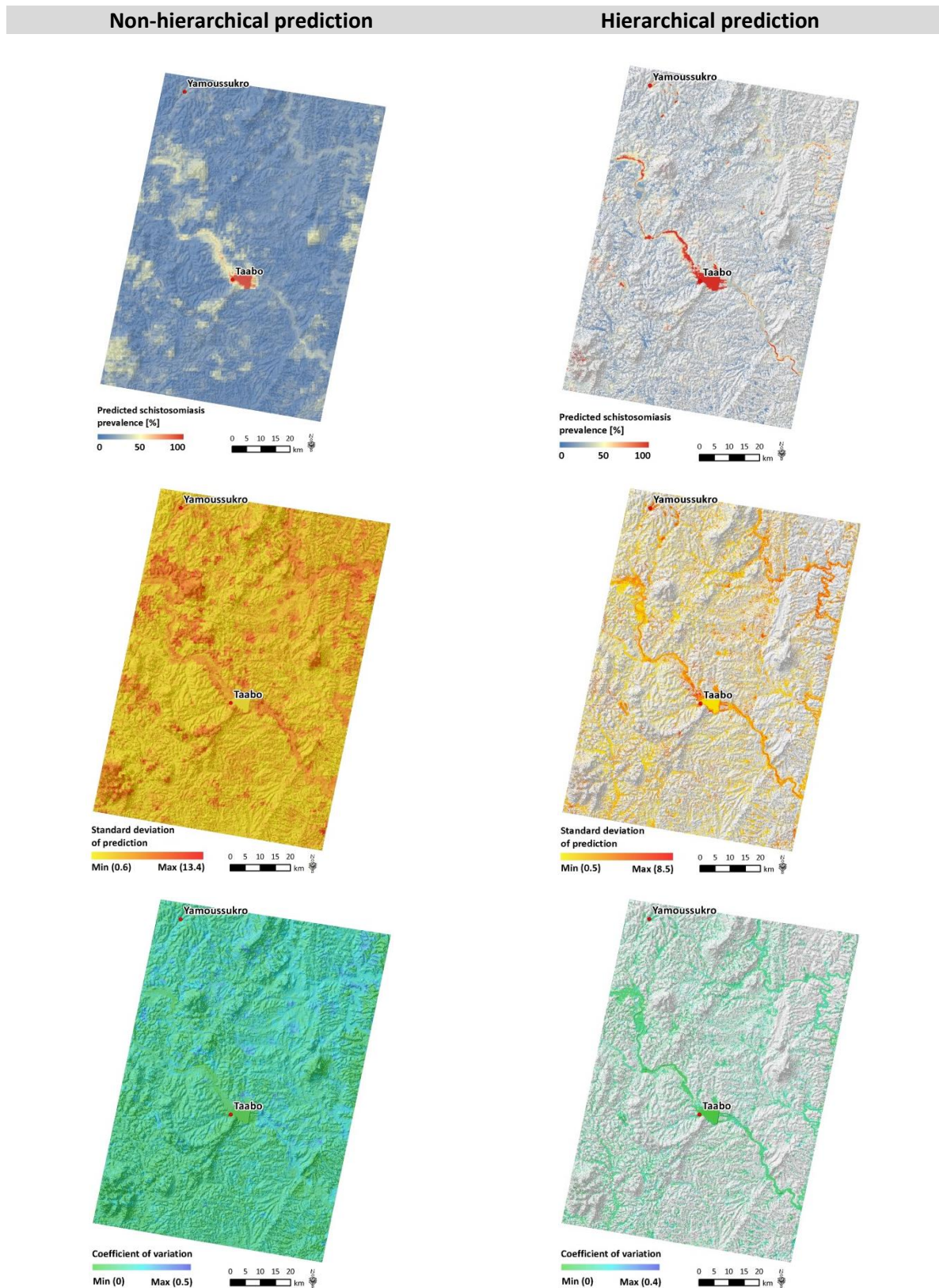


Figure 6-13: Mean prevalence (top row), standard deviation (middle row) and coefficient of variation (bottom row) derived from buffer extents between 0 and 5 km using the Random Forest model.

The improvement of model performance with the observation of an increasing school catchment area confirms the spatial conflict of schistosomiasis risk modelling based on RS data. It indicates that in a distance of 2 km and larger, the school catchment area integrates environmental conditions that represent the disease transmission conditions corresponding to the school-based measure of prevalence. This impact of different scales of aggregation on model performance has also been shown for the example of aggregating on different administrative levels, where treatment needs were over- or underestimated depending on the level of spatial aggregation and focality of disease distribution (Schur et al., 2012: 11). However, so far no spatial analysis of schistosomiasis risk models based on divergent buffer extents could be identified in the literature. In contrast, the spatial resolution of RS data did not show any impact on model performance, which is a direct consequence of the approach taken in this study, where RS measurements are summarised by their mean within a buffer zone around the school catchment area.

RS variables from different sensors showed that the explanatory power of the Landsat 5 TM-based model increased, when the full set of sensor specific bands and indices were used, whereas this was not the case for models based on RapidEye and Terra MODIS data. This superior performance of Landsat 5 TM could be explained by the additional information provided by shortwave infrared and thermal bands of the Landsat 5 TM, which are not provided by the RapidEye sensor but highly sensitive to water (see Section 4.2.2). In contrast, the rededge band from RapidEye did not provide any further information for modelling schistosomiasis risk. In this analysis, corresponding Terra MODIS reflectance data provided no valuable information for schistosomiasis risk modelling, which could be a consequence of poor data quality due to cloud contamination. Despite the application of the product specific cloud mask, cloud contamination was present in the data.

This study showed that few RS variables were highly important to explain the variation of disease prevalence. However, the importance of RS variables varied strongly between model algorithms and study sites, which confirms the hypothesis that different ecological regions require different RS variables. In the mountainous region of MAN, schistosomiasis risk has predominantly been explained by the distance to water bodies and topographic variables. This fits to the strong topographic gradient and confirms the positive correlation between disease prevalence and stream order of the nearest water body or altitude as derived by Beck-Wörner et al. (2007) for this study region. In contrast, the high importance of slope does not reflect the missing correlation between slope and disease prevalence in the study by Beck-Wörner et al. (2007). The variable "sinks" was identified as a new important RS variable from the non-hierarchical model approach. Topographic variables are considered as proxies for relevant ecological conditions for schistosomiasis-related parasite and snail species such as flow velocity of water or temperature conditions and were also found useful for schistosomiasis risk modelling in Côte d'Ivoire (Beck-Wörner et al., 2007; Vounatsou et al., 2009) and Kenya (Clennon et al., 2007). In the tropical lake-side region of TAB, LST and temperature difference between day and night-time summarised as yearly mean and median together with distance to water bodies and stream order, were the most important variables to predict the spatial schistosomiasis risk. As already demonstrated by Malone et al. (1994) in the Nile delta in Egypt, remotely sensed temperature is a useful proxy to model hydrological conditions at this scale of observation. Due to the poor performance of the models, variable importance in BUF could not

be evaluated. It is assumed, that the importance of variables is attributed rather to noise than information in this case. The fact that each ecological region resulted in dissimilar important variables, provides a further explanation for the poor performance of the cross-ecozonal model presented in Figure 6-9. The different importance of variables per ecozone indicates that valuable information from RS data for one specific ecozone is lost when modelled across different ecozones where this relation is not given.

## 6.4.2 Modelling schistosomiasis risk

In this research, a hierarchical model approach was developed to bridge the spatial gap between school prevalence and relevant environmental conditions for schistosomiasis transmission based on the two different algorithms Random Forest and PLSR. It was shown that the hierarchical model approach improved schistosomiasis risk models in regions where the environment contributes significantly to explain the spatial pattern of the disease. Analysis in the study site TAB, demonstrated that close to 70% of the variance could be explained by hierarchical models based on RS data. Given the complex social-ecological interaction that underlies schistosomiasis transmission, RS is assumed to provide a highly valuable contribution for schistosomiasis risk assessment in this case. However, models derived for the study site BUF have shown that in some regions environmental conditions explain the distribution of disease prevalence poorly. Both very high and very low disease prevalence has been measured in the direct neighbourhood of schools surrounded by sites of high environmental suitability for disease transmission (Figure 5-12). Despite the chronic character of the disease, the time-lag of surveys conducted in the mid-1980s could to some extent explain the missing relation between schistosomiasis prevalence and environmental data derived for the year 2010. In MAN, the hierarchical model approach improved its performance compared to the non-hierarchical only for the PLSR algorithm within school catchment radii up to 1km. In this region, it was generally very difficult to detect water bodies due to its small-scale heterogeneity and frequent coverage by riparian forests. The poor model performance across all ecozones despite an increase in sample size from the minimum of 38 in the TAB site to 184 in all ecozones confirmed the result of Brooker et al. (2001) in Tanzania, where only models fitted within the same ecological zone performed reasonably.

It has to be stated here that the outcome of all spatial models of schistosomiasis risk may be affected by spatial autocorrelation. As analysed in Section 4.1, the response variable of schistosomiasis prevalence is – like most spatial data – affected by spatial autocorrelation until a specific distance between points. This common phenomenon of spatial autocorrelation is on the one hand a relevant information from an ecological point of view, e.g. whether high or low disease prevalence is spatially clustered or shows a trend from rivers to mountainous areas. However, spatial autocorrelation at the same time impacts the predictive performance of spatial models and may lead to mis-estimation of up to 25% (Dormann, 2007: 135). Unfortunately, the impact of spatial autocorrelation on model accuracy could not be evaluated in this study due to sample size constraints. Even for the study site MAN, where the sample size was highest, the reduction of school locations that were located closer than 12.1km according to the modelled range of spatial autocorrelation (Figure 4-3) would have led to only ten remaining school locations, which is an insufficient sample size for statistical modelling. However, the impact of

spatial autocorrelation is assumed to be small due to the typical focal distribution of the disease mentioned above.

In most cases, the Random Forest and PLSR models performed in a comparable way with respect to their predictive power, which indicates that model results are stable. However, the variation between variable importance that result from each model is considerable. Variables found to be highly important for the Random Forest model were not considered important by the PLSR model and *vice versa*. This could be attributed to the different procedure of assessing variable importance by each model approach as described in Section 6.2.3.

The spatial predictions of the study sites MAN and TAB reflected well the variables that were considered most important by the respective variable importance measure. In MAN, the Random Forest model showed a high impact of euclidean distance to water on the spatial prediction (Figure 6-11), whereas in TAB the LST data had a high impact (Figure 6-12). The superior model performance of the hierarchical models in the study site TAB was well reflected by an appropriate spatial prediction of disease risk according to the distribution of reference data. However, the predicted disease risk at those single hotspots away from the well investigated water sites around Lake Taabo can not be evaluated due to the lack of respective data on disease prevalence.

In general, the schistosomiasis risk modelling and prediction in this study is based on a relatively small sample size, which can impact the stability of model performance. Furthermore, the comparison between different study sites and ecozones, respectively, might be impacted by inhomogeneities of input data on schistosomiasis prevalence (Table 4-1) due to different sample sizes, different years of surveys and different distributions of prevalence of samples between the three investigated study sites. Both the  $R^2$  value and the NSE index, are not very sensitive to systematic over- or underprediction of a model (Krause et al., 2005: 90). Nevertheless, the RS based models of schistosomiasis risk resulted in useful spatial predictions of schistosomiasis risk, whereas the hierarchical approach explicitly models the potential disease transmission sites. This spatial prediction can support the planning of disease prevention and control measures in a spatially explicit way and supports the identification of new causal relations of disease transmission in different ecological regions.

## 6.5 Summary of schistosomiasis risk modelling

In summary, this study has demonstrated that RS data have a highly valuable contribution for schistosomiasis risk modelling explaining up to 70% of the variation in disease prevalence. However, it has also been shown in the study site BUF that environmental conditions do in some cases not explain the spatial distribution of the disease. The hierarchical model approach developed in this thesis is capable of bridging the spatial discrepancy between school-based measurements of prevalence and the disease relevant remotely sensed environmental conditions. However, only in the study site TAB, where environmental conditions strongly indicate the spatial distribution of measured disease risk, did this hierarchical approach improve the model considerably. The analysis between different ecological settings provided insights into the variation of importance of RS variables according to the ecozone under investigation. The most important RS variables identified in this study were distance to water bodies and topographic variables for the mountainous region in MAN and LST and the stream order for the

tropical lakeside region in TAB. This underlines further that modelling across ecozones is challenging due to changing relationships between disease prevalence and the environment. Ignoring ecozonal differences strongly deteriorates the performance of schistosomiasis risk modelling. Whereas the spatial resolution of RS data *per se* did not have any impact with the approach taken, the scale of observation around surveyed school locations substantially impacted model accuracy. Best model performances were obtained at larger observational units with radii of 2 km and larger, although this depended on the study site and model approach. The combination of spectral properties and spatial resolution of the Landsat 5 TM data resulted in the best performance of schistosomiasis risk models compared to RapidEye and Terra MODIS data.



## 7 Synthesis and outlook

The overarching goal of this thesis was to bridge relevant disciplines and bring together several strands of scientific inquiry in order to investigate the potential of RS data for profiling disease risk. Emphasis was placed on schistosomiasis. A hierarchical model approach was developed to optimise the application of RS data and to overcome the spatial discrepancy between environmental habitats of disease-related parasites and intermediate host snails and measurements of human infections. In this chapter, the strengths and limitations of RS data are discussed according to the specific research questions (Section 7.1). The transferability of the hierarchical model approach to optimise spatial risk profiling of other environment-related diseases is highlighted in Section 7.2. Future research needs are outlined in Section 7.3.

### 7.1 Strengths and limitations of remote sensing data for schistosomiasis risk profiling

#### ***(1) Which RS data and variables are most useful to model environmental suitability and disease risk?***

Findings presented in this thesis revealed that RS data were highly useful to spatially delineate and quantitatively evaluate environmental suitability for transmission of schistosomiasis and model disease risk. Based on RapidEye and Landsat 5 TM data, the small-scale heterogeneity of water bodies could be well detected, as long as water was not considerably covered by trees. Despite its lower spatial resolution compared to RapidEye, Landsat 5 TM data thereby performed especially well, presumably due to its sensor configuration in the shortwave and thermal infrared spectrum, which is highly suitable to detect water. In contrast, Terra MODIS reflectance data performed unsuitably in detecting smaller water bodies and modelling disease risk, which reflects the limited spatial detail at the expense of the sensors well-known capacity to monitor large-scale temporal dynamics. Despite all three sensors having sufficient nominal temporal resolution for this application, cloud coverage has shown to be a strong limitation when establishing information on seasonal dynamics, especially in the tropical regions of Côte d'Ivoire.

The topographic model derived from ASTER data was highly useful to model disease risk, specifically in the mountainous region of western Côte d'Ivoire. However, the spatial resolution of 30 m provided a significant constraint to characterise the small-scale topographic features to quantify environmental suitability for schistosomiasis-related snails and parasites as proposed in this study. To better assess habitat suitability for freshwater snails, the content of submerged aquatic vegetation is of high ecological relevance. Yet, this measure would require hyperspectral RS data together with comprehensive *in situ* measurements, which was beyond the scope of this thesis.

In the schistosomiasis risk model, some 60 RS variables were investigated, of which only a handful were of relevance in the final model specifications. The most important RS variable identified in this study was the measure of distance to water bodies, which reflects well the overall importance of the spatial location of water sites for the explanation of schistosomiasis risk. For the mountainous region in the study site MAN, the topographic variables "altitude", "slope", "streams" and "sinks" were found of highest importance, which confirms the strong impact of the topographic gradient on schistosomiasis risk, as previously shown by Beck-Wörner et al. (2007). In contrast, the model of the tropical lakeside region of TAB resulted in mean and median temperature difference between day and night-time, LST and stream order as most important variables. Other than for the model of environmental suitability mentioned above, LST contributed significantly to modelling disease risk as has been shown two decades ago (Malone et al., 1994). In this study, the importance of RS variables varied strongly between model algorithms and study sites, which confirms that a single statistical model algorithm does not generate representative results and different ecological regions require different RS variables for modelling schistosomiasis risk. Furthermore, the widely used NDVI for modelling schistosomiasis risk, as well as other vegetation indices investigated in this thesis, did not provide a noteworthy contribution to explain the spatial variation of schistosomiasis prevalence.

The most suitable RS variables to model environmental suitability for schistosomiasis transmission consisted of the multi-temporal derivation of water bodies and the assessment of their riparian vegetation coverage based on high-resolution RS data from the RapidEye and Landsat 5 TM sensors. The regions of permanent water coverage and high vegetation coverage during the dry season outlined the hotspots of environmental suitability for schistosomiasis transmission. The topographic sinks and streams, which could potentially become habitats as a result of rainfall or flooding, complemented the spatial delineation of potential disease transmission sites. However, the accuracy of modelled topographic streams was very poor. Regarding the ecological context of disease-related parasites and snails, RS measurements of water surface temperature are of limited use as surface temperature does not directly indicate the water temperature that impacts parasite and snail development and critical limits of long-term variation and extreme values cannot appropriately be measured and monitored by currently available remote sensors. Additionally, the dynamic changes of water flow velocity between dry and wet seasons cannot be derived based on topographic RS data.

Taken together, the satisfactory performance in delineating environmental suitability for schistosomiasis transmission is a promising feature upon which one might establish an operational monitoring of environmental changes with focus on the suitability of disease transmission in the near future. Hence, this issue warrants further scientific inquiry, as discussed in Section 7.3.

**(2) How can the spatial discrepancy between environmental suitability for schistosomiasis transmission and the measure of disease risk be resolved?**

In this thesis, a hierarchical model approach was developed to overcome the spatial mismatch between school-based disease prevalence data and remotely sensed environmental information, which is of relevance regarding the ecological process of disease transmission. Therefore, the delineation of potential disease transmission sites derived by the mechanistic model of environmental suitability provided the spatial basis to model schistosomiasis risk within the catchment area of the school-based measure of disease prevalence providing the response variable for the model. Particularly for the study site TAB, where environmental conditions contributed strongly to model schistosomiasis risk, the hierarchical model approach improved model performance considerably in comparison to the non-hierarchical model approach, as indicated by the explanatory power close to 70% *versus* lower than 60%, respectively. However, both values are considered high against the background that a complex social-ecological interaction underlies the process of schistosomiasis transmission, which cannot be fully described by means of RS. Limitations of disease risk modelling using either a hierarchical or a non-hierarchical model approach were revealed for the study site BUF, where environmental conditions in general explained the spatial distribution of school-based disease prevalence very poorly. However, keeping the ecological process of disease transmission in mind and observing this process from a geographical and spatially explicit perspective, a non-hierarchical approach would not fully exploit the potential of RS data for profiling schistosomiasis risk as the relevant RS signal would be averaged between relevant and non-relevant sites.

The superior performance of the hierarchical model approach in the study site TAB was well reflected by the spatial prediction of disease risk, which is in agreement with the spatial distribution of school-based measures of disease prevalence. Furthermore, the issue of anisotropy inherent in the spatial structure of the data (Chammartin et al., 2013) is directly addressed by the hierarchical model approach. However, insufficient data were available to validate this prediction with independent data of school-based prevalence.

**(3) Which scale is most appropriate for spatial modelling of schistosomiasis risk?**

The impact of scale on schistosomiasis risk modelling was investigated in this thesis observing the scale of RS data, hence its spatial resolution, and the scale of observation regarding the environment within the catchment area of modelled school locations. The results confirmed the hypothesis that suitable environments for disease transmission rarely occur at the location of the school but within its larger catchment area up to a distance of 5 km, which is reflected by the increasing model performance with increasing extent of the school catchment area considered in the model. From a distance of 2 km from school location and above, the explanatory power of the schistosomiasis risk models either peaked at around 3 km or levelled out at its highest score for analysis within a buffer extent between 2 and 4 km. Due to the aggregation of RS measurements within the school catchment area investigated, the spatial resolution *per se* did not impact the model performance. However, based on these findings it can be assumed that spatial resolution of RS data greater than 500 m has an impact on model performance if no buffer is used for the spatial risk modelling as done in previous studies.

#### **(4) How do different ecozones impact the performance of schistosomiasis risk models in West Africa?**

Based on the three selected study sites in different ecological regions of West Africa that range from dry savannah (study site BUF in Burkina Faso) to tropical rainforest including flat and mountainous regions (study sites TAB and MAN in Côte d'Ivoire, respectively), the impact of different ecozones on model performance was investigated. The first step of disease risk profiling - modelling environmental suitability – already showed that the constitution and characteristics of specific sites for potential schistosomiasis transmission reflect an obvious impact of different ecological regions. To underscore this point, the riparian vegetation coverage during the dry season in the savannah region of Burkina Faso revealed environmental suitability as it detects irrigated agriculture as hotspot for disease transmission. In contrast, the densely forested riparian zones in the study sites of Côte d'Ivoire did not represent such specifically suitable environments, although the RS measures were comparable. The crucial role of divergent ecological regions is further emphasised by the results obtained from the statistical model of schistosomiasis risk, highlighting specific environmental features. Here, the model of each ecological region identified different sets of variables, which indicate that valuable information of RS data for one specific ecozone would be lost when models are developed across different ecozones, where such prior relations are not given. In view of the poor performance of the model across all ecozones, this study confirms the result of Brooker et al. (2001) from the United Republic of Tanzania, emphasising that only models fitted within the same ecological zone perform with reasonable fidelity.

## 7.2 Transferability of the hierarchical model approach to other environment-related diseases

Within this PhD thesis, a hierarchical model approach has been developed to bridge the spatial discrepancy between environmental suitability for disease-related parasites and snails and human infection measurements. Spatial processes of disease transmission are complex, this was illustrated and conceptualised for the water-based disease schistosomiasis that is the focus of the current work (see Section 3.3.1). However, schistosomiasis is not the only environment-related disease, where RS data are useful for spatial risk profiling. Hence, the question of transferability arises with regard to the optimised application developed in this thesis to other environment-related diseases, which is offered for discussion.

As outlined in the introduction (see Section 1.1), there are several categories of environment-related diseases, such as vector-borne diseases (e.g. malaria and dengue), aerosol-borne diseases (e.g. avian influenza due to H5N1 virus), soil-borne diseases (e.g. hookworm infection) or food-borne diseases (e.g. salmonellosis). In order to evaluate the transferability of the hierarchical model approach, the spatial processes of disease transmission have to be reviewed for the different diseases. Even if exposure pathways are very similar such as schistosomiasis referring to water-based and cholera to water-borne diseases, spatial processes of transmission are different. Whereas human infection with *Schistosoma* spp. results from penetration of the

parasite through the intact skin at a suitable freshwater body where intermediate hosts have released cercariae, an infection with the *Vibrio cholerae* bacterium occurs during consumption of contaminated water. With regards to the spatial processes of schistosomiasis transmission, human infection takes place at the infested water body, while, in the case of cholera, transmission can happen anywhere, as it is governed by contaminated water that is consumed, most often due to sub-standard hygienic conditions. Nevertheless, RS data have been used to monitor temporal and spatial variations of chlorophyll abundance and sea surface temperature as proxy for dynamics of cholera (Lobitz et al., 2000; Jutla et al., 2010). In this context, the linkage between human infection and environmental suitability for the respective disease agent is affected by a spatial discrepancy of measurements. For this application, the hierarchical model approach developed in this thesis would need some specific modifications to link respective regions beyond a 5 km buffer zone.

For the case of vector-borne diseases such as malaria and dengue, environmental data derived from RS aims at evaluating habitat conditions of the respective vector species. Of note, there is an inherent spatial discrepancy between human infection and relevant environmental conditions for vectors. However, in this case, the spatial relation is modified by the species-specific flight and drift range. In contrast to schistosomiasis, the delineation of potential habitats for vector species is rather fuzzy, especially if these are well adapted to often man-made microhabitats, such as old tyres stored around households that collect rain water. For the time being at this small scale, vector habitats cannot be detected by means of RS.

