Exploring the Association Between Remotely Sensed Environmental Parameters and Surveillance Disease Data: An Application to the Spatiotemporal Modelling of Schistosomiasis in Ghana

A thesis submitted by

Madeline Wrable

In partial fulfillment of the requirements for the degree of

Master of Science in

Civil and Environmental Engineering

Tufts University

May 2017

Advisor: David M. Gute, PhD, MPH, FACE

Committee Members:

Elena N. Naumova, PhD; Karen C. Kosinski, PhD, MSPH; Magaly Koch, PhD

ACKNOWLEDGEMENTS

I would first like to thank Tufts for creating a graduate program that helped me achieve a childhood dream. Sometime in middle school I stumbled upon a Wikipedia page describing the disease Schistosomiasis, which I found fascinating, and prompted me to spend the rest of the afternoon dreaming and drawing up sketches for my future as a water/humanitarian/infectious disease/inventor-engineer. This dream came true when searching for graduate schools and realizing that Tufts was one of few schools where you could be a "water/humanitarian/infectious disease/inventor-engineer", and it felt like fate when I discovered that schistosomiasis was a research topic in the Department of Civil and Environmental Engineering at Tufts University.

I next would like to thank Dr. David Gute for always being supportive of my passion to pursue further his research on schistosomiasis, and for inspiring me to pursue a Master's program in Environmental Health. It was in his classes that I made lasting friendships with those in my cohort, and through his research that I based my thesis. He has provided kindly guidance throughout my time and endless revisions of countless documents. He was there through achievements and setbacks, and I am thankful to have been able to spend time with him as an incredibly elegant person and to pleasure to have had him as my advisor.

Dr. Karen Kosinski the Ph.D. student of Dr. David Gute was the main driver of the schistosomiasis work, having traveled to Ghana extensively during her own studies and a "water/humanitarian/infectious disease/inventor-engineer" in her own right. Her work ethic, high standards, and empathy for others have made her my role model. Her generosity with time and

research solidified my plan to base my thesis on schistosomiasis in Ghana. I endeavored to craft a thesis that would benefit the future work of my colleagues.

My thesis would not have reached the same spatiotemporal proportions if not for a recommendation of another member of the schistosomiasis research group, Dr. Alexandra (Sasha) Kulinkina, who suggested I take a remote sensing course (RS) and a friend and TA, Carolyn Talmadge who suggested I take a course in geographic information systems (GIS). I admire both of these women for the rigor and passion with which they work. If not for their recommendations I would not have discovered my love for RS/GIS, and has culminated in a position as the GIS Specialist at MIT Libraries.

Dr. Magaly Koch taught me RS as well as allowed me to be her TA for two years. I treasure the knowledge she shared and the time I was able to spend learning from her. She is always enjoyable to spend time with and is interesting to talk to about remote sensing projects all over the world. From her I learned the invaluable skills involved with teaching which helped me in my role as a GIS lab assistant during two of my years at Tufts. I look forward to working with her on other RS endeavors.

Dr. Elena Naumova provided and was herself an invaluable resource. I want to thank her for allowing me to take part in her research group and all the opportunities it provided from traveling to Ghana for data collection to presenting my work in Dubai. I admire her drive and passion, which have mentored a research group of strong capable women. I think that not only has she inadvertently instilled in each of us a love for statistics but also strength of character. I would

like to thank members of our research group, Aishwarya Venkat, Tania Alarcon, and Melissa Cruz for their excellence and friendship during the time we shared.

I would like to thank the MIT Libraries for supporting me to travel to Edinburgh and Prague to present my work. The experiences I had at these statistics and remote sensing conferences will be treasured for a long time still, but nothing is as blissful is returning to begin a job you love. A special thanks to the GIS team supervisor, Daniel Sheehan and team members, Jennie Murack and Anne Graham for their warmth and support as I finished my masters. I am so happy to be able to work with you.

Thank you to my mother Mary Rose and father Les Wrable. They have taken many a phone call of woe and worry and never failed to help me assuage my fears and get back after my goals. Also precious to me is my partner Matthew Webber. I thank you for walking along with me in life and love.

0. Abstract

Schistosomiasis control in sub-Saharan Africa is enacted primarily through mass drug administration, where predictive modeling plays an important role in filling knowledge gaps in the distribution of disease burden. Remote sensing (RS) satellite imagery is used to predictively model infectious disease transmission in schistosomiasis, since transmission requires environmental conditions to sustain specific freshwater snail species. Surveys are commonly used to obtain health outcome data, and while they provide accurate estimates of disease in a specific time and place, the resources required make performing surveys at large spatiotemporal scales impractical. Ongoing national surveillance data in the form of reported counts from health centers is conceptually better suited to utilizing the full spatiotemporal capabilities of publically available RS data, as most open source satellite products can be utilized as global continuous surfaces with historical (in some cases 40-year) timespans. In addition RS data is often in the public domain and takes at most a few days to order. Therefore, the use of surveillance data as an initial descriptive approach of mapping areas of high disease prevalence (often with large focal variation present) could then be followed up with more resource intensive methods such as health surveys paired with commercial, high spatial resolution imagery. Utilization of datasets and technologies more cost effectively would lead to sustainable control, a precursor to eradication (Rollinson et al. 2013).

In this study, environmental parameters were chosen for their historical use as proxies for climate. They were used as predictors and as inputs to a novel climate classification technique. This allowed for qualitative and quantitative analysis of broad climatic trends, and were regressed on 8 years of Ghanaian national surveillance health data. Mixed effect modeling was

used to assess the relationship between reported disease counts and remote sensing data over space and time. A downward trend was observed in the reported disease rates (~1% per month). Seasonality was present, with two peaks (March and September) in the north of the country, a single peak (July) in the middle of the country, and lows consistently observed in December/January. Trend and seasonal patterns of the environmental variables and their associations with reported incidence varied across the defined climate zones. Environmental predictors explained little of the variance and did not improve model fit significantly, unlike district level effects which explained most of the variance. Use of climate zones showed potential and should be explored further. Overall, surveillance of neglected tropical diseases in low-income countries often suffers from incomplete records or missing observations. However, with systematic improvements, these data could potentially offer opportunities to more comprehensively analyze disease patterns by combining wide geographic coverage and varying levels of spatial and temporal aggregation. The approach can serve as a decision support tool and offers the potential for use with other climate-sensitive diseases in low-income settings.

ABBREVIATIONS

AP – Accumulated Precipitation

CDC – Centers for Disease Control and Prevention

CVI – Cluster Validity Index

dT – Diurnal Temperature

GAHI – Global Atlas of Helminth Infections

GHS – Ghana Health Service

GIS – Geographic Information Systems

GNTD – Global Neglected Tropical Disease

ICOSA – Integrated Control of Schistosomiasis in Sub Saharan Africa

KG – Köppen-Geiger

LST – Land Surface Temperature

MDA – Mass Drug Administration

MODIS – Moderate Resolution Imaging Spectroradiometer

NDVI – Normalized Difference Vegetation Index

NFSC - National Freshwater Snail Collection

NTD – Neglected Tropical Disease

PCA – Principal Components Analysis

RS – Remote Sensing

SCORE – Schistosomiasis Consortium for Operational Research and Evaluation

Tcap - three thermal-hydrological domains

TRMM – Tropical Rainfall Monitoring Mission

WHO – World Health Organization

LIST OF EQUATIONS

LIST OF TABLES

LIST OF FIGURES

1. Introduction

1.1 Schistosomiasis

1.1.1 Health Impact

Schistosomiasis is caused by parasitic blood flukes of the genus *Schistosoma*, and is acquired from skin contact with contaminated freshwater bodies. Schistosomiasis affects over 200 million people worldwide, with approximately 779 million at risk of infection (Steinmann et al. 2006), although these numbers likely substantially underestimate the true disease burden (King 2010).

Schistosomiasis is more likely to cause morbidity than mortality, and leads to the loss of around 4.5 million disability adjusted life years of people worldwide (Danso-Appiah et al. 2004). It is considered second only to malaria in terms of impact of a parasitic disease, and is one of the WHO's neglected tropical diseases (NTDs) (WHO 2011). Presently there is evidence to suggest that schistosomiasis, "is arguably the most important cofactor in Africa's AIDS epidemic," as stated by the dean of the National School of Tropical Medicine at Baylor College of Medicine, Dr. Peter J. Hotez (Zinyama-Gutsire et al. 2015; Ndeffo Mbah et al. 2013; Brodish and Singh 2016).

1.1.2 Schistosome Lifecycle

Schistosomiasis is spread through contact with water. When the eggs of the parasite contact a water body, they hatch into miracidia. The miracidia seek out certain species of snails to act as intermediate hosts. Over 350 snail species are suitable hosts for schistosomes; however, three genera of snails are most relevant for public health: Biomphalaria, Bulinus, and Oncomelania, because they serve as intermediate hosts for the three parasite species that most commonly infect

humans: S. haematobium, S. mansoni, and S. japonicum, respectively (Gryseels et al. 2006). The parasites multiply asexually inside the snail and then leave the snails in search of their final hosts. When the parasite comes in contact with a human host, they burrow through the skin and gain entry to the circulatory system. There they mature into adult worms, pair up, and then travel to the mesenteric veins surrounding the bladder (*S. haematobium*) and/or intestine (*S. mansoni* and *S. japonicum*). They then lay eggs, which burrow through the tissue to make their way into the urine or stool. If an infected person then urinates or defecates into fresh water or in soil proximal to bodies of surface water, the life cycle of schistosomiasis continues (CDC 2012). A detailed resource for learning more about the schistosome lifecycle and schistosomiasis pathology can be found in the review article, "Hepatobiliary Schistosomiasis" by Yehia et al. 2014. This thesis focuses specifically on *S. haematobium* and the intermediate host snail, *Bulinus spp*., which is the predominant cause of schistosomiasis in Ghana (Lai et al. 2015).

1.1.3 Distribution

Schistosomiasis is endemic to locations in South America, the Middle East, and Asia, and Africa. However, sub-Saharan Africa has an estimated <90% of the worldwide cases (Chitsulo et al., 2000; Gryseels et al., 2006; Hürlimann et al., 2011). Ghana is one of the most heavily affected countries with an estimated prevalence ≥50% (Utzinger et al. 2009), as reported by the WHO (Figure 1). An estimated 30-35% of the Ghanaian population requires preventative chemotherapy, but only 2–8% receive it annually (WHO 2010). Preventive chemotherapy with praziquantel is the predominant disease control strategy in Ghana. Ghana Health Service (GHS) currently uses a combination of limited field survey results, data from the national surveillance

system, and historical knowledge of endemnicity to determine which districts receive praziquantel.

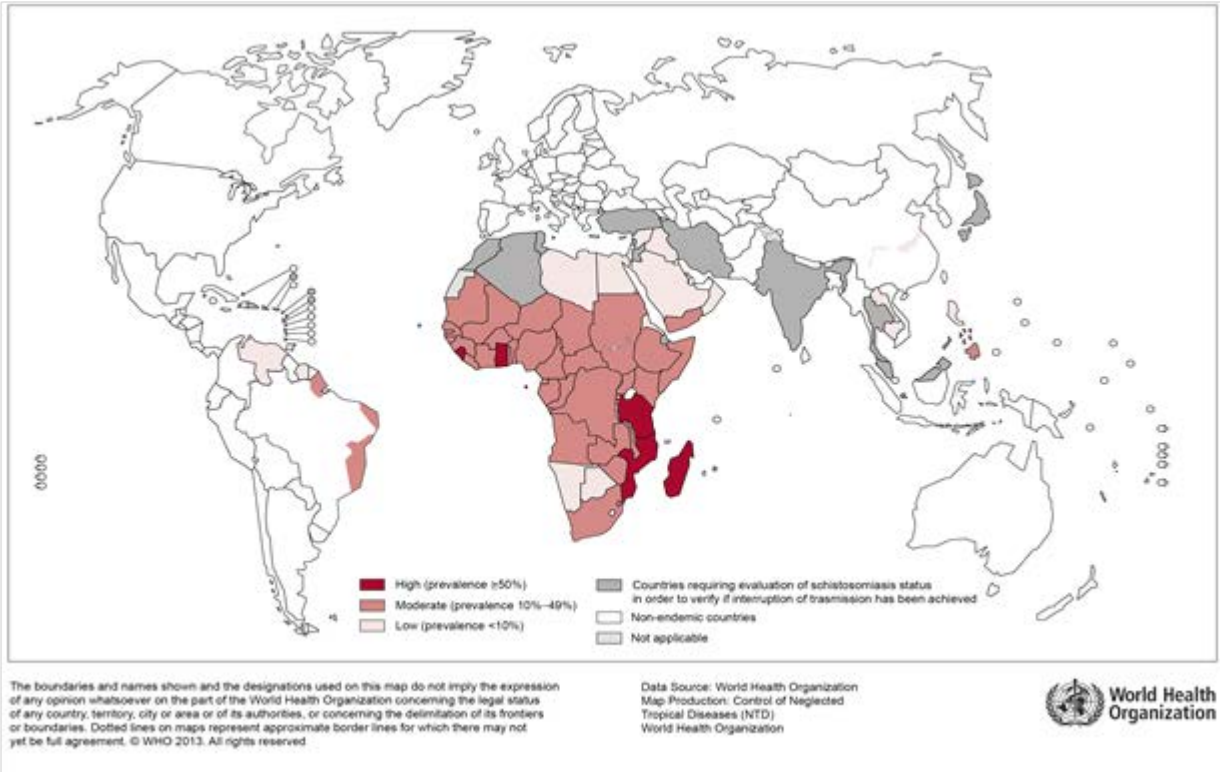


Figure 1: *Schistosomiasis prevalence per country (WHO 2012)*

1.1.4 Prevalence

Estimates of schistosomiasis prevalence in Africa have been made for the last 70 years, but it has been known to plague humans as far back as Egyptian times (Yehia et al. 2014). In 1947, Stoll's paper, "This Wormy World", summarized the state of parasitic worm diseases globally (Stoll 1947). He provided an overview of worm-based diseases along with a set of intervention strategies. Advice for public health professionals on disease mitigation was broken into three parts: 1) proper parasitology education, so that educated decisions could be made; 2) proper application, so that decisions are tailored to specific situations; and 3) use of updated control measures. These recommendations are still useful today. Stoll's estimates for worm prevalence

in sub-Saharan Africa, which was the first attempt to calculate this type of disease burden on a continental scale (Brooker et al., 2000).

In 1965, the global estimate of infected persons was around 200 million (Mandahl-Barth 1965). Around the same time, the distribution of *S. haematobium* and *S. mansoni* in Ghana was investigated through ground surveys (Onori et al., 1963; Odei, 1964; McCullough, 1965). Geographically these surveys (aggregated) spanned over two thirds of the country, leaving out districts in the South and Southwest. It estimated a national prevalence of between 15-20% and 1,000,000+ people having been infected with *S. haematobium* at one point in their lifetime (McCullough and Ali 1965). In the surveyed areas, *S. haematobium* was found to have the highest prevalence, 41-100%, in the north, and with small focal pockets ranging from 1-15%, 16-40%, and 41-100%, throughout the country. The authors noted that schistosomiasis was also an occupational disease associated with fishing and weed clearing around the Kumasi reservoir (McCullough and Ali 1965). They listed the major factors responsible for the distribution of the disease as consisting of human population density and presence of surface water supplies, with the minor factors being migration, settlement type, water-body characteristics and location, water body behavior and location, water behaviors, and extent of the forest (McCullough and Ali 1965). The authors also noted that snails were distributed much more widely than the disease, which puts large areas at risk for expansion of schistosomiasis. This work is one of the earliest to draw conclusions regarding the distribution of schistosomiasis within Ghana on a national scale.

In 1972, global schistosomiasis estimates dropped from 200 to a more conservative 125 million infected and 500 million at risk, with 90 million of the infected (72%) being attributed to Africa

alone (Wright 1972). In 1976 the WHO conducted its own data collection activity via a questionnaire distributed to 121 countries. 103 replies were received, and 59 countries self-reported as endemic with the number rising to 73 at the time of publishing these data (WHO 1981). The questionnaire estimated that in Ghana the prevalence was around 50% primarily due to S. *haematobium* (WHO 1981).

In 1987, the WHO published the "Atlas of the Global Distribution of Schistosomiasis", which provided a brief overview of each country's prevalence distribution as derived from aggregated surveys, climatic and physical conditions, and human activities affecting transmission. (Atlas in the title implies maps) The information specific to Ghana covers the distribution of *S. haematobium* and *S. mansoni* infection (WHO 1987). It confirms that at that time *S. haematobium* was the most prevalent species of schistosome, based on reports dating back to 1956. The Atlas also provided information on snail distribution, snail food/vegetation, water level rise in the rainy season, and that transmission occurs often during periods of flooding/heavy rainfall. The survey information used to create this atlas was aggregated to the district level, tabulated, and mapped. 30 years later, the Atlas was considered to be "the most complete global resource [for schistosomiasis] remains the 1987 *Atlas of the Global Distribution of Schistosomiasis*" (Brooker et al. 2010).

In 2000, schistosomiasis prevalence was mapped in 76 countries with results being tabulated and stored via the dataportal, Global Atlas of Helminth Infections (GAHI), which at the time contained data for 33% of all global district level administration units with a population density of >5 people per km^2 (Brooker et al., 2000). At the time, this atlas was the most up to date aggregation of all the published surveys as well as derived from a substantial grey literature.

However, it only used surveys conducted after 1970. A map produced from the database for Ghana at the time of creation had 11 published references, which provided 57 surveys mapped by region, thus covering 30% of the 10 regions making up the country (Brooker et al., 2000). Prevalence was found to be >50% near Accra and the Northeast border, and 25-50% in the center of Ghana through the south. Masked areas due to low population density (<20 people/km^2), were located in the Northern half of the country. No data was recorded in parts of the Northern half of Ghana, the Southwest corner, and a pocket along the Southeast coast. This digital global atlas was an improvement upon Stoll's more static mode of display, "This Wormy World" prevalence estimates, and provided a glimpse of possible future dynamic digital repositories. The GAHI repository created in 2000, is still active and as of March, 2017 lists 250 records, 173 surveys, and 11 reports and publications for schistosomiasis in Ghana (GAHI 2017).

Subsequently, the Global Neglected Tropical Disease (GNTD) database was created, which allowed for data to be accessed directly and as of 2017 has 35 survey-based datasets for schistosomiasis in Ghana (GNTD 2011). This database was created based on a systematic review of the literature from sources such as PubMed, Institute for Scientific Information (ISI) Web of Knowledge, and the African Journal Online. It built upon the previous database but went further by not putting restrictions on date or language, and expanding the scope of diseases (Hürlimann et al. 2011). The data could also be manipulated through an interactive mapping software, HealthMapper, which was created in conjunction with the WHO (Thomson et al. 2000)

1.1.5 Control

As prevalence information have been gathered and enhanced technologies created so too have control methods progressed. In 1976 the previously mentioned questionnaire administered by WHO estimated that Ghana's budget for healthcare was just over $100 million, making up just shy of 10% of the national budget, with only $200,000 or 0.2% of the national budget allocated to schistosomiasis (WHO 1981). In terms of manpower there were 480 governmental staff and 60 non-governmental staff working on schistosomiasis control. Hospital records analyzed in 1975, showed that no hospitalizations were recorded but that almost 7,000 outpatients were treated. In 1976 the most commonly reported anti-schistosomal drug was niridazole, used in 41 of 55 (75%) of endemic countries, and the most common pesticide against mollusks (molluscicide) was niclosamide, applied in 28 of 55 (51%) of endemic countries (WHO 1981). Control methods were recorded and ordered by frequency of use, as found among the 55 endemic countries. In order of highest to least frequency of use the list was as follows: chemotherapy, health education, installation of water supply, improvement of existing water supply, molluscicide, provision of sanitary facilities, environmental modification by engineering methods, modification of agricultural environment, biological control, and the protection of water sites (WHO 1981). Use of more than three of these control methods was common in roughly half the endemic countries.

Thirty years later in 2006, chemotherapy was the dominant control method (Gryseels et al. 2006). Gryseels attributed this to a, 'fundamental shift over the past few decades', in the availability of the drug praziquantel. Praziquantel is a broad spectrum deworming drug, known for low toxicity and chemical stability. It was originally developed in the parasitological research

laboratories of Bayer AG and Merck KGaA in Germany in the mid-1970s (Greenwood 2008). Originally used in veterinary medicine, it was approved for use in humans in the 1980s, after extensive testing (Frohberg and Schulze Schencking 1981). Additional human testing followed (EMEA 1996; Dollery 1999) with the results indicating an excellent therapeutic index, which led to praziquantel's progressively frequent use in mass treatment campaigns (Seubert et al. 1977). The World Health Organization (WHO) includes it on its Model List of Essential Medicines.

Established in 2002, the Schistosomiasis Control Initiative (SCI) is an organization focused on the control of schistosomiasis morbidity through chemotherapy. SCI funds prevalence surveys, thereby identifying areas requiring mass drug administration (MDA) with the deworming drug praziquantel (Brooker et al. 2010). It is coordinated by the Imperial College of London and focused on specific African countries. Within these countries, SCI promotes two major projects, the Integrated Control of Schistosomiasis in Sub Saharan Africa (ICOSA) and the Schistosomiasis Consortium for Operational Research and Evaluation (SCORE). ICOSA, began in 2010, focuses on schistosomiasis and soil transmitted helminths. It is active in eight countries split into three groups based on the status of their control programs. Group One countries have little or no control organizations, Group Two nations have control programs established but not on a national scale, and Group Three reports successful control programs on the national level and as well as the intention to move from control of schistosomiasis towards elimination. Based on this classification, Ghana exhibits the characteristics of Group Two.

Another area of focus for SCI is mapping. The SCI website promotes the use of mapping to direct disease control efforts, and on its website describes how to utilize the GAHI database. It recommends the use of the following environmental variables: altitude, temperature, and surface

water, as well as population distribution (London 2016). When there are not already available datasets for particular locations, the SCI carries out their own surveys. Their mapping often has the focus of tracking prevalence pre- and post- mass drug administration (MDA).

The WHO media center fact sheet on schistosomiasis has a section on prevention and control updated on January 2017. It states that control is administered on broad scales mainly achieved using praziquantel, which it calls effective, safe, and low-cost. It mentions the efficacy of control efforts dating back 40 years (likely in reference to its 1976 questionnaire), and highlights the recent 10-year trend towards larger geographic scales of treatment (WHO 2017).

1.2 Schistosomiasis and Climate

Schistosomiasis transmission requires the presence humans, snails, and parasites, all of which have varying degrees of sensitivity to climate (Gryseels et al. 2006). Environmental variables can serve as a proxy to climate, and therefore are useful in modeling schistosomiasis (Brooker et al. 2001; Walz et al. 2015a; Walz et al. 2015b). There are a number of specific parasite, snail, and human characteristics that affect schistosome transmission; these characteristics have been reviewed in detail (Walz et al. 2015a).