The direct transferability of this hierarchical model approach is limited due to the specific ecological requirements of environment-related diseases for transmission of the pathogen from human to human and its consequences on spatial processes. As soon as disease transmission from human to human occurs, which is the case for the current epidemic of the Ebola in West Africa (Butler, 2014), the potential of RS data to spatially model disease risk is highly limited. Despite the fact that this environment-related disease has its origin in areas adjacent to pristine habitats of primates or *Chiroptera*, which can - similar to other species - be monitored by RS data, this environmental information could then rather provide a basis for a dispersion model of the fuzzy spread of the disease among humans. Nevertheless, what can be drawn from the hierarchical model approach developed in this thesis is the crucial step to think spatially and question the specific ecological context when using RS data for disease risk profiling to fully exploit its potential.

## 7.3 Future research needs

Against the background of this systematic investigation of optimised RS applications for risk profiling of schistosomiasis, there remains the fundamental need to establish a generalised framework that synthesises the spatial relations of environment-related diseases and its implications on RS-based analyses. Such a framework could substantially contribute to bridge the disciplines of geographic RS and epidemiology through a common spatial perspective. Additionally, the yet limited transferability of the hierarchical model approach developed in this research needs to be further investigated and integrated in this generalised conceptual underpinning.

The model of environmental suitability needs to be complemented with ground-truthed field data, which could validate not only the RS variables themselves but also parameterise the derived variable and index suitability in reference to the measured prevalence of schistosomiasis-related parasite and snail species at transmission sites. Furthermore, this model can also be tested and adjusted with respect to similar disease ecologies as given for fascioliasis in wildlife, livestock and humans (Mas-Coma et al., 2009; Quayle et al., 2010). Moreover, suitable environments for schistosomiasis transmission are closely related to suitable habitats for mosquito breeding sites (Keiser et al., 2005) and could therefore provide useful information to further assess the environmental suitability for vector-borne diseases such as malaria, dengue or Rift Valley fever (Martens et al., 1997; Linthicum et al., 1999)

Future models of disease risk based on RS data need to consider the ecotonal transitions in their model approach. Based on the results of this thesis, it is suggested to establish separate models for each ecological region and consider the variation in landscape structure as well as the vertical structure of the modelled area. Diseases have, however, more triggers than just the environment. Hence, RS is only one building block in the research on the complex and multi-faceted phenomenon of disease transmission. Thus, in a next step, RS-based schistosomiasis risk models need to be complemented by demographic, educational, socio-economic and political data to model the comprehensive social-ecological process of disease transmission. To draw causal conclusions on the force of transmission from RS data, these comprehensive social-ecological models again need data on parasite prevalence both in humans and snails.

This work has shown that in order to fully exploit the potential of RS data, there are still several challenges imposing the spatial profiling of risk for a single disease. In practice however, polyparasitism is widespread in the developing world and resources for disease control and prevention are limited. Thus, the potential of RS data needs further exploration to determine the scope and limits within a modelling framework to predict co-endemic areas, where different parasitic diseases coexist (Brooker and Utzinger, 2007). In this regard, the spatial approach of Raso et al. (2006) has already shown that elevation was an important variable to predict the prevalence of *S. mansoni*-hookworm co-infection. Building upon the findings of this study, RS data are expected to fundamentally contribute to an integrative model of polyparasitic infection risk if the application of data and modelling respects the specific ecological requirements of the targeted diseases.

# References

- Abdel-Malek, E. 1958. Factors conditioning the habitat of bilharziasis intermediate hosts of the family Planorbidae. *Bull World Health Organ*, 18, 785-818.
- Abdel-Rahman, M.S., El-Bahy, M.M., Malone, J.B., Thompson, R.A. & El Bahy, N.M. 2001. Geographic information systems as a tool for control program management for schistosomiasis in Egypt. *Acta Trop*, 79, 49-57.
- Adamson, P.B. 1976. Schistosomiasis in antiquity. *Med Hist*, 20, 176-88.
- Albertson, M.L. & Simons, D.B. 1964. Fluid Mechanics. In: Chow, V.T. (ed.) *Handbook of applied hydrology: A compendium of water-resources technology*. New York, USA: McGraw-Hill.
- Albertz, J. 2001. *Einführung in die Fernerkundung - Grundlagen der Interpretation von Luft- und Satellitenbildern*, Darmstadt, Wissenschaftliche Buchgesellschaft.
- Andersen, C.M. & Bro, R. 2010. Variable selection in regression—a tutorial. *Journal of Chemometrics*, 24, 728–737.
- Appleton, C.C. 1978. Review of the literature on abiotic factors influencing the distribution and life-cycles of bilharziasis intermediate host snails. *Malacological Review*, 11, 1-25.
- Arcement, G.J. 1989. Guide for selecting Manning's roughness coefficients for natural channels and flood plains. *Water-Supply Paper*. US Department of the Interior - United States Geological Survey.
- Arino, O., Kalogirou, V., Perez, J.R., Bontemps, S., Defourny, P. & Van Bogaert, E. 2011. GlobCover 2009 - Products Description and Validation Report. European Space Agency (ESA) & Université Catholique de Louvain.
- Ayala, D., Costantini, C., Ose, K., Kamdem, G., Antonio-Nkondjio, C., Agbor, J.-P., Awono-Ambene, P., Fontenille, D. & Simard, F. 2009. Habitat suitability and ecological niche profile of major malaria vectors in Cameroon. *Malaria Journal*, 8, 307.
- Barsi, J.A., Hook, S.J., Schott, J.R., Raqueno, N.G. & Markham, B.L. 2007. Landsat-5 Thematic Mapper Thermal Band Calibration Update. *IEEE Geoscience and Remote Sensing Letters*, 4, 552-555.
- Bashford, A. & Tracy, S.W. 2012. Introduction: modern airs, waters, and places. *Bull Hist Med*, 86, 495-514.
- Beck-Wörner, C., Raso, G., Vounatsou, P., N'Goran, E.K., Rigo, G., Parlow, E. & Utzinger, J. 2007. Bayesian spatial risk prediction of *Schistosoma mansoni* infection in western Cote d'Ivoire using a remotely-sensed digital elevation model. *Am J Trop Med Hyg*, 76, 956-963.
- Beck, L.R., Lobitz, B.M. & Wood, B.L. 2000. Remote Sensing and Human Health: New Sensors and New Opportunities. *Emerging Infectious Diseases*, 6, 217-226.
- Beck, L.R., Rodriguez, M.H., Dister, S.W., Rodriguez, A.D., Rejmankova, E., Ulloa, A., Meza, R.A., Roberts, D.R., Paris, J.F., Spanner, M.A., et al. 1994. Remote sensing as a landscape epidemiologic tool to identify villages at high risk for malaria transmission. *Am J Trop Med Hyg*, 51, 271-280.

- Bella, H., de, C.M.T.F., Omer, A.H. & Vaughan, J.P. 1980. Migrant workers and schistosomiasis in the Gezira, Sudan. *Trans R Soc Trop Med Hyg*, 74, 36-39.
- Belton, S. & Stewart, T.S. 2002. *Multiple Criteria Decision Analysis. An Integrated Approach.*, Massachusetts, Kluwer Academic Publishers.
- Bergquist, R., Malone, J.B. & Kristensen, T.K. 2000. Schistosomiasis information systems and control of snail-borne diseases. *Parasitol Today*, 16, 363-364.
- Bierwirth, P.N., Lee, T.J. & Burne, R.V. 1993. Shallow SeamFloor Reflectance and Water Depth Derived by Unmixing Multispectral Imagery. *Photogrammetric Engineering & Remote Sensing*, 59, 331-338.
- Bilharz, T.M. 1852. Fernere Beobachtungen über das die Pfortader des Menschen bewohnende Distomum haematobium und sein Verhältniss zu gewissen pathologischen Bildungen aus brieflichen Mittheilungen an Professor v.Siebold vom 29 Marz 1852. *Zeitschrift für Wissenschaftliche Zoologie Leipzig*, 4, 72-76.
- Blackbridge. 2013. Satellite Imagery Product Specifications.
- Boelee, E., Cecchi, P. & Kone, A. 2009. Health impacts of small reservoirs in Burkina Faso. Colombo, Sri Lanka: International Water Management Institute.
- Boslaugh, S. & Watters, P.A. 2008. *Statistics in a nutshell*, O`Reilly Media, Inc.
- Breiman, L. 1996. Bagging Predictors. *Machine Learning*, 24, 123-140.
- Breiman, L. 2001. Random Forests. *Machine Learning*, 45, 5-32.
- Breiman, L., Friedman, J.H., Olshen, R.A. & Stone, C.J. 1984. *Classification and Regression Trees*, Belmont, California, Wadsworth International Group.
- Bright, P.R., Buxton, H.T., Balistreri, L.S., Barber, L.B., Chapelle, F.H., Cross, P.C., Krabbenhoft, D.P., Plumlee, G.S., Sleeman, J.M., Tillitt, D.E., et al. 2013. U.S. Geological Survey Environmental Health Science Strategy - Providing Environmental Health Science for a Changing World. U.S. Department of the Interior, U.S. Geological Survey.
- Brooker, S. 2002. Schistosomes, snails and satellites. *Acta Tropica*, 82, 207-214.
- Brooker, S. 2007. Spatial epidemiology of human schistosomiasis in Africa: risk models, transmission dynamics and control. *Trans R Soc Trop Med Hyg*, 101, 1-8.
- Brooker, S. & Clements, A.C. 2009. Spatial heterogeneity of parasite co-infection: Determinants and geostatistical prediction at regional scales. *Int J Parasitol*, 39, 591-597.
- Brooker, S., Clements, A.C. & Bundy, D.A. 2006. Global epidemiology, ecology and control of soil-transmitted helminth infections. *Adv Parasitol*, 62, 221-261.
- Brooker, S., Donnelly, C.A. & Guyatt, H.L. 2000a. Estimating the number of helminthic infections in the Republic of Cameroon from data on infection prevalence in schoolchildren. *Bull World Health Organ*, 78, 1456-1465.
- Brooker, S., Hay, S.I., Issae, W., Hall, A., Kihamia, C.M., Lwambo, N.J., Wint, W., Rogers, D.J. & Bundy, D.A. 2001. Predicting the distribution of urinary schistosomiasis in Tanzania using satellite sensor data. *Trop Med Int Health*, 6, 998-1007.
- Brooker, S., Hay, S.I., Tchuem Tchuenté, L.A. & Ratard, R. 2002. Using NOAA-AVHRR data to model human helminth distributions in planning disease control in Cameroon, West Africa. *Photogrammetric Engineering & Remote Sensing*, 68, 175-179.
- Brooker, S., Kabatereine, N.B., Smith, J.L., Mupfasoni, D., Mwanje, M.T., Ndayishimiye, O., Lwambo, N.J., Mbotha, D., Karanja, P., Mwandawiro, C., et al. 2009. An updated atlas of human helminth infections: the example of East Africa. *Int J Health Geogr*, 8, 42.
- Brooker, S. & Michael, E. 2000. The potential of geographic information systems and remote sensing in epidemiology and control of human helminth infections. In: Hay, S.I., Randolph, S.E., Rogers, D.J.,



- Baker, J.R., Muller, R. & Rollinson, D. (eds.) *Advances in Parasitology - Remote Sensing and Geographical Information Systems in Epidemiology*. Oxford: Academic Press.
- Brooker, S., Rowlands, M., Haller, L., Savioli, L. & Bundy, D.A.P. 2000b. Towards an Atlas of Human Helminth Infection in sub-Saharan Africa: The Use of Geographical Information Systems (GIS). *Parasitology Today*, 16, 303-307.
- Brooker, S. & Utzinger, J. 2007. Integrated disease mapping in a polyparasitic world. *Geospat Health*, 1, 141-146.
- Brooks, R.P. 1997. Improving Habitat Suitability Index Models. *Wildlife Society Bulletin*, 25, 163-167.
- Brown, S.K., Buja, K.R., Jury, S.H., Monaco, M.E. & Banner, A. 2000. Habitat Suitability Index Models for Eight Fish and Invertebrate Species in Casco and Sheepscot Bays, Maine. *North American Journal of Fisheries Management*, 20, 408-435.
- Bruun, B. & Aagaard-Hansen, J. 2008. The social context of schistosomiasis and its control: an introduction and annotated bibliography. *Special Programme for Research & Training in Tropical Diseases (TDR) sponsored by UNICEF / UNDP / World Bank / WHO*. Geneva: World Health Organization.
- Bundy, D.A.P. & Blumenthal, U. 1990. Human behaviour and the epidemiology of helminth infection. In: Barnard, C. & Behnke, J.M. (eds.) *Parasitism and Host Behaviour*. London: Taylor and Francis.
- Butler, D. 2014. Global Ebola response kicks into gear at last. *Nature*, 513, 469.
- Butterworth, A.E. 1993. Immunology of Schistosomiasis. In: Jordan, P., Webbe, G. & Sturrock, R.F. (eds.) *Human Schistosomiasis*. Wallingford: CAB International.
- Chabasse, D., Bertrand, G., Leroux, J.P., Gauthey, N. & Hocquet, P. 1985. [Developmental bilharziasis caused by *Schistosoma mansoni* discovered 37 years after infestation]. *Bull Soc Pathol Exot Filiales*, 78, 643-647.
- Chammartin, F., Hurlimann, E., Raso, G., N'Goran, E.K., Utzinger, J. & Vounatsou, P. 2013. Statistical methodological issues in mapping historical schistosomiasis survey data. *Acta Trop*, 128, 345-352.
- Chandiwana, S.K. 1987. Community water-contact patterns and the transmission of *Schistosoma haematobium* in the highveld region of Zimbabwe. *Soc Sci Med*, 25, 495-505.
- Chapin III, F.S., Zavaleta, E.S., Eviner, V.T., Naylor, R.L., Vitousek, P.M., Reynolds, H.L., Hooper, D.U., Lavorel, S., Sala, O.E., Hobbie, S.E., et al. 2000. Consequences of changing biodiversity. *Nature*, 405, 234-242.
- Chatelain, C., Dao, H., Gautier, L. & Spichiger, R. 2004. Forest cover changes in Cote d'Ivoire and Upper Guinea. In: Poorter, L., Bongers, F., Kouamé, F.Y.N. & Hawthorne, W.D. (eds.) *Biodiversity of West African Forests - An Ecological Atlas of Woody Plant Species*. Oxon, UK: CAB International.
- Cheesmond, A. 1980. Migrant workers and schistosomiasis in the Gezira, Sudan. *Trans R Soc Trop Med Hyg*, 74, 691-692.
- Chmielewski, F.-M., Hupfer, P., Kuttler, W. & Pethe, H. 1998. *Witterung un Klima: Eine Einführung in die Meteorologie und Klimatologie*, Stuttgart & Leipzig, Germany, B.G. Teubner.
- Clark, P.J. & Evans, F.C. 1954. Distance to Nearest Neighbor as a Measure of Spatial Relationships in Populations. *Ecology*, 35, 445-453.
- Clegg, J.A. 1965. In Vitro Cultivation of *Schistosoma Mansoni*. *Exp Parasitol*, 16, 133-147.
- Clements, A., Firth, S., Dembele, R., Garba, A., Toure, S., Sacko, M., Landoure, A., Bosque-Oliva, E., Barnett, A., Brooker, S., et al. 2009a. Use of Bayesian geostatistical prediction to estimate local variations in *Schistosoma haematobium* infection in western Africa. *Bull World Health Organ*, 87, 921-929.
- Clements, A.C., Bosque-Oliva, E., Sacko, M., Landoure, A., Dembele, R., Traore, M., Coulibaly, G., Gabrielli, A.F., Fenwick, A. & Brooker, S. 2009b. A comparative study of the spatial distribution of schistosomiasis in Mali in 1984-1989 and 2004-2006. *PLoS Negl Trop Dis*, 3, 1-11.

## References

---

- Clements, A.C., Garba, A., Sacko, M., Toure, S., Dembele, R., Landoure, A., Bosque-Oliva, E., Gabrielli, A.F. & Fenwick, A. 2008a. Mapping the probability of schistosomiasis and associated uncertainty, West Africa. *Emerg Infect Dis*, 14, 1629-1632.
- Clements, A.C., Lwambo, N.J., Blair, L., Nyandindi, U., Kaatano, G., Kinung'hi, S., Webster, J.P., Fenwick, A. & Brooker, S. 2006a. Bayesian spatial analysis and disease mapping: tools to enhance planning and implementation of a schistosomiasis control programme in Tanzania. *Trop Med Int Health*, 11, 490-503.
- Clements, A.C., Moyeed, R. & Brooker, S. 2006b. Bayesian geostatistical prediction of the intensity of infection with *Schistosoma mansoni* in East Africa. *Parasitology*, 133, 711-719.
- Clements, A.C.A., Brooker, S., Nyandindi, U., Fenwick, A. & Blair, L. 2008b. Bayesian spatial analysis of a national urinary schistosomiasis questionnaire to assist geographic targeting of schistosomiasis control in Tanzania, East Africa. *International Journal for Parasitology*, 38, 401-415.
- Clennon, J.A., King, C.H., Muchiri, E.M., Kariuki, H.C., Ouma, J.H., 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, 443-448.
- Clennon, J.A., King, C.H., Muchiri, E.M. & Kitron, U. 2007. Hydrological modelling of snail dispersal patterns in Msambweni, Kenya and potential resurgence of *Schistosoma haematobium* transmission. *Parasitology*, 134, 683-693.
- Clennon, J.A., Mungai, P.L., Muchiri, E.M., King, C.H. & Kitron, U. 2006. Spatial and temporal variations in local transmission of *Schistosoma haematobium* in Msambweni, Kenya. *Am J Trop Med Hyg*, 75, 1034-1041.
- Cline, B.L. 1970. New eyes for epidemiologists: aerial photography and other remote sensing techniques. *Am J Epidemiol*, 92, 85-89.
- Colley, D.G., Bustinduy, A.L., Secor, W.E. & King, C.H. 2014. Human schistosomiasis. *Lancet*.
- Confalonieri, U. & McMichael, A. 2006. Global Environmental Change and Human Health - Science Plan and Implementation Strategy. Earth System Science Partnership.
- Confalonieri, U., Menne, B., Akhtar, R., Ebi, K.L., Hauengue, M., Kovats, R.S., Revich, B. & Woodward, A. 2007. Human health. In: Parry, M.L., Canziani, O.F., Palutikof, J.P., Van Der Linden, P.J. & Hanson, C.E. (eds.) *Climate Change 2007: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge: Cambridge University Press.
- Corvalan, C., Hales, S. & McMichael, A. 2005. Ecosystems and human well-being : health synthesis. In: Sarukhan, J. & Whyte, A. (eds.) *Millennium Ecosystem Assessment*. World Health Organization.
- Cross, E.R., Sheffield, C., Perrine, R. & Pazzaglia, G. 1984. Predicting areas endemic for schistosomiasis using weather variables and a Landsat data base. *Mil Med*, 149, 542-544.
- Curran, P.J., Atkinson, P.M., Foody, G.M. & Milton, E.J. 2000. Linking remote sensing, land cover and disease. In: Hay, S.I., Randolph, S.E., Rogers, D.J., Baker, J.R., Muller, R. & Rollinson, D. (eds.) *Advances in Parasitology - Remote Sensing and Geographical Information Systems in Epidemiology*. Oxford: Academic Press.
- Czajkowskia, K.P., Goward, S.N., Shirey, D. & Walz, A. 2002. Thermal remote sensing of near-surface water vapor. *Remote Sensing of Environment*, 79, 253- 265.
- Dadjoari, D.M. 26.03.2011. National Schistosomiasis Control Program, Burkina Faso. *personal communication*.
- Dayal, B.S. & MacGregor, J.F. 1997. Improved PLS algorithms. *Journal of Chemometrics*, 11, 73-85.
- DCLG 2009. Multi-criteria analysis: a manual. London: Department for Communities and Local Government.

- De Jong, S. & Ter Braak, C.J.F. 1994. Comments on the PLS kernel algorithm. *Journal of Chemometrics*, 8, 169-174.
- De Vlas, S.J., Van Oortmarssen, G.J., Gryseels, B., Polderman, A.M., Plaisier, A.P. & Habbema, J.D. 1996. SCHISTOSIM: a microsimulation model for the epidemiology and control of schistosomiasis. *Am J Trop Med Hyg*, 55, 170-175.
- Deschiens, R. 1954. Incidence de la minéralisation de l'eau sur les mollusques vecteurs des bilharzioses. Consequences pratiques. *Bulletin de la Société de pathologie exotique*, 47, 915-929.
- Descloux, E., Mangeas, M., Menkes, C.E., Lengaigne, M., Leroy, A., Tehei, T., Guillaumot, L., Teurlai, M., Gourinat, A.-C., Benzler, J., et al. 2012. Climate-Based Models for Understanding and Forecasting Dengue Epidemics. *PLoS Negl Trop Dis*, 6.
- Di Gregorio, A. & Jansen, L.J.M. 2005. Land Cover Classification System Classification concepts and user manual Software version (2). Rome: Food and Agricultural Organization of the United Nations (FAO).
- Dianou, D., Poda, J.N., Sorgho, H., Wango, S.P. & Sondo, K.B. 2003. Hydraulic Plannings and Schistosomiasis: Case of Sourou in Burkina Faso. *The International Journal of Applied Research in Veterinary Medicine*, 1, 105-111.
- Dietterich, T.G. 1999. An Experimental Comparison of Three Methods for Constructing Ensembles of Decision Trees: Bagging, Boosting, and Randomization. *Machine Learning*, 1-22.
- Dipama, J.-M. & Anne, C.A.T. 2010. The Biophysical Environment: Geology and geomorphology. In: Thiombiano, A. & Kampmann, D. (eds.) *Biodiversity Atlas of West Africa, Volume II: Burkina Faso*. Ouagadougou & Frankfurt/Main.
- Dipama, J.M. 2010a. The Biophysical Environment: Climate. In: Thiombiano, A. & Kampmann, D. (eds.) *Biodiversity Atlas of West Africa, Volume II: Burkina Faso*. Ouagadougou & Frankfurt/Main.
- Dipama, J.M. 2010b. The Biophysical Environment: Hydrology. In: Thiombiano, A. & Kampmann, D. (eds.) *Biodiversity Atlas of West Africa, Volume II: Burkina Faso*. Ouagadougou & Frankfurt/Main.
- DLR. 2014a. CATENA [Online]. Weßling, Germany: German Aerospace Center (DLR). Available: [http://www.dlr.de/eoc/desktopdefault.aspx/tabid-5444/9113\\_read-17838/](http://www.dlr.de/eoc/desktopdefault.aspx/tabid-5444/9113_read-17838/) Accessed 2014-02-25.
- DLR. 2014b. RESA - RapidEye Science Archive [Online]. Deutsches Zentrum für Luft- und Raumfahrt (DLR) e.V. Available: <http://resaweb.dlr.de/> Accessed 2014-02-24.
- DLR. 2014c. TerraSAR-X - Deutschlands Radar-Auge im All [Online]. Available: [http://www.dlr.de/dlr/desktopdefault.aspx/tabid-10377/565\\_read-436/#gallery/350](http://www.dlr.de/dlr/desktopdefault.aspx/tabid-10377/565_read-436/#gallery/350) Accessed 2014-06-13.
- Doenhoff, M.J., Hagan, P., Cioli, D., Southgate, V., Pica-Mattoccia, L., Botros, S., Coles, G., Tchuem Tchuente, L.A., Mbaye, A. & Engels, D. 2009. Praziquantel: its use in control of schistosomiasis in sub-Saharan Africa and current research needs. *Parasitology*, 136, 1825-1835.
- Dormann, C.F. 2007. Effects of incorporating spatial autocorrelation into the analysis of species distribution data. *Global Ecology and Biogeography*, 16, 129-138.
- Doumbia, M. 2010. The Human Environment: Economic organization. In: Konaté, S. & Kampmann, D. (eds.) *Biodiversity Atlas of West Africa, Volume III: Cote d'Ivoire*. Abidjan & Frankfurt/Main.
- Doumenge, J.P. & Mott, K.E. 1987. Global distribution of schistosomiasis: CEGET/WHO atlas. *World Health Stat Q*, 37, 186-199.
- Dray, S. & Dufour, A.-B. 2007. The ade4 Package: Implementing the Duality Diagram for Ecologists. *Journal of Statistical Software*, 22, 1-20.
- Duncan, B.W., Breininger, D.R., Schmalzer, P.A. & Larson, V.L. 1995. Validating Florida Scrub Jay Habitat Suitability Model, Using Demography Data on Kennedy Space Center. *Photogrammetric Engineering and Remote Sensing*, 11, 1361-1370.