Information on snails comes from field surveys, modeling, and laboratory testing. WHO reports that roughly 350 species of snail can act as intermediate hosts for schistosomes and can survive from temperatures ranging from 10˚C to 35˚C. Snails are found in fresh water such as small ponds, streams, large lakes, and rivers with most found in perennial water sources (WHO 1987). Within water bodies the snails tend to prefer shallow waters near the shores as they feed on water plants. Ghana has one of the largest man-made lakes in the world. Lake Volta's 5,000 km of

shoreline is an ideal habitat for snails (WHO 1987). Other parts of Ghana (e.g. Eastern Region) have an abundance of small rivers and streams (Kulinkina et al. 2017) that may also contribute to the country's schistosomiasis burden. *Bulinus* is the snail species responsible for S. haematobium, the most prevalent Schistosoma in Ghana. There are 37 known species of *Bulinus* (Brown 1994), which are subdivided amongst B. *africanus,* B. *forskalii,* B. *reticulatus,* and B. *tuncatus/tropicus (*Rollinson et al. 2001; Mkize et al. 2016). A 2004 report based on 1,543 *Bulinus* snail samples found that temperature followed by water body type best predicted the geographical distribution amongst *Bulinus* sub-species, taken from snail samples in the National Freshwater Snail Collection (NFSC) of South Africa (De Kock et al. 2004). WHO documented that snail reproduction is commonly associated with the temperature range of 22˚C to 26˚C; however, the range for *Bulinus* snails in equatorial nations such as Ghana is wider. Climate has been shown to have an impact on the distribution of the intermediate host for schistosomiasis (Appleton 1978; Brown 1994; Brooker and Michael 2000).

Climate is an important factor in modeling diseases with strong links to environmental parameters. The predominant climate classification system for the past 100 years, the Köppen-Geiger (KG) climate classification system, is loosely based on the assumption that vegetation is the best proxy for climate, and that temperature and precipitation are the best proxies for vegetation (Kottek et al. 2006). Later, it was realized that evapotranspiration was better suited for use with temperature in defining climate, due to its role in aridity and soil moisture. However, evapotranspiration is not always collected by ground-based climate stations and so a complex temperature-precipitation formula was used in its place. Ghana specifically has two climates Af and Aw, where Af has the criteria of having at least 60mm of precipitation in the driest month,

and Aw has the criteria of having less than 60 mm and 100-(r/25) precipitation in the driest month, where r is the average annual precipitation total (mm) (Kottek et al. 2006). Af is considered a "tropical wet-dry climate" and covers most of Ghana. Aw is a "wet equatorial climate" and covers only a small portion of the Southwest corner of Ghana. More broadly, both of these subclimates Af (no dry season) and Aw (winter dry season), fall under the major climate division A, which is considered to be the warmest of the five major climate divisions (A, B, C, D, & E).

In addition existing climate classification schemes have not been designed for public health applications (Liss et al. 2014). Their broad scales are far too large to be useful in national disease control campaigns. They are also dependent on terrestrial climate stations, which cannot be located in sufficient numbers to achieve optimal coverage due to forbidding logistics and costs. The interpolations used to generate the climate grids are complex when made public or unreproducible when proprietary. They are also often lacking in fine temporal resolutions, due to their resource requirements and making them better suited to static applications. These challenges contribute to difficultly in adapting these systems to public health needs, which are usually analyzed at fine spatiotemporal scales, require reproducibility, and dynamic in nature.

1.3 Schistosomiasis and Remote Sensing (RS)

Fortunately, with the introduction of satellites, there is another way to quantify climate data; thereby utilizing climate characteristics which are favorable for snails to serve as proxies for schistosomiasis transmission.

1.3.1 Overview

Developed in the 1950s, remote sensing (RS) is defined as the gathering of information without direct contact. The ways in which this can be done are varied, but commonly are by aircraft or satellite. In this thesis, RS refers to satellite data. Collecting information on the ground can often be time consuming and costly. RS allows for a relatively low-cost tool that can cover a large spatial and temporal scale, making it ideal for use in resource-constrained countries (Walz et al. 2015a), and a more sustainable solution to ground-based data (Gryseels et al. 2006).

RS applications have steadily progressed, and since the 1970s has been commonly used for schistosomiasis modeling and predictive mapping (Simoonga et al. 2009). A number of factors that influence schistosomiasis transmission lend themselves to remotely sensed environmental variables. Two reviews of the primary literature show that the most commonly used environmental variables to predict schistosomiasis transmission are temperature, vegetation, rainfall, water chemistry, distance to water bodies, and elevation (Simoonga et al. 2009; Walz et al. 2015a). The use of ecological zones as opposed to geo-political administrative units has been recommended and is shown to be valuable in predicting the occurrence of schistosomiasis (Simoonga et al. 2008; Grosse 1993; Walz et al. 2015a). Recent papers written on the subject of RS applied to schistosomiasis have explored the environmental variables further so as to better understand their function as transmission proxies (Walz et al. 2015b).

1.3.2 Review Papers

Most studies on the application of RS to schistosomiasis take place in Africa. The use of RS with regards to schistosomiasis was reviewed recently (Simoonga et al. 2009; Walz et al. 2015a). Simoonga et al. analyzed studies from 1996-2008, while Walz et al. focused on studies from

2001-2014. This thesis will give priority to exploring studies included in these reviews, but effort has been made to investigate applicable studies (RS applications to schistosomiasis, preferably in Africa) since 2014 to present.

Simoonga et al. (2009) reviewed studies temporally and spatially. For the temporal analysis, the authors used a PubMed search in which the following terms and Boolean operators were entered: "remote sensing" OR "mapping" OR "prediction" AND "schistosomiasis" AND "Africa", the authors found 41 studies, 32 of which were deemed to be relevant. The number of studies on this subject has been increasing (Simoonga, Utzinger, Brooker, Vounatsou, Appleton, Stensgaard, et al. 2009). The spatial distribution of studies differentiated between snail- or human-applications, and by 2009 most African countries had RS data applied to human schistosomiasis. An adapted version of this search, that takes all the returned results from 1996-2015 using the PubMed input: ((((remote sensing) OR mapping) OR prediction) AND schistosomiasis) AND Africa has been created (Figure 2).

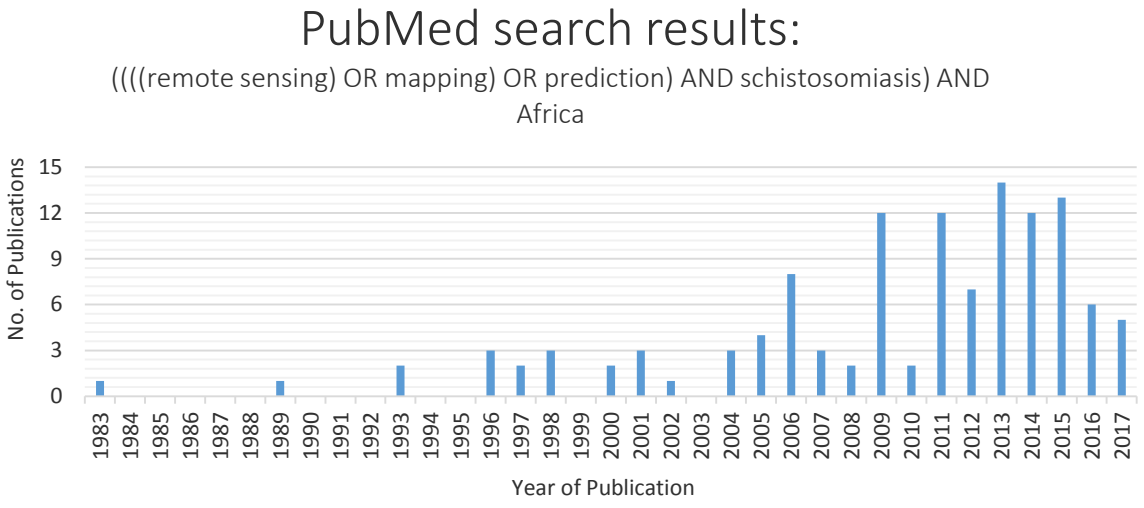


Figure 2: *Pubmed search results of publications with RS/GIS applied to schistosomiasis in Africa adapted from (Simoonga et al., 2009)*

Walz et al. (2015) tabulated the characteristics of studies based on satellite sensor. Prior to analyzing the tabulated studies the review discusses the connection between snail habitat, RS technology, and disease transmission as well as the factors that govern risk in the parasite, snail, and human. It then devotes attention to exploring in depth the historical variables used for modeling schistosomiasis, honing in on differences in primary or secondary proxies of transmission. Walz et al. saw this review as a bridge linking knowledge of different disciplines: epidemiology, disease ecology, and RS risk profiling.

These reviews benefitted from an extensive literature. The tabulated studies from both reviews totaled 41, with about 10 shared between the two reviews. The characteristics of these studies have been tabulated and color-coded by variable type (Table 1). The studies are presented in chronological order beginning in 1994 and ending in 2013. Descriptive analysis of this table demonstrates that up until the millennium, the main variables in use were temperature and vegetation. After 2000 rainfall and elevation were commonly incorporated into studies. Table 1 references that are in bold font are described in detail in the following section.

Table 1: *Studies of RS applied to schistosomiasis, adapted from Simoonga et al., (2009) and Walz et al., (2015).*

RS Environmental Variables	Reference	Review
Temperature, Vegetation	**Malone et. al (1994)**	S
Temperature, Vegetation	Mukaratirwa et al. (1999)	S
Temperature, Vegetation	**Brooker et al. (2000)**	S
Temperature, Vegetation	**Abdel-Rahman et. al (2001)**	S, W
Temperature, Vegetation	Kristensen et. al (2001)	S, W
Temperature, Vegetation, Elevation, Rainfall	**Brooker et al. (2001)**	S, W
Temperature, Vegetation	Malone et. al (2001a, 2001b)	W
Vegetation	Stothard et. al (2002)	S
Temperature, Vegetation, Rainfall	**Brooker et. al (2002b)**	S, W
Temperature	**Moodley et al. (2003)**	S
	Utzinger et al. (2003)	S
Temperature, Vegetation, Elevation	Kabatereine et. al (2004)	W
Temperature, Vegetation	Malone et. al (2004)	W
Temperature, Vegetation, Elevation, Rainfall, Landcover	**Raso et. al (**2005, **2006**, 2007**)**	W
Temperature, Vegetation	**Stensgaard et. al (2005**,2006)	S, W
Temperature, Vegetation, Elevation, Rainfall	Clements et. al (2006a, 2006b, 2009)	S, W
Elevation	Clennon et. al (2007)	W
Landcover	**Clennon et. al (**2004, **2006**, 2007**)**	S, W
Elevation	Beck-Worner et. al (2007)	W
Temperature, Vegetation	**Simoonga et al. (2008)**	W
Temperature, Vegetation, Elevation	Clements et. al (2008a, 2008b)	W
Temperature, Vegetation	Ekpo et. al (2008)	W
Temperature, Vegetation	Brooker et. al (2009)	W
Temperature, Vegetation, Elevation, Rainfall	**Vounatsou et. al (2009)**	W
	Clements et. al (2009)	W
Vegetation	Sturrock et. al (2009)	W
Temperature, Vegetation, Elevation	Koroma et. al (2010)	W
Temperature, Vegetation	Soares et. al (2011)	W
Temperature, Vegetation, Elevation, Landcover	**Schur et. al (2011a**, 2011b, 2013**)**	W
Temperature, Vegetation	Hodges et. al 2012)	W

1.3.3 Applications

The Kafr El-Sheik governorate of Egypt is potentially the first location to apply RS data to schistosomiasis (Malone et al. 1994; Abdel-Rahman et al. 2001). This location is ideal for RS/GIS work because of the high prevalence of schistosomiasis. As the research in this area progressed so has the understanding of the relationships between schistosomiasis and climate.

In 1994, Malone et al. found a positive association between day to night temperature difference (dT) and the risk of schistosomiasis. The variable dT was thought to define moisture regimes, which was thought to be a proxy for a variety of geologic, hydrologic, and vegetative characteristics (Malone et al. 1994). Median dT values were calculated for 28km^2 areas atop 41 survey locations, and in terms of statistical analysis a Spearman rank correlation coefficient was used to assess the relationship between dT and the historical schistosomiasis prevalence survey data collected in 1937, 1983, and 1990 (Malone et al., 1994; Scott, 1937; Cline et al., 1989; Michelson et al. 1993). An inverse positive association was found between higher dT values and S. *haematobium*. This has biological plausibility because the snail species most often associated with S. *haematobium*, B. *truncatus,* is able to withstand drought and high temperatures for a long period of time (Malone et al. 1994). It was noted that persistent control measures such as chemotherapy, molluscicide, and education may have dampened the prevalence in such ways as to obscure its relationship with the environment (Malone et al. 1994). The positive association with dT indicated higher disease risk with a wetter moisture regime (Malone et al. 1994). Annual and seasonal dT composite maps were proposed to reflect stable landform, soils, and climate-irrigation-water table factors influencing the environmental suitability of schistosomiasis (Thompson et al. 1996). Future work has been proposed to use higher resolution satellite

imagery, environmental databases, snail and human data to improve control efforts on the local level (Malone et al. 1994).

Malone et al.'s (1994) work directly potentiated a four-year study into the development of RS/GIS environmental risk assessment methods (Malone et al. 1997). Three years later, this study utilized extraction of information from the aggregation of static ground surveys, derived from 41 survey sites as conducted in 1935, 1983, and 1990, and compared to three thermal-hydrological domains produced from the tasseled cap (Tcap) classification: wet, moist, and dry (Abdel-Rahman et al. 2001). Temporally, thermal patterns were explored individually, seasonally, and annually. Spatially, the study used 25km^2 buffers around the survey points, and looked for relationships with the environment at the local and regional levels. The local/village scale model used higher resolution, Landsat, reflectance data to create a Tcap classification (Malone et al. 1998). The regional field validation studies found 13 villages representative of the wet, moist, dry thermal-moisture domains and focused on these climate differences. Waterways within 1km of these villages had collection stations 50-100m apart for snail surveys.

Results differed by spatial scale. The regional model confirmed that dT was related to risk of schistosomiasis (Malone et al. 1998). Snails were present more often in sites with low (wetter) dT values. The local model was also consistent with the observation of an inverse relationship between dT values and prevalence ($P<0.05$). A linear regression of five years of prevalence data, Tcap data alone and in combination with dT values, developed from 51 villages and randomly tested on 25, produced a model that explained 74% of the variation, suggesting that Tcap & dT can predict risk. The study recommended using defined agricultural zones for Egypt

(Aboukhaled 1975), which reflect variations in temp (usually only a 4-6 ^{o}C degree difference from North and South), soil type, elevation, water table, and other factors to promote the use of agro-ecological and climatic driven suitability on a national scale (Abdel-Rahman et al. 2001).

In 2000, Brooker and Michael reviewed RS/GIS applications to soil transmitted helminths, particularly schistosomes, geohelminths, and lymphatic filarial worms. It found RS/GIS tools capable of both acquiring and analyzing data spatiotemporally, modeling, and providing results useful to directing control efforts. Limitations of this work ranged from the physical lack of high quality data to a lack of theoretical knowledge in areas such as: the impact of using different spatial scales, different spatial modeling techniques, and how environmental variables affect the lifecycle stages of the snail and parasite (Brooker & Michael, 2000)**.**

In 2002, Brooker et al. used RS data to model schistosomiasis in Cameroon. These researchers concluded that the focality of schistosomiasis requires a more refined approach (Brooker et al. 1999), and that RS can be used to "fill the gap in empirical data" (Malone et al. 2001). This study uses the environmental variables: minimum, mean, maximum land surface temperature (LST) and Normalized Difference Vegetation Index (NDVI), elevation, and interpolated rainfall. When using environmental variables, intercorrelation can often pose a problem (Morgenstern 1998) To account for this in a logic regression model, variables likely to have the highest biological significance were included first and the remaining variables were added in a stepwise fashion (Brooker & Michael, 2000). The resulting model's residual deviance and chi squared distributions were compared (Venables and Ripley 1999). Regression models used (max, min, mean) LST, (min, max, mean) NDVI, total annual rainfall, and altitude, listed in order of

addition to the model. Max LST was found to have a positive effect on S. *haematobium,* and thought to be a proxy for the habitat of the snail spp. B. *Senegalese* (Brooker et al., 2002). The final model predicted the probability of prevalence >50%, and was used to estimate the number of school aged children requiring treatment. This was then used to calculate the cost of praziquantel drug regimens. These cost estimates were similar to what had been previously done in Ghana and Tanzania (PCD 1999). These yielded values of US$0.67 and US$0.21 per treatment respectively, and were used as upper and lower bound cost estimates. The predictive risk map generated from the final model estimated that no districts warranted MDA at the >50% threshold, so the model was rerun at a >20% threshold and an estimated 1.8 million children in 9 districts warranted treatment. The total treatment cost ranged from US$ 0.39-1.24 million, based on the bounds of the cost estimates.

In 2003, Moodley et al. created temperature-suitability maps for schistosomiasis in South Africa. The authors reviewed previous studies, which demonstrated that snails are sensitive to high and low water temperatures (De Kock et al. 1986; Pflüger 1980; Pflüger et al. 1984; Shiff et al. 1967; Brown 1994; Pitchford et al. 1969). The analysis of air temperature regimes, (Pitchford 1981) also formed a fundamental basis of this work (Moodley et al., 2003). Three temperature indices were used: monthly mean daily max (Mdx), mean daily min (Mdn), and the resulting range (R). Temperature maps were created and tailored to the individual species of schistosomes. They found that *S. mansoni* covers a wider area than *S. haematobium*, because *Bi. pfeifferi's* optimal temp range is 22-27 °C, while *Bu. africanus* is 25 °C (Appleton 1977; De Kock et al. 1986). This study's unique use of temperature regimes it is not directly comparable to other RS/GIS methods previously implemented in Egypt (Malone et al. 1994) and Brazil (Bavia et al. 1999). Moodley

was not able to affirm Malone et al.'s (1994) positive association with dT, because this study's final models did not include elevated and fluctuating temperatures. Instead, it concluded that mapping on temperature alone was not sufficient to explain prevalence.

In contrast to Moodley et al. (2003)'s sub-continental focus, Utzinger et al. (2003) honed in on the transmission dynamics occurring within a single village. This micro scale study explored the focal distribution of *S. mansoni* within households, unlike most studies, which would explore prevalence and infection intensity at the district level. What they found were differences in prevalence and intensity of infection for students living inside and outside of the village, as well as random spatial patterns amongst households. Utzinger et al. (2003) found a random/homogenous distribution of infection intensity. This is in contrast to their other studies of villages in the region, which showed persistent heterogeneity. This demonstrates the marked spatiotemporal nature of the disease (Utzinger et al. 2003; Woolhouse 1998). Such results emphasize the need for uniform, community-wide control efforts, and the importance of noting the GPS of household locations as used in later studies (Brooker & Michael, 2000; Brooker 2002).

In 2005, Stensgaard et al. modeled the distribution of S. *mansoni* and its host snails, Bi. *pfeifferi* and Bi. *sudanica*, in Uganda. The specific environmental variables used were: rescaled NDVI (0-200), Tday, Tnight, dT, long-term normal environmental data, monthly rainfall in mm, monthly potential, and actual evapotranspiration for 1931-1960 (Malone et al. 2001). Temporally, seasonal and annual models were used and iteratively compared to parasite and snail distributions in order to find the specific ranges for each variable that best fit with the health

data. The results suggested that the strongest ecological determinants were a combination of remotely-sensed LST and standard environmental data (ground station) rainfall. Low nighttime temperature was found to be a significant factor inhibiting transmission in parts of the Southwestern highlands of Uganda. It also found that species-specific ecologic preferences existed inland. The analysis utilized NDVI as a surrogate for soil moisture (Malone et al. 2001), and as Uganda exhibits two principal rainy seasons, both wet and dry seasons were analyzed. Averages were calculated for wet, dry, and annual months of precipitation (PRE) and rain/potential evaporation (PPE), the latter representing availability of water in a given time period. For the annual composite model, the investigators defined the Tday, Tnight, and NDVI value ranges consistent with S. *mansoni* and snail endemic areas and analyzed scatter plots, and then used the ranges to create predictive maps. These were then repeated iteratively using incremental changes in each variable to find the best fit, as well as the narrowest ranges of each that associated with >5% prevalence for each variable and then saw where they overlaid each other. This operation was repeated for the dry models (Malone et al. 2001), as well as the snail data aggregated both annually and seasonally.

Statistical evaluations included the Spearman rank correlation and logistic regression. Independent variables were tested for linearity and correlation. Multi-collinearity was seen between LST and NDVI, and PPE and PPE. To overcome this some models were developed separately, and developed using backwards elimination. This resulted in no significant correlation being found with NDVI composite annual and wet season, a weak negative association with dry, and none with wet season PPE or dT. The two final models were 1) elevation, Tdaywet, PREwet 2) Tnightwet, PREwet. Both models had R^2 values of around 0.44

and 0.42 respectively. Only small differences were seen between annual and seasonal maps, because Uganda does not offer much climactic variation due to its equatorial location. This climatic variation is even less than the Ethiopian study it based many of its analyses upon (Malone et al. 2001). However, some of these differences could be attributed to the use of different RS satellites and time periods. Despite these extreme location differences, an RS based approach can be used to predict risk of schistosomiasis at the regional scale (Stensgaard et al., 2005). It is wise to apply caution in the interpretation of results even if other countries have similar ecological zones because this approach might not be transferrable due to many factors involved in schistosomiasis transmission (Brooker et al. 2001).

In 2006, Raso et al. studied co-infection of schistosomiasis and hookworm using demographic, socioeconomic, and environmental data via a Bayesian geostatistical models. The non-spatial model results showed sex, age group, socioeconomic status, land cover, elevation, slope, rainfall, LST, NDVI, soil type, and distance to health care facility were significant covariates, and distance to permanent rivers showed a significant association with infection status. The binomial spatial model results found that age, sex, socioeconomic status, and elevation were statistically significant. The multinomial spatial model results were the same as the binomial except for gender, which was not found to be significant. Overall the rate of spatial correlation <5% took place at 4.1km, which had decreased 7.5km found in a previous study (Raso et al. 2005). Smooth risk maps were created using Bayesian kriging, in which 71.4% of the data was used for establishing the model, and 28.6% for prediction. This resulted in a sensitivity of 93.8% and a specificity of 87.5% (Raso et al. 2006). This work differs from that of previous investigators in that the spatial statistics were analyzed using Bayesian modeling instead of GIS.

Also in 2006, Clennon et al. studied the spatiotemporal associations of S. *haematobium* in 10 villages in the Msambweni Division, Kwale District, and Coast Province, Kenya. This was a scaled up study from previous work describing the spatial clustering of S. *haematobium* infection around an infested pond in a single rural village in Msambweni, Kenya, in which households and water sources were mapped using a high resolution imagery (1m^2 panchromatic, 4 m^2 multi spectral) (Clennon et al. 2004). The 2006 work integrated historical and present spatial patterns of human household and snail habitat data. It found snails dispersed among ponds, a rice field, a river, and a stream, with the river being the least favorable habitat. Most snail shedding was recorded in areas near dams, with only 2% of the snail's population shedding at any one time. Differences in prevalence between age groups was tested using the homogeneity chi-square test, correlation, and logistic regression (forward conditional). Spatial statistics were assessed on the global, local, focal, and directional scale using Global Ripley's K-function, point pattern analysis, and cluster seer. Spatial patterns were able to be classified as random, clustered, or uniformly dispersed with significance determined by Monte Carlo Simulations, an alternative to spatial analysis using GIS or Bayesian modeling.

The results of this work differed according to the regression model employed. The logistic regression adequately fitted the data with a chi sq. of 9.7, $p > 2.29$, but explained only 12% of variance, with age as the only significant covariate among sex, distance to nearest alternative water source, and school attendance. The linear regression explained a greater percent of the variance, <10%, and also found age to be the only significant covariate. The spatial model found significant clustering at different distances based on age. At one pond, high infection clustering

persisted up to 550m for ages 6-9, and >1,000m for ages 10+, while in another area it was >1,500m for ages 10+. Infection was found to be clustered as a function of spatial distribution of water bodies, the effects of which were significant up to 1,500m, with strong anisotropy (directional clustering) in some areas. Clustering was also found to have shifted over time (Clennon et al. 2004; Clennon et al. 2006).

Clennon et al. (2006) posited that snails are not the rate limiting factor in the "habitat" of schistosomiasis because snails are found to be more widely distributed than the disease. It is instead human attributes that spread schistosomiasis based on where they go and what they do, with some impact from climatic and environmental changes. El Nino likely affected the study area by causing soil erosion thereby decreasing some water habitats, and creating small temporary ponds thereby offering an increase in snail habitats. Meanwhile, conversion of sugar cane plots to rice fields increased surface water bodies. Overall age-related water contact behaviors was offered as the most probable explanation for differences among age groups, and the local focus of the study allowed for overall water usage to be explored. A risk map was not created because the model was too unstable, and the study concluded by stating that a range of spatiotemporal scales need to be considered to understand transmission.

Around the same time Simoonga et al. (2009) published a review of RS/GIS applications addressing schistosomiasis, and prior to that Simoonga et al. (2008) published a small-scale study in Zambia. The risk factors analyzed included geographical location, elevation, NDVI, max LST, age, sex, and intermediate host snail abundance. Three logistic models were employed. The first was an ordinary logistic model, the second included random effects, and the

third assumed spatially correlated random effects. The results of which found elevation, NDVI, and snail population to be significant, with the spatial random effects outperforming the other models. Small-scale heterogeneity was observed and thought to be based on the dispersion of human settlements.

In 2009, Vounatsou et al. developed a novel Bayesian geostatistical model ideal for count data and non-stationary spatial modeling and validated using many of the datasets to be gathered in Côte d'Ivoire (Beck-Wörner et al. 2007; Raso et al. 2005; Raso et al. 2006; Raso et al. 2007). It was only the second study to base predictions off of schistosomiasis intensity levels (Clements et al. 2006), and argued that man-made ecological effects often resulted in non-stationarity (Vounatsou et al. 2009). The author's Bayesian geostatistical zero-inflated (ZI) regression model outperformed all others tested, which included negative binomial, ZI Poisson (ZIP), and ZI negative binomial (ZINB). This work relied on ecological zones based off of water catchment areas rather than administrative boundaries, which is better to use over large areas (Beck-Wörner et al. 2007). Vounatsou et al. (2009) recommend continued study of non-spatiality in addition to anisotropy and polyparasitism infection intensity.

In 2011, Shur et al. compiled open access survey data and created smooth empirical prevalence maps in order to determine country specific prevalence estimates. This interpolation of existing data created prevalence estimates on a scale not before analyzed. This work predicted risk for West Africa, and later for East Africa (Schur et al. 2013). The environmental covariates used were LST, NDVI, and rainfall, all of which were found to be statistically significant covariates. The S. *haematobium* survey data available for Ghana indicated that as of 2011 there had been 8

unique survey locations, most of which took place in the 1980s and one in the 2000s. The results specific for Ghana predicted a combined prevalence of S. *haematobium* and S. *mansoni* of 53.7%, which matched the 2012 WHO prediction of 50% (Figure 1).

Also in 2011, Soares Magalhães et al. created the most up to date predictive map for Ghana on a national scale. The results of this study provided baseline information on national prevalence prior to the implementation of mass drug administration (MDA) by the schistosomiasis control initiative (SCI), and before a national cross-sectional school-based parasitological survey. Data was collected on soil transmitted helminths and schistosomiasis (*S. haematobium* and *S. mansoni*), from 77 schools distributed across the country. It used both environmental and human variables, which were LST, NDVI, distance to nearest perennial inland water body (PIWB), sex, and age respectively. They did not find statistical evidence for including rainfall in any of their models. The health data included intensity and prevalence. A fixed-effects multinomial regression model was used, and found distance to PIWB, LST, age, and sex to be significant. Bayesian geostatistical modeling was used to create a predictive risk map for the country. S. *haematobium* mono-infection was estimated to have an average prevalence around 10% with some lower pockets in the North and higher pockets in the South, especially near Kumasi, Accra, and Lake Volta. Soares Magalhães et al. (2011) state that environmental variables may be the drivers of human variables, even though human variables have been shown to explain about a quarter of the variance (Pullan et al. 2010).

1.4 Findings & Gaps in Literature

In the two recent reviews of RS applications to schistosomiasis modeling, all 32 studies utilized health data that was derived from either a school, village, community, or town survey (Simoonga

et al. 2009; Walz et al. 2015a). Surveys provide accurate estimates of disease in a specific time and place but come at a cost. The resources that surveys require render performing them at large spatiotemporal scales impractical. The studies that do create predictions at national or continental scale do so using interpolated survey results (Walz et al. 2009a). Although, interpolation of micro-scale survey data to estimate macro-scale trends is attractive for the purpose of setting national control strategies, concerns persist as to whether this is the most sustainable or cost effective route for resource poor countries.

1.4.1 Mismatch between the spatiotemporal capabilities of dataset pairings

Just as it is impractical to interpolate survey data to national scales it is inefficient to restrict publically available RS data to school, village, community, or town scales. This underutilizes the full spatiotemporal capabilities of RS technology, as most open source satellite products can be utilized as global continuous surfaces with historical (in some cases 40-year) timespans. In addition RS data is often available free to the public and takes at most a few days to order. Yet, only 25 out of the 32 studies used RS data that had been extracted from the areas immediately surrounding the health survey sites (at most 5km) and at the specific times of their collection (Malone et al. 2001; Malone et al. 2004; Walz et al. 2015b). At these micro-scales, the finest spatial resolution available for publically available RS data is 30m (~100ft) using the Landsat satellite series going back to 1982 to the launch of Landsat 4, or 20m (~66ft) using the Sentinel satellite launched in 2015. These spatial scales are often too large to capture the details of rivers or streams and around survey sites, and environmental variables are unlikely to vary over school, village, community, or town scales.

Commercial satellite imagery is well suited to the detailed spatiotemporal resolution obtained through health surveys. The spatial resolution commonly available using high spatial resolution imagery all are less than 1m for satellites operated by DigitalGlobe, GeoEye, and ImageSat International, but are prohibitively expensive, \$10-25/km^2 and date back to only 1999 (AAAS 2016). Research grants in the form of free commercial satellite imagery are available, but are limited in the size of imagery and number of "scenes" requested. Due to these restrictions and the inherent computational price associated with processing commercial quality satellite imagery, high resolution RS data makes for a complementary source of environmental data for survey data.

Resource intensive outcome data is not well suited for control of NTD's, "diseases of poverty" (Gryseels et al. 2006). "In our view, strong district health systems that are nationally-owned are the main prerequisite for successful, cost-effective and sustainable control of neglected tropical diseases" (Utzinger et al. 2009). Ongoing national surveillance data in the form of reported counts from health centers is fundamentally better suited to utilizing the full spatiotemporal capabilities of publically available RS data.

National surveillance systems clearly require support, and the reward for their success is sustainable control, the precursor to eradication (Rollinson et al. 2013). There is an obvious reluctance to use national surveillance data in conjunction with RS for schistosomiasis. This leaves a dearth of literature in its wake. This becomes a missed opportunity to make use of the many administrative spatial scales and the continuous weekly and monthly temporal scales, which can be aggregated for seasonal and annual analyses. These aspects of national surveillance

data make it an ideal addition to publically available RS data, where the intention in combining these datasets is to produce cost-effective, sustainable solutions for broad predictive mapping.

There is the opportunity to combine the strengths of each health outcome and RS data type. National surveillance data records individual information such as gender and age, which can be used to correct for reporting bias among age groups through comparison with aggregated health surveys, which are often biased towards school age children. Using both health outcomes a clearer picture of the true population affected by schistosomiasis can emerge. This would also allow for data validation, where dynamic spatiotemporal predictive models based off surveillance data can be verified at finite points in space and time leading to more efficient predictions at larger scales.

This naturally progresses into consideration about how best to use each pairing. According to a review of remote sensing in arthropod (e.g. mosquitoes, ticks, flies) vector-borne diseases, it is recommended that a combination of high and course resolution satellite data, abiotic and biotic factors, and both environmental and human factors be utilized (Kalluri et al. 2007). It would appear that outcomes modeled by publically available RS data could be used to provide broad predictive maps as a "front line approach", which could highlight areas requiring more detailed and accurate data collection in the form of more resource intensive but more accuracy health surveys modeled using commercial RS datasets and human variables. This comparison in approaches would result in a better understanding of scale, uncertainty, environmental predictors, and the dynamics between disease exposure, transmission, and reporting; topics echoed

throughout the future direction sections of multiple literature reviews (Gryseels et al. 2006; Utzinger et al. 2009; Simoonga et al. 2009; Walz et al. 2015).

1.4.2 Ghana health surveillance

The reasons for not using surveillance data has not been previously addressed in the reviews of RS applications to schistosomiasis (Simoonga et al. 2009; Walz et al. 2015a). However, an observational study, "Evaluation of the integrated disease surveillance and response system for infectious diseases control in northern Ghana" sheds light on the utility of surveillance data (Adokiya et al. 2015). This study explains that the national surveillance system was created in response to major disease outbreaks in the 1990s. The Ghana Health System (GHS) began recording reported schistosomiasis case counts in 2008, but in 2012 there was a push towards improving disease surveillance in sub-Saharan Africa, and this strategy was called the Integrated Disease Surveillance and Response (IDSR). In 2009, GHS in conjunction with the University of Oslo in Norway developed a software called the district health information management system (DHIMS), which took the form of mailed paper submission of weekly and monthly reportable disease counts. On April 1st, 2012 a web-based version was implemented for the majority of health facilities and was called DHIMS2 (GHS 2012; Adokiya et al. 2015). In addition to conversion to digital reporting, in 2012 the previously 171 districts were redistributed into 216 districts. How GHS reconciled reported counts from the previous to new administrative boundaries was not mentioned. The study concluded that although the surveillance system beneficially contributes to disease control it currently has many challenges.

Most of the challenges for the surveillance system were related to a lack of resources, which affects every stage of data collection. For those who visit a health clinic it is possible that the health staff will not have sufficient training to properly identify schistosomiasis. GHS stipulates

that 20 notifiable diseases are to be reported immediately with 23 other diseases that are recorded daily and reported in either a weekly or monthly basis (WHO-AFRO; CDC 2010). Schistosomiasis is not a notifiable disease but is recorded nonetheless. Diagnoses can be difficult for health care workers to verify when there is no sufficient medical equipment or technicians available. When verification is possible, the turn/around time can vary and so recorded counts can be forgotten or mistakenly attributed to the wrong week or even month. Assuming that a patients diagnosis is accurately diagnosed, and verified there are biases for reporting certain diseases over others due to the amount or ease with which some reimbursements occur. Even though DHIMS2 is online, health center reports can be either mailed, texted, or emailed to respective districts, which can lead to confusions especially when accounting for retroactive corrections. Recording and aggregation of reported counts up the administrative hierarchy from health clinic, community, sub-district, district, region, and finally the national level can lead to data input errors especially since the data is entered by hand at multiple stages. The authors of the observational study that focused generally on infectious diseases discussed here found discrepancies between weekly and monthly counts (Adokiya et al. 2015). Albeit challenging, these issues are understandable with the augmentation and scaling up of a national surveillance system.

A slightly different challenge in interpreting surveillance data is found in exploring the differences between those who actually seek care and the general population. A study on the factors influencing women to seek care at health centers for labor and delivery across three climate zones sheds light on health seeking behavior (Enuameh et al. 2016). Aside from the previously known predictors associated with health facility delivery, place of residence, socio-

economic status, and possession of valid health insurance, there was an additional factor identified, healthcare provider's influence. The study also found significant differences across geographic locations, namely the northern, central, and southern belts of Ghana (Enuameh et al. 2016). Another study on health seeking behavior looked at differences based on enrollment in health insurance, and found a significant difference among those in the lowest wealth quintiles as compared to the highest, "suggesting that [insurance] has not succeeded in bridging inequalities in health services utilization between the poor and rich" (Kuuire et al. 2015). It is all too likely that health care seeking behavior will likely vary greatly and thus suppress disease prevalence estimates.

Other challenges of national surveillance systems are easier to fix and fall under the category of design flaws. As previously mentioned, the current reimbursement structure is leading to bias in the reporting of some diseases over others. If systematic, this bias can be adjusted for. An additional design flaw in the DHIMS2 system is that it is unable to display zeros that have been input into the system, thereby creating uncertainty as to the nature of missing values (Adokiya et al. 2015). It stands that if this structural error could be fixed, then the result would be an entire national surveillance system going from presence-only data to presence-absence data. Despite the fact that "challenges regarding accuracy, reliability and soundness …may lead to low utilization of health system data for planning and decision-making" (Adokiya et al. 2015), these structural shortcomings restrict the ability to pursue and measure the impact of sustainable control measures implemented at scale.

Improving public health surveillance systems in low income countries can come in many forms. One study, proposes focusing on what the WHO considers health system building blocks: "service delivery, financing, governance, the health workforce, information systems, and supply management systems" (Nsubuga et al. 2010). The topic of the health workforce was explored in detail. The study suggests use of the Field Epidemiology Training Program (FETP) and the Field Epidemiology and Laboratory Training Program (FELTP). FETP/FELTP programs are effectively at the level of master's degrees and these graduates are capable of leading sub-/national public health surveillance and response systems. The study suggests that there be 3-5 FETP/FELTP graduates per million inhabitants. Ghana's current population of ~26 million would necessitate between 78-130 FETP/FELTP personnel (Nsubuga et al. 2010). If this many highly trained personnel are required, then an even greater number of health workers with basic training are needed to run national public health surveillance systems.

An organization focused on reaching one million community health workers (1mCHW), has been moving forward in Ghana with an ultimate goal of having 1 CHW per 500 people in 2023 (Kim 2014). This campaign, worth US$2.5 billion was the first standardized approach to recruiting CHWs in Ghana. A 2016 study on the CHWs in Ghana has found that meeting this goal will not be without challenges. CHWs vary in the roles they serve, services they provide, the funding used to support them, their connection with the larger health system, and the names by which they are called (Baatiema et al. 2016). The, "lack of clarity on the operational mandates of CHWs in Ghana has often undermined the effectiveness and efficiency of their roles in healthcare dates of CHWs in Ghana has often undermined the effectiveness and efficiency of their roles in healthcare delivery at the community level" (Baatiema et al. 2016). Proper

compensation to reduce attrition, training, recognition, structure, and oversight among other measures were recommended (Baatiema et al. 2016). Interest in improving public health surveillance systems in low and middle income countries (LMICs) is partly due to concerns about LMIC susceptibility and potential role in global pandemics (Nsubuga et al. 2010; Kim 2014).

1.5 Thesis Structure

This thesis is designed with sustainability in mind. For this initial portion of the work it avoided "human" variables such as socio-economics and demographics, and sought instead to see what could be gained from looking at purely environmental RS-based data. Of particular interest was investigating the association of environmental variables with reported schistosomiasis. It is understood that even with strong correlation between selected environmental variables with the distribution of schistosomiasis does not confer a clear picture of etiology. An eventual aspirational goal for this thesis is to contribute to yielding a "front line" predictive mapping capability of schistosomiasis at large spatiotemporal scales. This work was focused on identifying proxies for climate when deciding what environmental predictors to utilize in the modeling of national surveillance data.

This study aimed to assess associations between monthly rates of schistosomiasis cases obtained from Ghana's national surveillance and reporting system, aggregated by administrative district, and three RS-based environmental predictors: vegetation, temperature, and precipitation, arranged as time series. The analysis was stratified by three major and nine minor climate zones, defined according to a new classification method using multiple satellite data streams (Liss et al. 2014). It was hypothesized that there may be spatial and temporal patterns in schistosomiasis

incidence that can be partially explained by environmental parameters, and that these patterns may vary across climate zones (Brooker et al. 2001; Walz et al. 2015a).

This work has the potential to be validated and evaluated against the more typical interpolation approach used by Soares-Magalhães et al. (2011), where a predictive risk map was created from an interpolation of data from 77 individual surveys distributed across Ghana. Overall, Ghana having documented high schistosomiasis prevalence, ongoing national reporting of the disease, and historical prevalence prediction maps makes it an ideal location to test a rapid mapping approach.

2. Data and Methods

2.1 Health outcome: schistosomiasis cases

GHS provided monthly counts of schistosomiasis cases aggregated to the level of administrative district (n = 216 districts) as reported into the District Health Information Management System (DHIMS). Data were acquired in January 2016 for an 8-year period (96 months) from January 2008 through December 2015. District-level population estimates for 2010 were obtained from the population census and projected for each study year using intercensal population growth rates estimated for each of Ghana's 10 administrative regions (GSS 2013). Disease counts were divided by the district population, expressed as rates per million people, and $\log_e$ transformed to achieve a distribution close to normal, as tested with the coefficients of skewness and kurtosis. The original dataset consisted of 20,736 monthly observations (216 districts x 96 months), yet

due to incomplete, missing, and extreme values, monthly disease records were reduced to achieve reliable modeling. Data processing steps are described and summarized (Figure 3).

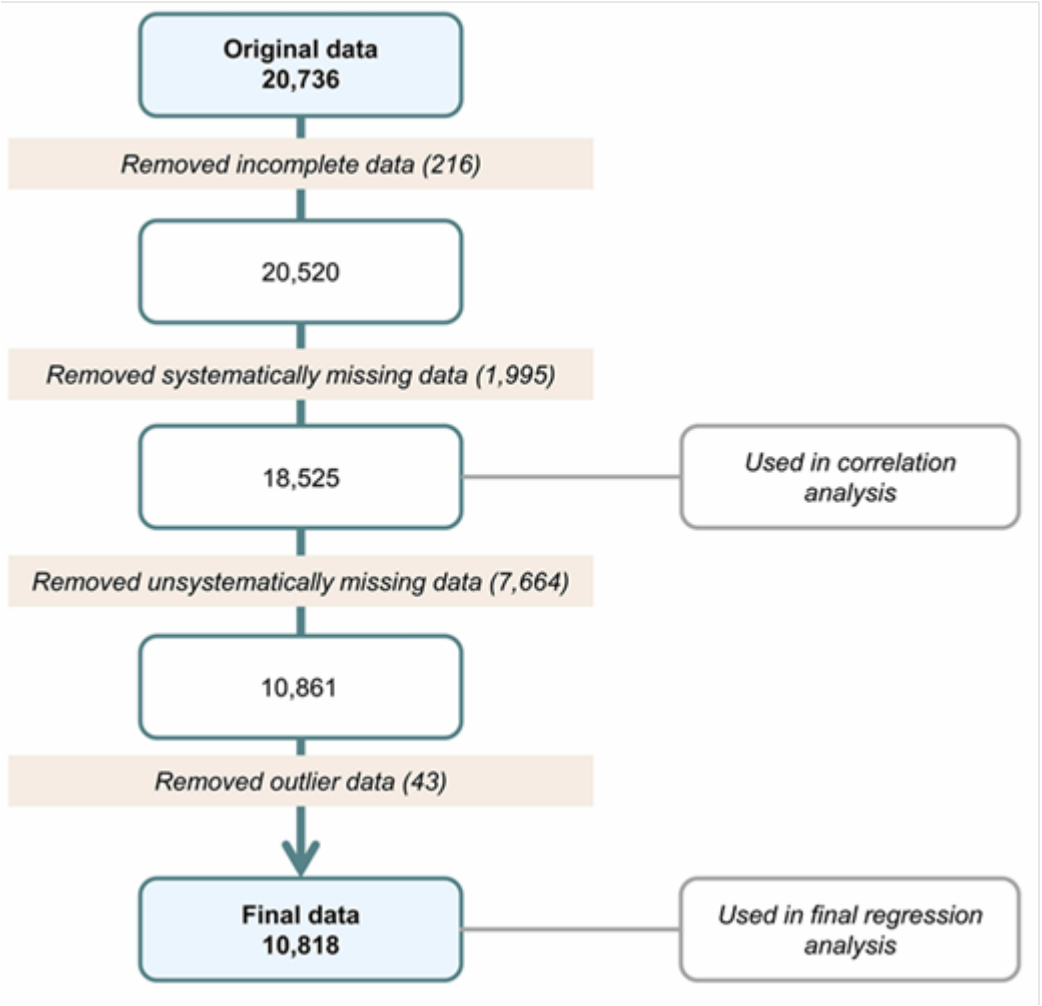


Figure 3: *Data processing steps and subsequent sample size reduction*

Incomplete data: Examination of the monthly time series of total disease counts (Figure 4) revealed that the last month had substantially lower counts due to probable delay in reporting of data acquired in January of 2016. Thus, data from December 2015 (216 observations) were excluded from the analysis.