- Dussault, C., Courtois, R. & Ouellet, J.-P. 2006. A habitat suitability index model to assess moose habitat selection at multiple spatial scales. *Canadian Journal of Forest Research*, 36, 1097-1107.
- Dye, C., Boerma, T., Evans, D., Harries, A., Lienhardt, C., McManus, J., Pang, T., Terry, R. & Zachariah, R. 2013. Research for Universal Health Coverage. *The World Health Report 2013*. Luxembourg: WHO.
- Ekpo, U.F., Mafiana, C.F., Adeofun, C.O., Solarin, A.R. & Idowu, A.B. 2008. Geographical information system and predictive risk maps of urinary schistosomiasis in Ogun State, Nigeria. *BMC Infect Dis*, 8, 74.
- Elliott, P., Wakefield, J.C., Best, N. & Briggs, D.J. 2006. Spatial epidemiology: methods and applications. In: Elliott, P., Wakefield, J.C., Best, N. & Briggs, D.J. (eds.) *Spatial Epidemiology: Methods and Applications*. New York: Oxford University Press.
- Elliott, P. & Wartenberg, D. 2004. Spatial epidemiology: current approaches and future challenges. *Environ Health Perspect*, 112, 998-1006.
- Esch, T., Marconcini, M., Felbier, A., Roth, A., Heldens, W., Huber, M., Schwinger, M., Taubenböck, H., Müller, A. & Dech, S. 2013. Urban Footprint Processor - Fully Automated Processing Chain Generating Settlement Masks From Global Data of the TanDEM-X Mission. *Geoscience and Remote Sensing Letters, IEEE*, 10, 1617-1621.
- Esch, T., Taubenböck, H., Roth, A., Heldens, W., Felbier, A., Thiel, M., Schmidt, M., Müller, A. & Dech, S. 2012. TanDEM-X mission—new perspectives for the inventory and monitoring of global settlement patterns. *Journal of Applied Remote Sensing*, 6.
- Esch, T., Thiel, M., Schenk, A., Roth, A., Müller, A. & Dech, S. 2010. Delineation of Urban Footprints From TerraSAR-X Data by Analyzing Speckle Characteristics and Intensity Information. *Geoscience and Remote Sensing, IEEE Transactions on*, 48, 905-916.
- FAO 2014a. aquastat. Food and Agricultural Organization of the United Nations.
- FAO. 2014b. *GeoNetwork: Find and analyze geo-spatial data* [Online]. Food and Agricultural Organization of the United Nations. Available: <http://www.fao.org/geonetwork/srv/en/main.home> Accessed January 14 2014.
- FAO, IIASA, ISRIC, ISS-CAS & JRC. 2012. Harmonized World Soil Database (version 1.2).
- Farooq, M. 1973. Historical Development. In: Ansari, N. (ed.) *Epidemiology and Control of Schistosomiasis (Bilharziasis)*. Basel and Baltimore: S. Karger and University Park Press.
- Farooq, M., Nielsen, J., Samaan, S.A., Mallah, M.B. & Allam, A.A. 1966. The epidemiology of *Schistosoma haematobium* and *S. mansoni* infections in the Egypt-49 project area. 2. Prevalence of bilharziasis in relation to personal attributes and habits. *Bull World Health Organ*, 35, 293-318.
- Fenwick, A. 2006. Waterborne infectious diseases--could they be consigned to history? *Science*, 313, 1077-81.
- Fenwick, A., Webster, J.P., Bosque-Oliva, E., Blair, L., Fleming, F.M., Zhang, Y., Garba, A., Stothard, J.R., Gabrielli, A.F., Clements, A.C., et al. 2009. The Schistosomiasis Control Initiative (SCI): rationale, development and implementation from 2002-2008. *Parasitology*, 136, 1719-1730.
- Fisher, J. & Mustard, J.F. 2004. High spatial resolution sea surface climatology from Landsat thermal infrared data. *Remote Sensing of Environment*, 90, 293-307.
- Foley, J.A., DeFries, R., Asner, G.P., Barford, C., Bonan, G., Carpenter, S.R., Chapin, F.S., Coe, M.T., Daily, G.C., Gibbs, H.K., et al. 2005. Global Consequences of Land Use. *Science*, 309, 570-574.
- Fortin, M.J. & Dale, M. 2005. *Spatial Analysis: A Guide for Ecologists.*, Cambridge, UK, Cambridge University Press.
- Franklin, J. 2010. *Mapping Species Distribution*, Cambridge University Press.
- Gagnon, A.S., Smoyer-Tomic, K.E. & Bush, A.B. 2002. The El Niño southern oscillation and malaria epidemics in South America. *Int J Biometeorol*, 46, 81-89.

- Gao, B.-c. 1996. NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sensing of Environment*, 58, 257-266.
- Garthwaite, P.H. 1994. An Interpretation of Partial Least Squares. *Journal of the American Statistical Association*, 89, 122-127.
- Geladi, P. & Kowalski, B.R. 1986. Partial least-squares regression: a tutorial. *Analytica Chimica Acta*, 185, 1-17.
- Gibson, M. & Warren, K.S. 1970. Capture of *Schistosoma mansoni* miracidia and cercariae by carnivorous aquatic vascular plants of the genus *Utricularia*. *Bull World Health Organ*, 42, 833-835.
- Gillespie, T.W., Foody, G.M., Rocchini, D., Giorgi, A.P. & Saatchi, S. 2008. Measuring and modelling biodiversity from space. *Progress in Physical Geography*, 32, 203-221.
- Gilvear, D.J., Waters, T.M. & Milner, A.M. 1995. Image analysis of aerial photography to quantify changes in channel morphology and instream habitat following placer mining in interior Alaska. *Freshwater Biology*, 34, 389-398.
- Githeko, A.K., Lindsay, S.W., Confalonieri, U.E. & Patz, J.A. 2000. Climate change and vector-borne diseases: a regional analysis. *Bull World Health Organ*, 78, 1136-1147.
- Goetz, S.J., Prince, S.D. & Small, J. 2000. Advances in satellite remote sensing of environmental variables for epidemiological applications. In: Hay, S.I., Randolph, S.E., Rogers, D.J., Baker, J.R., Muller, R. & Rollinson, D. (eds.) *Advances in Parasitology - Remote Sensing and Geographical Information Systems in Epidemiology*. Oxford: Academic Press.
- GoogleInc. 2010. Google Earth (Version). Google.
- Gornitz, V. 1985. A survey of anthropogenic vegetation changes in West Africa during the last century — climatic implications. *Climatic Change*, 7, 285-325.
- Gratz, N.G. 1999. Emerging and resurging vector-borne diseases. *Annu Rev Entomol*, 44, 51-75.
- Grimes, J.E.T., Croll, D., Harrison, W.E., Utzinger, J., Freeman, M.C. & Templeton, M.R. 2014 - under review. Relationship between Water and Sanitation, and Schistosomiasis: A Systematic Review and Meta-Analysis. *PLoS Negl Trop Dis*.
- Grinnell, J. 1917. Field Tests of Theories Concerning Distributional Control. *The American Naturalist*, 51, 115-128.
- Gryseels, B., Polman, K., Clerinx, J. & Kestens, L. 2006. Human schistosomiasis. *The Lancet*, 368, 1106-1118.
- Guernier, V., Hochberg, M.E. & Guegan, J.F. 2004. Ecology drives the worldwide distribution of human diseases. *PLoS Biol*, 2, 740-746.
- Guimaraes, R.J., Freitas, C.C., Dutra, L.V., Scholte, R.G., Amaral, R.S., Drummond, S.C., Shimabukuro, Y.E., Oliveira, G.C. & Carvalho, O.S. 2010. Evaluation of a linear spectral mixture model and vegetation indices (NDVI and EVI) in a study of schistosomiasis mansoni and *Biomphalaria glabrata* distribution in the state of Minas Gerais, Brazil. *Mem Inst Oswaldo Cruz*, 105, 512-518.
- Guisan, A. & Zimmermann, N.E. 2000. Predictive habitat distribution models in ecology. *Ecological Modelling*, 135, 147-186.
- Guto, S.N., Pypers, P., Vanlauwe, B., de Ridder, N. & Giller\*, K.E. 2012. Socio-Ecological Niches for Minimum Tillage and Crop-Residue Retention in Continuous Maize Cropping Systems in Smallholder Farms of Central Kenya. *Agron. J.*, 104, 188-198.
- Haas, W. & Schmitt, R. 1982. Characterization of chemical stimuli for the penetration of *Schistosoma mansoni* cercariae. I. Effective substances, host specificity. *Z Parasitenkd*, 66, 293-307.
- Hagen, L.J. 2004. Evaluation of the Wind Erosion Prediction System (WEPS) erosion submodel on cropland fields. *Environmental Modelling & Software*, 19, 171-176.

- Hairston, N.G. 1973. The Dynamics of Transmission. In: Ansari, N. (ed.) *Epidemiology and Control of Schistosomiasis (Bilharziasis)*. Basel and Baltimore: S. Karger and University Park Press.
- Harrison, A.D. & Farina, T.D.W. 1965. A naturally turbid water with deleterious effects on egg capsules of planorbis snails. *Annals of Tropical Medicine and Parasitology*, 59, 327-330.
- Hastie, T., Tibshirani, R. & Friedman, J. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer.
- Hay, S.I. 2000. An overview of remote sensing and geodesy for epidemiology and public health application. In: Hay, S.I., Randolph, S.E., Rogers, D.J., Baker, J.R., Muller, R. & Rollinson, D. (eds.) *Advances in Parasitology - Remote Sensing and Geographical Information Systems in Epidemiology*. Oxford: Academic Press.
- Hay, S.I., Packer, M.J. & Rogers, D.J. 1997. The impact of remote sensing on the study and control of invertebrate intermediate hosts and vectors for disease. *International Journal of Remote Sensing*, 18, 2899-2930.
- Hay, S.I., Randolph, S.E. & Rogers, D.J. 2000a. Guest Editors' Preface. In: Hay, S.I., Randolph, S.E., Rogers, D.J., Baker, J.R., Muller, R. & Rollinson, D. (eds.) *Advances in Parasitology - Remote Sensing and Geographical Information Systems in Epidemiology*. Oxford: Academic Press.
- Hay, S.I., Randolph, S.E. & Rogers, D.J. 2000b. *Remote sensing and geographical information systems in epidemiology*, Oxford, Academic Press.
- Hay, S.I., Tatem, A.J., Graham, A.J., Goetz, S.J. & Rogers, D.J. 2006. Global environmental data for mapping infectious disease distribution. *Adv Parasitol*, 62, 37-77.
- Herbreteau, V., Salem, G., Souris, M., Hugot, J.P. & Gonzalez, J.P. 2007. Thirty years of use and improvement of remote sensing, applied to epidemiology: from early promises to lasting frustration. *Health Place*, 13, 400-403.
- Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G. & Jarvis, A. 2005. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, 25, 1965-1978.
- Hirzel, A.H., Hausser, J., Chessel, D. & Perrin, N. 2002. Ecological niche factor analysis: how to compute habitat-suitability maps without absence data? *Ecology*, 83, 2027-2036.
- Ho, T.K. 1998. The random subspace method for constructing decision forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20, 832-844.
- Hodges, M., Dada, N., Wamsley, A., Paye, J., Nyorkor, E., Sonnie, M., Barnish, G., Bockarie, M. & Zhang, Y. 2011. Improved mapping strategy to better inform policy on the control of schistosomiasis and soil-transmitted helminthiasis in Sierra Leone. *Parasit Vectors*, 4, 1-7.
- Hodges, M.H., Soares Magalhaes, R.J., Paye, J., Koroma, J.B., Sonnie, M., Clements, A. & Zhang, Y. 2012. Combined spatial prediction of schistosomiasis and soil-transmitted helminthiasis in Sierra Leone: a tool for integrated disease control. *PLoS Negl Trop Dis*, 6.
- Höskuldsson, A. 1988. PLS Regression Methods. *Journal of Chemometrics*, 2, 211-228.
- Höskuldsson, A. 1995. A combined theory for PCA and PLS. *Journal of Chemometrics*, 9, 91-123.
- Hotez, P.J. & Fenwick, A. 2009. Schistosomiasis in Africa: an emerging tragedy in our new global health decade. *PLoS Negl Trop Dis*, 3, 1-3.
- Huang, Y. & Manderson, L. 1992. Schistosomiasis and the social patterning of infection. *Acta Trop*, 51, 175-194.
- Huete, A., Didan, K., Miura, T., Rodriguez, E.P., Gao, X. & Ferreira, L.G. 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment*, 83, 195-213.
- Huete, A., Justice, C. & Liu, H. 1994. Development of vegetation and soil indices for MODIS-EOS. *Remote Sensing of Environment*, 49, 224-234.

- Huete, A., Justice, C. & van Leeuwen, W.J.D. 1999. MODIS Vegetation Index (MOD 13): Algorithm Theoretical Basis Document. Version 3.
- Huete, A.R. 1988. A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment*, 25, 295-309.
- Huete, A.R. & Liu, H.Q. 1994. An error and sensitivity analysis of the atmospheric- and soil-correcting variants of the NDVI for the MODIS-EOS. *Geoscience and Remote Sensing, IEEE Transactions on*, 32, 897-905.
- Hugh-Jones, M. 1989. Applications of remote sensing to the identification of the habitats of parasites and disease vectors. *Parasitology Today*, 5, 244-251.
- Hugh-Jones, M. 1991. Satellite imaging as a technique for obtaining disease-related data. *Rev Sci Tech*, 10, 197-204.
- Hunter, J.M., Rey, L., Chu, K.Y., Adekolu-John, E.O. & Mott, K.E. 1993. Parasitic diseases in water resources development. The need for intersectoral negotiation. Geneva: WHO.
- Hürlimann, E., Schur, N., Boutsika, K., Stensgaard, A.S., Laserna de Himpsl, M., Ziegelbauer, K., Laizer, N., Camenzind, L., Di Pasquale, A., Ekpo, U.F., *et al.* 2011. Toward an open-access global database for mapping, control, and surveillance of neglected tropical diseases. *PLoS Negl Trop Dis*, 5.
- Husting, E.L. 1983. Human water contact activities related to the transmission of bilharziasis (schistosomiasis). *J Trop Med Hyg*, 86, 23-35.
- Hutchinson, G.E. 1957. Concluding remarks. *Cold Spring Harbor Symposia on Quantitative Biology*, 22, 415 – 427.
- ICL. 2013. *Schistosomiasis Control Initiative* [Online]. London: Imperial College London - School of Public Health. Available: [http://www1.imperial.ac.uk/publichealth/departments/ide/research\\_groups/thesci/](http://www1.imperial.ac.uk/publichealth/departments/ide/research_groups/thesci/) Accessed 2014-03-10.
- ICL. 2014. *Schistosomiasis Control Initiative: Strategy in Ivory Coast* [Online]. London: Imperial College London - School of Public Health. Available: <http://www3.imperial.ac.uk/schisto/wherewework/cotedivoire/cotedivoirestrategy> Accessed 2014-03-26.
- IEA 1995. *A Dictionary of Epidemiology*, New York, Oxford, Toronto, Oxford University Press.
- IPCC 2013. Summary for Policymakers. In: Stocker, T.F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S.K., Boschung, J., Nauels, A., Xia, Y., Bex, V. & Midgley, P.M. (eds.) *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge, United Kingdom and New York, NY, USA.
- Janssen, M., Deng, Z., Mulindabigwi, V. & Röhrig, J. 2010. Agriculture and food. In: Speth, P., Christoph, M. & Diekkrüger, B. (eds.) *Impacts of Global Change on the Hydrological Cycle in West and Northwest Africa*. Heidelberg, Germany: Springer.
- Jensen, J.R. 2000. *Remote Sensing of the Environment - An Earth Resource Perspective*, Upper Saddle River, Prentice-Hall.
- Jordan, P. 2000. From Katayama to the Dakhla Oasis: the beginning of epidemiology and control of bilharzia. *Acta Tropica*, 77, 9-40.
- Jordan, P. & Webbe, G. 1993. Epidemiology. In: Jordan, P., Webbe, G. & Sturrock, R.F. (eds.) *Human Schistosomiasis*. Wallingford: CAB International.
- JSS. 2014. *ASTER GDEM* [Online]. Japan Space Systems (J-spacesystems). Available: <http://gdem.ersdac.jspacesystems.or.jp/search.jsp> Accessed 2014-03-21.
- Justice, C.O., Townshend, J.R.G., Vermote, E.F., Masuoka, E., Wolfe, R.E., Saleous, N., Roy, D.P. & Morisette, J.T. 2002. An overview of MODIS Land data processing and product status. *Remote Sensing of Environment*, 83, 3-15.
- Jutla, A.S., Akanda, A.S. & Islam, S. 2010. Tracking Cholera in Coastal Regions using Satellite Observations. *J Am Water Resour Assoc*, 46, 651-662.

- Kabatereine, N.B., Brooker, S., Tukahebwa, E.M., Kazibwe, F. & Onapa, A.W. 2004. Epidemiology and geography of *Schistosoma mansoni* in Uganda: implications for planning control. *Trop Med Int Health*, 9, 372-380.
- Kabatereine, N.B., Standley, C.J., Sousa-Figueiredo, J.C., Fleming, F.M., Stothard, J.R., Talisuna, A. & Fenwick, A. 2011. Integrated prevalence mapping of schistosomiasis, soil-transmitted helminthiasis and malaria in lakeside and island communities in Lake Victoria, Uganda. *Parasit Vectors*, 4, 232.
- Kalluri, S., Gilruth, P., Rogers, D. & Szczer, M. 2007. Surveillance of Arthropod Vector-Borne Infectious Diseases Using Remote Sensing Techniques: A Review. *PLoS Pathog*, 3.
- Karl, J.W. & Maurer, B.A. 2010. Spatial dependence of predictions from image segmentation: A variogram-based method to determine appropriate scales for producing land-management information. *Ecological Informatics*, 5, 194-202.
- Kearney, M. 2006. Habitat, environment and niche: what are we modelling? *Oikos*, 115, 186-191.
- Kearney, M. & Porter, W. 2009. Mechanistic niche modelling: combining physiological and spatial data to predict species' ranges. *Ecology Letters*, 12, 334-350.
- Keiser, J., De Castro, M.C., Maltese, M.F., Bos, R., Tanner, M., Singer, B.H. & Utzinger, J. 2005. Effect of irrigation and large dams on the burden of malaria on a global and regional scale. *Am J Trop Med Hyg*, 72, 392-406.
- Kerr, J.T. & Ostrovsky, M. 2003. From space to species: ecological applications for remote sensing. *Trends in ecology & evolution (Personal edition)*, 18, 299-305.
- Kiel, B., Alsdorf, D. & LeFavour, G. 2006. Capability of SRTM C- and X-band DEM Data to Measure Water Elevations in Ohio and the Amazon. *Photogrammetric Engineering & Remote Sensing*, 72, 313-320.
- Kinanpara, K., Yves, B.K., Félix, K.K., Edia, E.O., Théophile, G. & Germain, G. 2013. Freshwater snail dynamics focused on potential risk of using urine as fertilizer in Katiola, an endemic area of Schistosomiasis (Ivory Coast; West Africa). *Journal of Entomology and Zoology Studies*, 1, 110-115.
- King, C.H. 2009. Toward the elimination of schistosomiasis. *N Engl J Med*, 360, 106-109.
- King, C.H. 2010. Parasites and poverty: the case of schistosomiasis. *Acta Trop*, 113, 95-104.
- King, C.H. & Dangerfield-Cha, M. 2008. The unacknowledged impact of chronic schistosomiasis. *Chronic Illn*, 4, 65-79.
- King, C.H., Dickman, K. & Tisch, D.J. 2005. Reassessment of the cost of chronic helminth infection: a meta-analysis of disability-related outcomes in endemic schistosomiasis. *Lancet*, 365, 1561-1569.
- Kitron, U. 1998. Landscape ecology and epidemiology of vector-borne diseases: tools for spatial analysis. *J Med Entomol*, 35, 435-445.
- Kliskey, A.D., Lofroth, E.C., Thompson, W.A., Brown, S. & Schreier, H. 1999. Simulating and evaluating alternative resource-use strategies using GIS-based habitat suitability indices. *Landscape and Urban Planning*, 45, 163-175.
- Koenig, W.D. 1999. Spatial autocorrelation of ecological phenomena. *Trends Ecol Evol*, 14, 22-26.
- Koroma, J.B., Peterson, J., Gbakima, A.A., Nylander, F.E., Sahr, F., Soares Magalhaes, R.J., Zhang, Y. & Hodges, M.H. 2010. Geographical distribution of intestinal schistosomiasis and soil-transmitted helminthiasis and preventive chemotherapy strategies in Sierra Leone. *PLoS Negl Trop Dis*, 4.
- Kouassi, N.F. 2010. The Human Environment: Socio-cultural organization. In: Konaté, S. & Kampmann, D. (eds.) *Biodiversity Atlas of West Africa, Volume III: Cote d'Ivoire*. Abidjan & Frankfurt/Main.
- Kouassi, N.F. & Ahoussi, J.M.-S. 2010. Fire as an Environmental Factor: Fire as an anthropogenic factor. In: Konaté, S. & Kampmann, D. (eds.) *Biodiversity Atlas of West Africa, Volume III: Cote d'Ivoire*. Abidjan & Frankfurt/Main.



- Koukounari, A., Gabrielli, A.F., Toure, S., Bosque-Oliva, E., Zhang, Y., Sellin, B., Donnelly, C.A., Fenwick, A. & Webster, J.P. 2007. Schistosoma haematobium infection and morbidity before and after large-scale administration of praziquantel in Burkina Faso. *J Infect Dis*, 196, 659-669.
- Krause, P., Boyle, D.P. & Bäse, F. 2005. Comparison of different efficiency criteria for hydrological model assessment. *Advances in Geosciences*, 5, 89-97.
- Kristensen, T.K., Malone, J.B. & McCarroll, J.C. 2001. Use of satellite remote sensing and geographic information systems to model the distribution and abundance of snail intermediate hosts in Africa: a preliminary model for Biomphalaria pfeifferi in Ethiopia. *Acta Trop*, 79, 73-78.
- Kuenzer, C. & Dech, S. 2013. Theoretical Background of Thermal Infrared Remote Sensing. In: Kuenzer, C. & Dech, S. (eds.) *Thermal Infrared Remote Sensing*. Dordrecht: Springer.
- Kuhn, M. 2008. Building Predictive Models in R Using the caret Package. *Journal of Statistical Software*, 28, 1-26.
- Kumar, K.V., Palit, A. & Bhan, S.K. 1997. Cover Bathymetric mapping in Rupnarayan-Hooghly river confluence using Indian remote sensing satellite data. *International Journal of Remote Sensing*, 18, 2269-2270.
- Legendre, P. 1993. Spatial Autocorrelation: Trouble or New Paradigm? *Ecology*, 74, 1659-1673.
- Liaw, A. & Wiener, M. 2002. Classification and regression by randomForest. *R News*, 2, 18-22.
- Lillesand, T.M. & Kiefer, R.W. 2000. *Remote sensing and image interpretation*, John Wiley & Sons.
- Lima e Costa, M.F., Magalhaes, M.H., Rocha, R.S., Antunes, C.M. & Katz, N. 1987. Water-contact patterns and socioeconomic variables in the epidemiology of schistosomiasis mansoni in an endemic area in Brazil. *Bull World Health Organ*, 65, 57-66.
- Linard, C., Gilbert, M., Snow, R.W., Noor, A.M. & Tatem, A.J. 2012. Population distribution, settlement patterns and accessibility across Africa in 2010. *PLoS One*, 7.
- Lindgren, F., Geladi, P. & Wold, S. 1993. The kernel algorithm for PLS. *Journal of Chemometrics*, 7, 45-59.
- Linthicum, K.J., Anyamba, A., Tucker, C.J., Kelley, P.W., Myers, M.F. & Peters, C.J. 1999. Climate and satellite indicators to forecast Rift Valley fever epidemics in Kenya. *Science*, 285, 397-400.
- Lobitz, B., Beck, L., Huq, A., Wood, B., Fuchs, G., Faruque, A.S.G. & Colwell, R.R. 2000. Climate and infectious disease: Use of remote sensing for detection of Vibrio cholerae by indirect measurement. *PNAS*, 97, 1438-1443.
- Lozano, R., Naghavi, M., Foreman, K., Lim, S., Shibuya, K., Aboyans, V., Abraham, J., Adair, T., Aggarwal, R., Ahn, S.Y., et al. 2012. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet*, 380, 2095-2128.
- M.E.F. 2008. Recensement général de la population et de l'habitat. Ouagadougou: RGPH 2006.
- Mak, S., Morshed, M. & Henry, B. 2010. Ecological niche modeling of lyme disease in British Columbia, Canada. *J Med Entomol*, 47, 99-105.
- Malczewski, J. 2006. GIS-based multicriteria decision analysis: a survey of the literature. *International Journal of Geographical Information Science*, 20, 703-726.
- Malone, J.B. 2005. Biology-based mapping of vector-borne parasites by Geographic Information Systems and Remote Sensing. *Parassitologia*, 47, 27-50.
- Malone, J.B., Huh, O.K., Fehler, D.P., Wilson, P.A., Wilensky, D.E., Holmes, R.A. & Elmagdoub, A.I. 1994. Temperature data from satellite imagery and the distribution of schistosomiasis in Egypt. *Am J Trop Med Hyg*, 50, 714-722.
- Malone, J.B., McNally, K.L., McCarroll, J.C., Corbett, J.D. & Mkoji, G. 2004. Modeling the biocoenose of parasitic diseases using remote sensing and geographic information systems. *Parassitologia*, 46, 59-61.

- Malone, J.B., Yilma, J.M., McCarroll, J.C., Erko, B., Mukaratirwa, S. & Zhou, X. 2001. Satellite climatology and the environmental risk of *Schistosoma mansoni* in Ethiopia and east Africa. *Acta Trop*, 79, 59-72.
- Mantel, N. 1967. The Detection of Disease Clustering and a Generalized Regression Approach. *Cancer Research*, 27, 209-220.
- Marcus, W.A. & Fonstad, M.A. 2008. Optical remote mapping of rivers at sub-meter resolutions and watershed extents. *Earth Surface Processes and Landforms*, 33, 4-24.
- Marshall, R.J. 1991. A Review of Methods for the Statistical Analysis of Spatial Patterns of Disease. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 154, 421-441.
- Martens, W.M., Jetten, T. & Focks, D. 1997. Sensitivity of malaria, schistosomiasis and dengue to global warming. *Climatic Change*, 35, 145-156.
- Martinuzzi, S., Gould, W.A. & Ramos Gonzalez, O.M. 2007. Creating Cloud-Free Landsat ETM+ Data Sets in Tropical Landscapes: CLOUD and Cloud-Shadow Removal. United States Department of Agriculture - Forest Service: International Institute of Tropical Forestry.
- Mas-Coma, S., Valero, M.A. & Bargues, M.D. 2009. Climate change effects on trematodiasis, with emphasis on zoonotic fascioliasis and schistosomiasis. *Vet Parasitol*, 163, 264-280.
- Mayer, J.D. 1983. The role of spatial analysis and geographic data in the detection of disease causation. *Soc Sci Med*, 17, 1213-1221.
- McFeeters, S.K. 1996. The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *International Journal of Remote Sensing*, 17, 1425-1432.
- McMichael, A.J. 2013. Globalization, Climate Change, and Human Health. *New England Journal of Medicine*, 368, 1335-1343.
- Mendoza, G.A. & Martins, H. 2006. Multi-criteria decision analysis in natural resource management: A critical review of methods and new modelling paradigms. *Forest Ecology and Management*, 230, 1-22.
- Mevik, B.-H. & Wehrens, R. 2007. The pls Package: Principal Component and Partial Least Squares Regression in R. *Journal of Statistical Software*, 18, 1-24.
- Meyer, D. 2011. ASTER Global Digital Elevation Model Version 2—Summary of Validation Result. Available: [http://www.jspacesystems.or.jp/ersdac/GDEM/ver2Validation/Summary\\_GDEM2\\_validation\\_report\\_final.pdf](http://www.jspacesystems.or.jp/ersdac/GDEM/ver2Validation/Summary_GDEM2_validation_report_final.pdf).
- Mitchell, M.S., Zimmerman, J.W. & Powell, R.A. 2002. Test of a Habitat Suitability Index for Black Bears in the Southern Appalachians. *Wildlife Society Bulletin*, 30, 794-808.
- MoH & PNLSc 2010. Plan d'action 2011 du programme national de lutte contre les schistosomoses. Ouagadougou: Ministère de la Santé Burkina Faso - Programme National de Lutte Contre les schistosomoses.
- Moodley, I., Kleinschmidt, I., Sharp, B., Craig, M. & Appleton, C. 2003. Temperature-suitability maps for schistosomiasis in South Africa. *Ann Trop Med Parasitol*, 97, 617-627.
- Morgan, J.A. & Messenger, R.C. 1973. *THAID, a sequential analysis program for the analysis of nominal scale dependent variables*, Survey Research Center, Institute for Social Research, University of Michigan.
- Morgan, J.N. & Sonquist, J.A. 1963. Problems in the Analysis of Survey Data, and a Proposal. *Journal of the American Statistical Association*, 58, 415-434.
- Mota, E. & Sleigh, A.C. 1987. Water-contact patterns and *Schistosoma mansoni* infection in a rural community in northeast Brazil. *Rev Inst Med Trop Sao Paulo*, 29, 1-8.
- Mtei, K.M., Ngome, A.F. & Becker, M. 2013. Socio-ecological niches for targeting technology options to improve agricultural production in smallholder systems of western Kenya. *International Journal of AgriScience*, 3, 280-297.

- Mulvey, M. & Vrijenhoek, R.C. 1982. Population structure in *Biomphalaria glabrata* examination of an hypothesis for the patchy distribution of susceptibility to schistosomes. *Am J Trop Med Hyg*, 31, 1195-1200.
- Murray, C.J.L., Vos, T., Lozano, R., Naghavi, M., Flaxman, A.D., Michaud, C., Ezzati, M., Shibuya, K., Salomon, J.A., Abdalla, S., *et al.* 2012. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. *The Lancet*, 380, 2197-2223.
- Myers, M.F., Rogers, D.J., Cox, J., Flahault, A. & Hay, S.I. 2000a. Forecasting disease risk for increased epidemic preparedness in public health. *In*: Hay, S.I., Randolph, S.E., Rogers, D.J., Baker, J.R., Muller, R. & Rollinson, D. (eds.) *Advances in Parasitology - Remote Sensing and Geographical Information Systems in Epidemiology*. Oxford: Academic Press.
- Myers, N., Mittermeier, R.A., Mittermeier, C.G., da Fonseca, G.A.B. & Kent, J. 2000b. Biodiversity hotspots for conservation priorities. *Nature*, 403, 853-858.
- N'Goran, E.K., Diabate, S., Utzinger, J. & Sellin, B. 1997. Changes in human schistosomiasis levels after the construction of two large hydroelectric dams in central Côte d'Ivoire. *Bull World Health Organ.*, 75, 541-545.
- NASA. 1998. *Global Monitoring and Human Health Program* [Online]. Available: <http://geo.arc.nasa.gov/sge/health/gmhh/gmhh.html> Accessed 02 January 2014.
- NASA. 2014. *MODIS Web* [Online]. Available: <http://modis.gsfc.nasa.gov/> Accessed 2014-09-02.
- Nash, J.E. & Sutcliffe, J.V. 1970. River flow forecasting through conceptual models part I — A discussion of principles. *Journal of Hydrology*, 10, 282-290.
- Norman, J.M. & Becker, F. 1995. Terminology in thermal infrared remote sensing of natural surfaces. *Agricultural and Forest Meteorology*, 77, 153-166.
- Odum, E.P. & Barrett, G.W. 2005. *Fundamentals of Ecology*, Brooks/Cole, Canada, Thomson Learning, Inc.
- Ojiem, J.O., de Ridder, N., Vanlauwe, B. & Giller, K.E. 2006. Socio-ecological niche: a conceptual framework for integration of legumes in smallholder farming systems. *International Journal of Agricultural Sustainability*, 4, 79-93.
- Openshaw, S. 1984. *The Modifiable Areal Unit Problem*, Norwich.
- Ortigosa, G.R., De Leo, G.A. & Gatto, M. 2000. VVF: integrating modelling and GIS in a software tool for habitat suitability assessment. *Environmental Modelling & Software*, 15, 1-12.
- Ostfeld, R.S., Glass, G.E. & Keesing, F. 2005. Spatial epidemiology: an emerging (or re-emerging) discipline. *Trends in Ecology & Evolution*, 20, 328-336.
- Ottaviani, D., Lasinio, G.J. & Boitani, L. 2004. Two statistical methods to validate habitat suitability models using presence-only data. *Ecological Modelling*, 179, 417-443.
- Patz, J.A., Campbell-Lendrum, D., Holloway, T. & Foley, J.A. 2005. Impact of regional climate change on human health. *Nature*, 438, 310-317.
- Patz, J.A., Graczyk, T.K., Geller, N. & Vittor, A.Y. 2000. Effects of environmental change on emerging parasitic diseases. *International Journal for Parasitology*, 30, 1395-1405.
- Pavlovsky, E.N. 1966. *Natural nidity of transmissible diseases*, Urbana, Illinois, University of Illinois Press.
- Pebesma, E.J. 2004. Multivariable geostatistics in S: the gstat package. *Computers & Geosciences*, 30, 683-691.
- Peterson, A.T. 2006. Ecologic niche modeling and spatial patterns of disease transmission. *Emerg Infect Dis*, 12, 1822-1826.
- Peterson, A.T. & Holt, R.D. 2003. Niche differentiation in Mexican birds: using point occurrences to detect ecological innovation. *Ecology Letters*, 6, 774-782.

- Peterson, A.T., Sanchez-Cordero, V., Beard, C.B. & Ramsey, J.M. 2002. Ecologic niche modeling and potential reservoirs for Chagas disease, Mexico. *Emerg Infect Dis*, 8, 662-667.
- Pflüger, W. 1980. Experimental epidemiology of schistosomiasis. I. The prepatent period and cercarial production of *Schistosoma mansoni* in *Biomphalaria* snails at various constant temperatures. *Z Parasitenkd*, 63, 159-169.
- Pflüger, W., Roushdy, M.Z. & El Emam, M. 1984. The prepatent period and cercarial production of *Schistosoma haematobium* in *Bulinus truncatus* (Egyptian field strains) at different constant temperatures. *Z Parasitenkd*, 70, 95-103.
- Pitchford, R.J., Meyling, A.H., Meyling, J. & Du Toit, J.F. 1969. Cercarial shedding patterns of various schistosome species under outdoor conditions in the Transvaal. *Ann Trop Med Parasitol*, 63, 359-371.
- Poda, J.N. 1996. *Spatial distribution of intermediate hosts of schistosomes in Burkina Faso: factors influencing the dynamics of *Bulinus truncatus rohlfsi* and *Bulinus senegalensis**. PhD.
- Poda, J.N., Dianou, D., Kambou, T., Sawadogo, B. & Sondo, B. 2001. Étude comparative de trois foyers bilharziens à *Schistosoma haematobium* au Burkina Faso. *Bull Soc Pathol Exot*, 94, 25-28.
- Poda, J.N., Sawadogo, L.L., Sellin, B. & Sanogo, S. 1996. Dynamique des Populations de *Bulinus truncatus rohlfsi* Clessin, 1886, dans le barrage de Dyoro en Zone Nord Soudanienne du Burkina Faso. *Agronomie Africaine*, 8, 61-68.
- Poda, J.N., Sellin, B., Sawadogo, L. & Sanogo, S. 1994. Distribution spatiale des mollusques hotes intermediaires potentiels des schistosomes et de leurs biotopes au Burkina Faso. *Bulletin de Liaison de l'Organisation de Coordination et de Cooperation pour la Lutte Contre les Grandes Endemies*, 101, 12-19.
- Poda, J.N., Traore, A. & Sondo, B.K. 2004. [Schistosomiasis endemic in Burkina Faso]. *Bull Soc Pathol Exot*, 97, 47-52.
- Poorter, L., Bongers, F. & Lemmens, R.H.M.J. 2004. West African forests: introduction. In: Poorter, L., Bongers, F., Kouamé, F.Y.N. & Hawthorne, W.D. (eds.) *Biodiversity of West African Forests - An Ecological Atlas of Woody Plant Species*. Oxon, UK: CAB International.
- Porembski, S., Finckh, M. & Orthmann, B. 2010. Flora and vegetation. In: Speth, P., Christoph, M. & Diekkrüger, B. (eds.) *Impacts of Global Change on the Hydrological Cycle in West and Northwest Africa*. Heidelberg, Germany: Springer.
- Prüss-Üstün, A. & Corvalan, C. 2006. Preventing diseases through healthy environments: Towards an estimate of the environmental burden of disease. WHO.
- Pullan, R.L., Gething, P.W., Smith, J.L., Mwandawiro, C.S., Sturrock, H.J.W., Gitonga, C.W., Hay, S.I. & Brooker, S. 2011. Spatial Modelling of Soil-Transmitted Helminth Infections in Kenya: A Disease Control Planning Tool. *PLoS Negl Trop Dis*, 5.
- Qi, J., Chehbouni, A., Huete, A.R., Kerr, Y.H. & Sorooshian, S. 1994. A modified soil adjusted vegetation index. *Remote Sensing of Environment*, 48, 119-126.
- Quayle, L.M., Appleton, C.C. & Dickens, C.W.S. 2010. The effects of stream flow manipulation on the invertebrate hosts of malaria, bilharzia and liver fluke disease. Water Research Commission: Institute of Natural Resources.
- Raso, G., Matthys, B., N'Goran, E.K., Tanner, M., Vounatsou, P. & Utzinger, J. 2005. Spatial risk prediction and mapping of *Schistosoma mansoni* infections among schoolchildren living in western Cote d'Ivoire. *Parasitology*, 131, 97-108.
- Raso, G., Vounatsou, P., McManus, D.P. & Utzinger, J. 2007. Bayesian risk maps for *Schistosoma mansoni* and hookworm mono-infections in a setting where both parasites co-exist. *Geospat Health*, 2, 85-96.

- Raso, G., Vounatsou, P., Singer, B.H., N'Goran, E.K., Tanner, M. & Utzinger, J. 2006. An integrated approach for risk profiling and spatial prediction of *Schistosoma mansoni*-hookworm coinfection. *Proc Natl Acad Sci U S A*, 103, 6934-6939.
- Ray, N. & Burgman, M.A. 2006. Subjective uncertainties in habitat suitability maps. *Ecological Modelling*, 195, 172-186.
- Reichert, B., Klose, S. & Kocher, A. 2010. Geology. In: Speth, P., Christoph, M. & Diekkrüger, B. (eds.) *Impacts of Global Change on the Hydrological Cycle in West and Northwest Africa*. Heidelberg, Germany: Springer.
- Reisen, W.K. 2010. Landscape epidemiology of vector-borne diseases. *Annu Rev Entomol*, 55, 461-483.
- Richter, R. 1996. A spatially adaptive fast atmospheric correction algorithm. *International Journal of Remote Sensing*, 17, 1201-1214.
- Richter, R. & Schlöpfer, D. 2012. Atmospheric / Topographic Correction for Airborne Imagery. ATCOR-4 User Guide, Version 6.2 BETA. Available: [http://www.dlr.de/eoc/Portaldata/60/Resources/dokumente/5\\_tech\\_mod/atcor4\\_manual\\_2012.pdf](http://www.dlr.de/eoc/Portaldata/60/Resources/dokumente/5_tech_mod/atcor4_manual_2012.pdf) Accessed 2014-02-25.
- Rinaldi, L., Musella, V., Biggeri, A. & Cringoli, G. 2006. New insights into the application of geographical information systems and remote sensing in veterinary parasitology. *Geospat Health*, 1, 33-47.
- Roberts, D., Rodriguez, M., Rejmankova, E., Pope, K., Savage, H., Rodriguez-Ramirez, A., Wood, B., Salute, J. & Legters, L. 1991. Overview of field studies for the application of remote sensing to the study of malaria transmission in Tapachula, Mexico. *Preventive Veterinary Medicine*, 11, 269-275.
- Robinson, T.P. 2000. Spatial statistics and geographical information systems in epidemiology and public health. In: Hay, S.I., Randolph, S.E., Rogers, D.J., Baker, J.R., Muller, R. & Rollinson, D. (eds.) *Advances in Parasitology - Remote Sensing and Geographical Information Systems in Epidemiology*. Oxford: Academic Press.
- Rollinson, D., Knopp, S., Levitz, S., Stothard, J.R., Tchuem Tchuenté, L.-A., Garba, A., Mohammed, K.A., Schur, N., Person, B., Colley, D.G., et al. 2013. Time to set the agenda for schistosomiasis elimination. *Acta Tropica*, 128, 423-440.
- Rosenberg, C.E. 2012. Epilogue: Airs, Waters, Places. A status report. *Bull Hist Med*, 86, 661-670.
- Ruffer, M.A. 1910. NOTE ON THE PRESENCE OF "BILHARZIA HAEMATOBIA" IN EGYPTIAN MUMMIES OF THE TWENTIETH DYNASTY [1250-1000 B.C.]. *Br Med J*, 1, 16.
- Rüger, N., Schlüter, M. & Matthies, M. 2005. A fuzzy habitat suitability index for *Populus euphratica* in the Northern Amudarya delta (Uzbekistan). *Ecological Modelling*, 184, 313-328.
- Savane, I. 2010a. The Biophysical Environment: Climate. In: Konaté, S. & Kampmann, D. (eds.) *Biodiversity Atlas of West Africa, Volume III: Cote d'Ivoire*. Abidjan & Frankfurt/Main.
- Savane, I. 2010b. The Biophysical Environment: Physical Geography. In: Konaté, S. & Kampmann, D. (eds.) *Biodiversity Atlas of West Africa, Volume III: Cote d'Ivoire*. Abidjan & Frankfurt/Main.
- Savane, I. 2010c. The Biophysical Environment: Water resources. In: Konaté, S. & Kampmann, D. (eds.) *Biodiversity Atlas of West Africa, Volume III: Cote d'Ivoire*. Abidjan & Frankfurt/Main.
- Schamberger, M.L. & O'Neil, L.J. 1986. Concepts and constraints of habitat-model testing. In: Verner, J., Morrison, M.L. & Ralph, C.J. (eds.) *Wildlife 2000: Modelling habitat relationships of terrestrial vertebrates*. Wisconsin, USA: The University of Wisconsin Press.
- Schmidlin, T., Hürlimann, E., Silue, K.D., Yapi, R.B., Hougbedji, C., Kouadio, B.A., Acka-Douabele, C.A., Kouassi, D., Ouattara, M., Zouzou, F., et al. 2013. Effects of hygiene and defecation behavior on helminths and intestinal protozoa infections in Taabo, Cote d'Ivoire. *PLoS One*, 8.
- Schmugge, T., Hook, S.J. & Coll, C. 1998. Recovering Surface Temperature and Emissivity from Thermal Infrared Multispectral Data. *Remote Sensing of Environment*, 65, 121-131.
- Schneider, A., Friedl, M.A. & Potere, D. 2009. A new map of global urban extent from MODIS satellite data. *Environmental Research Letters*, 4.

## References

---

- Schur, N., Gosoni, L., Raso, G., Utzinger, J. & Vounatsou, P. 2011a. Modelling the geographical distribution of co-infection risk from single-disease surveys. *Stat Med*, 30, 1761-1776.
- Schur, N., Hürlimann, E., Garba, A., Traoré, M.S., Ndir, O., Ratard, R.C., Tchuem Tchuenté, L.-A., Kristensen, T.K., Utzinger, J. & Vounatsou, P. 2011b. Geostatistical Model-Based Estimates of Schistosomiasis Prevalence among Individuals Aged  $\leq 20$  Years in West Africa. *PLoS Negl Trop Dis*, 5.
- Schur, N., Hürlimann, E., Stensgaard, A.S., Chimfwembe, K., Mushinge, G., Simoonga, C., Kabatereine, N.B., Kristensen, T.K., Utzinger, J. & Vounatsou, P. 2013. Spatially explicit Schistosoma infection risk in eastern Africa using Bayesian geostatistical modelling. *Acta Trop*, 128, 365-377.
- Schur, N., Utzinger, J. & Vounatsou, P. 2011c. Modelling age-heterogeneous Schistosoma haematobium and S. mansoni survey data via alignment factors. *Parasit Vectors*, 4, 142.
- Schur, N., Vounatsou, P. & Utzinger, J. 2012. Determining treatment needs at different spatial scales using geostatistical model-based risk estimates of schistosomiasis. *PLoS Negl Trop Dis*, 6.
- SCI. 2014. *Schistosomiasis Control Initiative* [Online]. Imperial College London. Available: <http://www3.imperial.ac.uk/schisto> Accessed 2014-09-19.
- Scorza, J.V., Silva, J., Gonzalez, L. & Machado, R. 1961. Stream velocity as a gradient in *Australorbis glabratus* (say, 1818). *Z Tropenmed Parasitol*, 12, 191-196.
- Senghor, A. 2010a. The Human Environment: Demographic data. In: Thiombiano, A. & Kampmann, D. (eds.) *Biodiversity Atlas of West Africa, Volume II: Burkina Faso*. Ouagadougou & Frankfurt/Main.
- Senghor, A. 2010b. The Human Environment: Sociocultural and socioeconomic indicators. In: Thiombiano, A. & Kampmann, D. (eds.) *Biodiversity Atlas of West Africa, Volume II: Burkina Faso*. Ouagadougou & Frankfurt/Main.
- 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. *Photogrammetric Engineering & Remote Sensing*, 68, 167-174.
- Shiff, C.J. 1964. STUDIES ON BULINUS (PHYSOPSIS) GLOBOSUS IN RHODESIA. I. THE INFLUENCE OF TEMPERATURE ON THE INTRINSIC RATE OF NATURAL INCREASE. *Ann Trop Med Parasitol*, 58, 94-105.
- Shiff, C.J. & Garnett, B. 1967. The influence of temperature on the intrinsic rate of natural increase of the freshwater snail *B. pfeifferi* *Archiv für Hydrobiologie*, 62, 429-438.
- Simoonga, C., Kazembe, L.N., Kristensen, T.K., Olsen, A., Appleton, C.C., Mubita, P. & Mubila, L. 2008. The epidemiology and small-scale spatial heterogeneity of urinary schistosomiasis in Lusaka province, Zambia. *Geospat Health*, 3, 57-67.
- Simoonga, C., Utzinger, J., Brooker, S., Vounatsou, P., Appleton, C.C., Stensgaard, A.S., Olsen, A. & Kristensen, T.K. 2009. Remote sensing, geographical information system and spatial analysis for schistosomiasis epidemiology and ecology in Africa. *Parasitology*, 136, 1683-1693.
- Smith, M., Clegg, J.A. & Webbe, G. 1976. Culture of *Schistosoma haematobium* in vivo and in vitro. *Ann Trop Med Parasitol*, 70, 101-107.
- Snow, J. 1855. *On the mode of communication of cholera* John Churchill.
- Soares Magalhães, R.J., Biritwum, N.K., Gyapong, J.O., Brooker, S., Zhang, Y., Blair, L., Fenwick, A. & Clements, A.C. 2011. Mapping helminth co-infection and co-intensity: geostatistical prediction in Ghana. *PLoS Negl Trop Dis*, 5.
- Soberon, J. & Peterson, A.T. 2005. Interpretation of models of fundamental ecological niches and species' distributional areas. *Biodiversity Informatics*, 2, 1-10.
- Standley, C.J., Adriko, M., Alinaitwe, M., Kazibwe, F., Kabatereine, N.B. & Stothard, J.R. 2009. Intestinal schistosomiasis and soil-transmitted helminthiasis in Ugandan schoolchildren: a rapid mapping assessment. *Geospat Health*, 4, 39-53.