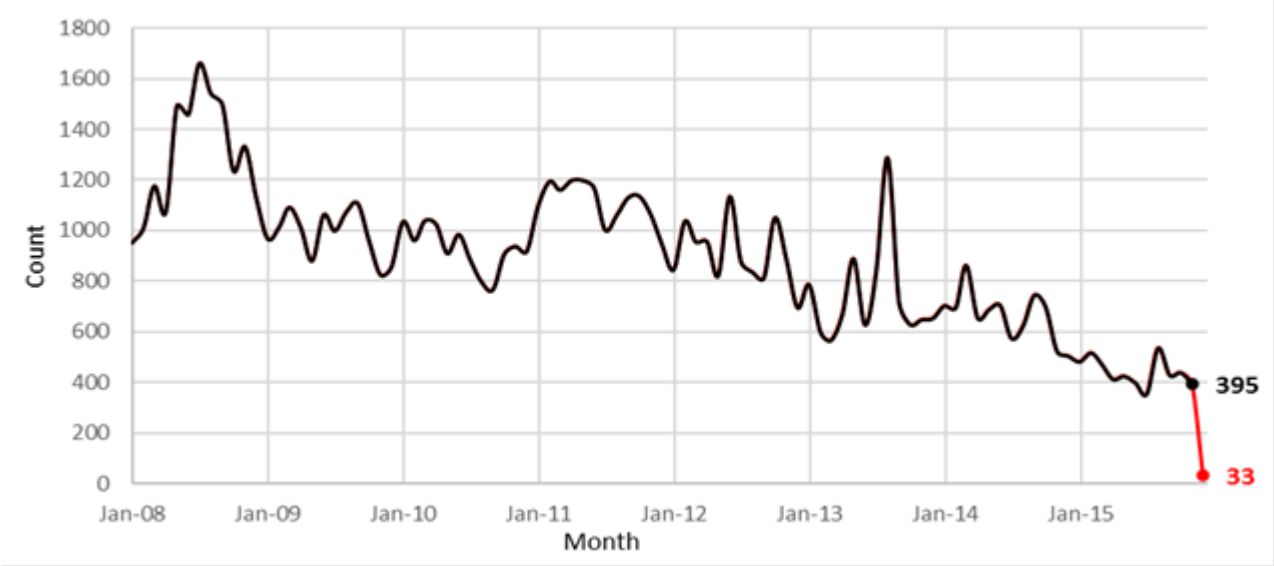


Figure 4: *Total monthly schistosomiasis case counts for all districts*

Missing data: A high percentage of observations in DHIMS were blank values, and it was unclear as to whether these observations represented the true absence of events or the lack of reporting. If the former, the blanks should be treated as zeros; if the latter, the blanks should be treated as missing values. Since it was not possible to determine the reasons behind the coding scheme, all blanks were treated as missing values. Four districts that had no reported cases and an additional seventeen districts that had >95% missing values (1,995 systematically missing observations) were removed from the dataset, and analyzed for bias using skewness and kurtosis statistics (Table 2). We defined "systematically missing" as belonging to districts that never or almost never reported cases. An additional 7,664 cells represented unsystematically missing observations. Missing values were visualized based on their district, month, and whether they occurred before/after the first/last count or in between reported counts (Figure 5). Vertical lines denote districts, ordered by percentage of missing values along the x-axis. Horizontal lines denote months, ordered from January 2008 (top) through December, 2015 (bottom). The color of

each observations represents its value and when it occurred. White was used for missing values prior to the first count or after the last count, to reflect potential differences in the start/end of district reporting. Grey was used to reflect all missing values between the first and last count, and black was used for all observations with recorded counts. The bottom row of the plot (Figure 5), December, 2015, shows more missing values than any other month as described previously (Figure 4). The pervasiveness of missing values means that how they are treated greatly influences the model. The decision to treat them as missing values, will likely lead to a conservative estimate, since models will be restricted to districts and months with at least one reported case of schistosomiasis.

Outlier data: Exceptionally high monthly counts were explored as potential data input errors. To help differentiate naturally occurring vs. unlikely high count values, we reviewed the skewness and kurtosis of the data distributions for raw counts (Table 3 & Figure 6) and for $\log_e$ transformed rates. Among the districts adjusted for possible outliers, the skew and kurtosis dropped from 5.59 to 2.91 and 36.29 to 6.78 respectively. The districts affected most by cleaning based on missing values were primarily in the northeast, while districts affected by cleaning based on outliers were primarily in the north and south.

Table 2: *Descriptive statistics of the 21 districts removed for missing values*

#	ID	Z3	Z9	%NA	Max.	Med.	Mean	Stdev.	Q1	Q3	Skew	Kurt.
1	2	1	2	100	0	NA	NA	NA	NA	NA	NA	NA
2	184	1	9	100	0	NA	NA	NA	NA	NA	NA	NA
3	20	3	5	100	0	NA	NA	NA	NA	NA	NA	NA
4	124	3	5	100	0	NA	NA	NA	NA	NA	NA	NA
5	192	2	9	99	1	1	1.000	NA	NA	NA	NA	NA
6	143	2	7	98	1	1	1.000	0.000	NA	NA	NA	NA
7	170	3	5	98	3	2.5	2.500	0.707	NA	NA	NA	NA
8	18	3	5	98	400	297	297.000	145.664	NA	NA	NA	NA
9	142	2	7	97	1	1	1.000	0.000	1.500	1.500	NA	NA
10	112	2	8	97	51	2	18.333	28.290	3.000	76.500	1.732	NA
11	190	1	3	96	29	7	11.000	13.266	1.500	37.500	1.096	-0.050
12	105	1	3	96	9	3.5	4.250	3.594	1.875	12.000	0.889	-0.582
13	109	1	3	96	3	1.5	1.750	0.957	1.500	4.125	0.855	-1.289
14	111	1	8	96	4	2	2.250	1.500	1.500	5.625	0.370	-3.901
15	191	1	9	96	7	1.5	2.750	2.872	1.500	8.625	1.846	3.412
16	182	2	9	96	2	1	1.250	0.500	1.500	2.625	2.000	4.000
17	203	3	7	96	5	2.5	2.750	2.062	1.500	7.125	0.200	-4.858
18	14	1	1	95	41	1	9.800	17.527	1.500	34.500	2.184	4.797
19	100	2	4	95	3	2	2.000	1.000	1.500	4.500	0.000	-3.000
20	155	3	8	95	2	1	1.400	0.548	1.500	3.000	0.609	-3.333
21	185	1	3	95	24	9.5	11.500	8.961	6.000	31.500	0.435	-1.475
Avg	NA			97	30	25	26.679	16.618	1.708	17.292	1.123	-0.467

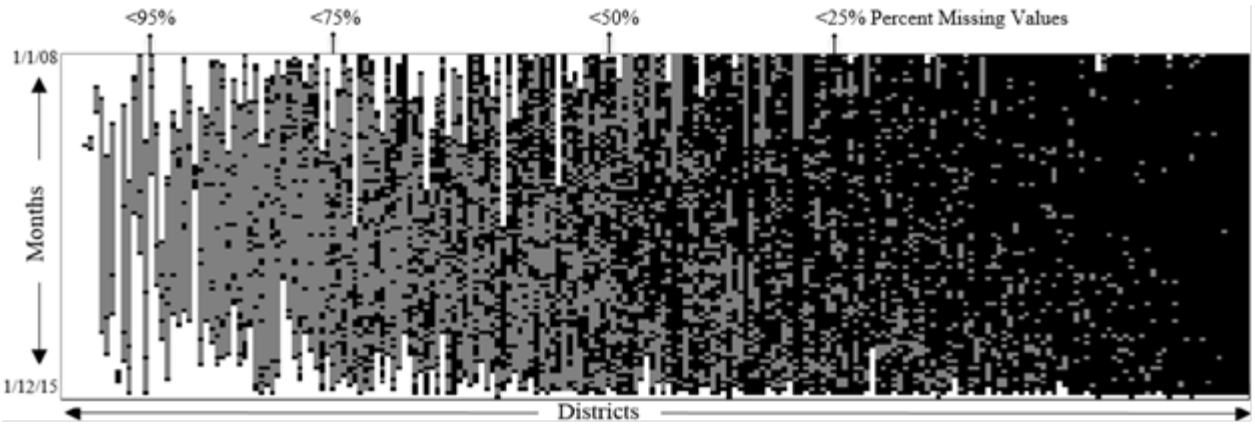


Figure 5: *Visualization of missing values in disease counts, with reported counts colored black, missing values within the timeframe of the first and last reported count colored gray, and missing values leading up to and after the last reported count colored white. The y-axis represents consecutive months; the x-axis represents individual districts, organized by the percentage of missing values per district*

Table 3: *Descriptive statistics of the disease counts prior to any replacements, after replacing the first highest value with a blank, and after replacing the second highest value with a blank. The descriptive statistics include the maximum, median, mean, standard deviation, skewness, and kurtosis. Cells with gray shading indicate a skewness >3 and/or a kurtosis >10, which were used to flag districts requiring further attention*

	Original Data						After 1 Replacement						After 2 Replacements					
#	Max	Med	Mean	Stdv	Skew	Kurt	Max	Med	Mean	Stdv	Skew	Kurt	Max	Med	Mean	Stdv	Skew	Kurt
1	158	2.5	18.40	49.10	3.15	9.95	7	2	2.89	2.26	0.84	-0.71						
2	23	1	2.50	4.65	4.48	20.54	4	1	1.52	0.81	1.76	3.22						
3	63	4	6.07	7.77	4.99	33.00	30	4	5.43	4.88	2.34	7.72						
4	46	3	3.89	5.62	6.14	44.72	16	3	3.31	2.60	2.17	7.28						
5	94	2	4.09	13.63	6.66	44.88	7	2	2.09	1.47	1.53	1.91						
6	72	8	11.32	11.41	3.20	13.91	38	8	10.23	8.03	1.64	2.63						
7	11	1	1.88	1.97	4.28	20.14	3	1	1.52	0.65	0.90	-0.15						
8	50	1	4.82	10.50	4.15	18.25	11	1	2.67	2.99	2.00	3.24						
9	115	1	5.06	18.93	5.92	35.30	8	1	1.91	1.82	2.55	6.10						
10	40	3	4.58	5.65	4.38	24.73	17	3	4.02	3.44	2.02	4.43						
11	63	1	4.21	12.58	4.82	23.47	6	1	1.65	1.27	2.35	5.80						
12	67	2	3.03	7.90	8.02	65.80	6	2	2.09	1.19	1.24	1.26						
13	50	2.5	3.69	7.04	6.32	42.13	9	2	2.70	1.73	1.38	2.52						
14	68	4	4.49	7.68	7.72	64.36	10	4	3.64	2.14	0.66	0.06						
15	56	1	4.28	12.92	4.23	17.95	2	1	1.24	0.44	1.37	-0.15						
16	18	2	2.58	3.20	3.85	17.38	8	1.5	2.09	1.63	2.02	4.51						
17	79	4	5.58	8.84	7.44	61.96	12	4	4.65	3.02	0.78	-0.36						
18	865	4	15.43	92.80	9.24	85.59	25	4	5.44	4.47	1.77	3.86						
19	33	4	4.51	4.36	3.70	21.16	15	3.5	4.17	3.06	1.24	1.62						
20	92	3	4.93	11.22	7.30	57.07	13	3	3.61	3.07	1.60	2.01						
21	90	4	5.87	10.38	7.39	60.08	17	4	4.73	3.32	1.32	1.85						
22	36	2	3.36	4.93	5.61	36.60	11	2	2.76	2.08	1.61	3.43						
23	29	3	3.78	4.44	3.91	19.47	15	3	3.31	2.81	2.13	5.44						
24	25	2	3.04	3.42	4.29	24.19	15	2	2.74	2.27	2.53	10.51						
25	94	2	3.63	11.61	7.74	61.07	11	2	2.19	1.74	2.79	10.42						
26	209	6	33.60	224.65	9.24	85.53	64	6	9.41	11.57	3.02	10.85						
27	26	2	2.54	3.58	5.35	32.96	12	2	2.14	1.80	3.43	15.78						
28	47	2	4.29	8.44	4.44	21.25	22	2	3.00	3.84	4.03	19.38						
29	233	2	7.25	32.82	6.80	47.32	45	1.5	2.74	6.20	6.73	46.71						
30	150	2	5.14	17.53	8.17	68.30	20	2	3.10	2.79	3.67	18.94						
31	52	4	5.95	7.31	4.68	26.93	28	4	5.20	4.29	2.79	12.36	13	4	4.82	3.12	0.90	0.09
32	127	1	7.04	24.18	5.06	25.99	17	1	2.42	3.20	4.07	18.51	5	1	1.84	1.21	1.39	0.87
33	62	4	5.83	9.34	4.73	25.34	40	3.5	4.86	5.70	4.35	25.41	14	3	4.25	3.25	1.04	0.67
34	51	5	7.11	7.90	4.00	19.39	50	5	6.62	6.40	4.09	24.09	25	5	6.13	4.42	1.80	4.42
35	69	2	4.84	11.46	5.03	25.61	63	2	3.83	8.13	6.55	47.05	19	2	2.87	2.97	3.59	15.81
36	350	2	11.20	52.56	6.52	42.94	33	2	3.33	5.66	4.33	20.10	21	2	2.62	3.30	4.53	24.26
37	138	1	8.00	26.79	4.77	23.54	35	1	3.00	6.67	4.78	23.58	7	1	1.72	1.40	2.63	7.97
38	22	2	2.27	3.39	5.28	30.30	8	1.5	1.78	1.25	3.43	15.60	4	1	1.62	0.75	1.18	1.34
39	127	3	7.26	19.95	5.38	28.95	99	3	5.27	12.57	7.27	54.95	12	3	3.68	2.56	1.17	1.29
40	164	6	11.96	22.85	5.25	30.98	114	6	9.93	14.58	5.17	35.21	35	6	8.53	8.07	1.26	0.99
41	67	2	4.34	8.51	6.42	45.78	27	2	3.40	3.61	4.47	27.40	10	2	3.05	2.14	1.13	0.72
42	13	1	1.84	2.28	4.35	20.45	6	1	1.47	0.97	3.71	16.71	2	1	1.31	0.47	0.87	-1.35
43	220	1.5	9.08	35.25	5.82	35.16	50	1	3.67	8.63	4.76	23.77	26	1	2.45	4.11	5.38	31.04
Avg	**104**	**3**	**6.38**	**19.80**	**5.59**	**36.29**	**24**	**3**	**3.67**	**3.98**	**2.86**	**12.67**	**15**	**2**	**3.45**	**2.91**	**2.07**	**6.78**

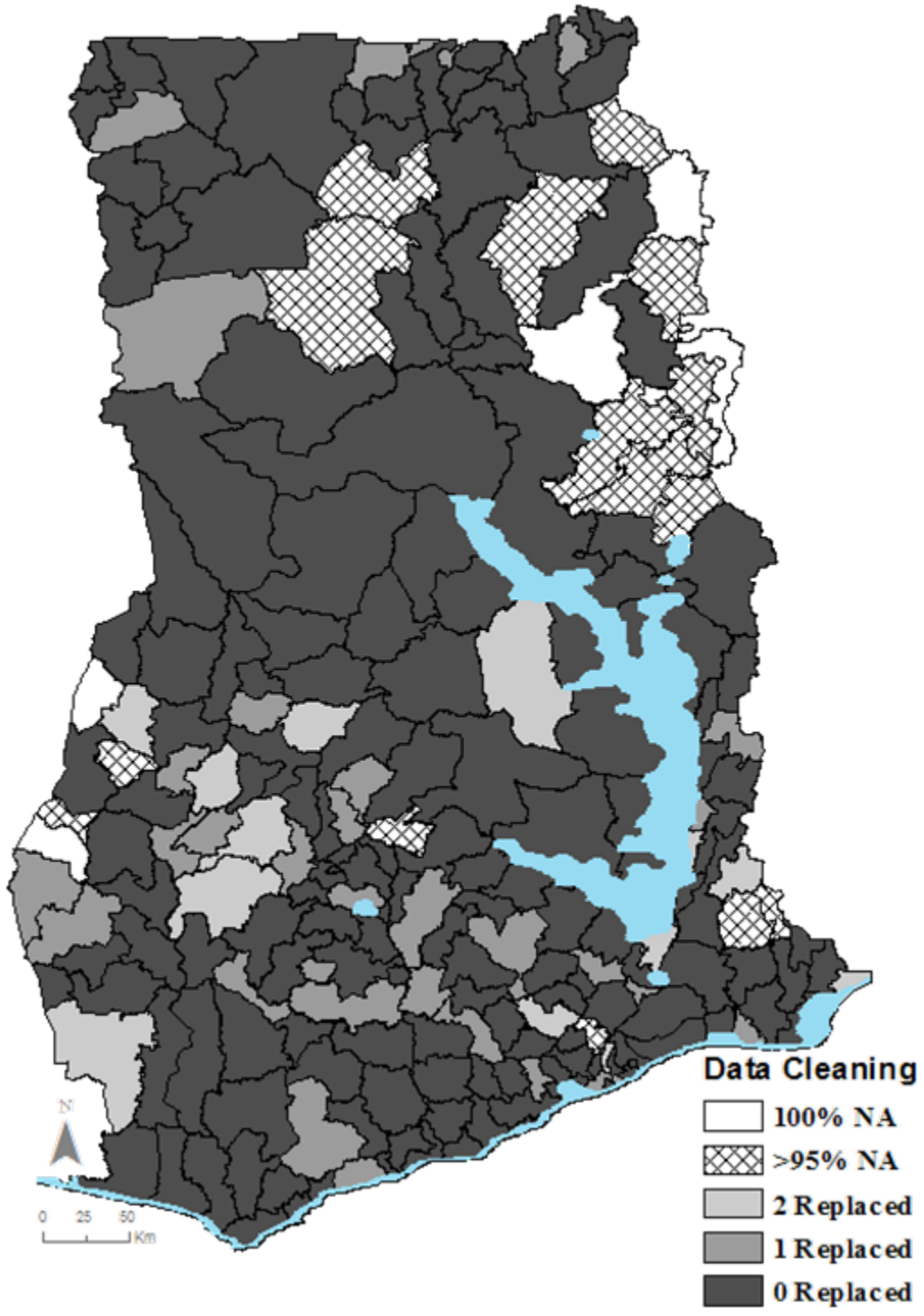


Figure 6: *Map of districts affected by data cleaning. White and crosshatched areas represent districts removed based on the number of missing values. The shade of gray reflects the amount of high values that were replaced with missing values*

Health data cleaning resulted in a reduction in the number of observations (Figure 5). The final dataset consisted of 18,525 observations or month/district combinations, of which 10,817 (58%) had at least one reported case of schistosomiasis.

2.2 Environmental predictors

Three environmental variables were derived from publically available remote sensing data streams: land surface temperature (LST), normalized difference vegetation index (NDVI), and accumulated precipitation (AP) (Table 4). LST and NDVI were downloaded from the online Data Pool at the NASA Land Processes Distributed Active Archive Center (LP DAAC), USGS/Earth Resources Observation and Science (EROS) Center, Sioux Falls, South Dakota. Their data came from the Moderate Resolution Imaging Spectroradiometer (MODIS) sensor aboard the Aqua and Terra satellites, which were combined to produce 8-day composites. AP was downloaded from Goddard Earth Sciences Data and Information Services Center (GES DISC)'s data visualization tool, GIOVANNI, and utilized data from the Tropical Rainfall Monitoring Mission (TRMM). All three datasets were mosaicked to cover the extent of Ghana, and aggregated to monthly mean values per district in order to match the temporal and spatial aggregation of the health outcome data. Spatial aggregation was performed using cell statistics and zonal statistics tools in ArcGIS (Version 10.4.1). Where resampling was required, the cubic convolution technique was used because it more realistically reflected the smooth transitions of environmental data across terrain.

Table 4: *Data sources, temporal and spatial resolution of environmental parameters*

Parameter	Source	Data Product	Temporal Resolution	Spatial Resolution
NDVI	MODIS	MOD/MYD13A2	Calculated 8-day composites	1 km²
LST (ºC)	MODIS	MOD/MYD11A2	Calculated 8-day composites	1 km²
AP (m)	TRMM	3B43 v7	Monthly	28 km²

2.3 Defining the climate zones

Ghana has a diverse climate, ranging from hot and dry savannah in the north, tropical forest in the middle, and coastal savannah in the south of the country (Frenken 2005). We wanted to explore temporal and spatial patterns in disease counts across a range of climatic conditions. As an alternative to the commonly used Köppen–Geiger (KG) climate classification, which would have resulted in 2 distinct zones for Ghana, we used a new "Limiting, K-means, Nomination" (LKN) method to define climate zones. The LKN method is based on a k-means clustering algorithm over space and time (Liss et al. 2014). The datasets used were NDVI and LST data from the MODIS sensor, collected for 15 years (2000-2015). The 8-day composites were mosaicked to cover the extent of Ghana, and aggregated in a layered space-time series. After masking the water bodies, the multiple layers of 8-day composite NDVI and LST images were pixel-averaged and principal component decomposition was applied to reduce dimensionality of the time series. The first 4 and 8 principal components retained 90% and 95% of the original information, respectively, and composite images of these components showed high spatial separation and a large signal to noise ratio. Multiple k-means unsupervised classifications were performed using varying classes, principal components, and distance measures, which were analyzed using cluster validity indexes. The most compact clustering solution exhibited the highest degree of homogeneity within each cluster and the highest degree of heterogeneity across

different clusters. Out of 600 candidate partitions, 3 major zones (Z3:1 to Z3:3) and 9 minor (non-hierarchical) zones (Z9:1 to Z9:9) were produced that were entirely data-driven and specific to Ghana. We used zonal statistics tools in ArcGIS (Version 10.4.1) to determine the major and minor climate zones assigned to each district.

2.4 Statistical modeling

Exploratory analyses included histograms, maps, plots of trend and seasonality, and descriptive statistics for the outcome and environmental predictors, stratified by climate zone. The associations among variables were examined using Spearman's rank correlation and regression models. Generalized linear mixed effects regression models with a random intercept term were used to assess temporal features and associations between environmental predictors and the outcome, accounting for district-level clustering. Temporal features included trend and two seasonal harmonic terms (Jagai et al. 2012; Kulinkina et al. 2016; Naumova et al. 2007). Models were repeated for major and minor climate zones.

Equation 1: *Complete mixed effects model*

$$Y_{tj} = \beta_0 + \beta_1 t + \beta_L S + \beta_M RS_{tj} + \alpha District_j + \epsilon_{tj}, \quad (1)$$

where Y_{tj} is the $\log_e$ transformed disease rate per million for t-month and j-district; β_o is the intercept; β_1 represents the regression coefficient for trend represented by continuous value for the month of the study period t ranging from 1 to 95; β_L represents the set of four regression coefficients for seasonality, S measured by four harmonic terms.

Equation 2: *Seasonality mixed effects model*

$$S = \beta_2 \sin(2\pi\omega t) + \beta_3 \cos(2\pi\omega t) + \beta_4 \sin(4\pi\omega t) + \beta_5 \cos(4\pi\omega t). \quad (2)$$

Seasonality in disease counts was assessed based on the significance of the four harmonic terms, capturing up to two annual peaks, with ω = 1/12; we considered seasonality to be present if at least one harmonic term was statistically significant. The effects of remotely sensed environmental variables on the health outcome were represented by β_M.

Equation 3: *Remotely sensed environmental variables mixed effects model*

$$RS = \beta_6 LST_{tj} + \beta_7 NDVI_{tj} + \beta_8 AP_{tj} \quad (3)$$

We built the model sequentially from three partial models to a complete final model. Model 1 included only the temporal trend, Model 2 only the seasonal component (Equation 2), Model 3 only the environmental variables (Equation 3), and Model 4 contained all components (Equation 1). The estimates of the predicted percent change in monthly rates of reported cases (%R) per one unit increase for each environmental variable along with their 95% confidence interval limits (CI95%) were calculated by exponentiation of the regression coefficients and converting to % form: %R = (exp{β_M} – 1)*100% and CI95% = (exp{β_M ±1.96 SEβm} – 1)*100%, respectively. Similarly, the estimates for trend were obtained using the β_1 coefficient. Predicted temporal curves were plotted using partial model results. All models were fitted by the restricted maximum likelihood (REML) method, using the glmer function of the R package [lme4] (version 3.3.1). Model fit was assessed using R^2, or percent variability explained.

3. Results

3.1 Spatial and temporal distribution of variables

Histograms showed the frequency distribution of average monthly values for the 8-years of reporting for all analysis variables (Figure 7). Log_e transformed disease rate had a peak around 3.5-4.0 (33-55 cases per million people). NDVI showed a bimodal distribution with a major peak around 0.7 and a minor peak around 0.2. LST also exhibited two peaks, one around 27˚C and another around 37˚C, which contributed to a long right skew. AP had a high frequency of low values and a long tail indicating frequency of high AP values during rainy seasons.