- 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. *The Lancet Infectious Diseases*, 6, 411-425.
- Stensgaard, A., Jorgensen, A., Kabatereine, N.B., Malone, J.B. & Kristensen, T.K. 2005. Modeling the distribution of *Schistosoma mansoni* and host snails in Uganda using satellite sensor data and Geographical Information Systems. *Parasitologia*, 47, 115-125.
- Stensgaard, A.S., Jorgensen, A., Kabatereine, N.B., Rahbek, C. & Kristensen, T.K. 2006. Modeling freshwater snail habitat suitability and areas of potential snail-borne disease transmission in Uganda. *Geospat Health*, 1, 93-104.
- Stensgaard, A.S., Utzinger, J., Vounatsou, P., Hurlimann, E., Schur, N., Saarnak, C.F., Simoonga, C., Mubita, P., Kabatereine, N.B., Tchuem Tchuente, L.A., *et al.* 2013. Large-scale determinants of intestinal schistosomiasis and intermediate host snail distribution across Africa: does climate matter? *Acta Trop*, 128, 378-390.
- Stirewalt, M.A. 1973. Important Features of the Schistosomes. In: Ansari, N. (ed.) *Epidemiology and Control of Schistosomiasis (Bilharziasis)*. Basel and Baltimore: S. Karger and University Park Press.
- Stirewalt, M.A., Cousin, C.E. & Dorsey, C.H. 1983. *Schistosoma mansoni*: stimulus and transformation of cercariae into schistosomules. *Exp Parasitol*, 56, 358-368.
- Store, R. & Kangas, J. 2001. Integrating spatial multi-criteria evaluation and expert knowledge for GIS-based habitat suitability modelling. *Landscape and Urban Planning*, 55, 79-93.
- Stothard, J.R., Chitsulo, L., Kristensen, T.K. & Utzinger, J. 2009. Control of schistosomiasis in sub-Saharan Africa: progress made, new opportunities and remaining challenges. *Parasitology*, 136, 1665-1675.
- Stothard, J.R., Mgeni, A.F., Khamis, S., Seto, E., Ramsan, M., Hubbard, S.J., Kristensen, T.K. & Rollinson, D. 2002. New insights into the transmission biology of urinary schistosomiasis in Zanzibar. *Trans R Soc Trop Med Hyg*, 96, 470-475.
- Strahler, A., Muchoney, D., Borak, J., Friedl, M., Gopal, S., Lambin, E. & Moody, A. 1999. MODIS Land Cover Product Algorithm Theoretical Basis Document (ATBD). Version 5.0.
- Strahler, A.N. 1957. Quantitative Analysis of Watershed Geomorphology. *Transactions, American Geophysical Union*, 38, 913-920.
- Sturrock, H.J., Picon, D., Sabasio, A., Oguttu, D., Robinson, E., Lado, M., Rumunu, J., Brooker, S. & Kolaczinski, J.H. 2009. Integrated mapping of neglected tropical diseases: epidemiological findings and control implications for northern Bahr-el-Ghazal State, Southern Sudan. *PLoS Negl Trop Dis*, 3.
- Sturrock, R.F. 1993a. The Intermediate Hosts and Host-Parasite Relationships. In: Jordan, P., Webbe, G. & Sturrock, R.F. (eds.) *Human Schistosomiasis*. Wallingford: CAB International.
- Sturrock, R.F. 1993b. The Parasites and Their Life Cycle. In: Jordan, P., Webbe, G. & Sturrock, R.F. (eds.) *Human Schistosomiasis*. Wallingford: CAB International.
- Sutton, C. 2005. Classification and Regression Trees, Bagging, and Boosting. In: Rao, C.R., Wegman, E.J. & Solka, J.L. (eds.) *Handbook of Statistics 24. Data Mining and Data Visualization*. Amsterdam, The Netherlands: ELSEVIER Inc.
- Tchuem Tchuente, L.A., Kamwa Ngassam, R.I., Sumo, L., Ngassam, P., Dongmo Noumedem, C., Nzu, D.D., Dankoni, E., Kenfack, C.M., Gipwe, N.F., Akame, J., *et al.* 2012. Mapping of schistosomiasis and soil-transmitted helminthiasis in the regions of centre, East and West Cameroon. *PLoS Negl Trop Dis*, 6.
- Tchuem Tchuente, L.A. & N'Goran E, K. 2009. Schistosomiasis and soil-transmitted helminthiasis control in Cameroon and Cote d'Ivoire: implementing control on a limited budget. *Parasitology*, 136, 1739-1745.

## References

---

- Thomasma, L.E., Drummer, T.D. & Peterson, R.O. 1991. Testing the Habitat Suitability Index Model for the Fisher. *Wildlife Society Bulletin*, 19, 291-297.
- Tobler, W.R. 1970. A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography*, 46, 234-240.
- Torgersen, C.E., Faux, R.N., McIntosh, B.A., Poage, N.J. & Norton, D.J. 2001. Airborne thermal remote sensing for water temperature assessment in rivers and streams. *Remote Sensing of Environment*, 76, 386-398.
- Tran, A., Goutard, F., Chamailé, L., Baghdadi, N. & Lo Seen, D. 2010. Remote sensing and avian influenza: A review of image processing methods for extracting key variables affecting avian influenza virus survival in water from Earth Observation satellites. *International Journal of Applied Earth Observation and Geoinformation*, 12, 1-8.
- Traore, S. & Anne, C.A.T. 2010. The Biophysical Environment: Soils. In: Thiombiano, A. & Kampmann, D. (eds.) *Biodiversity Atlas of West Africa, Volume II: Burkina Faso*. Ouagadougou & Frankfurt/Main.
- UN 2012a. The future we want. New York: United Nations General Assembly.
- UN 2012b. World Urbanization Prospects: The 2011 Revision; File 13: Population of capital cities in 2011. United Nations, Department of Economic and Social Affairs, Population Division.
- UN 2013. World Population Prospects: The 2012 Revision, Highlights and Advance Tables. United Nations Department of Economic and Social Affairs, Population Division.
- UNESCO 2011. Data centre: Country and Regional Profiles. Montréal, Canada: United Nations Educational, Scientific and Cultural Organization - Institute for Statistics.
- Upatham, E.S. 1972. Rapidity and duration of hatching of St. Lucian *Schistosoma mansoni* eggs in outdoor habitats. *J Helminthol*, 46, 271-276.
- Upatham, E.S. 1973. Location of *Biomphalaria glabrata* (Say) by miracidia of *Schistosoma mansoni* Sambon in natural standing and running waters on the West Indian Island of St. Lucia. *Int J Parasitol*, 3, 289-297.
- USDA-NRCS. 1998. *Assessment of water holding capacity of soils map* [Online]. Washington D.C.: United States Department of Agriculture - Natural Resources Conservation Service. Available: [http://www.nrcs.usda.gov/wps/portal/nrcs/detail/soils/use/worldsoils/?cid=nrcs142p2\\_054022](http://www.nrcs.usda.gov/wps/portal/nrcs/detail/soils/use/worldsoils/?cid=nrcs142p2_054022) Accessed 2014-03-04.
- USFWS 1981. Standards for the development of habitat suitability index models. *Ecological Services Manual (ESM)*. Washington D.C.: United States Fish and Wildlife Service.
- USFWS. 2014. *USGS National Wetlands Research Center Digital Library: Habitat Suitability Index Models* [Online]. U.S. Department of the Interior, U.S. Geological Survey. Available: <http://www.nwrc.usgs.gov/wdb/pub/hsi/hsiintro.htm> Accessed 2014-04-02.
- USGS. 2013. *Land Processes Distributed Active Archive Center: Data Access* [Online]. United States Geological Survey. Available: [https://lpdaac.usgs.gov/data\\_access](https://lpdaac.usgs.gov/data_access) Accessed 2014-03-21.
- USGS. 2014. *Landsat Missions* [Online]. Available: <http://landsat.usgs.gov/> Accessed 2014-09-02.
- Utzinger, J., Bergquist, R., Shu-Hua, X., Singer, B.H. & Tanner, M. 2003. Sustainable schistosomiasis control--the way forward. *Lancet*, 362, 1932-1934.
- Utzinger, J., N'Goran E, K., Caffrey, C.R. & Keiser, J. 2011a. From innovation to application: social-ecological context, diagnostics, drugs and integrated control of schistosomiasis. *Acta Trop*, 120 Suppl 1, 121-137.
- Utzinger, J., Raso, G., Brooker, S., De Savigny, D., Tanner, M., Ornbjerg, N., Singer, B.H. & N'Goran E, K. 2009. Schistosomiasis and neglected tropical diseases: towards integrated and sustainable control and a word of caution. *Parasitology*, 136, 1859-1874.
- Utzinger, J., Raso, G., Steinmann, P., Zhou, X.N., Vounatsou, P., Brooker, S. & N'Goran, E.K. 2011b. Schistosomiasis. In: O. Nriagu, J. (ed.) *Encyclopedia of Environmental Health*. Burlington: Elsevier.



- Vahtmäe, E., Kutser, T., Martin, G. & Kotta, J. 2006. Feasibility of hyperspectral remote sensing for mapping benthic macroalgal cover in turbid coastal waters—a Baltic Sea case study. *Remote Sensing of Environment*, 101, 342-351.
- Vermote, E.F. & Vermeulen, A. 1999. MODIS Algorithm Technical Background Document: Atmospheric Correction Algorithm: Spectral Reflectances (MOD09). Version 4.0.
- Vinagre, C., Fonseca, V., Cabral, H. & Costa, M.J. 2006. Habitat suitability index models for the juvenile soles, *Solea solea* and *Solea senegalensis*, in the Tagus estuary: Defining variables for species management. *Fisheries Research*, 82, 140-149.
- von Liebig, J. 1840. Die organische Chemie in ihrer Anwendung auf Agrikultur und Physiologie. Braunschweig: Vieweg.
- Vounatsou, P., Raso, G., Tanner, M., N'Goran E, K. & Utzinger, J. 2009. Bayesian geostatistical modelling for mapping schistosomiasis transmission. *Parasitology*, 136, 1695-1705.
- Walz, Y., Wegmann, M. & Dech, S. Schistosomiasis risk assessment from space using high resolution Rapid Eye data. Geoscience and Remote Sensing Symposium (IGARSS), 2012 IEEE International, 22-27 July 2012 2012. 7224-7227.
- Wan, Z. 1999. MODIS Land-Surface Temperature Algorithm Theoretical Basis Document (LST ATBD). Version 3.3.
- Wan, Z., Zhang, Y., Zhang, Q. & Li, Z.L. 2004. Quality assessment and validation of the MODIS global land surface temperature. Taylor & Francis.
- Webbe, G. & Jordan, P. 1993. Control. In: Jordan, P., Webbe, G. & Sturrock, R.F. (eds.) *Human Schistosomiasis*. Wallingford: CAB International.
- Weiss, R.A. & McMichael, A.J. 2004. Social and environmental risk factors in the emergence of infectious diseases. *Nat Med.*, 10, 70-76.
- WHA 2001. Schistosomiasis and soil-transmitted helminth infections. Geneva: World Health Assembly.
- White, F. 1983. Vegetation of Africa - a descriptive memoir to accompany the UNESCO/AETFAT/UNSO vegetation map of Africa. *Natural Resources Research Report*. Paris, France: U.N. Educational, Scientific and Cultural Organisation.
- WHO 1957. Study group on the ecology of intermediate snail hosts of bilharziasis. *Technical Report Series*. Geneva: World Health Organization.
- WHO 1999. The World Health Report 1999: Making a difference. Geneva.
- WHO 2002a. Prevention and control of schistosomiasis and soil-transmitted helminthiasis. *World Health Organ Tech Rep Ser*, 912, 1-57.
- WHO 2002b. The World Health Report 2002: Reducing risks, promoting healthy life. Geneva.
- WHO 2010a. Country profile: Cote d'Ivoire. Geneva, Switzerland: World Health Organization - Department of Neglected Tropical Diseases.
- WHO 2010b. Schistosomiasis Control at a glance. World Health Organization - Regional Office for Africa.
- WHO. 2013. *Schistosomiasis* [Online]. Available: <http://www.who.int/schistosomiasis/en/> Accessed 2013-10-18.
- William, D.J., Rybicki, N.B., Lombana, A.V., O'Brien, T.M. & Gomez, R.B. 2003. Preliminary investigation of submerged aquatic vegetation mapping using hyperspectral remote sensing. *Environmental monitoring and assessment*, 81, 383-392.
- Williams, N.S.G., Hahs, A.K. & Morgan, J.W. 2008. A dispersal-constrained habitat suitability model for predicting invasion of alpine vegetation. *Ecological Applications*, 18, 347-359.
- Winterbottom, S.J. & Gilvear, D.J. 1997. Quantification of channel bed morphology in gravel-bed rivers using airborne multispectral imagery and aerial photography. *Regulated Rivers: Research & Management*, 13, 489-499.

- Wold, H. 1975. Path models with latent variables: The NIPALS approach. *In: Blalock, H.M., Aganbegian, A., Borodkin, F.M., Boudon, R. & Capecchi, V. (eds.) Quantitative Sociology: International perspectives on mathematical and statistical model building.* New York: Academic Press.
- Wold, S., Ruhe, A., Wold, H. & Dunn, I.W. 1984. The Collinearity Problem in Linear Regression. The Partial Least Squares (PLS) Approach to Generalized Inverses. *SIAM Journal on Scientific and Statistical Computing*, 5, 735-743.
- Wold, S., Sjöström, M. & Eriksson, L. 2001. PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems*, 58, 109-130.
- Wong, S.L., Wan, K.K.W. & Lam, T.N.T. 2010. Artificial neural networks for energy analysis of office buildings with daylighting. *Applied Energy*, 87, 551-557.
- World Bank. 2014. *Data by country* [Online]. The World Bank. Available: <http://data.worldbank.org/country>.
- Xu, H. 2006. Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. *International Journal of Remote Sensing*, 27, 3025-3033.
- Yang, G.J., Vounatsou, P., Zhou, X.N., Tanner, M. & Utzinger, J. 2005a. A Bayesian-based approach for spatio-temporal modeling of county level prevalence of *Schistosoma japonicum* infection in Jiangsu province, China. *Int J Parasitol*, 35, 155-62.
- Yang, G.J., Vounatsou, P., Zhou, X.N., Tanner, M. & Utzinger, J. 2005b. A potential impact of climate change and water resource development on the transmission of *Schistosoma japonicum* in China. *Parassitologia*, 47, 127-34.
- Yang, G.J., Vounatsou, P., Zhou, X.N., Utzinger, J. & Tanner, M. 2005c. A review of geographic information system and remote sensing with applications to the epidemiology and control of schistosomiasis in China. *Acta Trop*, 96, 117-129.
- Yapi, R.B., Hürlimann, E., Hounghbedji, C.A., Ndri, P.B., Silué, K.D., Soro, G., Kouamé, F.N., Vounatsou, P., Fürst, T., N'Goran, E.K., *et al.* 2014. Infection and Co-infection with Helminths and Plasmodium among School Children in Côte d'Ivoire: Results from a National Cross-Sectional Survey. *PLoS Negl Trop Dis*, 8.
- Zhang, Z., Zhu, R., Ward, M.P., Xu, W., Zhang, L., Guo, J., Zhao, F. & Jiang, Q. 2012. Long-term impact of the World Bank Loan Project for schistosomiasis control: a comparison of the spatial distribution of schistosomiasis risk in China. *PLoS Negl Trop Dis*, 6, 1-10.
- Zhou, X.N., Yang, G.J., Yang, K., Wang, X.H., Hong, Q.B., Sun, L.P., Malone, J.B., Kristensen, T.K., Bergquist, N.R. & Utzinger, J. 2008. Potential impact of climate change on schistosomiasis transmission in China. *Am J Trop Med Hyg*, 78, 188-194.

# Appendix

Table A 1: Overview of all RS variables available for the three study sites MAN, TAB and BUF. The nomenclature indicates the abbreviation of the RS variable as used in this study.

RS data	RS variable	Nomenclature	MAN	TAB	BUF
RapidEye	Blue reflectance	RE_blue	X	X	
	Red reflectance	RE_green	X	X	
	Green reflectance	RE_red	X	X	
	Rededge reflectance	RE_rededge	X	X	
	Near infrared reflectance	RE_nir	X	X	
	NDVI	RE_ndvi	X	X	
	SAVI	RE_savi	X	X	
	MSAVI	RE_msavi	X	X	
	EVI	RE_evi	X	X	
	NDWI	RE_ndwif	X	X	
Landsat 5 TM (dry season image)	Blue reflectance	lsd_blue	X		X
	Red reflectance	lsd_blue	X		X
	Green reflectance	lsd_green	X		X
	Red reflectance	lsd_red	X		X
	Near infrared reflectance	lsd_nir	X		X
	Middle infrared reflectance	lsd_mir	X		X
	Shortwave infrared reflectance	lsd_swir	X		X
	Thermal infrared emissivity	lsd_tir	X		X
	NDVI	lsd_ndvi	X		X
	SAVI	lsd_savi	X		X
	MSAVI	lsd_msavi	X		X
	EVI	lsd_evi	X		X
	NDWI	lsd_ndwif	X		X
	MNDWI	lsd_mndwi	X		X
Landsat 5 TM (wet season image)	Blue reflectance	lsw_blue			X
	Red reflectance	lsw_blue			X
	Green reflectance	lsw_green			X
	Red reflectance	lsw_red			X
	Near infrared reflectance	lsw_nir			X
	Middle infrared reflectance	lsw_mir			X

RS data	RS variable	Nomenclature	MAN	TAB	BUF
Landsat 5 TM (wet season image)	Shortwave infrared reflectance	lsw_swir			X
	Thermal infrared emissivity	lsw_tir			X
	NDVI	lsw_ndvi			X
	SAVI	lsw_savi			X
	MSAVI	lsw_msavi			X
	EVI	lsw_evi			X
	NDWI	lsw_ndwif			X
	MNDWI	lsw_mndwi			X
Terra MODIS MOD13Q1	Mean NDVI of 2010	ndvi_mean	X	X	X
	Median NDVI of 2010	ndvi_median	X	X	X
	Maximum NDVI of 2010	ndvi_max	X	X	X
	Minimum NDVI of 2010	ndvi_min	X	X	X
	Mean EVI of 2010	evi_mean	X	X	X
	Median EVI of 2010	evi_median	X	X	X
	Maximum EVI of 2010	evi_max	X	X	X
	Minimum EVI of 2010	evi_min	X	X	X
Terra MODIS MOD11A2	Mean LST of 2010	lst_mean	X	X	X
	Median LST of 2010	lst_median	X	X	X
	Maximum LST of 2010	lst_max	X	X	X
	Minimum LST of 2010	lst_min	X	X	X
	Mean of difference between day and night-time LST of 2010	dtemp_mean	X	X	X
	Median of difference between day and night-time LST of 2010	dtemp_median	X	X	X
	Maximum of difference between day and night-time LST of 2010	dtemp_max	X	X	X
	Minimum of difference between day and night-time LST of 2010	dtemp_min	X	X	X
ASTER GDEM	Altitude	alt	X	X	X
	Slope	slope	X	X	X
	Sink depth	sinks	X	X	X
	Stream order	streams	X	X	X
RapidEye / Landsat 5 TM	Water distance	water_dist	X	X	X
RapidEye / TerraSARX	Settlement area	settl_area	X	X	X

Field Verification Form - Field Campaign 03/11 – Burkina Faso (WalzY)

**A. General Information**

Waypoint ..... Date .....

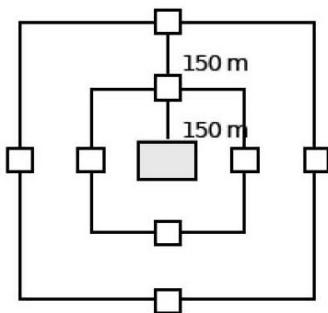
Location ..... Lat .....

Observer ..... Lon .....

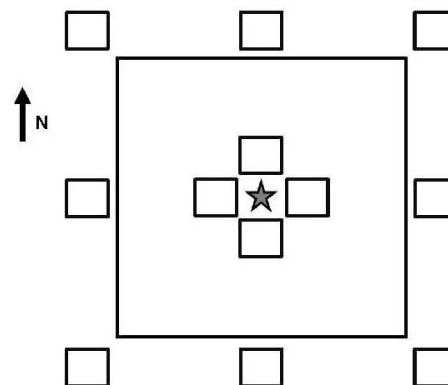
**B. General Land Cover Information**

**B.1 Area Land Cover Homogeneity**

Land Cover Homogeneous for more than 300m around the sample area: **yes**  **no**



**Position of field photographs**



**B.2 General Landform:**

- Slope**
- Flat to Gently Sloping Terrain (0 – 7 %; ca 0 – 4 °)
  - Gently Sloping to Moderately Sloping (8 – 13 %; ca 4 – 7 °)
  - Sloping to Moderately Steep, Undulating to Rolling terrain (14 – 20 %; ca 8 – 11 °)
  - Steep to Very Steep, Rolling to Hilly Terrain (21 – 55 %; 11 – 29 °)
  - Extremely Steep Terrain, Steeply Dissected Hilly and Mountainous Terrain (56 – 140%; 30 – 55°)

**B.3 Land Cover**

- General Land Cover**
- vegetated
  - non-vegetated
  - terrestrial
  - aquatic or regularly flooded land (+ WADY)
- 
- Specific Land Cover**
- cultivated
  - natural / semi-natural
  - built-up
  - bare
  - artificial water body
  - inland water

**B.4 Seasonal Aspect**

	Natural / Semi-Natural Vegetation				Cultivated Fields			
	dry	green	flowering	fruits	ploughed	initial stage	full mat. stage	harvested
TREES	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
SHRUBS	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
HERBS	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>



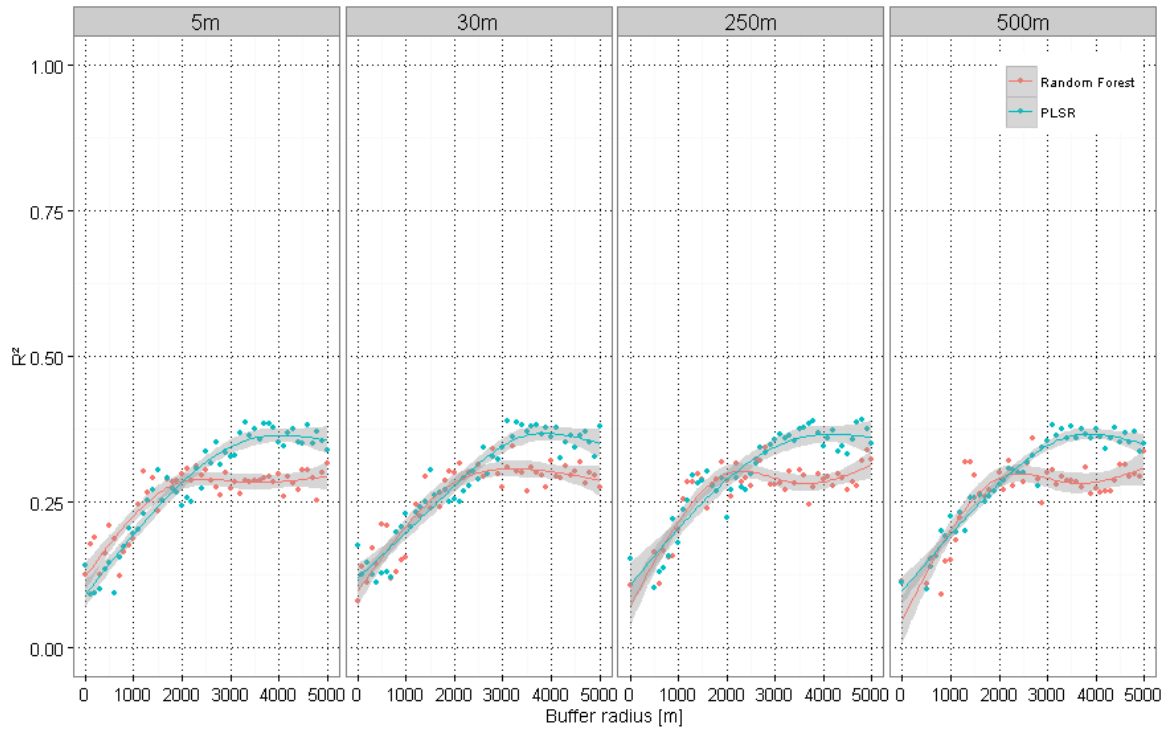


Figure A 2: Impact of spatial resolution on model accuracy derived from the RapidEye database, which was aggregated from 5 m to 30 m, 250 m and 500 m. The red dots represent models results ( $R^2$ ) derived from the Random Forest model and blue dots refer to the results of the PLSR model. The lines reflect the results of spline interpolation and the grey bar represents the confidence interval. Note that cases where the buffer radius is smaller than the spatial resolution are not shown except for the extraction of the pixel value with no buffer (corresponds to 0m buffer radius).

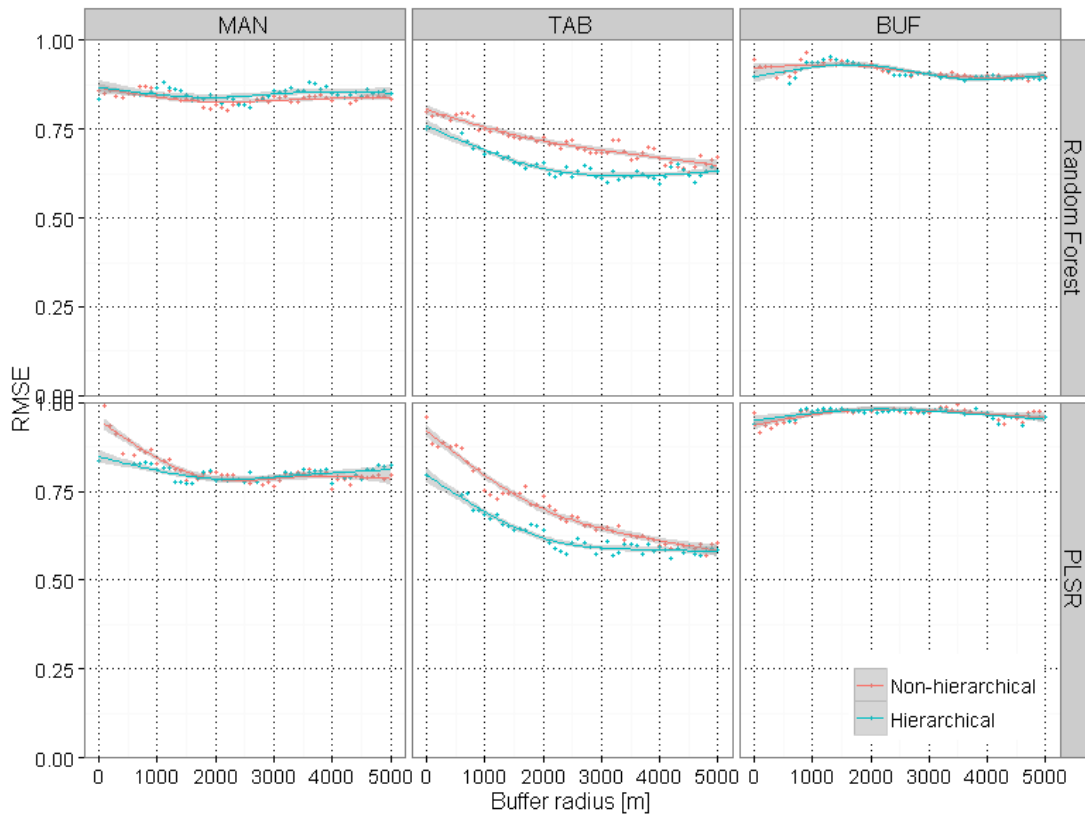


Figure A 3: RMSE values of the non-hierarchical (red) and the hierarchical (blue) models for the three study sites MAN, TAB and BUF based on the Random Forest (upper row) and PLSR model (lower row).

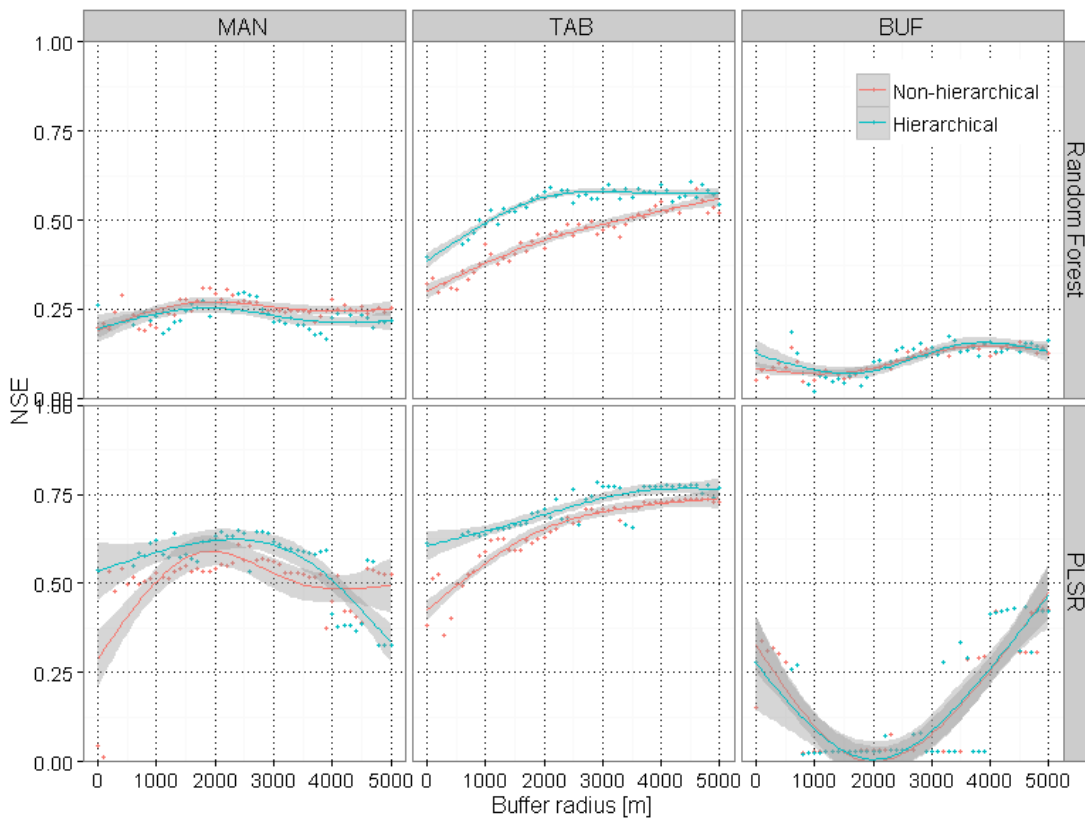


Figure A 4: NSE of the non-hierarchical (red) and the hierarchical (blue) models for the three study sites MAN, TAB and BUF based on the Random Forest (upper row) and PLSR model (lower row).



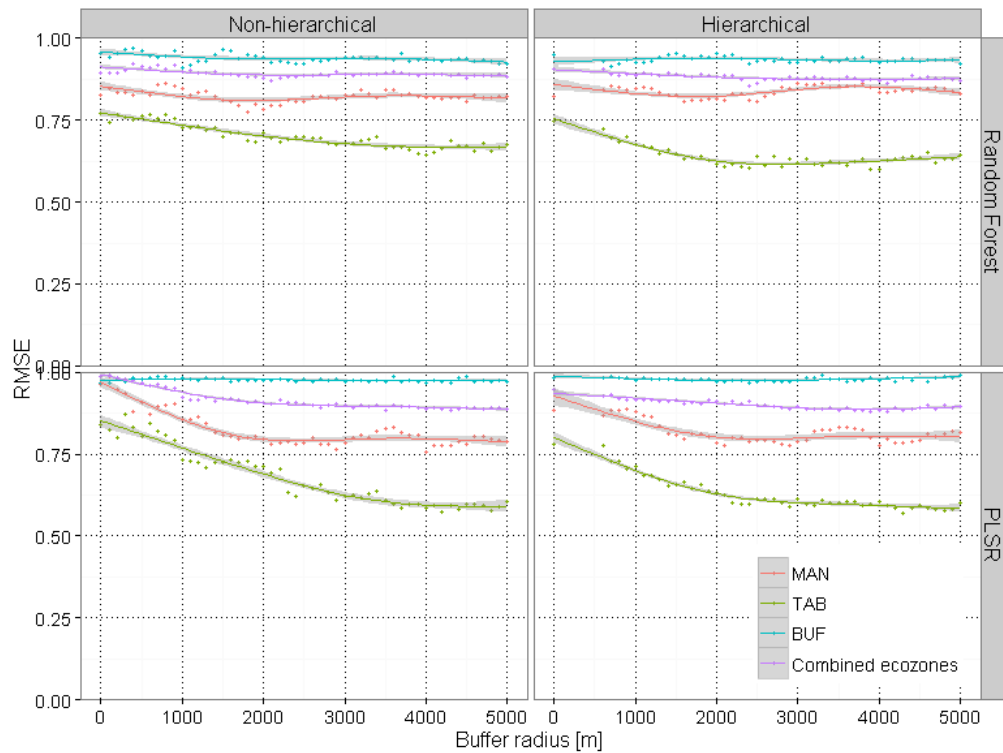


Figure A 5: RMSE of the non-hierarchical (left) and the hierarchical (right) models of the combined ecozones in comparison to the respective model result from the three study sites MAN, TAB and BUF based on the Random Forest (upper row) and PLSR model (lower row). The models per site are derived from the cut-set of data that match all three zones (see Table A 1).

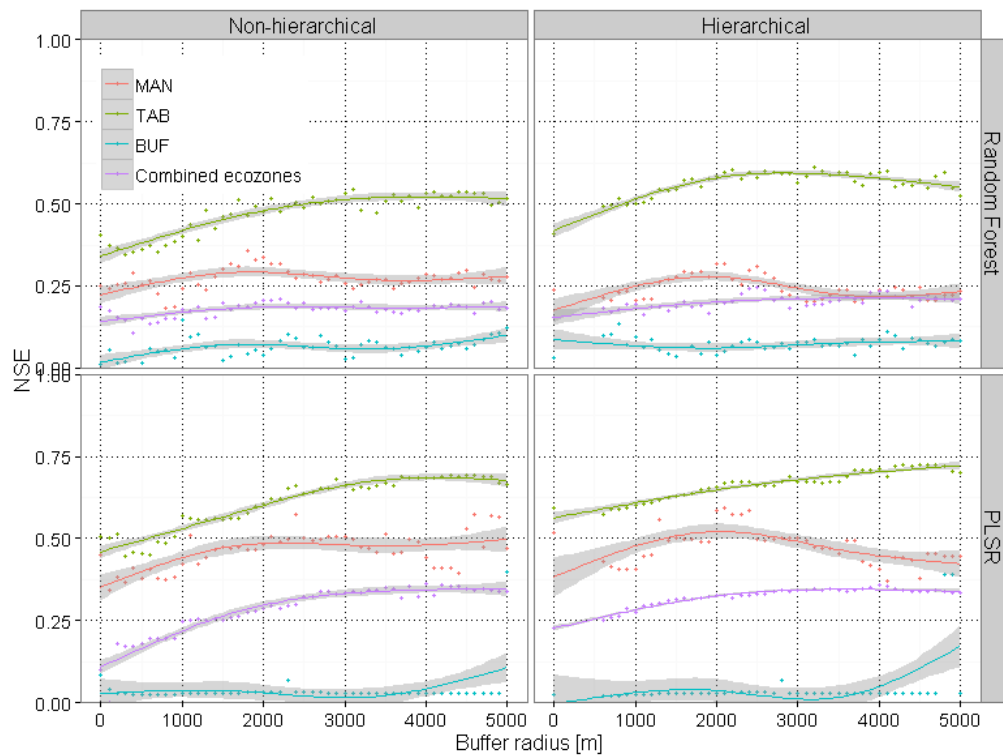


Figure A 6: NSE of the non-hierarchical (left) and the hierarchical (right) models of the combined ecozones in comparison to the respective model result from the three study sites MAN, TAB and BUF based on the Random Forest (upper row) and PLSR model (lower row). The models per site are derived from the cut-set of data that match all three zones (see Table A 1).

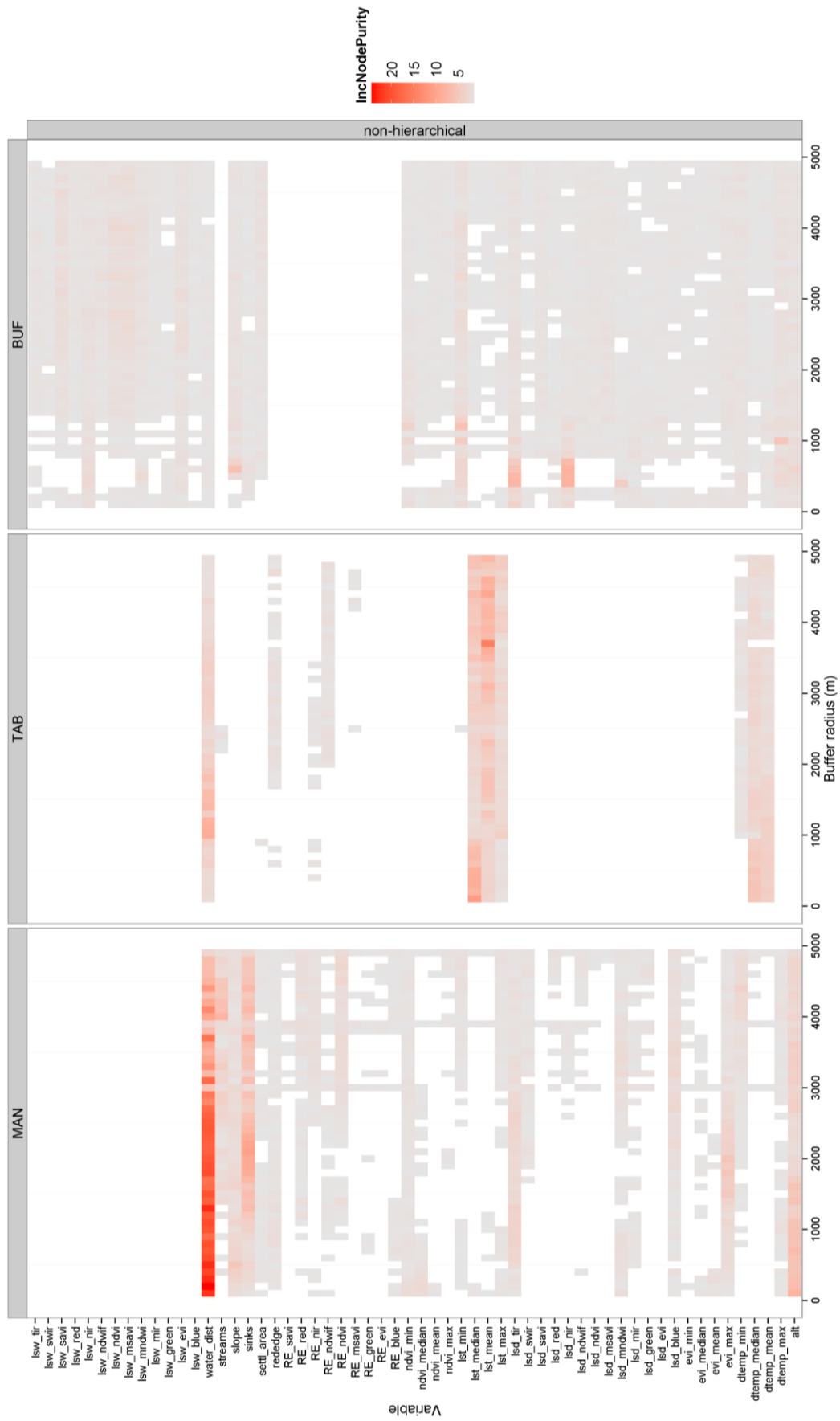


Figure A 7: Variable importance of each study site from the measure of increase in node purity provided by the Random Forest model for the non-hierarchical model approach. A dark red cell represents highest increase in node purity (variable importance) and grey lowest. The white cells represent missing variables and variables that are below the threshold of 1.

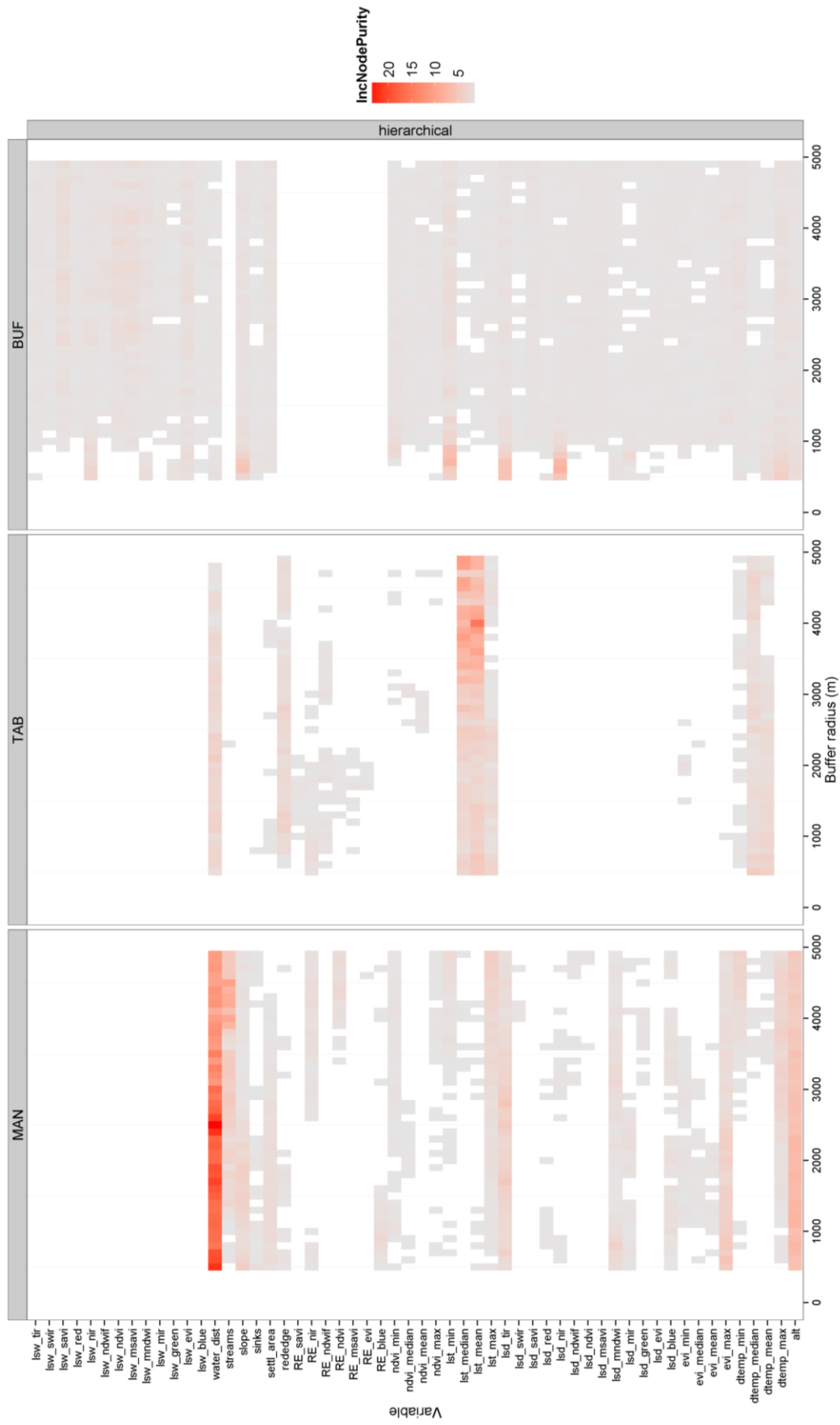


Figure A 8: Variable importance of each study site from the measure of increase in node purity provided by the Random Forest model for the hierarchical model approach. A dark red cell represents highest increase in node purity (variable importance) and grey lowest. The white cells represent missing variables and variables that are below the threshold of 1.

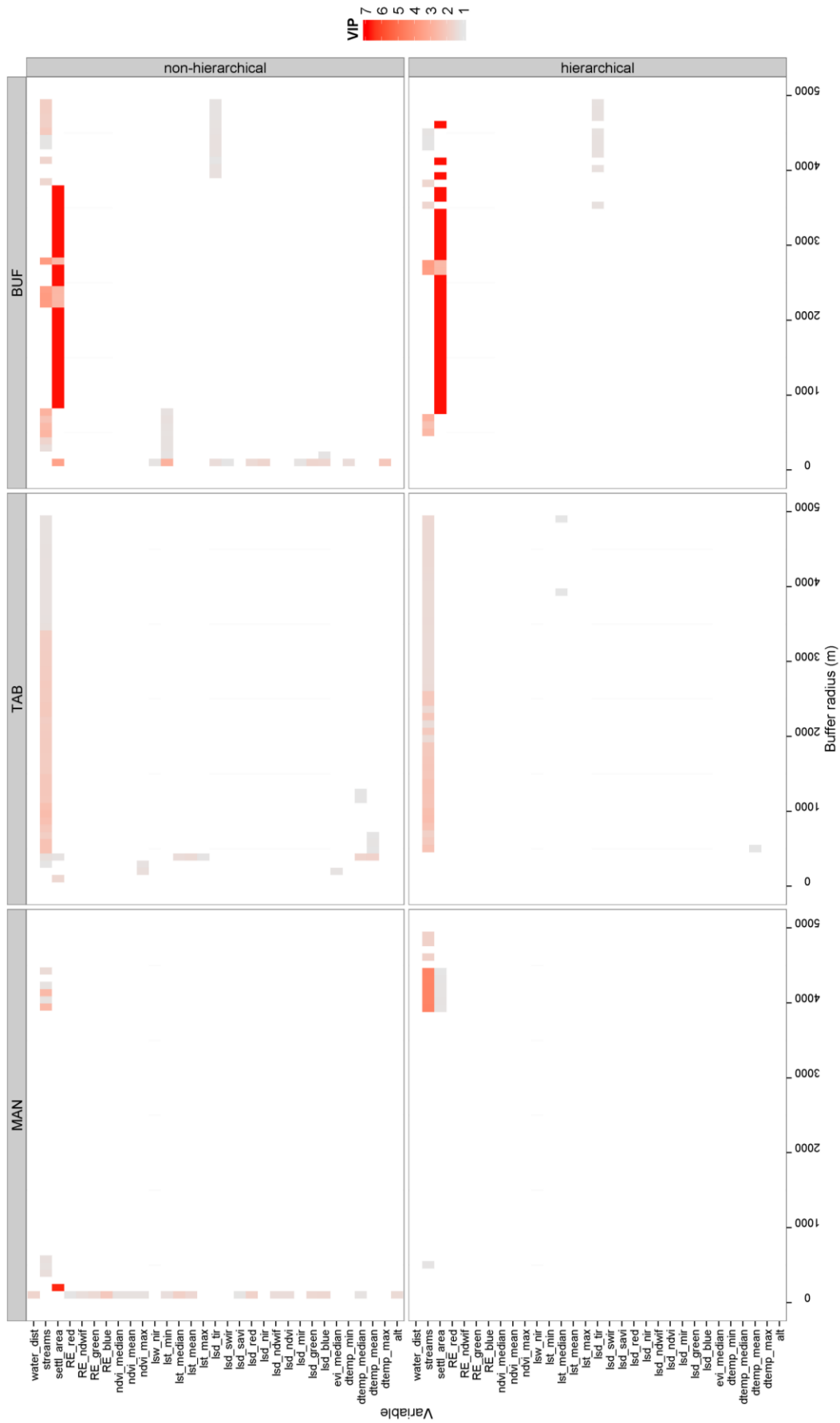


Figure 9: Variable importance of each study site from VIP index calculated from the PLSR model, for the non-hierarchical (upper row) and the hierarchical (lower row) model approach. A dark red cell represents highest variable importance and light blue lowest. The white cells represent missing variables and variables that are below the VIP threshold of 1.