To examine the temporal patterns, monthly values for all variables were plotted as aggregates (Figure 7, bottom row) and consecutively over the 8-year period (Figure 8). Log_e transformed disease rates did not appear to exhibit seasonality (Figure 7, bottom row), and showed a slight decline over the study period (Figure 8). Environmental predictors exhibited seasonality (Figure 7, bottom row and Figure 8), with NDVI and AP having two peaks per year and LST having one peak per year. Peaks in NDVI occurred in March and September (0.70), with a dip in June (0.55). LST peaked around February (27 ºC) with lower values in July (25 ºC). The highest AP values occured in June and September (20 cm), with little to no precipitation in January and August (Figure 7, bottom row).

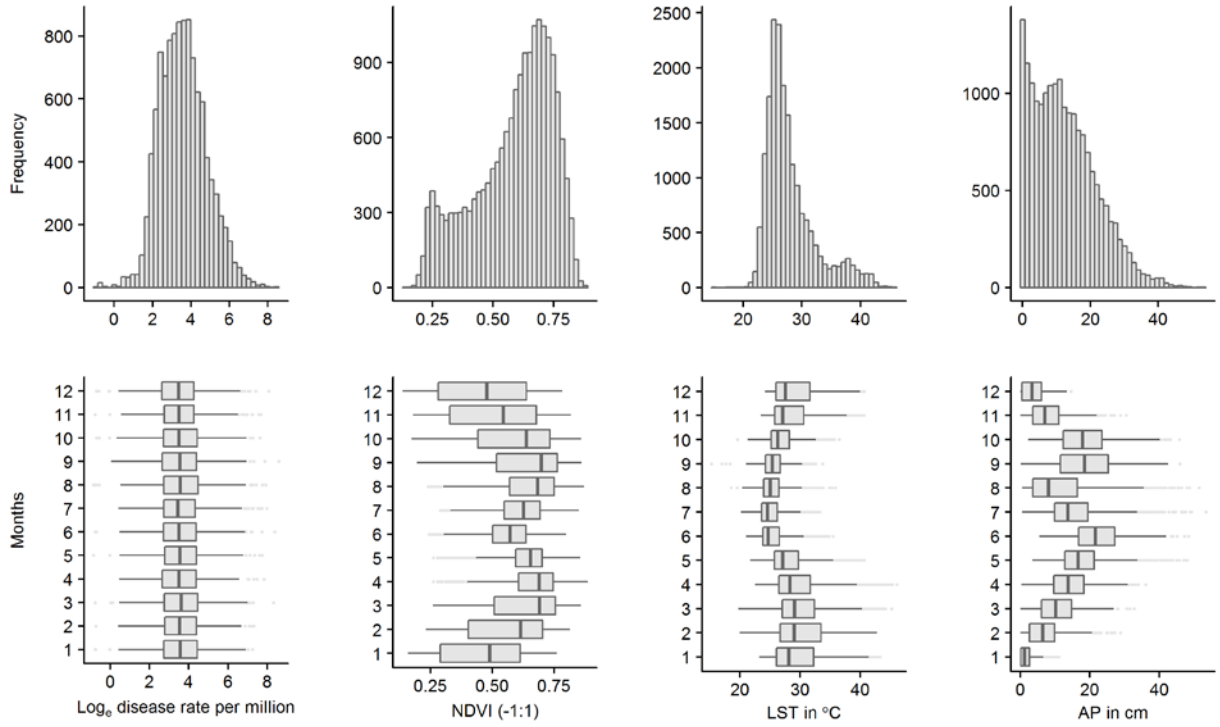


Figure 7: *Top row: histograms of health outcome and environmental parameters; Bottom row: annual seasonal patterns based on monthly boxplots representing distribution of health outcome and environmental parameters across 195 districts*

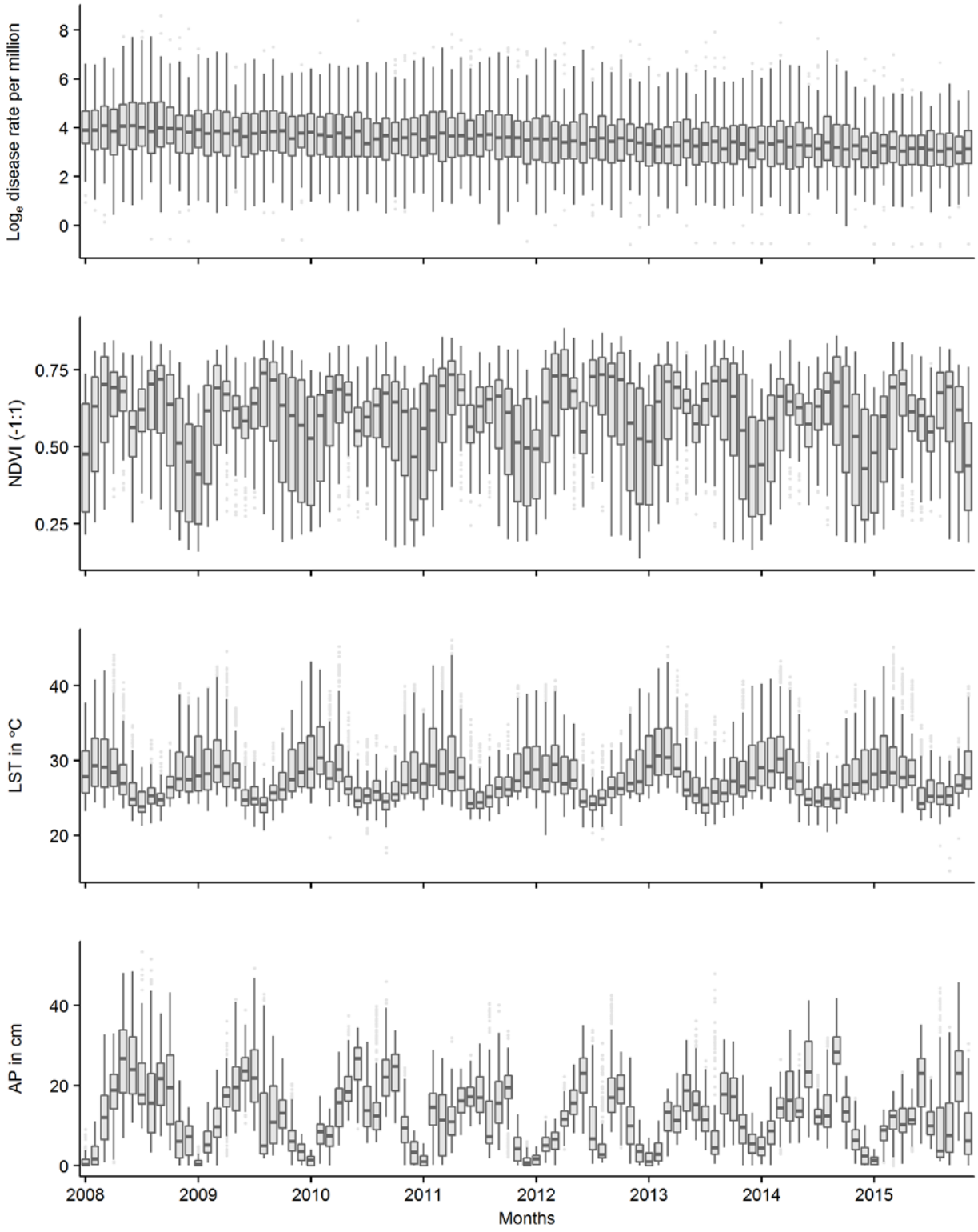


Figure 8: *Time series of monthly boxplots representing distribution of disease outcome and three environmental parameters across 195 districts*

To examine the spatial distribution of values, variables were mapped based on their 8-year district mean values (Figure 9). Average $\log_e$ transformed disease rate had a spatially heterogeneous pattern that ranged from 0.72 to 5.91 (2 to 369 cases per million people, respectively). The environmental variables showed trends along a southwest to northeast diagonal. NDVI decreased along this diagonal, with slightly higher values along the eastern shore of Lake Volta. LST increased along the diagonal, except for higher temperatures along the heavily urbanized Southeast coast and the peri-urban area surrounding Kumasi, the second largest city. AP showed trends along a diagonal that extended from the southwest corner to the center east portion of the country. Along this diagonal AP was high and decreased to either side, declining rapidly near the coast and more gradually towards the north.

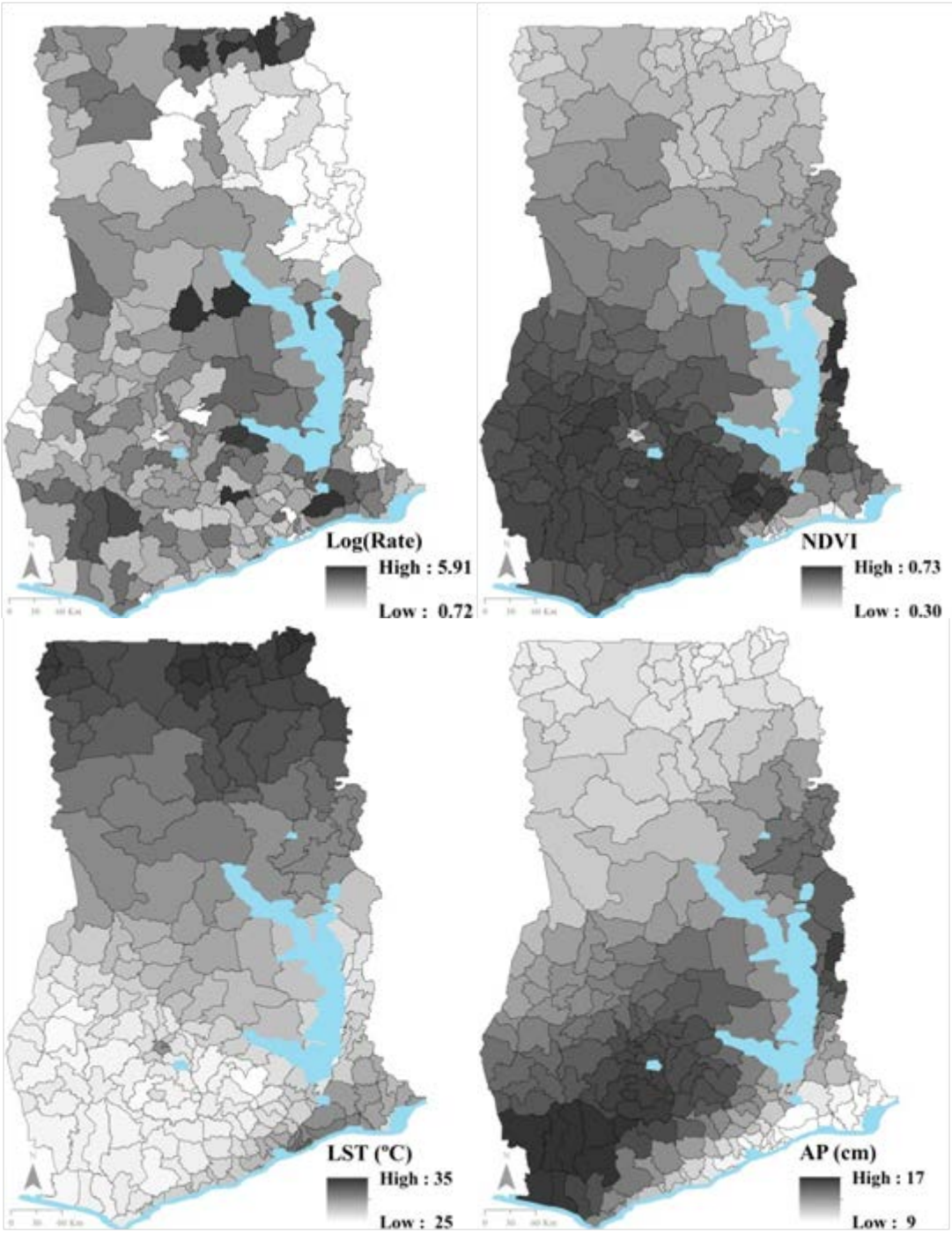


Figure 9: *Maps of health outcome (loge transformed disease rates per million) and three environmental predictors (8-year average aggregated at the district level). In the map of health outcome, districts with >95% of missing values are colored white*

3.2 Climate zones

The LKN climate regionalization method resulted in three major and nine minor climate zones (Figure 10). The mean monthly values for NDVI, LST, and AP on the national level were 0.58±0.16, 28.01±4.26 ºC, and 12.47±9.22 cm, respectively (Table 5). For the major zones, the northern part of the country (Z3:1) had the lowest amount of rainfall, lowest vegetation index, and highest temperature. Moderate precipitation, vegetation, and temperature values were observed in the middle of the country (Z3:2). The south (Z3:3) had the highest precipitation, vegetation, and temperature values. The mean values along the minor zones revealed further North to South trends within all major zones (except for areas represented by zones Z9:8 and Z9:9). These coastal areas are likely to be the most urban, which resulted in less vegetation and higher temperatures. Minor zone Z9:9 specifically, contains the capital, Accra, and Kumasi the second largest city. Histograms and boxplots for all variables stratified by climate zone were also explored (Figure 10 & Figure 11), and demonstrate pronounced diversity of seasonal patterns for all three environmental parameters. The seasonal patterns of NDVI and AP exhibited a single annual peak in the north transitioning to two annual peaks in the southern part of the country, while the seasonal peaks in LST became less pronounced.

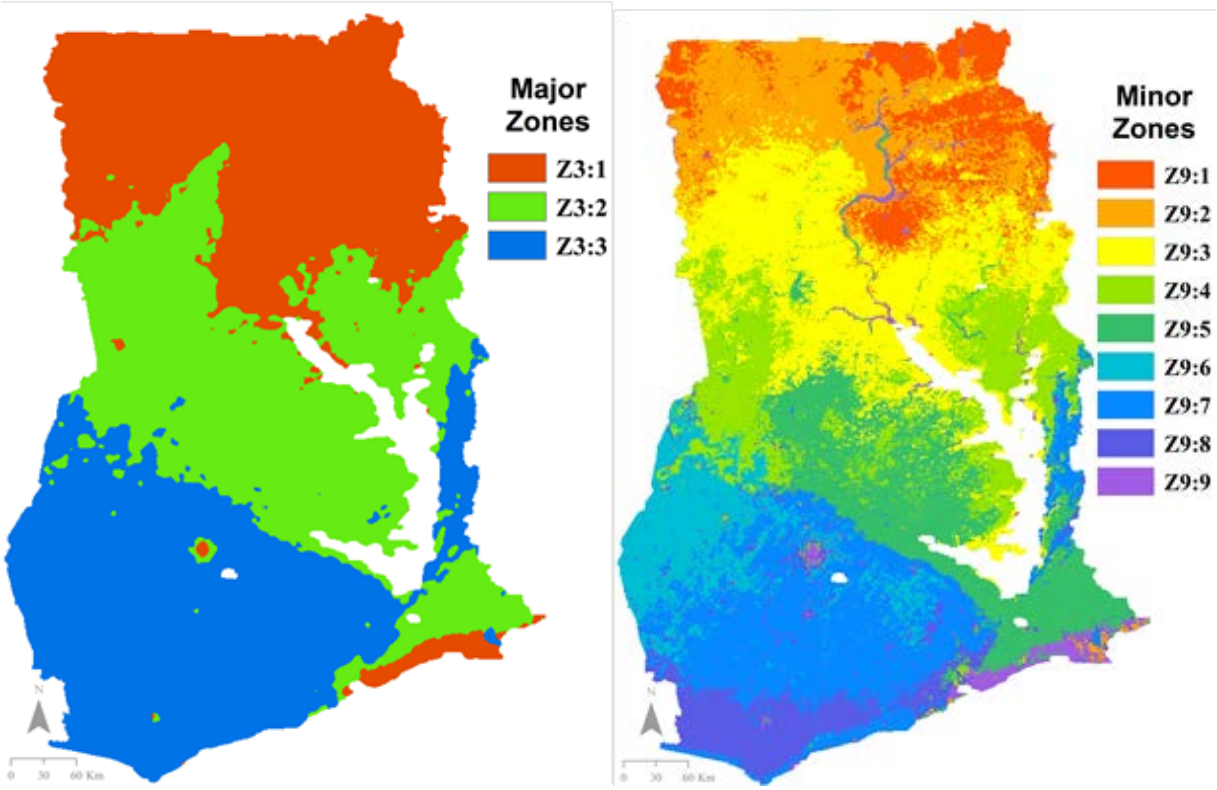


Figure 10: *LKN climate classification for Ghana resulted in 3 major and 9 minor zones*

Overall, the descriptive statistics for each variable were analyzed, and stratified by climate zone (Table 5). We present the total number of districts, total number of observations (including NA values), and population density per zone (per million people).

Table 5: *Descriptive statistics of the health outcome and predictors stratified by climate zones*

Descriptive Statistics		CLIMATE ZONES												
		ALL	MAJOR			MINOR								
			Z3:1	Z3:2	Z3:3	Z9:1	Z9:2	Z9:3	Z9:4	Z9:5	Z9:6	Z9:7	Z9:8	Z9:9
Districts		195	50	43	102	13	18	8	13	31	15	62	18	17
Observations		18525	4750	4085	5690	1235	1710	760	1235	2945	1425	5890	1710	1615
Median pop. density		123	134	64	143	142	62	19	50	97	98	159	234	1775
Log_e rate per million	min	-0.853	-0.853	0.745	0.466	1.410	0.848	1.856	1.379	0.745	1.691	1.196	0.466	-0.853
	max	8.577	8.054	8.577	7.827	8.054	6.969	6.817	8.577	8.308	7.003	7.586	7.143	6.558
	median	3.523	3.697	3.605	3.445	4.687	3.583	3.455	3.503	3.817	2.969	3.495	3.136	2.983
	mean	3.591	3.667	3.705	3.513	4.638	3.717	3.540	3.613	3.926	3.108	3.551	3.226	2.885
	std dev	1.216	1.429	1.206	1.104	1.172	1.213	1.008	1.122	1.236	0.894	1.071	1.135	1.313
NDVI	min	0.137	0.137	0.165	0.216	0.174	0.205	0.165	0.175	0.245	0.267	0.216	0.353	0.137
	max	0.885	0.751	0.807	0.885	0.695	0.720	0.772	0.815	0.807	0.871	0.885	0.858	0.705
	median	0.622	0.400	0.591	0.691	0.345	0.429	0.522	0.580	0.627	0.698	0.700	0.664	0.402
	mean	0.584	0.417	0.567	0.674	0.386	0.439	0.504	0.560	0.603	0.675	0.681	0.655	0.407
	std dev	0.162	0.143	0.135	0.102	0.150	0.157	0.156	0.148	0.120	0.113	0.102	0.090	0.104
LST	min	15.27	15.27	20.30	19.50	23.56	21.99	21.27	20.30	21.39	19.50	20.53	21.92	15.27
	max	46.06	46.06	41.69	42.40	46.06	45.21	42.13	40.36	41.69	42.40	41.51	33.61	42.07
	median	26.73	31.59	27.95	25.63	34.90	34.25	28.60	27.16	27.62	25.43	25.56	25.78	29.80
	mean	28.01	32.35	28.26	25.78	34.06	33.61	29.91	27.70	27.92	25.66	25.69	26.16	29.50
	std dev	4.260	5.167	3.078	1.843	5.178	5.553	4.680	3.298	2.723	2.117	1.793	1.912	3.155
AP	min	0.000	0.000	0.000	0.000	0.000	0.000	0.003	0.000	0.000	0.000	0.000	0.128	0.000
	max	53.39	51.49	46.87	53.39	47.94	51.49	41.85	45.95	46.87	42.66	53.39	48.47	41.36
	median	11.10	8.168	10.98	12.62	7.655	8.105	10.84	12.32	11.09	13.14	12.96	10.78	7.997
	mean	12.47	10.52	12.12	13.57	10.73	11.00	12.01	12.85	12.23	13.33	13.86	12.68	9.643
	std dev	9.218	10.10	8.810	8.753	11.27	11.22	9.915	9.402	8.231	8.582	8.855	8.798	7.320

Figure 11 : *Histograms of loge transformed disease rate, NDVI, LST, and AP for climate zones*