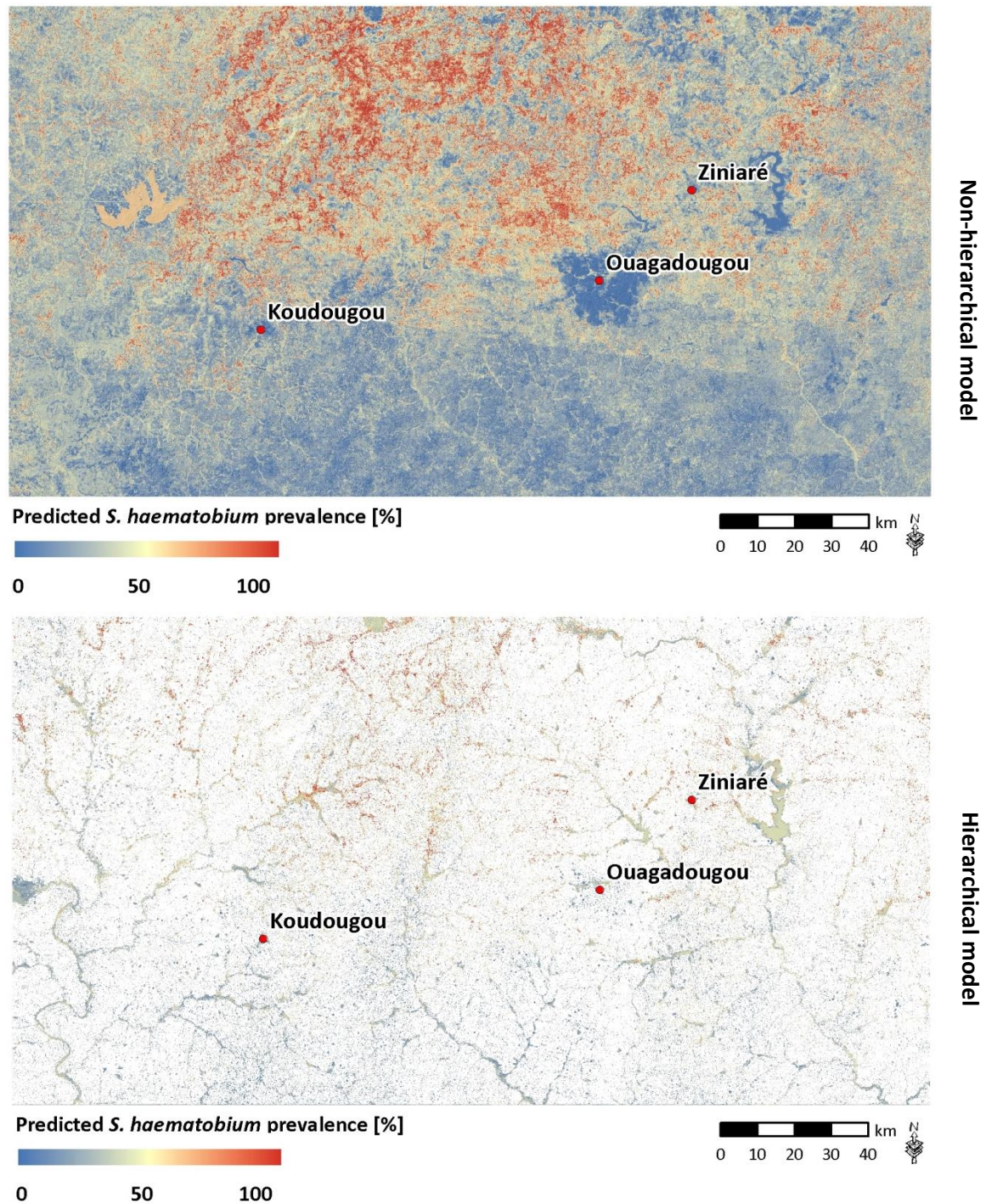


Figure A 10: Comparison between non-hierarchical (top) and hierarchical (bottom) prediction of *S. haematobium* prevalence for the study site BUF. This example is based on spatial modelling within a buffer extent of 3 km around school location.

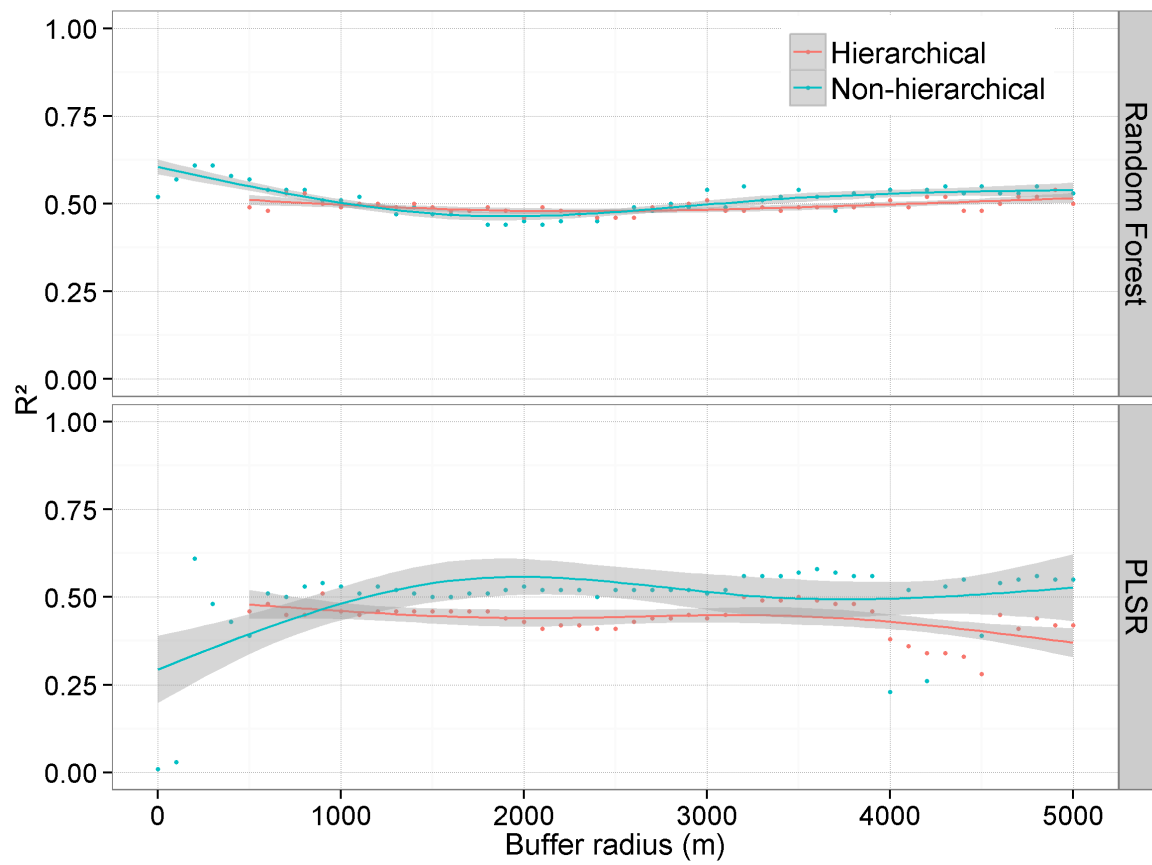


Figure A 11:  $R^2$  values resulting from a linear model of the validation of spatial predictions based on an independent external dataset at the study site MAN.

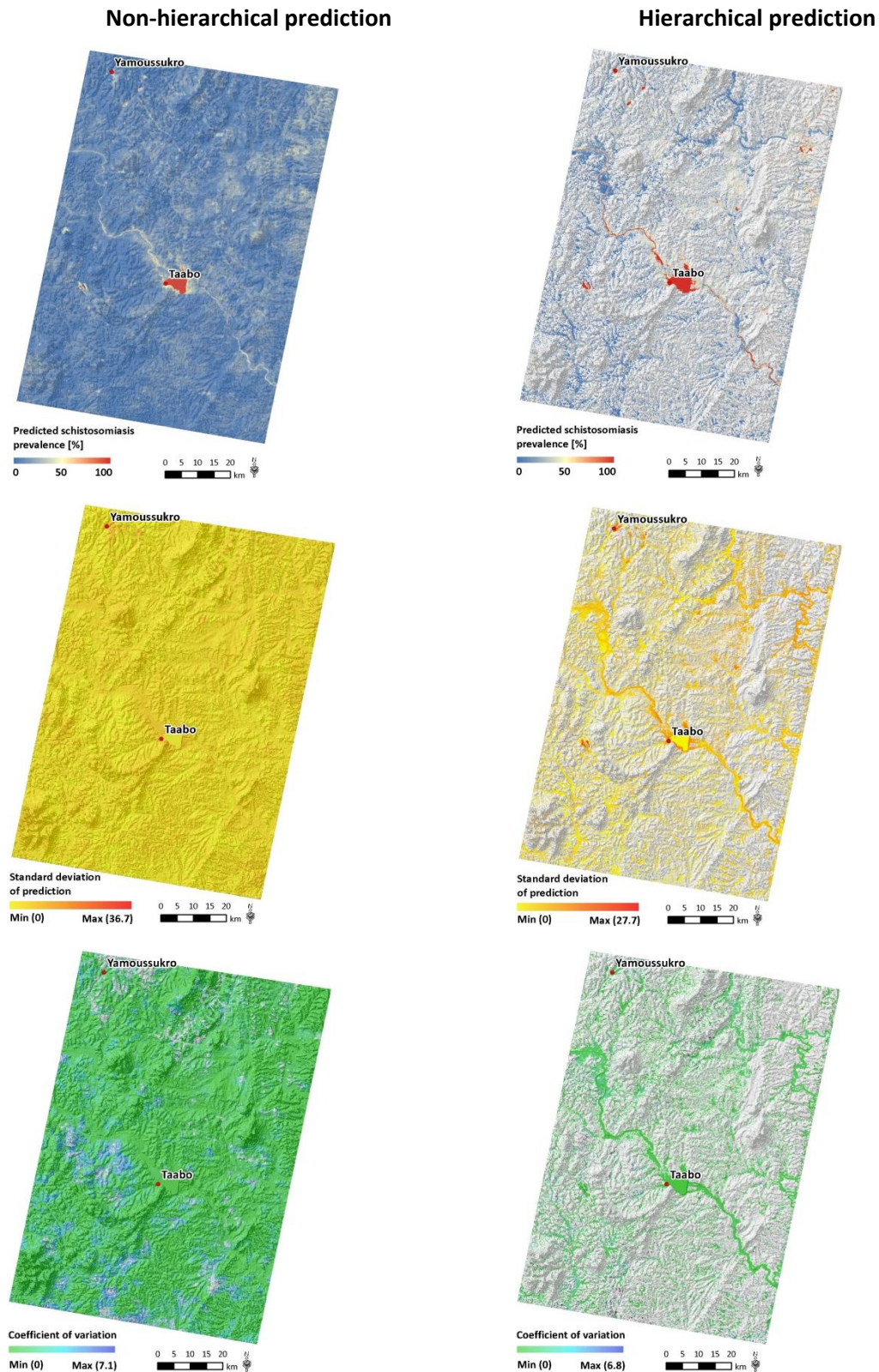


Figure A 12: Mean prevalence (top row), standard deviation (middle row) and coefficient of variation (bottom row) derived from buffer extents between 0 and 5 km using the PLSR model in TAB.

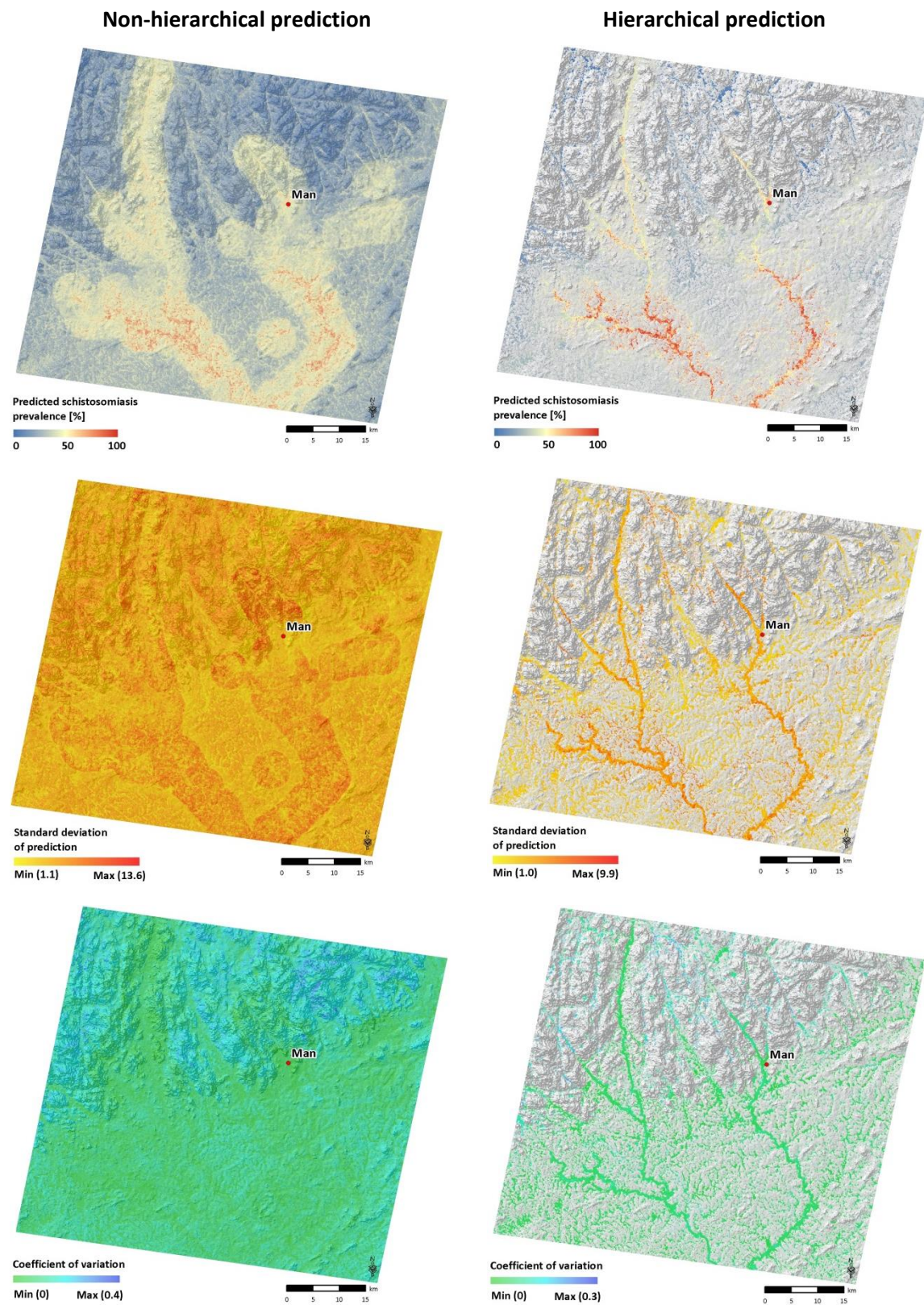


Figure A 13: Mean prevalence (top row), standard deviation (middle row) and coefficient of variation (bottom row) derived from buffer extents between 0 and 5 km using the Random Forest mode in MAN.



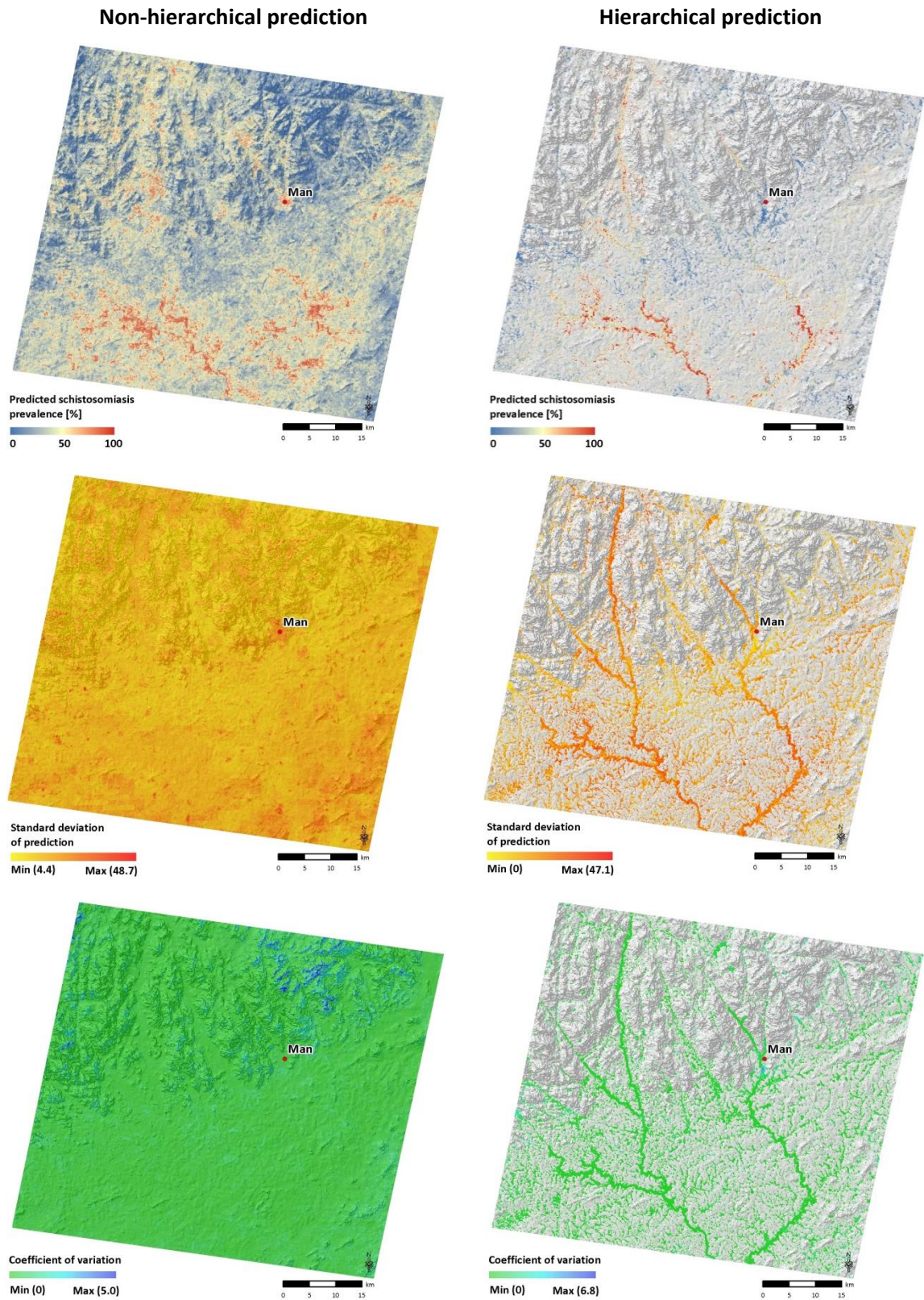


Figure A 14: Mean prevalence (top row), standard deviation (middle row) and coefficient of variation (bottom row) derived from buffer extents between 0 and 5 km using the PLSR mode in MAN.



# Abbreviations

AID	Automatic Interaction Detection
ASI	Italian Space Agency
ASTER	Advanced Spaceborne Thermal Emission and Reflection Radiometer
AVHRR	Advanced Very High Resolution Radiometer
BRDF	Bi-directional reflectance distribution function
BUF	Study site around Ouagadougou in Burkina Faso
CART	Classification and Regression Trees
DALY	Disability-adjusted life year
DEM	Digital elevation model
DFD	German Remote Sensing Data Center
DLR	German Aerospace Center
EADS	European Aeronautic Defence and Space Company
ECOWAS	Economic Community of West African Nations
EMR	Electromagnetic radiation
EOS	Earth Observing System
EROS	Earth Resources Observation and Science
EVI	Enhanced vegetation index
FAO	Food and Agricultural Organization of the United Nations
GCP	Ground control point
GDD	Growing degree days
GDEM	Global digital elevation model
GIS	Geographic information systems
GNTD	Global Neglected Tropical Disease database
GPS	Global Positioning System
GTOPO30	Global 30 Arc-Second Elevation
GUF	Global urban footprint
HSI	Habitat suitability index
ICOSA	Integrated Control of Schistosomiasis in Sub Saharan Africa
IncMSE	Increase of the mean squared error
IncNodePurity	Increase in node purity

JPL	Jet Propulsion Laboratory
JSS	Japan Space Systems
L1B	Level 1B (level of RS data preprocessing including radiometric and geometric sensor corrections)
Landsat 5 TM	Landsat 5 Thematic Mapper
Landsat 7 ETM+	Landsat 7 Enhanced Thematic Mapper Plus
LCCS	Land Cover Classification System
LP DAAC	Land Processes Distributed Active Archive Centre
LPGS	Level 1 Product Generation System
LST	Land surface temperature
LV	Latent variable
MAN	Study site around the city of Man in Côte d'Ivoire
MCDCA	Multi-criteria decision analysis
MDG	Millenium Development Goal
MLR	Multiple linear regression
MNDWI	Modified normalised difference water index
MODIS	Moderate Resolution Imaging Spectroradiometer
MOD09GA	Surface reflectance daily L2G global 1km and 500m
MOD09GQ	Surface reflectance daily L2G global 250m
MOD11A2	Land surface temperature & emissivity 8-day L3 global 1km
MOD12	MODIS global land cover product
MOD13Q1	Vegetation indices 16-day L3 global 250m
MSAVI	Modified soil-adjusted vegetation index
$m_{try}$	Subset of predictor variables (Random Forest)
NASA	National Aeronautics and Space Administration
NDVI	Normalized difference vegetation index
NDWI	Normalized difference water index
NIPALS	Nonlinear iterative partial least squares
NOAA	National Oceanic Atmospheric Administration
NSE	Nash-Sutcliffe efficiency
OLS	Ordinary least-squares
PCA	Principal component analysis
PLSR	Partial least squares regression
$R^2$	Coefficient of determination
RESA	RapidEye Science Archive
RMSE	Root mean square error
RS	Remote sensing
RSS	Residual sum of squares
SAVI	Soil-adjusted vegetation index
SCI	Schistosomiasis Control Initiative
SLC	Scan Line Corrector
SRTM	Shuttle Radar Topography Mission

TAB	Study site around Lake Taabo in Côte d'Ivoire
TPH	Swiss Tropical and Public Health Institute
TRMM	Tropical Rainfall Measuring Mission
USAID	United States Agency for International Development
USFWS	United States Fish and Wildlife Service
USGS	United States Geological Survey
UTM	Universal Transverse Mercator
VIP	Variable importance measure (PLSR)
WGS84	World Geodetic System 1984
WHA	World Health Assembly
WHO	World Health Organisation



# Glossary

<b>Acrisol</b>	Acidic soils with a layer of clay accumulation. This class consists only of clays with low cation exchange capacity.
<b>Aestivation</b>	The ability of intermediate host snails to survive under dry conditions for a certain period of time.
<b>Anti-helminthic drug</b>	Pharmaceutical that takes effect to kill parasitic worms that live within human bodies.
<b>Arenosol</b>	Sandy soils with little profile development.
<b>Cambisol</b>	Soils with slight profile development that is not dark in colour.
<b>Chiroptera</b>	Scientific name for the order of bats.
<b>Cercaria, -ae</b>	Larval stage of the <i>Schistosoma</i> parasites that are free swimming in water and able to penetrate through the intact skin of humans.
<b>Disease vector</b>	A vector is any living agent (animal or microorganism) that carries and transmits an infectious pathogen into another living organism.
<b>Endemic regions</b>	Regions, where the pathogen or parasite is present.
<b>Endorheic</b>	Endorheic waterbodies do not drain into the sea but pertain in the interior drainage basin.
<b>Exorheic</b>	Exorheic waterbodies do not drain into the sea.
<b>Ferralsol</b>	Highly weathered soils rich in sesquioxide clays and with low cation exchange capacities.
<b>Gleysol</b>	Freshwater saturated soils.
<b>Haematuria</b>	Symptom of infection with <i>S haematobium</i> . Red blood cells can be found in the urine.
<b>Intermediate host</b>	An intermediate host is any living agent that harbours the parasite only for a short transition period during which usually some developmental stage is completed.
<b>Leptosol</b>	Shallow soil over hard rock or highly calcareous material or a deeper soil that is extremely gravelly and/or stony.

<b>Lixisol</b>	Soils with subsurface accumulation of low activity clays and high base saturation that develop under intensive tropical weathering conditions.
<b>Miracidium, -a</b>	Larval stage of the <i>Schistosoma</i> parasites that are free swimming in water and need to find a suitable intermediate host snail to further develop to the next larval stage of cercariae.
<b>Plinthosol</b>	Soil type defined by a subsurface layer containing an iron-rich mixture of clay minerals (chiefly kaolinite) and silica that hardens on exposure into ironstone concretions known as plinthite. The impenetrability of the hardened plinthite layer, as well as the fluctuating water table that produces it, restrict the use of these soils to grazing or forestry.
<b>Prepatent period</b>	Period of time, which is necessary for a species to develop (e.g. from egg to adult parasite).
<b>Prevalence rate</b>	Proportion of a population found to have a disease.
<b>Pulmonate snail</b>	Snails that belong to an informal group of snails that have the ability to breathe air.
<b>Regosol</b>	Surface layer of rocky material.
<b>Vector competence</b>	Vector competence refers to the ability of arthropods to acquire, maintain, and transmit a pathogen or microbial agent to the final host.
<b>Vertisol</b>	Clayey soils that form deep and wide cracks when dry.



# List of figures

Figure 1-1: Conceptual framework of environment-related diseases .....	3
Figure 1-2: Conceptual framework linking remotely sensed images with diseases .....	5
Figure 1-3: Overview of the structure of this thesis.....	9
Figure 2-1: Overview of the study area Burkina Faso and Côte d'Ivoire in West Africa .....	12
Figure 2-2: Topographic contrast in the study area .....	13
Figure 2-3: Soil types of the study area.....	14
Figure 2-4: Climate zones of the study area and Walther Lieth climate diagrams of selected sites .....	16
Figure 2-5: Vegetation zones in the study area.....	17
Figure 2-6: Hydrology of the study area.....	19
Figure 2-7: Population density in the study area .....	21
Figure 2-8: Geographic distribution of schistosomiasis prevalence in the study area on district level .....	24
Figure 3-1: Global distribution of schistosomiasis .....	28
Figure 3-2: Consecutive steps of the parasite life cycle of schistosomiasis transmission.....	30
Figure 3-3: Electromagnetic spectrum .....	39
Figure 3-4: Spectral signature of vegetation, soil and water .....	40
Figure 3-5: Energy emitted from black body radiators for Sun and Earth as a function of wavelength .....	40
Figure 3-6: Theoretical concept of ecological and social-ecological disease niche.....	52
Figure 3-7: Conceptual framework of the model approaches used in this study .....	54
Figure 4-1: Steps in pre-selection of epidemiological data .....	60
Figure 4-2: Spatial distribution of epidemiological data from the GNTD for Burkina Faso and Côte d'Ivoire.....	61

Figure 4-3: Patterns of the spatial structure of schistosomiasis prevalence in the three study sites .....	63
Figure 4-4: Spectral bands of multi-scale RS data from the sensors RapidEye, Landsat 5 TM, and Terra MODIS.....	64
Figure 4-5: Processing chain to derive topographic variables from the ASTER GDEM.....	72
Figure 4-6: Main types of potential disease transmission sites within the study site BUF .....	74
Figure 5-1: Suitability of potential transmission sites for schistosomiasis in Burkina Faso.....	76
Figure 5-2: Overview of a mechanistic model procedure to derive environmental suitability for schistosomiasis transmission using RS data .....	78
Figure 5-3: Sub-priority functions of relative suitability of habitat stability for <i>S. mansoni</i> and <i>S. haematobium</i> .....	80
Figure 5-4: Sub-priority functions of relative suitability of water temperature for <i>S. mansoni</i> and <i>S. haematobium</i> parasites (left) and <i>Bio. glabrata</i> and <i>Bu. truncatus</i> (right).....	81
Figure 5-5: Sub-priority function of relative suitability of water flow velocity.....	82
Figure 5-6: Sub-priority function of relative suitability of water depth .....	83
Figure 5-7: Sub-priority function of relative suitability of vegetation coverage .....	84
Figure 5-8: Sub-priority function of relative suitability of stream order. ....	84
Figure 5-9: Sub-priority function of relative suitability of sink depth.....	85
Figure 5-10: Composition of habitat suitability variables to model environmental suitability for schistosomiasis transmission .....	86
Figure 5-11: Overview of single habitat variable suitability and result of a mechanistic model of environmental suitability for schistosomiasis transmission .....	91
Figure 5-12: Environmental suitability for potential schistosomiasis transmission in the study site BUF .....	97
Figure 5-13: Environmental suitability for potential schistosomiasis transmission in the study site MAN .....	98
Figure 5-14: Environmental suitability for potential schistosomiasis transmission in the study site TAB .....	99
Figure 6-1: Illustration of the spatial discrepancy between the measurement of schistosomiasis prevalence at a school (surrounded in red) and the location where disease transmission has potentially occurred.....	106
Figure 6-2: The impact of different ecological regions on RS based schistosomiasis risk models .....	107
Figure 6-3: Schematic representation of a Random Forest regression tree.....	109
Figure 6-4: Schematic representation of the PLSR approach .....	111
Figure 6-5: Multi-scale analysis of RS environmental factors.....	113

Figure 6-6: Flowchart of the hierarchical model approach .....	114
Figure 6-7: Comparison between model results from RapidEye, Landsat 5 TM, and Terra MODIS data for the study site MAN .....	118
Figure 6-8: R <sup>2</sup> values per selected buffer radius of the non-hierarchical (red) and the hierarchical (blue) models.....	119
Figure 6-9: Hierarchical model performance in different ecological regions and for the cross-ecozonal model .....	120
Figure 6-10: Variable importance measured for selected variables in comparison between the three study sites .....	121
Figure 6-11: Comparison of spatial risk prediction between non-hierarchical and hierarchical model approach at the study site MAN .....	123
Figure 6-12: Comparison of spatial risk prediction between non-hierarchical and hierarchical model approach at the study site TAB .....	124
Figure 6-13: Mean prevalence (top row), standard deviation (middle row) and coefficient of variation (bottom row) derived from buffer extents between 0 and 5km using the Random Forest model.....	126
 <b>Appendix</b>	
Figure A 1: Field verification form to guide the sampling of environmental in-situ data .....	160
Figure A 2: Impact of spatial resolution on model accuracy .....	161
Figure A 3: RMSE values of the non-hierarchical (red) and the hierarchical (blue) models .....	162
Figure A 4: NSE of the non-hierarchical (red) and the hierarchical (blue) models.....	162
Figure A 5: RMSE of the non-hierarchical (left) and the hierarchical (right) models of the combined ecozones.....	163
Figure A 6: NSE of the non-hierarchical (left) and the hierarchical (right) models of the combined ecozones.....	163
Figure A 7: Variable importance of each study site from the measure of increase in node purity provided by the Random Forest model for the non-hierarchical model approach .....	164
Figure A 8: Variable importance of each study site from the measure of increase in node purity provided by the Random Forest model for the hierarchical model approach.....	165
Figure A 9: Variable importance of each study site from VIP index calculated from the PLSR model.....	166
Figure A 10: Comparison between non-hierarchical (top) and hierarchical (bottom) prediction of <i>S. haematobium</i> prevalence for the study site BUF.....	167
Figure A 11: R <sup>2</sup> values resulting from a linear model of the validation of spatial predictions based on an independent external dataset.....	168

Figure A 12: Mean prevalence (top row), standard deviation (middle row) and coefficient of variation (bottom row) derived from buffer extents between 0 and 5km using the PLSR model in TAB ..... 169

Figure A 13: Mean prevalence (top row), standard deviation (middle row) and coefficient of variation (bottom row) derived from buffer extents between 0 and 5km using the Random Forest mode in MAN ..... 170

Figure A 14: Mean prevalence (top row), standard deviation (middle row) and coefficient of variation (bottom row) derived from buffer extents between 0 and 5km using the PLSR mode in MAN ..... 171

# List of tables

Table 2-1: Characteristics of climate types in the study area .....	15
Table 2-2: Selected socio-cultural and socio-economic indicators for the study area .....	22
Table 3-1: Overview of parasite-, snail- and human-related factors that modify, retain or intensify the cycle of schistosomiasis transmission .....	32
Table 3-2: Overview of RS data and environmental variable investigated for spatial analyses of schistosomiasis .....	46
Table 3-3: Overview of the contribution of RS for schistosomiasis risk assessment .....	56
Table 4-1: Description of the epidemiological database for each of the three study sites .....	62
Table 4-2: Technical details of the multi-scale RS data used in this study .....	65
Table 4-3: Overview of RS environmental variables used in this study and its availability for the three selected study sites .....	69
Table 5-1: Typical habitat types of schistosomiasis transmission in Burkina Faso .....	88
Table 5-2: Modelled suitability of single habitat variables and the HSI in comparison to field-based expert judgement for water sites .....	92
Table 5-3: Modelled suitability of single habitat variables and the HSI in comparison to field-based expert judgement for potential water sites .....	93
Table 5-4: Spearman rank correlation coefficients for human Schistosoma prevalence and modelled environmental suitability and its corresponding confidence intervals .....	95
Table 6-1: Overview of bands and indices used for multi-scale and multi-sensor analysis .....	112



# Curriculum Vitae

## Personal Information

Name	Yvonne Walz
Nationality	German
Date of birth	February 20, 1976
Marital status	Married (two children)

## Education

since 10/2010	Ph. D. Student, University of Würzburg Thesis: Remote sensing for disease risk profiling: a spatial analysis of schistosomiasis in West Africa Supervisors: Prof. Dr. S. Dech, Prof. Dr. Jürg Utzinger
01/2008 - 04/2010	UNIGIS Master of Geographical Information Science and Systems (GIS) Correspondence course at the University of Salzburg (Grade: 1.0) Master thesis: Challenges in using remotely sensed temperature for epidemiology Supervisor: Prof. Dr. J. Strobl
10/1999 - 07/2005	Diploma in Geography (side subjects: Geology, Botany) Studies at the University of Würzburg (Grade: 1.2) Diploma thesis: Measuring Burn Severity in Forests of South-West Western Australia Using MODIS Supervisors: Prof. Dr. S. Dech, Dr. S. Maier
04/1996 - 03/1999	Vocational training as hospital nurse
07/1995	Valentin-Heider-Gymnasium (Grammar School), Lindau, Germany Abitur (A-levels equivalent)

## Work Experience

Since 05/2014	Research Associate, United Nations University – Institute for Environment and Human Security (UNU-EHS), Bonn: EU FP7 Project “Enabling knowledge for disaster risk reduction in integration to climate change adaptation” (KNOW-4-DRR)
---------------	---

10/2010 – 06/2014	PhD scholar, Department of Remote Sensing, University of Würzburg, in cooperation with the German Aerospace Center (DLR) and the Swiss Tropical and Public Health Institute (TPH), Basel, Switzerland
10/2010 - 09/2012	Scientific assistant of the head of the Department of Remote Sensing, University of Würzburg: Research coordination and administration
06/2010 - 09/2010	Scientific staff at the Center for Remote Sensing of Land Surfaces (ZFL) at the University of Bonn: Establishment of project MON-ALB (monitoring the impact of a plant pathologic beetle)
10/2006 - 04/2009	Scientific staff at the Institute for Hygiene and Public Health (IHPH) at the University of Bonn – WHO Collaborating Centre for Health Promoting Water Management and Risk Communication: GIS and Web-GIS applications to analyse and visualise health data, EU-funded project on impact of climate change on water- and food-borne diseases, editor of the WHO-CC Newsletter
03/2006 - 09/2006	Scientific staff at the DLR – German Remote Sensing Data Center (DFD) – Center for satellite-based crisis intervention: Responsibility for Health Mapping in RESPOND GMES Service Elements
05/2005 - 12/2005	Scientific assistant at the Department of Remote Sensing, University of Würzburg: Data pre-processing and analysis within the BIOTA and Khorezm project
03/2004 - 11/2004	Scholar (financed by DAAD) at the Department of Land Information (DLI) – Satellite Remote Sensing Services (SRSS), Perth, Australia: Diploma thesis, field validation in cooperation with Department of Conservation and Land Management (CALM)
11/01 - 12/01	Internship at DED (German Development Service), GIZ (German institute for international cooperation) and CONAF (Chilean National Forest Corporation), Castro/Chiloé, Chile: Project „Conservation and sustainable management of native forests“, regional approaches of environmental education
10/01 - 11/01	Internship at the Ministry of Agriculture of Chile - CONAF, Temuco, Chile: Eco-tourism concepts and management in the National Park Villarrica

## Computer Skills

Software: Arc-GIS, ERDAS Imagine, ENVI, ATCOR, TiSeG, ER Mapper, Intergraph Geomedia, SPSS, Adobe InDesign, Adobe Photoshop, Corel Draw, EndNote, MS Office  
Programming skills: R



## Languages

German: native speaker  
 English: fluent – spoken and written  
 Spanish: advanced skills (certificate Escuela Sakribal, Quetzaltenango, Guatemala 06/99 - 07/99)  
 French: advanced skills (certificate of Ifalpes Institute Francais, Chambéry, France 01/06 - 02/06)  
 Latin: Latinum

## Publications

### Journal article

Walz, Y.; Maier, S.W.; Dech, S.W.; Conrad, C.; Colditz, R.R. (2007): Classification of burn severity using MODIS: A case study in the jarrah-marri forest of southwest Western Australia. In: *International Journal of Geophysical Research - Biogeosciences*, Special Edition: Remote sensing contributions to forest fire effects assessment, 14 pp

### Book contribution

Walz, Y.; Wegmann, M.; Dech, S.W. (2012): Beitrag von hochaufgelösten Rapid Eye Daten zur räumlichen Risikoanalyse in der Gesundheitsforschung am Beispiel von Schistosomiasis. In: Proceedings 4. RESA Nutzerworkshop, DLR Rapid Eye Science Archive (RESA), DLR Neustrelitz, pp. 325-332

### Conferences

Walz, Y.; Wegmann, M.; Dech, S.; Raso, G.; Utzinger, J. (2014): Anwendung der Fernerkundung für die räumliche Risikoabschätzung der Schistosomiasis. Jahrestagung AK Medizinische Geographie: „Krankheitsrisiken und Gesundheitschancen: Der räumliche Blick auf die Gesundheit“, 9.9.2014 - 11.9.2014 in Remagen (Germany), oral presentation

Walz, Y.; Wegmann, M.; Dech, S. (2012): Schistosomiasis risk assessment from space using high resolution Rapid Eye data. In: *IEEE Geoscience and Remote Sensing Symposium - Remote Sensing for a Dynamic Earth*, 22-27 Jul 2012, S. 7224-7227

Walz, Y.; Wegmann, M.; Dech, S.W. (2012): Transmission of schistosomiasis in Africa and its vulnerability to climate change. 32nd International Geographical Congress (IGC), 26.08.2012 - 30.08.2012 in Cologne (Germany), oral presentation

Walz, Y.; Wegmann, M.; Dech, S.W. (2011): Disease risk assessment from space: case study of human schistosomiasis. Workshop "Biodiversity and Infectious Diseases" (DIVERSITAS Germany e.V.), 28.11.2011 - 29.11.2011 in Frankfurt (Germany), poster presentation

Walz, Y.; Maier, S.; Dech, S. (2005): Measuring burn severity in forests of South-West Western Australia using MODIS. In: *Proceedings of the 5th International Workshop on Remote Sensing and GIS on Forest Fire Management: Fire Effects Assessment*, Zaragoza, Spain, 5 pp

Walz, Y. (2004): Measuring burn severity in forests of South-West Western Australia using MODIS. In: *Conference Proceedings of the 12th Australasian Remote Sensing and Photogrammetry Conference*, Fremantle, Western Australia, 9 pp

### Miscellaneous

Walz, Y. (2010): Challenges in using remotely sensed temperature for epidemiology - illustrated by a spatio-temporal analysis of dengue fever and MODIS land surface temperature in the Vietnamese Mekong Delta. Master Thesis within the postgraduate course "Geographical

Information Science & Systems" (UNIGIS MSc) at the Centre for Geoinformatics (Z\_GIS) at the Paris Lodron University of Salzburg, Austria, 116 pp.

Herbst, S.; Kistemann, T.; Lajoie, L.; Walz, Y.; Wienand, I. (2007): Medical Geography 600 - Centre for International Health. Curtin University of Technology, Elective Unit in the Master of International Health at the Curtin University of Technology, Perth, Australia

Walz, Y.; Scholte, K.; Kemper, T.; Radestock, C.; Voigt, S. (2006): Health mapping service for Ivory Coast - RESPOND GMES Services Supporting Humanitarian Relief. Disaster Reduction & Reconstruction, DLR, 10 pp.

Walz, Y. (2004): Measuring burn severity in forests of south-west Western Australia using MODIS. Diploma thesis at the Department of Remote Sensing, Institute for Geography and Geology, University of Würzburg, 127 pp.

# Acknowledgements

This thesis was accomplished with the help of numerous people that were involved and contributed in many ways to the realisation of this work, which is most deeply acknowledged in the following:

First and foremost, I would like to thank my first supervisor Prof. Dr. Stefan Dech, who provided the unique opportunity to conduct this research at the Department of Remote Sensing, University of Würzburg, in close collaboration with the German Remote Sensing Data Centre (DFD) of the DLR. During all stages of this thesis, he provided critical and constructive scientific advice and above all, continuous personal support, guidance, and encouragement in every aspect.

Furthermore, this interdisciplinary research focus would have never been feasible without the strong scientific support from the discipline of epidemiology. I wish to express my sincerest gratitude to my second supervisor Prof. Dr. Jürg Utzinger (Swiss Tropical and Public Health Institute (TPH) in Basel) for this fruitful collaboration, his constructive inputs and sharing his expertise on schistosomiasis throughout this work. In addition to his highly valuable assistance and the inspiring discussions, his enthusiasm and motivation to conduct research on this disease supported me exceptionally.

I would like to give special acknowledgement also to my scientific mentor Dr. Martin Wegmann for his guidance, the company at seminal meetings and on the field trip to Burkina Faso. His support and encouragement to conduct this research throughout the last years was boundless. I always found his door open for spontaneous and stimulating discussions and instructive comments, which helped me to sort out many of the technical details of my work.

Furthermore, I would like to thank Prof. Dr. Christopher Conrad for his critical but constructive view on my research activities. He encouraged me at the right time to stop and think rather than to proceed with continuous trying and testing, which was the moment when I fully understood the central theme of my research.

I am deeply indebted to my office colleagues Benjamin Leutner and Julian Zeidler, who provided a magnificent support on technical issues and were patient enough to accompany my efforts to learn programming in R for this thesis. In addition, I am very grateful to my other office colleagues Gerald Forkuor and Carina Kübert for their companionship, their valuable help with a number of contextual problems, and the moments of joy we had. For their support regarding any sort of minor and major hardware and software problem, I would like to express my special

thanks to Christian Bäuerlein and Simon Sebold. Furthermore, I want to extend my great appreciation for the unique working atmosphere I experienced with numerous past and present members of the Department of Remote Sensing. Whenever I had a need for discussion or a question to ask, the team was available in a kind and helpful way. Warmest thanks go to: Sarah Asam, Dr. Anna Cord, Dr. Sebastian Fritsch, Vitus Himmler, Dr. Christian Hüttich, Patrick Knöfel, Dr. Tobias Landmann, Dr. Hooman Latifi, Sylvia Lex, Dr. Fabian Löw, Dr. Miriam Machwitz, Dr. Noellie Rüth, Gunther Schorcht, Martin Schmidt, Dr. Matthias Schramm, and Dr. Michael Thiel. I thank Eva-Maria Obermaier and Konstantin Serfas for supporting my research during their Master and Bachelor thesis, respectively.

My colleagues at the DFD, specifically Dr. Thomas Esch, Dr. Doris Klein, Dr. Hannes Taubenböck and Dr. Ursula Geßner, provided highly valuable support for my work through intensive discussion of many specific technical and methodological details. Thank you very much for your view on my analyses and the provision of the GUF data for my study area.

I want to extend my special thanks to several colleagues at the TPH in Basel. Special thanks go to Dr. Giovanna Raso for sharing her expertise on schistosomiasis epidemiology in Côte d'Ivoire and the highly valuable guidance and support throughout this research. Furthermore, I am grateful to PD Dr. Penelope Vounatsou, Dr. Nadine Schur, and Dr. Eveline Hürlimann for the fruitful discussions on schistosomiasis and statistical modelling issues and their warm welcome and support during several productive meetings and my extended visit at the institute. I would like to express my special thanks for their readiness to share the data on schistosomiasis prevalence comprised in the GNTD. At this point, I would also like to acknowledge the several discussions and reflections of Prof. Dr. August Stich from the Tropical Institute of the Missionsärztliche Klinik in Würzburg, which highly supported my epidemiological understanding for this thesis.

During a field trip to Burkina Faso, I had the unique opportunity to experience the local conditions relevant for schistosomiasis transmission risk on the ground. This would not have been possible without the help of various people: Foremost, I would like to thank Prof. Dr. Adjima Thiombiano (University of Ouagadougou) for his warm welcome and providing the relevant papers and infrastructure to realise this field trip. My field assistant Hermann Ouoba participated in all field surveys and did a wonderful job. I shall never forget his great support to get in contact with and access to the variety of relevant institutions and people. I thank Prof. Jean Noel Poda from the Department of Biomedical and Public Health of the University of Ouagadougou for sharing his unique scientific expertise on schistosomiasis transmission in Burkina Faso within highly inspiring discussions. Furthermore, I express my gratitude to Dr. Moussa Dadjoari from the national schistosomiasis control program in Burkina Faso for his valuable information on the activities of schistosomiasis control and prevention in the country and for sharing the most recent database on disease epidemiology. I thank Dr. Arzouma Ouedrago (Head of the health district in Ziniaré) for his insight into schistosomiasis surveillance at the very local scale. Besides this, Compaore Louisi Dimitri from the national institute of statistics and demography provided access to a range of valuable data and information on the country, which I greatly appreciate.

Unfortunately, the planned field trip to study sites in Côte d'Ivoire had to be cancelled at the last minute due to the escalating political unrest during the time. However, the preparation of this trip was strongly supported by colleagues at TPH and the Swiss Centre of Scientific Research

(CSRS) in Abidjan. Here I would like to thank Prof. Dr. Eliézer N'Goran and Dr. Giovanna Raso for their great support. Furthermore, I would like to thank the RESA, which is hosted by the DLR, for providing high resolution RapidEye data for this thesis.

Proof-reading of this thesis was done by a number of people whom I owe my sincerest thanks for their valuable suggestions and critical comments that greatly improved the manuscript: Dr. Anna Klose, Dr. Kirsten Kienzler, Benjamin Leutner, PD Dr. Svenja Meierjohann and Dr. Giovanna Raso. Thereof, Benjamin Leutner was the one who dealt with the first version of my manuscript and was patient enough to go through all chapters. Words cannot express my appreciation for his critical view on my work and the extensive discussions we had. I furthermore would like to give my sincerest acknowledgement to Luke-Sebastian Boshoff for his speedy and thorough check of the English language in the manuscript.

I furthermore would like to give my special thanks to Dr. Valentin Curtef, Johannes Hain, Dr. Jens Jordan, and Benjamin Leutner to support my mathematical procedures and statistical analysis with their expertise. They never failed to answer my questions with the highest of detail.

The last year of this thesis was fully financed by a scholarship received from the "Equal Opportunities for Women in Research and Training" programme at the University of Würzburg. This scholarship gave me the unique opportunity to put this research in the forefront focus of my work. Looking back, I doubt to have been able to accomplish this work with the desired quality and in reasonable time without this financial support.

In addition, I would like to address my sincerest thanks to all of my friends for their patience and encouragement during the last years. Many thanks are due to my parents for their unrestricted support and for believing in me. Most importantly, I would like to express my profound gratefulness to my family. I thank Alex for his enduring patience, his strong commitment on family issues on the week-ends and encouragement whenever I needed it. Our wonderful children Ravi and Yara took my mind off science and back to basic things that matter. They fulfill my every-day life with so much joy. Thank you!

Yvonne Walz



# Eidesstattliche Erklärung

Ich erkläre hiermit, dass die von mir eingereichte Dissertation zum Thema „*Remote sensing for disease risk profiling: a spatial analysis of schistosomiasis in West Africa*“ selbstständig und ausschließlich unter Verwendung der angegebenen Literatur und Hilfsmittel verfasst wurde. Alle den angeführten Quellen wörtlich oder sinngemäß entnommenen Stellen habe ich als solche kenntlich gemacht.

Diese Arbeit ist keiner anderen Prüfungsbehörde vorgelegt worden.

Würzburg, den 13.10.2014

Yvonne Walz