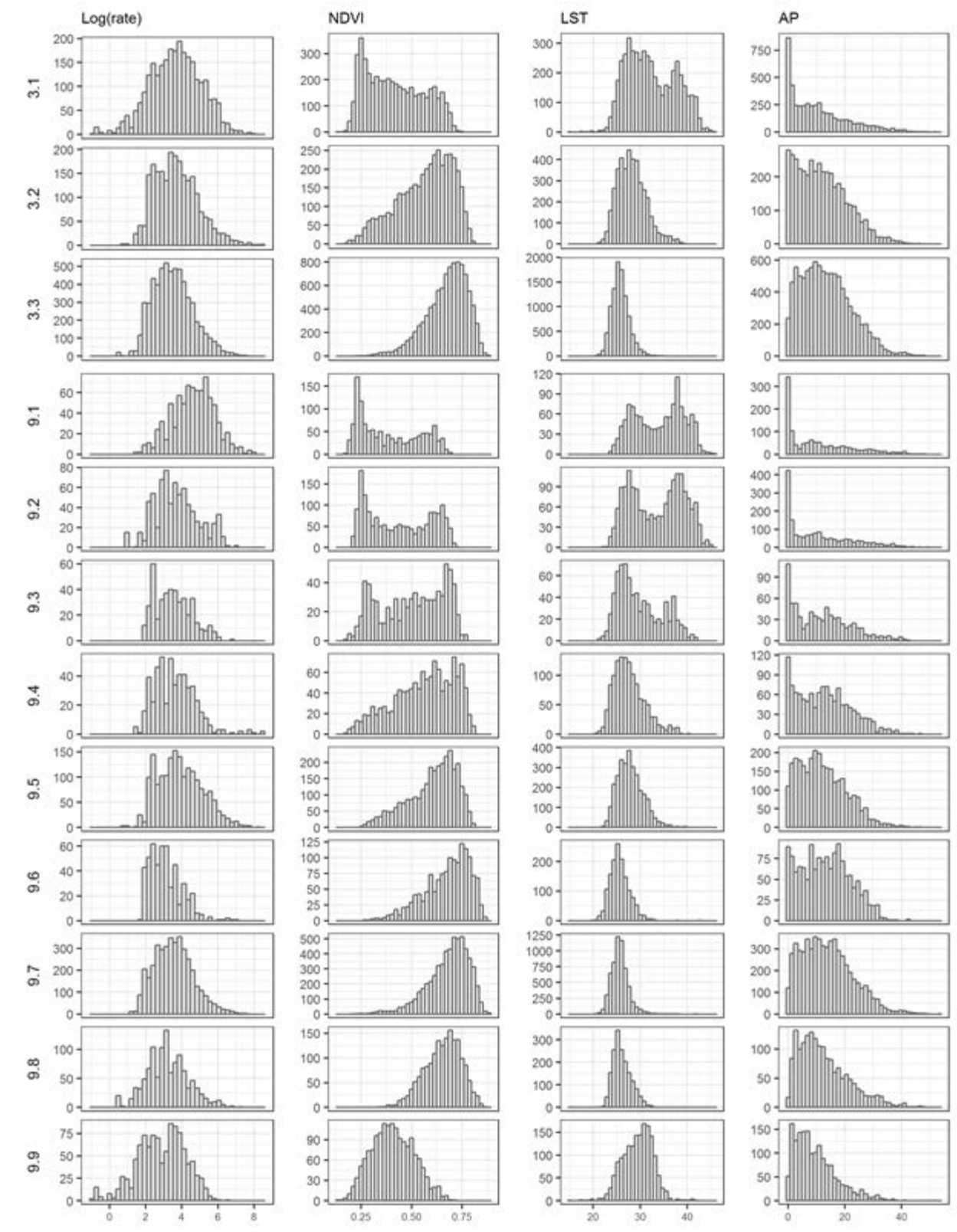
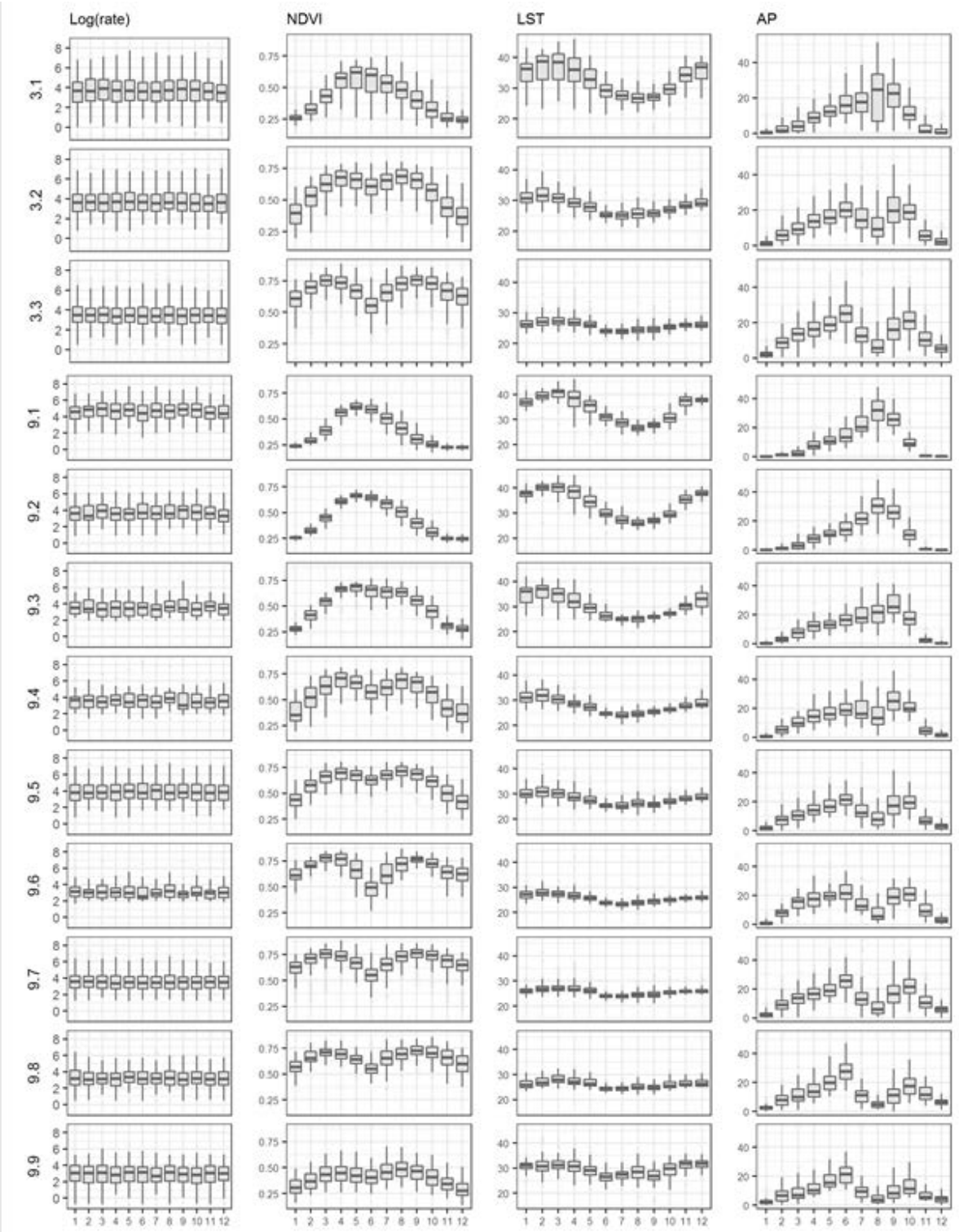


Figure 12: *Annual seasonal patterns shown with boxplots of monthly values across 195 districts for $\log_e$ transformed disease rate, NDVI, LST, and AP for major and minor climate zones*



3.3 Associations between health outcome and environmental predictors

Spearman's rank correlation was used to analyze pairwise relationships among variables, stratified by climate zone (Table 6). The association between NDVI and LST was primarily negative; it was weak at the national level (rs = -0.13) and moderate in major zones Z3:1 and Z3:2 (rs ~ -0.30) and in minor zones Z9:1 through Z9:5 (rs ~ -0.35). The association was weak or non-existent in major zone Z3:1 and minor zones Z9:6 through Z9:8 and moderate in peri-urban zone Z9:9 (rs = -0.31). The association between LST and AP was consistently negative; it was moderate at the national level (rs = -0.37) and strong in major zones Z3:1 and Z3:2 (rs ~ -0.55) and minor zones Z9:1 through Z9:4 (rs ~ -0.70). The association between NDVI and AP was primarily positive; it was weak at the national level (rs = 0.25), moderate in major zones Z3:1 and Z3:2 (rs ~ 0.45), and strong in minor zones Z9:1 through Z9:4 (rs ~ 0.55). Correlations among all environmental parameters were lowest in major zone Z3:3 and minor zones Z9:7 and Z9:8. Correlation coefficients between disease rates and environmental parameters were negligible at all levels and varied in direction and magnitude between -0.01 and 0.12.

Table 6: *Spearman's rank correlation coefficients for monthly values calculated for each district and averaged (mean ± standard deviation) across each climate zone.*

Correlations	CLIMATE ZONES												
	ALL	MAJOR			MINOR								
		Z3:1	Z3:2	Z3:3	Z9:1	Z9:2	Z9:3	Z9:4	Z9:5	Z9:6	Z9:7	Z9:8	Z9:9
Districts	195	50	43	102	13	18	8	13	31	15	62	18	17
n	18525	4750	4085	9690	1235	1710	760	1235	2945	1425	5890	1710	1615
NDVI	-0.126	-0.278	-0.365	0.050	-0.208	-0.277	-0.472	-0.352	-0.352	0.154	0.075	0.037	-0.307
LST	±0.253	±0.102	±0.175	±0.196	±0.085	±0.049	±0.123	±0.248	±0.136	±0.160	±0.170	±0.088	±0.114
LST	-0.367	-0.599	-0.515	-0.192	-0.747	-0.771	-0.693	-0.608	-0.452	-0.286	-0.172	-0.045	-0.282
AP	±0.279	±0.257	±0.174	±0.194	±0.041	±0.016	±0.177	±0.126	±0.146	±0.148	±0.179	±0.129	±0.155
NDVI	0.251	0.462	0.444	0.067	0.535	0.606	0.637	0.528	0.390	0.109	0.034	-0.021	0.218
AP	±0.268	±0.192	±0.183	±0.183	±0.062	±0.048	±0.032	±0.127	±0.137	±0.160	±0.167	±0.090	±0.108

n	10873	2677	2281	5915	811	755	434	507	1811	496	3914	1090	1055
Log_e rate	0.031	0.034	0.001	0.042	0.073	0.039	-0.065	0.066	0.020	0.093	-0.005	0.096	0.030
NDVI	±0.173	±0.174	±0.184	±0.168	±0.159	±0.176	±0.145	±0.207	±0.200	±0.252	±0.141	±0.127	±0.156
Log_e rate	-0.005	-0.005	0.028	-0.018	0.000	-0.030	-0.007	0.123	-0.027	-0.028	-0.015	0.007	0.012
LST	±0.204	±0.218	±0.199	±0.198	±0.166	±0.299	±0.127	±0.292	±0.163	±0.339	±0.154	±0.190	±0.143
Log_e rate	0.042	0.046	0.043	0.039	0.054	0.051	0.036	-0.009	0.084	-0.008	0.042	0.057	0.012
AP	±0.160	±0.148	±0.165	±0.164	±0.132	±0.177	±0.121	±0.152	±0.156	±0.209	±0.151	±0.180	±0.156

3.4 Modeling

Following exploratory analyses, mixed effects regression models were conducted, stratified by major and minor zones. The first univariate model, using only trend as an explanatory variable, showed a significant decline in reported disease rates equivalent to approximately 1% per month at all levels (Table 7). District effects and trend explained 55% of the variability in the data at the national level, 46-70% in the major zones, and 31-65% in the minor zones (Table 8). The second model, using four seasonal harmonic variables revealed that seasonality in reported disease rates was present at the national level, in major zone Z3:1, and in minor zones Z9:1 and Z9:2. In the third model, using the three environmental predictors, AP showed a small in magnitude but statistically significant positive association with schistosomiasis rates at the national level, in major zones Z3:1 and Z3:2, and in minor zones Z9:3, Z9:4, Z9:5, and Z9:7. A 1-cm increase in rainfall was associated with a 0.3-1.6% increase in monthly disease rates (Table 7). Associations between disease rates and LST varied in magnitude and direction but were not significant. In major zone Z3:3, NDVI had a positive effect on the health outcome equivalent to a 3.4% increase in disease rates associated with a 0.1-unit increase in NDVI. In minor zone Z9:3, NDVI had a negative effect equivalent to a 7.7% decrease in disease rates corresponding to a 0.1-unit increase in NDVI.

Table 7: *Estimated trend in reported rates of schistosomiasis and associations with environmental parameters (p<0.05 bolded) for partial mixed effects regression models, shown as % change in monthly rates associated with 1 unit increase in each parameter and their 95% confidence limits*

Model parameters		CLIMATE ZONES												
		ALL	MAJOR					MINOR						
			Z3:1	Z3:2	Z3:3	Z9:1	Z9:2	Z9:3	Z9:4	Z9:5	Z9:6	Z9:7	Z9:8	Z9:9
Model 1: Univariate - trend only														
Trend (1 month)	% change	**-0.987**	**-0.892**	**-0.933**	**-1.052**	**-1.119**	**-0.826**	**-0.856**	**-1.332**	**-1.059**	**-0.317**	**-1.000**	**-1.153**	**-0.798**
	LCL	**-1.046**	**-1.004**	**-1.075**	**-1.131**	**-1.319**	**-1.029**	**-1.151**	**-1.641**	**-1.215**	**-0.593**	**-1.099**	**-1.321**	**-0.984**
	UCL	**-0.928**	**-0.781**	**-0.791**	**-0.973**	**-0.918**	**-0.623**	**-0.559**	**-1.021**	**-0.902**	**-0.040**	**-0.902**	**-0.985**	**-0.612**
Model 3: Multivariate - environmental variables only														
NDVI (0.1)	% change	1.530	1.157	-3.597	**3.389**	2.872	1.407	**-7.711**	-1.579	0.218	5.334	1.750	4.855	2.035
	LCL	-0.054	-1.412	-7.306	**0.862**	-1.467	-2.597	**-13.53**	-9.696	-4.297	-1.335	-1.407	-1.413	-4.778
	UCL	3.139	3.793	0.260	**5.980**	7.402	5.575	**-1.496**	7.269	4.946	12.454	5.008	11.522	9.335
LST (1 °C)	% change	0.354	0.811	0.322	-0.513	1.053	-0.841	1.449	1.401	-0.120	-0.126	0.068	-2.784	1.359
	LCL	-0.277	-0.155	-1.260	-1.824	-0.905	-2.478	-1.093	-2.272	-2.025	-3.820	-1.581	-5.778	-0.659
	UCL	0.988	1.787	1.930	0.816	3.050	0.823	4.057	5.211	1.822	3.710	1.744	0.305	3.417
AP (1 cm)	% change	**0.391**	**0.513**	**1.273**	0.193	0.875	-0.412	**1.637**	**1.345**	**0.727**	-0.336	**0.337**	0.130	0.487
	LCL	**0.183**	**0.046**	**0.700**	-0.064	-0.068	-1.309	**0.463**	**0.044**	**0.083**	-1.139	**0.018**	-0.440	-0.256
	UCL	**0.600**	**0.982**	**1.849**	0.451	1.827	0.493	**2.823**	**2.664**	**1.375**	0.473	**0.656**	0.702	1.236

Table 8: *Results of 3 partial mixed effects regression models with p-values <0.05 bolded*

Model parameters		CLIMATE ZONES												
		ALL	MAJOR			MINOR								
			Z3:1	Z3:2	Z3:3	Z9:1	Z9:2	Z9:3	Z9:4	Z9:5	Z9:6	Z9:7	Z9:8	Z9:9
Model 1: Trend only														
Trend	ß coef.	**-0.010**	**-0.009**	**-0.009**	**-0.011**	**-0.011**	**-0.008**	**-0.009**	**-0.013**	**-0.011**	**-0.003**	**-0.010**	**-0.012**	**-0.008**
	Std. error	**0.000**	**0.001**	**0.001**	**0.000**	**0.001**	**0.001**	**0.002**	**0.002**	**0.001**	**0.001**	**0.001**	**0.001**	**0.001**
	p-value	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**0.025**	**<0.001**	**<0.001**	**<0.001**
R^2		0.55	0.70	0.46	0.47	0.54	0.65	0.40	0.34	0.51	0.31	0.41	0.54	0.61
Model 2: Seasonality only														
sin(2π)	ß coef.	0.019	0.027	0.008	0.018	-0.012	0.017	-0.042	-0.001	0.012	-0.001	0.037	-0.021	0.067
	Std. error	0.012	0.022	0.027	0.016	0.042	0.039	0.055	0.062	0.030	0.048	0.020	0.036	0.037
	p-value	0.111	0.218	0.760	0.245	0.778	0.661	0.449	0.993	0.699	0.980	0.060	0.553	0.070
cos(2π)	ß coef.	**-0.023**	**-0.047**	-0.044	-0.005	**-0.113**	-0.025	0.032	-0.072	-0.055	0.002	0.011	-0.027	-0.034
	Std. error	**0.012**	**0.023**	0.028	0.016	**0.043**	0.040	0.056	0.063	0.031	0.049	0.020	0.036	0.037
	p-value	**0.049**	**0.038**	0.113	0.769	**0.008**	0.523	0.569	0.257	0.076	0.966	0.595	0.454	0.367
sin(4π)	ß coef.	**0.024**	0.002	0.034	0.031	-0.038	-0.005	0.061	0.042	0.025	0.014	0.032	0.043	0.019
	Std. error	**0.012**	0.022	0.027	0.016	0.042	0.039	0.055	0.063	0.030	0.050	0.020	0.035	0.037
	p-value	**0.037**	0.939	0.212	0.050	0.365	0.896	0.268	0.506	0.399	0.782	0.098	0.221	0.605
cos(4π)	ß coef.	**-0.024**	**-0.092**	-0.003	-0.001	**-0.147**	**-0.084**	-0.021	-0.010	0.000	0.003	-0.007	-0.005	-0.030
	Std. error	**0.012**	**0.023**	0.028	0.016	**0.043**	**0.039**	0.056	0.062	0.031	0.048	0.020	0.036	0.037
	p-value	**0.045**	**<0.001**	0.912	0.956	**0.001**	**0.034**	0.707	0.877	0.994	0.946	0.715	0.893	0.421
R^2		0.51	0.68	0.42	0.41	0.48	0.62	0.36	0.25	0.47	0.30	0.35	0.47	0.59
Model 3: Environmental variables only														
NDVI	ß coef.	0.152	0.115	-0.366	**0.333**	0.283	0.140	**-0.802**	-0.159	0.022	0.520	0.173	0.474	0.201
	Std. error	0.080	0.131	0.200	**0.126**	0.220	0.206	**0.333**	0.439	0.235	0.334	0.161	0.314	0.353
	p-value	0.058	0.381	0.067	**0.008**	0.198	0.497	**0.016**	0.717	0.926	0.119	0.281	0.132	0.568
LST	ß coef.	0.004	0.008	0.003	-0.005	0.010	-0.008	0.014	0.014	-0.001	-0.001	0.001	-0.028	0.013
	Std. error	0.003	0.005	0.008	0.007	0.010	0.008	0.013	0.019	0.010	0.019	0.008	0.016	0.010
	p-value	0.272	0.100	0.692	0.447	0.294	0.320	0.267	0.460	0.903	0.948	0.936	0.077	0.188
AP	ß coef.	**0.004**	**0.005**	**0.013**	0.002	0.009	-0.004	**0.016**	**0.013**	**0.007**	-0.003	**0.003**	0.001	0.005
	Std. error	**0.001**	**0.002**	**0.003**	0.001	0.005	0.005	**0.006**	**0.007**	**0.003**	0.004	**0.002**	0.003	0.004
	p-value	**<0.001**	**0.031**	**<0.001**	0.142	0.069	0.371	**0.006**	**0.043**	**0.027**	0.415	**0.038**	0.656	0.199
R^2		0.51	0.68	0.43	0.41	0.48	0.62	0.37	0.25	0.47	0.31	0.35	0.47	0.59

A visual representation of a model including trend and seasonality in reported schistosomiasis cases was created (Figure 13). At the national scale, both trend and seasonality were significant ($p<0.05$), with two relative peaks in reported disease rates observed around March and September with the lowest counts observed in December/January. For the major climate zones,

trend was significant in all zones and seasonality remained significant only in zones Z3:1 and Z3:2 ($p<0.05$). In zone Z3:1 (north), two peaks occurred in March and September and in zone Z3:2 (middle), a single peak occurred in July. For the minor zones, the declining trend remained; however, seasonality terms were only significant ($p<0.05$) for zones Z9:1, Z9:2 and Z9:8. Like the major zone Z3:1, zones Z9:1 and Z9:2 exhibited two peaks per year in March and September. Zones Z9:4, Z9:5 and Z9:8 resembled major zone Z3:2 and exhibited a single annual peak around June/July. The remaining minor zones did not show pronounced seasonality. The lowest counts in most zones were observed in December/January.

Figure 13: *Visualization of the fitted values produced by the trend and seasonality model. Line colors match those of Figure 9; x-axis represents months from January 2008 to November 2015 (vertical lines coincide with the month of January)*

In the final regression models, inclusive of all predictors, downward trend remained significant at all levels (Table 9 & Table 10) and seasonality was significant in zones Z3:1, and Z9:1 through Z9:3 (in the north of the country). Controlling for trend and seasonality, associations with environmental parameters varied substantially by climate zone. The association remained significant for NDVI only in zone Z9:3, exhibiting even stronger estimated percent change, and

for AP in zone Z3:2 at approximately the same magnitude of 1.1% increase in monthly disease rates for 1-cm increase in AP. In the most hot and dry zones Z9:1 and Z9:2, the associations were significant for LST (in zone Z9:1) and AP (in zone Z9:2). The R^2 values of the final models ranged between 0.32 and 0.71 (Table 10).

Table 9: *Estimated trend in reported rates of schistosomiasis and associations with environmental parameters (p<0.05 bolded) for complete mixed effects regression models, shown as % change in monthly rates associated with 1 unit increase in each parameter and their 95% confidence limits*

Model parameters		CLIMATE ZONES												
		ALL	MAJOR			MINOR								
			Z3:1	Z3:2	Z3:3	Z9:1	Z9:2	Z9:3	Z9:4	Z9:5	Z9:6	Z9:7	Z9:8	Z9:9
Trend (1 month)	% change	**-0.972**	**-0.896**	**-0.896**	**-1.046**	**-1.094**	**-0.896**	**-0.896**	**-1.292**	**-1.094**	**-0.300**	**-0.995**	**-1.193**	**-0.797**
	LCL	**-1.032**	**-1.090**	**-1.090**	**-1.126**	**-1.288**	**-1.090**	**-1.284**	**-1.678**	**-1.288**	**-0.495**	**-1.189**	**-1.386**	**-0.991**
	UCL	**-0.913**	**-0.702**	**-0.702**	**-0.966**	**-0.900**	**-0.702**	**-0.507**	**-0.904**	**-0.900**	**-0.104**	**-0.801**	**-0.999**	**-0.602**
NDVI (0.1)	% change	0.702	-2.107	-4.084	2.419	-1.636	-8.066	**-16.47**	0.692	0.713	9.111	0.341	1.633	2.655
	LCL	-1.097	-6.055	-9.915	-0.665	-15.37	-19.70	**-28.11**	-12.29	-6.754	-0.471	-3.554	-5.328	-6.910
	UCL	2.535	2.006	2.124	5.599	14.32	5.247	**-2.933**	15.59	8.777	19.617	4.393	9.106	13.20
LST (1 °C)	% change	0.300	0.602	-0.300	0.351	**2.429**	-1.686	2.425	0.602	-0.100	2.840	0.300	-2.078	0.803
	LCL	-0.483	-0.574	-2.426	-1.340	**0.048**	-4.347	-1.069	-4.584	-2.613	-2.843	-1.839	-5.842	-1.347
	UCL	1.090	1.792	1.873	2.072	**4.867**	1.049	6.043	6.069	2.478	8.854	2.486	1.837	3.000
AP (1 cm)	% change	**0.200**	0.031	**1.106**	0.100	-0.399	**-1.193**	0.702	0.501	0.501	-0.399	**0.401**	0.003	0.300
	LCL	**0.004**	-0.451	**0.513**	-0.292	-1.564	**-2.156**	-0.670	-0.868	-0.284	-1.564	**0.008**	-0.716	-0.678
	UCL	**0.397**	0.516	**1.702**	0.493	0.779	**-0.220**	2.094	1.890	1.292	0.779	**0.795**	0.726	1.288

Table 10: *Results of complete mixed effects regression model with p-values <0.05 bolded*

Model parameters		CLIMATE ZONES												
		ALL	MAJOR			MINOR								
			Z3:1	Z3:2	Z3:3	Z9:1	Z9:2	Z9:3	Z9:4	Z9:5	Z9:6	Z9:7	Z9:8	Z9:9
Trend	ß coef.	**-0.010**	**-0.009**	**-0.009**	**-0.011**	**-0.012**	**-0.009**	**-0.009**	**-0.013**	**-0.011**	**-0.003**	**-0.010**	**-0.012**	**-0.008**
	Std. error	**0.000**	**0.001**	**0.001**	**0.000**	**0.001**	**0.001**	**0.002**	**0.002**	**0.001**	**0.001**	**0.001**	**0.001**	**0.001**
	p-value	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**0.035**	**<0.001**	**<0.001**	**<0.001**
sin(2π)	ß coef.	-0.013	-0.023	0.014	-0.015	-0.225	0.020	-0.130	-0.049	-0.006	-0.065	0.007	-0.040	0.032
	Std. error	0.014	0.031	0.036	0.018	0.102	0.105	0.100	0.101	0.038	0.068	0.023	0.041	0.040
	p-value	0.353	0.467	0.702	0.402	0.027	0.846	0.195	0.627	0.870	0.339	0.750	0.329	0.425
cos(2π)	ß coef.	-0.009	**-0.090**	-0.033	0.003	-0.254	-0.257	**-0.341**	-0.049	-0.018	-0.060	0.032	-0.013	-0.019
	Std. error	0.015	**0.045**	0.052	0.018	0.160	0.159	**0.173**	0.136	0.056	0.066	0.023	0.040	0.059
	p-value	0.567	**0.044**	0.530	0.851	0.113	0.108	**0.048**	0.719	0.743	0.363	0.153	0.740	0.750
sin(4π)	ß coef.	0.017	-0.014	0.049	0.023	-0.006	-0.022	0.001	0.053	0.025	0.016	0.039	0.018	0.025
	Std. error	0.012	0.022	0.028	0.018	0.056	0.047	0.055	0.062	0.032	0.061	0.023	0.042	0.041
	p-value	0.164	0.529	0.077	0.217	0.921	0.637	0.982	0.391	0.440	0.793	0.100	0.671	0.551
cos(4π)	ß coef.	-0.018	**-0.095**	-0.016	0.019	**-0.180**	**-0.104**	-0.072	0.029	-0.001	0.094	-0.002	0.002	-0.016
	Std. error	0.013	**0.022**	0.036	0.020	**0.052**	**0.042**	0.073	0.084	0.042	0.071	0.025	0.043	0.043
	p-value	0.167	**<0.001**	0.645	0.332	**0.001**	**0.013**	0.325	0.728	0.990	0.187	0.936	0.957	0.704
NDVI	ß coef.	0.070	-0.200	-0.417	0.239	-0.118	-0.841	**-1.799**	0.069	0.071	0.872	0.034	0.162	0.262
	Std. error	0.092	0.211	0.320	0.156	0.787	0.690	**0.766**	0.704	0.393	0.469	0.202	0.362	0.499
	p-value	0.447	0.344	0.192	0.125	0.881	0.223	**0.019**	0.922	0.856	0.063	0.865	0.654	0.600
LST	ß coef.	0.003	0.006	-0.003	0.004	**0.026**	-0.017	0.024	0.006	-0.001	0.028	0.003	-0.021	0.008
	Std. error	0.004	0.006	0.011	0.009	**0.012**	0.014	0.018	0.027	0.013	0.029	0.011	0.020	0.011
	p-value	0.476	0.288	0.753	0.686	**0.032**	0.208	0.176	0.831	0.914	0.334	0.764	0.291	0.456
AP	ß coef.	**0.002**	0.000	**0.011**	0.001	-0.004	**-0.012**	0.007	0.005	0.005	-0.004	0.004	0.000	0.003
	Std. error	**0.001**	0.002	**0.003**	0.002	0.006	**0.005**	0.007	0.007	0.004	0.006	0.002	0.004	0.005
	p-value	**0.064**	0.923	**<0.001**	0.534	0.465	**0.030**	0.364	0.478	0.158	0.553	0.053	0.994	0.553
R^2		0.55	0.71	0.47	0.47	0.56	0.66	0.42	0.34	0.52	0.32	0.41	0.55	0.61

4. Discussion

4.1 Innovation

Our study was innovative in several ways. We used monthly records of schistosomiasis cases reported to a national surveillance system aggregated at the district level. We matched the health outcome to times series of remote sensing data. This approach allowed us to characterize the spatial and temporal variability of reported rates of schistosomiasis cases. Furthermore, we explored associations between disease incidence and environmental predictors across climatically homogeneous areas defined using a novel climate classification methodology.

To define the climate zones in our analysis, we used the LKN classification system and applied it specifically to Ghana, as compared to the predominant global KG climate classification system. The KG system is based on the assumptions that vegetation is the best proxy for climate and temperature and precipitation are the best proxies for vegetation (Kottek et al. 2006). The KG system divides the world into 6 major zones and 31 minor zones based on temperature, precipitation, and their seasonal variations. The KG classification partitions Ghana into only two climate zones; using the LKN method, finer divisions were possible. The major zones approximately corresponded to the agro-ecological zones, where the northern zone Z3:1 represents Guinea Savannah, zone Z3:2 the transitional zone, and zone Z3:3 represents a combination of deciduous forest and rainforest (Frenken 2005). Coastal Savannah was classified as a combination of zones Z3:1 and Z3:2. Using the minor zone divisions, urban areas were naturally separated into their own zone (Z9:9). An advantage of the LKN method is that it is fully automated and zone delineations can be updated over time.

To our knowledge, this is the first study using schistosomiasis surveillance data from a low-income African country to conduct a spatial and temporal analysis at the national scale. Using surveillance data offers many advantages over field data, such as expansive geographic coverage, temporal continuity, relatively low cost of data collection, and ability to aggregate data over various temporal and spatial scales. Furthermore, all predictors were publicly available satellite-remote sensing products, downloadable online. This methodology offers a way for public health officials in low-income countries to begin exploring patterns for climate-sensitive diseases using routinely collected data.

4.2 Major findings

There was a significant decline in reported disease rates over the study period, nationally and across all climate zones. Decreasing schistosomiasis reporting could be indicative of the success of MDA campaigns, which have increased in their frequency and geographic coverage in recent years. However, a limited exploration of the age distribution of cases reported between 2012 and 2015 (Figure 14) showed that a range of age groups contributed cases for both males and females, whereas MDA currently targets school children with limited community-based treatment of children and adults. Therefore, additional factors are likely contributing to the steady decline in reported cases and should be explored at various spatial and temporal scales.

Seasonality in reported disease rates was observed in several zones, with a consistent dip in December/January and varying patterns across climate zones. Two peaks in March and September in the dry northern areas (zones Z3:1, Z9:1 and Z9:2) correspond to periods of relatively low precipitation but high vegetation. The single peak in June/July in the middle of the

country (zones Z3:2, Z9:4, and Z9:5) correspond to the major rainy season. While the major reason for a winter dip is unknown, the Christmas holiday might potentially contribute to low reported counts around this time. It is unclear from the available data whether low reporting is a result of patient treatment-seeking behavior or internal reporting delays, which would be important to explore in future studies.

Based on the partial models (Table 7), the association between precipitation and schistosomiasis rates was positive in several zones. This finding is in agreement with prior studies that focused on *S. mansoni* (Scholte et al. 2014) and *S. haematobium* (Schur et al. 2011). The results of prior studies on the association between LST and schistosomiasis are somewhat contradictory, with positive association found with *S. haematobium* (Soares Magalhães et al. 2011); negative association with *S. mansoni* (Scholte et al. 2014); and no association with *S. haematobium* (Clements et al. 2006b). We did not find an association with LST and the association with NDVI was inconsistent. Only one minor zone Z9:3 with an extended annual peak in NDVI exhibited a strong association with the reported disease. After controlling for trend and seasonality, the associations between reported incidence and environmental variables remained for a few zones, yet largely not significant. Our findings suggest that the direction and strength of associations with remotely sensed parameters varied by climate zone; thus broad application of these parameters to countries or regions with heterogeneous climatic conditions should consider these properties.

4.3 Limitations

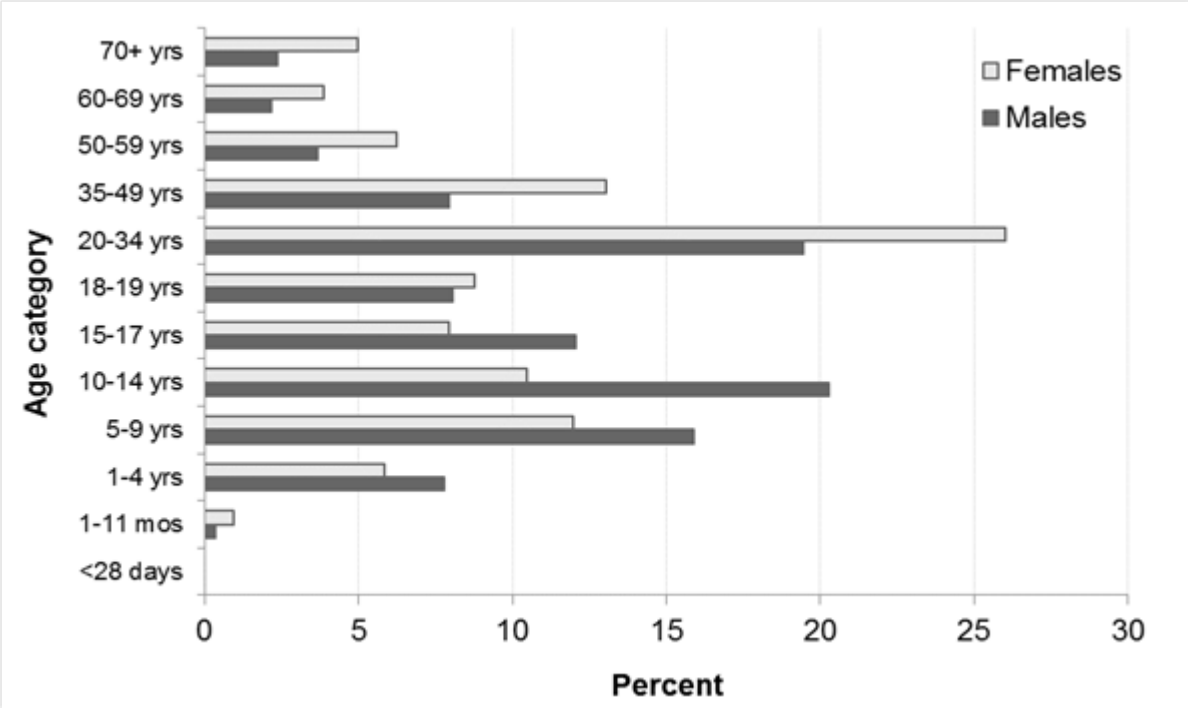
As previously mentioned, the dataset contained >50% of blank values, which could indicate either a lack of reporting or a lack of cases. During exploratory analysis, we found that the

majority of missing values occurred in between reported counts with fewer missing observations occurring prior to the first and/or after the last month with reported counts (Figure 5). In the present study, we assumed that the reporting started and ended at the same time (January 2008 and December 2015, respectively). However, it is possible that some of the districts began reporting later than others and/or stopped reporting prior to December 2015 for unknown reasons. Several outliers were also removed. The possible reason behind unusually high values may be reporting delays causing multiple months of data to be reported in a single month. Investigating the reasons behind missing observations and data inconsistencies was beyond the scope of this analysis but can be done in the future. Our data processing methodology likely produced conservative estimates, as the models were restricted to districts and months with at least one reported case of schistosomiasis. An exploration of various techniques to address blanks, missing data, and outliers deserves further study.

A second limitation is that the analysis used aggregated monthly disease counts recorded in the month they were reported and not necessarily as they occurred. Detailed information relevant to exposure, transmission, and reporting was not available. A third limitation is severe underreporting of disease counts due to limited resources and socio-economic conditions. Overall, the reported case numbers are extremely low, as compared to the estimated population at risk for Ghana. For example, in 2010, 24,996 cases were reported into DHIMS. If only children ≤15 years of age living in rural areas (5,128,118 individuals according to the 2010 census) are considered (i.e. the most at risk population), at the estimated 50% infection rate, ~2.5 million cases would be expected (Kulinkina 2017). These numbers could be used to suggest that DHIMS is capturing <1% of the expected cases. However thoughts along these lines must be

tempered with the fact that the actual schistosomiasis prevalence in Ghana expressed at the national level is estimated by interpolating focal health surveys. The reported schistosomiasis cases do however most likely represent the most serious cases from a subset of the population who are able to seek treatment at large government hospitals with diagnostic capability. Based on a visual analysis of the age distribution of average annual disease case counts reported into dhims2 from 2012-2015, it would appear that there is a skew towards an older patient subpopulation and that children are likely under-represented in our dataset (Figure 14). This is not a limitation of our analysis, but rather a reflection of the healthcare system in Ghana and other sub-Saharan African countries, dominated by routine school-based distribution of praziquantel, which results in a lower likelihood that cases among children would be reported into the surveillance system.

Figure 14: *Age distribution of average annual disease case counts reported during 2012-2015 (GHS, 2015)*



4.4 Future Directions

There are many options for future directions. One option is to include human parameters. Facebook in conjunction with the commercial satellite imagery provider, Digital Globe, the Center for International Earth Science Information Network (CIESIN) based at Columbia University, and the World Bank have produced publically available high-resolution (30m) population data for Ghana and a few other countries as of November, 2016. This is a byproduct of Facebook's ongoing project to spread internet to rural settlements around the world (CIESIN 2016). As stated previously the inclusion of human variables was purposefully omitted in this thesis to explore the utility of purely remotely-sensed environmental predictors. Moving forward this constraint could be relaxed so as to include both environmental and human variables. The additional variables that show potential for being explored are plentiful; however, sustainable data collection and analysis should be at the core of any methodology that intends to have the most impact in disease control within low income nations.

Another option would be to make use of previous survey data for verification purposes. Since the surveillance dataset used in this thesis only extended to 2015, there is a full year's worth of data from 2016 and future years that could be used to test the model's predictive power. However, it has been recommended that instead of testing with presence-only data, which will always have geographic and ecological biases, to instead use presence-absence data or artificial data (Elith et al. 2017). While finding or creating this data would be difficult, if made available it would allow for analysis of the sensitivity and specificity of the models.

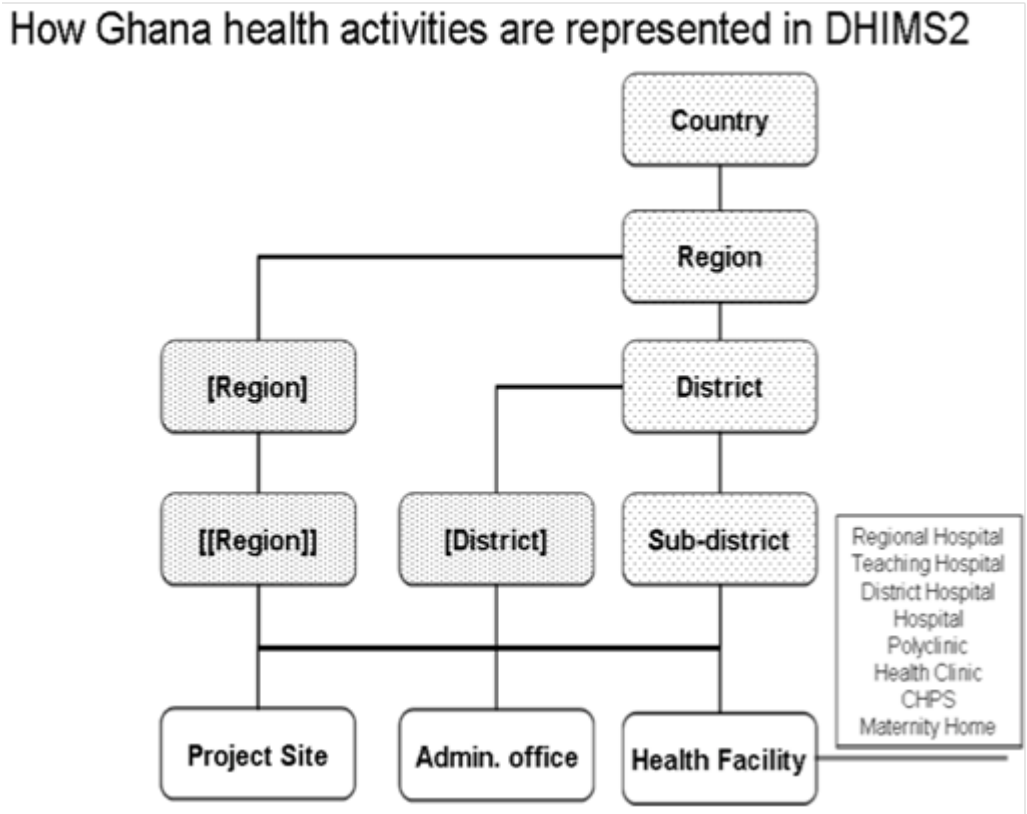
It would also seem practical to evaluate other methodologies such as outcome distributions, types of statistical analyses, and modeling frameworks. Fortunately a review of presence-only data modeling of 16 methods, 226 ecological species distribution, and 6 regions worldwide has been documented (Elith et al. 2017). The sophistication of the models utilized and detail with which the models were compared far exceeds what has been performed in the previous literature of RS applications to schistosomiasis. Thus, there is much to be gained from studying techniques and analyses used in other fields such as ecological modeling.

In our models, the majority of the explained variability was attributed to district-level effects, meaning that further analysis into what is causing these differences is needed. In keeping with the pursuit of sustainability, there are many human variables that could be explored that do not require much additional resources to collect, such as could be found in census data. Incorporating temporal lags and spatial population patterns and migration would better inform the relationship of exposure, transmission, and reporting of disease. Stratifying the associations by climate zones remains relevant as climate zones may serve as proxies of unstudied agricultural activities and socio-economic events that are closely linked to the environment.

Use of the full spatiotemporal capabilities of surveillance and publically available RS data is another target of opportunity. The GHS national surveillance data records individual information such as gender and age, which can be used to correct for reporting bias among age groups through comparison with aggregated health surveys. Reported cases are also sometimes available at the weekly level, which could be used in a more complete analysis of the effect of temporal scale on prediction and uncertainty. In terms of exploring spatial scales, if health clinic outcome

data can be extracted from the national surveillance system then publically available RS data capable of representing much smaller spatial scales can be potentially more fully utilized (Figure 15). This study was restricted to the district level, but the MODIS satellite sensor can achieve a 250km spatial resolution. Other publically available satellites such as the Landsat series can yield a 30m (~100ft) spatial resolution and has data going back 40 years, 20m (~66ft) data is available using the Sentinel satellite launched in 2015. This wealth of data sources and multiple spatiotemporal scales has great potential if combined with national surveillance and access to the needed computational resources.

***Figure 15**: Hierarchal structure of the GHS national surveillance system, DHIMS2*



5. Conclusions

Our analysis demonstrates that the numbers of reported schistosomiasis cases in Ghana are declining, which could be an artifact of the data capture gaps in the surveillance system or an actual decrease most likely attributed to successful deworming campaigns. Only the environmental parameters demonstrated consistent seasonality. The interpretation of the general lack of seasonality in the surveillance data is complicated by the lag between transmission and disease reporting and remains difficult to assess. Temporal trends of schistosomiasis as well as for other climate-sensitive diseases contained within the Ghanaian surveillance system require further analysis that should be performed. The importance of improving the surveillance of neglected tropical diseases in low-income countries should not be underestimated. National surveillance systems play a significant role in ensuring availability of vital health data; our analysis demonstrates its utility as a decision support tool and supports the benefits of local and national governments investing in data quality improvements. Predictive modelling should be an iterative process that undergoes progressive improvements in its methodology and data inputs (Kabore et al. 2013). The future steps in improving the predictive capacity of surveillance data for diseases like schistosomiasis are to focus on the location and timing of disease exposure, transmission, and reporting. Additionally, more work should be done in complementing surveillance data with field survey data and novel data streams to offer reliable and cost-effective disease monitoring tools.

6. References

AAAS. 2016. “High-Resolution Satellite Imagery Ordering and Analysis Handbook | AAAS - The World’s Largest General Scientific Society.” *American Association for the Advancement of Science*. https://www.aaas.org/page/high-resolution-satellite-imagery-ordering-and-analysis-handbook.

Abdel-Rahman, M.S., M.M. El-Bahy, J.B. Malone, R.a. Thompson, and N.M. El Bahy. 2001. “Geographic Information Systems as a Tool for Control Program Management for Schistosomiasis in Egypt.” *Acta Tropica* 79 (1): 49–57. doi:10.1016/S0001-706X(01)00102-4.

Aboukhaled, A. 1975. *Research on Crop Use, Salt Affecting Soils and Drainage in the Arab Republic of Egypt: A Review with Recommendations*.

Adokiya, Martin Nyaaba, John Koku Awoonor-Williams, Inuwa Yau Barau, Claudia Beiersmann, and Olaf Mueller. 2015. “Evaluation of the Integrated Disease Surveillance and Response System for Infectious Diseases Control in Northern Ghana.” *BMC Public Health* 15. doi:10.1186/s12889-015-1397-y.

Alexander Liss, Magaly Koch, Elena N. Naumova. n.d. “Redefining Climate Regions in the United States Using Remote Sensing and Machine Learning: Applications to the Public Health and Vulnerability Assessment.” Boston, MA.

Appleton, C.C. 1977. “The Influence of Temperature on the Life-Cycle and Distribution of Biomphalaria Pfeifferi (Krauss, 1948) in South-Eastern Africa.” *International Journal for Parasitology* 7 (5): 335–45. doi:10.1016/0020-7519(77)90057-1.

Appleton, C C. 1978. “Review of the Literature on Abiotic Factors Influencing the Distribution and Life-Cycles of Bilharziasis Intermediate Host Snails.” *Malacol Rev* 11.

Baatiema, Leonard, Anthony Mwinkaara Sumah, Prosper Naazumah Tang, and John Kuumuori Ganle. 2016. “Community Health Workers in Ghana: The Need for Greater Policy Attention.” *BMJ Global Health* 1 (4). http://gh.bmj.com/content/1/4/e000141.

Bavia, M E, L F Hale, J B Malone, D H Braud, and S M Shane. 1999. “Geographic Information Systems and the Environmental Risk of Schistosomiasis in Bahia, Brazil.” *The American Journal of Tropical Medicine and Hygiene* 60 (4). The American Society of Tropical Medicine and Hygiene: 566–72. doi:10.4269/AJTMH.1999.60.566.

Beck-Wörner, C, G Raso, P Vounatsou, E K N’Goran, G Rigo, and E Parlow. 2007. “Bayesian Spatial Risk Prediction of Schistosoma Mansoni Infection in Western C{ô}te d’Ivoire Using a Remotely-Sensed Digital Elevation Model.” *Am J Trop Med Hyg* 76.

Brodish, P. H., and K. Singh. 2016. “Association Between Schistosoma Haematobium Exposure and Human Immunodeficiency Virus Infection Among Females in Mozambique.” *American Journal of Tropical Medicine and Hygiene* 94 (5): 1040–44. doi:10.4269/ajtmh.15-0652.

Brooker, Simon J. 2002a. “Schistosomes, Snails and Satellites.” *Acta Trop* 82.

———. 2002b. “Schistosomes, Snails and Satellites.” *Acta Tropica* 82 (2): 207–14. http://www.ncbi.nlm.nih.gov/pubmed/12020894.

Brooker, Simon J, M Booth, and H Guyatt. 1999. “Comparisons of Schistosome and Geohelminth Infection Prevalences in School-Aged Children from Selected Areas of Africa: Implications for Rapid Assessment and Combined Control.” *Transactions of the Royal Society of Tropical Medicine and Hygiene* 93 (2): 125–26. http://www.ncbi.nlm.nih.gov/pubmed/10450431.

Brooker, Simon J, Simon I. Hay, Wahab Issae, Andrew Hall, Charles M. Kihamia, Nicholas J. S. Lwambo, William Wint, David J. Rogers, and Don A. P. Bundy. 2001. “Predicting the Distribution of Urinary Schistosomiasis in Tanzania Using Satellite Sensor Data.” *Tropical Medicine and International Health* 6 (12): 998–1007. doi:10.1046/j.1365-3156.2001.00798.x.

Brooker, Simon J, Simon I. Hay, Louis-Albert Tchuem Tchuente, and Raoult C. Ratard. 2002. “Using NOAA-AVHRR Data to Model Human Helminth Distributions in Planning Disease Control in Cameroon, West Africa.” *Photogrammetric Engineering and Remote Sensing* 68 (2): 175–79.

Brooker, Simon J, Peter J. Hotez, Donald A. P. Bundy, S Mariotti, D Mabey, and A Foster. 2010. “The Global Atlas of Helminth Infection: Mapping the Way Forward in Neglected Tropical Disease Control.” Edited by Serap Aksoy. *PLoS Neglected Tropical Diseases* 4 (7). International Trachoma Initiative: e779. doi:10.1371/journal.pntd.0000779.

Brooker, Simon J, and E Michael. 2000. “The Potential of Geographical Information Systems and Remote Sensing in the Epidemiology and Control of Human Helminth Infections.” *Advances in Parasitology* 47 (January): 245–88. http://www.ncbi.nlm.nih.gov/pubmed/10997209.

Brooker, Simon J, M. Rowlands, L. Haller, L. Savioli, and D.A.P. Bundy. 2000a. “Towards an Atlas of Human Helminth Infection in Sub-Saharan Africa: The Use of Geographical Information Systems (GIS).” *Parasitology Today* 16 (7): 303–7. doi:10.1016/S0169-4758(00)01687-2.

———. 2000b. “Towards an Atlas of Human Helminth Infection in Sub-Saharan Africa: The Use of Geographical Information Systems (GIS).” *Parasitology Today* 16 (7): 303–7. doi:10.1016/S0169-4758(00)01687-2.

Brown, David. 1994. *Freshwater Snails Of Africa And Their Medical Importance eBook: David S. Brown: Kindle Store*. 2nded.

CDC. 2012. “CDC - Schistosomiasis - Biology.” *Global Health - Division of Parasitic Diseases*. https://www.cdc.gov/parasites/schistosomiasis/biology.html.

Chitsulo, L, D Engels, a Montresor, and L Savioli. 2000. “The Global Status of Schistosomiasis and Its Control.” *Acta Tropica* 77 (1): 41–51. http://www.ncbi.nlm.nih.gov/pubmed/10996119.

CIESIN. 2016. “CIESIN Teams with Facebook to Develop Open, Improved Settlement Data.” *Columbia University*. http://blogs.ei.columbia.edu/2016/11/29/ciesin-teams-with-facebook-to-develop-open-improved-settlement-data/.

Clements, A C A, R Moyeed, and Simon J Brooker. 2006. “Bayesian Geostatistical Prediction of the Intensity of Infection with Schistosoma Mansoni in East Africa.” *Parasitology* 133 (Pt 6): 711–19. doi:10.1017/S0031182006001181.

Clennon, Julie a, Charles H King, Eric M Muchiri, H Curtis Kariuki, John H Ouma, Peter Mungai, and Uriel Kitron. 2004. “Spatial Patterns of Urinary Schistosomiasis Infection in a Highly Endemic Area of Coastal Kenya.” *The American Journal of Tropical Medicine and Hygiene* 70 (4): 443–48. http://www.ncbi.nlm.nih.gov/pubmed/15100462.

Clennon, Julie a, Peter L Mungai, Eric M Muchiri, Charles H King, and Uriel Kitron. 2006. “Spatial and Temporal Variations in Local Transmission of Schistosoma Haematobium in Msambweni, Kenya.” *The American Journal of Tropical Medicine and Hygiene* 75 (6): 1034–41. http://www.ncbi.nlm.nih.gov/pubmed/17172362.

Cline, B., Frank 0. Richards, M. A. El Alamy, S. El Hak, and David F. McNeeley Ernesto Ruiz-Tiben, Janet M. Hughes. 1989. “1983 Nile Delta Schistosomiasis Survey: 48 Years after Scott.” *American Journal of Tropical Medicine and Hygiene* 41 (1): 56–62.

Danso-Appiah, a, S J De Vlas, K M Bosompem, and J D F Habbema. 2004. “Determinants of Health-Seeking Behaviour for Schistosomiasis-Related Symptoms in the Context of Integrating Schistosomiasis Control within the Regular Health Services in Ghana.” *Tropical Medicine & International Health : TM & IH* 9 (7): 784–94. doi:10.1111/j.1365-3156.2004.01267.x.

Dollery, C.T. 1999. “Drug Discovery and Development in the Molecular Era.” *British Journal of Clinical Pharmacology* 47 (1): 5–6.

Elith, Jane, Catherine H Graham, Robert P Anderson, Miroslav Dudík, Simon Ferrier, Antoine Guisan, Robert J Hijmans, et al. 2017. “Novel Methods Improve Prediction of Species’ Distributions from Occurrence Data.” Accessed April 2. http://rob.schapire.net/papers/nceas.pdf.

EMEA. 1996. “Praziquantel: Summary Report (1).” *EMEA/MRL* 141 (96): 3. http://www.ema.europa.eu/docs/en_GB/document_library/Maximum_Residue_Limits_-_Report/2009/11/WC500015784.pdf.

Enuameh, Yeetey Akpe Kwesi, Sumiyo Okawa, Kwaku Poku Asante, Kimiyo Kikuchi, Emmanuel Mahama, Evelyn Ansah, Charlotte Tawiah, et al. 2016. “Factors Influencing Health Facility Delivery in Predominantly Rural Communities across the Three Ecological Zones in Ghana: A Cross-Sectional Study.” Edited by Massimo Ciccozzi. *PLOS ONE* 11 (3). Ghana Statistical Service: e0152235. doi:10.1371/journal.pone.0152235.

Frenken, Karen. 2005. “Irrigation in Africa in Figures: AQUASTAT Survey – 2005.” *FAO Water Reports*. Rome.

———. 2006. “Spatially Explicit Schistosoma Infection Risk in Eastern Africa Using Bayesian Geostatistical Modelling.” Edited by Ulrike Gertrud Munderloh. *Acta Tropica* 9 (1). Paris: Elsevier B.V.: 1–8. doi:10.1016/S0140-6736(03)14968-9.

Frohberg, H., and M. Schulze Schencking. 1981. “Toxicological Profile for Praziquantel, a New Drug against Cestode and Schistosome Infections, as Compared to Some Other Schistosomicides.” *Drug Research* 31 (3a): 555–65.

GAHI. 2017. “Data | Global Atlas of Helminth Infections.” *London Applied & Spatial Epidemiology Reseach Group (LASER)*. http://www.thiswormyworld.org/data-download.

GHS. 2012. “Rolling out a Nationwide Web-Based District Health Information System, DHIMS2- The Ghana Experience.” %0Adhis2.org/doc/snapshot/en/implementer/dhis2.

GNTD. 2011. “Global Neglected Tropical Disease (GNTD) Database.” *Swiss TPH*. http://www.gntd.org/index.html.

Greenwood, David. 2008. “Progress Against Parasites: Praziquantel.” In *Antimicrobial Drugs: Chronicale of a*

Twentieth Century Medical Triumph, 1sted., 368. Oxford University Press.
Grosse, Scott. 1993. "Schistosomiasis and Water Resources Development: A Re-Evaluation of an Important Environment-Health Linkage." Ann Arbor, Michigan. http://ageconsearch.umn.edu/bitstream/11881/1/twp2.pdf.
Gryseels, Bruno, Katja Polman, Jan Clerinx, and Luc Kestens. 2006. "Human Schistosomiasis." *Lancet* 368 (9541): 1106–18. doi:10.1016/S0140-6736(06)69440-3.
Hürlimann, Eveline, Nadine Schur, Konstantina Boutsika, Anna-Sofie Stensgaard, Maiti Laserna de Himpsl, Kathrin Ziegelbauer, Nassor Laizer, et al. 2011. "Toward an Open-Access Global Database for Mapping, Control, and Surveillance of Neglected Tropical Diseases." *PLoS Neglected Tropical Diseases* 5 (12). Public Library of Science: e1404. doi:10.1371/journal.pntd.0001404.
Jagai, Jyotsna S., Rajiv Sarkar, Denise Castronovo, Deepthi Kattula, Jesse McEntee, Honorine Ward, Gagandeep Kang, and Elena N. Naumova. 2012. "Seasonality of Rotavirus in South Asia: A Meta-Analysis Approach Assessing Associations with Temperature, Precipitation, and Vegetation Index." Edited by Cécile Viboud. *PLoS ONE* 7 (5). Birkhauser: e38168. doi:10.1371/journal.pone.0038168.
Kabore, Achille, Nana-Kwadwo Biritwum, Philip W. Downs, Ricardo J. Soares Magalhaes, Yaobi Zhang, and Eric A. Ottesen. 2013. "Predictive vs. Empiric Assessment of Schistosomiasis: Implications for Treatment Projections in Ghana." Edited by Simon Brooker. *PLoS Neglected Tropical Diseases* 7 (3). Public Library of Science: e2051. doi:10.1371/journal.pntd.0002051.
Kalluri, Satya, Peter Gilruth, David Rogers, Martha Szczur, JP Gonzalez, and M Boussinesq. 2007. "Surveillance of Arthropod Vector-Borne Infectious Diseases Using Remote Sensing Techniques: A Review." *PLoS Pathogens* 3 (10). UNDP/World Bank/WHO Special Programme for Research and Training in Tropical Diseases (TDR): e116. doi:10.1371/journal.ppat.0030116.
Kim, Sharon. 2014. "One Million Community Health Workers | Ghana Completes the 1st CHW Roadmap With the 1mCHW Campaign." http://1millionhealthworkers.org/2014/04/23/ghana-completes-the-1st-chw-roadmap-with-the-1mchw-campaign-2/.
King, Charles H. 2010. "2010 - King - Parasites and Poverty." *Acta Tropica* 113: 95–104.
Kock, K N De, and J A Van Eeden. 1986. "Effect of Programmed Circadian Temperature Fluctuations on Population Dynamics of Biomphalaria Pfeifferi (Krauss)." *South African Journal of Zoology S. Afr. J. Zool* 2121 (1): 28–32. doi:10.1080/02541858.1986.11447952.
Kock, KN De, CT Wolmarans, and M Bornman. 2004. "Distribution and Habitats of Biomphalaria Pfeifferi, Snail Intermediate Host of Schistosoma Mansoni, in South Africa." *Water SA* 30 (1). Water Research Commission (WRC): 29–36. doi:10.4314/wsa.v30i1.5023.
Kottek, M., J. Grieser, C. Beck, B. Rudolf, and F. Rubel. 2006. "World Map of the Köppen-Geiger Climate Classification Updated. Meteorologische Zeitschrift." *Meteorologische Zeitschrift*. http://www.srh.noaa.gov/jetstream/global/images/Koppen-Geiger.pdf.
Kulinkina, Alexandra V. 2017. "Community Based Methods for Schistosomiasis Prediction and Sustainable Control in Ghana." Tufts University.
Kulinkina, Alexandra V., Karen C. Kosinski, Jeanine D. Plummer, John L. Durant, Kwabena M. Bosompem, Michael N. Adjei, Jeffrey K. Griffiths, David M. Gute, and Elena N. Naumova. 2017. "Indicators of Improved Water Access in the Context of Schistosomiasis Transmission in Rural Eastern Region, Ghana." *Science of The Total Environment* 579 (February): 1745–55. doi:10.1016/j.scitotenv.2016.11.140.
Kulinkina, Alexandra V., Venkat R. Mohan, Mark R. Francis, Deepthi Kattula, Rajiv Sarkar, Jeanine D. Plummer, Honorine Ward, Gagandeep Kang, Vinohar Balraj, and Elena N. Naumova. 2016. "Seasonality of Water Quality and Diarrheal Disease Counts in Urban and Rural Settings in South India." *Scientific Reports* 6 (1). Nature Publishing Group: 20521. doi:10.1038/srep20521.
Kuuire, Vincent Z., Elijah Bisung, Andrea Rishworth, Jenna Dixon, and Isaac Luginaah. 2015. "Health-Seeking Behaviour during Times of Illness: A Study among Adults in a Resource Poor Setting in Ghana." *Journal of Public Health* 50 ((12)). Oxford University Press: fdv176. doi:10.1093/pubmed/fdv176.
Lai, Ying-Si, Patricia Biedermann, Uwem F Ekpo, Amadou Garba, Els Mathieu, Nicholas Midzi, Pauline Mwinzi, et al. 2015. "Spatial Distribution of Schistosomiasis and Treatment Needs in Sub-Saharan Africa: A Systematic Review and Geostatistical Analysis." *The Lancet Infectious Diseases* 15 (8): 927–40. doi:10.1016/S1473-3099(15)00066-3.
London, Imperial College. 2016. "Schistosomiasis Control Initiative."
Malone, J. B., O. K. Huh, D. P. Fehler, P. A. Wilson, D. E. Wilensky, R. A. Holmes, and A. I. Elmagdoub. 1994. "Temperature Data from Satellite Imagery and the Distribution of Schistosomiasis in Egypt." *American Journal of Tropical Medicine and Hygiene*.

Malone, J.B., N.R. Bergquist, O.K. Huh, M.E. Bavia, M. Bernardi, M.M. El Bahy, M.V. Fuentes, et al. 2001. "A Global Network for the Control of Snail-Borne Disease Using Satellite Surveillance and Geographic Information Systems." *Acta Tropica* 79 (1): 7–12. doi:10.1016/S0001-706X(01)00098-5.

Malone, J.B., J.M. Yilma, J.C. McCarroll, B. Erko, S. Mukaratirwa, and Xinyu Zhou. 2001. "Satellite Climatology and the Environmental Risk of Schistosoma Mansoni in Ethiopia and East Africa." *Acta Tropica* 79 (1): 59–72. doi:10.1016/S0001-706X(01)00103-6.

Malone, J B, M S Abdel-Rahman, M M El Bahy, O K Huh, M Shafik, and M Bavia. 1997. "Geographic Information Systems and the Distribution of Schistosoma Mansoni in the Nile Delta." *Parasitology Today (Personal Ed.)* 13 (3): 112–19. http://www.ncbi.nlm.nih.gov/pubmed/15275115.

Malone, JB, R Gommes, J Hansen, and JM Yilma. 1998. "A Geographic Information System on the Potential Distribution and Abundance of Fasciola Hepatica and F. Gigantica in East Africa Based on Food and Agriculture." *Veterinary*. http://www.sciencedirect.com/science/article/pii/S030440179800137X.

Mandahl-Barth, G. 1965. "The Species of the Genus Bulinus, Intermediate Hosts of Schistosoma." *Bulletin of the World Health Organization* 33 (1): 33–44. http://www.ncbi.nlm.nih.gov/pubmed/5294263.

McCullough, F.S., and Y.M. Ali. 1965. "The Distribution and Prevalence of Schistosoma Haematobium and Schistosoma Mansoni in Ghana." *Ghana Medical Journal: The Journal of the Ghana Medical Association* 4 (4): 83–84.

McCullough, F S. 1965. "Annual Reports of the Medical Field Units: A Note on Intestinal Schistosomiasis and the Snail Hosts in Ghana."

Michelson, M K, F A Azziz, F M Gamil, A A Wahid, F O Richards, D D Juranek, M A Habib, and H C Spencer. 1993. "Recent Trends in the Prevalence and Distribution of Schistosomiasis in the Nile Delta Region." *The American Journal of Tropical Medicine and Hygiene* 49 (1): 76–87. http://www.ncbi.nlm.nih.gov/pubmed/8352395.

Mkize, Lwamkelekile Sitshilelo, Samson Mukaratirwa, and Oliver Tendayi Zishiri. 2016. "Population Genetic Structure of the Freshwater Snail, Bulinus Globosus, (Gastropoda: Planorbidae) from Selected Habitats of KwaZulu-Natal, South Africa." *Acta Tropica* 161: 91–99. doi:10.1016/j.actatropica.2016.06.001.

Moodley, I, I Kleinschmidt, B Sharp, M Craig, and C Appleton. 2003. "2003 - Moodley - Temperature-Suitability Maps for Schistosomiasis in South Africa." *Annals of Tropical Medicine andParisitology* 97 (6): 617–27.

Morgenstern, Hal. 1998. "Ecologic Study." In *Encyclopedia of Biostatistics*. Chichester, UK: John Wiley & Sons, Ltd. doi:10.1002/0470011815.b2a03055.

Naumova, E N, J S Jagai, B Matyas, A DeMaria, I B MacNeill, J K Griffiths, and J. K. GRIFFITHS. 2007. "Seasonality in Six Enterically Transmitted Diseases and Ambient Temperature." *Epidemiology and Infection* 135 (2). Cambridge University Press: 281–92. doi:10.1017/S0950268806006698.

Ndeffo Mbah, Martial L., Eric M. Poolman, Paul K. Drain, Megan P. Coffee, Marieke J. van der Werf, and Alison P. Galvani. 2013. "HIV and *Schistosoma Haematobium* Prevalences Correlate in Sub-Saharan Africa." *Tropical Medicine & International Health* 18 (10): 1174–79. doi:10.1111/tmi.12165.

Norman R. Stoll. 1947. "This Wormy World." *The Journal of Parisitology* 33 (1): 1–18. http://www.jstor.org/stable/3273613.

Nsubuga, Peter, Okey Nwanyanwu, John N Nkengasong, David Mukanga, and Murray Trostle. 2010. "Strengthening Public Health Surveillance and Response Using the Health Systems Strengthening Agenda in Developing Countries." *BMC Public Health* 10 Suppl 1 (Suppl 1). BioMed Central: S5. doi:10.1186/1471-2458-10-S1-S5.

Odei, M. A. 1964. "Schistosomiasis in Ghana: Present Knowledge and Future Prospects." In *First International Congress of Parasitology*.

Onori, E., F S McCullough, and L. Rosei. 1963. "Schistosomiasis in Ghana: Present Knowledge and Future Prospects." *Annals of Tropical Medicine andParisitology*, 57–59.

PCD. 1999. "The Cost of Large-Scale School Health Programmes Which Deliver Anthelmintics to Children in Ghana and Tanzania. The Partnership for Child Development." *Acta Tropica* 73 (2): 183–204. http://www.ncbi.nlm.nih.gov/pubmed/10465058.

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.

Pflüger, W, M Z Roushdy, and M Emam. 1984. "The Prepatent Period and Cercarial Production of Schistosoma Haematobium in Bulinus Truncatus (Egyptian Field Strains) at Different Constant Temperatures." *Z Parasitenkd* 70.

Pitchford, R.J. 1981. *South African Journal of Science. South African Journal of Science*. Vol. 77. [South African

Association for the Advancement of Science]. https://journals.co.za/content/sajsci/77/6/AJA00382353_1714.
Pitchford, R J, A H Meyling, J Meyling, and J F Toit. 1969. “Cercarial Shedding Patterns of Various Schistosome Species under Outdoor Conditions in the Transvaal.” *Ann Trop Med Parasitol* 63.
Pullan, Rachel L., Jeffrey M. Bethony, Stefan M. Geiger, Rodrigo Correa-Oliveira, Simon J Brooker, and Rupert J. Quinnell. 2010. “Human Helminth Co-Infection: No Evidence of Common Genetic Control of Hookworm and Schistosoma Mansoni Infection Intensity in a Brazilian Community.” *International Journal for Parasitology* 40 (3): 299–306. doi:10.1016/j.ijpara.2009.08.002.
Raso, G., B. Matthys, E. K. N’Goran, M. Tanner, P. Vounatsou, and J. Utzinger. 2005. “Spatial Risk Prediction and Mapping of Schistosoma Mansoni Infections among Schoolchildren Living in Western Côte d’Ivoire.” *Parasitology* 131 (1): 97–108. doi:10.1017/S0031182005007432.
Raso, G, B Matthys, E K N’Goran, M Tanner, P Vounatsou, and J Utzinger. 2005. “Spatial Risk Prediction and Mapping of Schistosoma Mansoni Infections among Schoolchildren Living in Western C{ô}te d’Ivoire.” *Parasitology* 131.
Raso, G, P Vounatsou, D P McManus, E K N’Goran, and J Utzinger. 2007. “A Bayesian Approach to Estimate the Age-Specific Prevalence of Schistosoma Mansoni and Implications for Schistosomiasis Control.” *Int J Parasitol* 37.
Raso, G, P Vounatsou, B H Singer, E K N’Goran, M Tanner, and J Utzinger. 2006. “An Integrated Approach for Risk Profiling and Spatial Prediction of Schistosoma Mansoni-Hookworm Coinfection.” *Proc Natl Acad Sci U S A* 103.
Raso, Giovanna, Penelope Vounatsou, Burton H Singer, Eliézer K N’Goran, Marcel Tanner, and Jürg Utzinger. 2006. “An Integrated Approach for Risk Profiling and Spatial Prediction of Schistosoma Mansoni-Hookworm Coinfection.” *Proceedings of the National Academy of Sciences of the United States of America* 103 (18): 6934–39. doi:10.1073/pnas.0601559103.
Rollinson, David, Stefanie Knopp, Sarah Levitz, J Russell Stothard, Louis-Albert Tchuem Tchuenté, Amadou Garba, Khalfan A Mohammed, et al. 2013. “Time to Set the Agenda for Schistosomiasis Elimination.” *Acta Tropica* 128 (2): 423–40. doi:10.1016/j.actatropica.2012.04.013.
Rollison, D., J.R. Stothard, and V.R. Southgate. 2001. “Interactions between Intermediate Snail Hosts of the Genus Bulinus and Schistosomes of the Schistosoma Haematobium Group.” *Parasitology* 123 (7). Cambridge University Press: 245–60. doi:10.1017/S0031182001008046.
Scholte, Ronaldo G C, Laura Gosoniu, John B. Malone, Frederique Chammartin, Jurg Utzinger, and Penelope Vounatsou. 2014. “Predictive Risk Mapping of Schistosomiasis in Brazil Using Bayesian Geostatistical Models.” *Acta Tropica* 132: 57–63. doi:10.1016/j.actatropica.2013.12.007.
Schur, Nadine, Jürg Utzinger, and Penelope Vounatsou. 2011. “Modelling Age-Heterogeneous Schistosoma Haematobium and S. Mansoni Survey Data via Alignment Factors.” *Parasites & Vectors* 4 (January): 142. doi:10.1186/1756-3305-4-142.
Schur, N, E Hürlimann, A S Stensgaard, K Chimfwembe, G Mushinge, and C Simoonga. 2013. “Spatially Explicit Schistosoma Infection Risk in Eastern Africa Using Bayesian Geostatistical Modelling.” *Acta Trop* 128.
Scott, J. Allen. 1937. “The Incidence and Distribution of the Human Schistosomes in Egypt.” *American Journal of Epidemiology* 25 (3): 566–614.
Service, Ghana Statistical. 2013. “2010 Population and Housing Census: National Analytical Report.” Awusabo-Asare, Ghana.
Seubert, J., R. Pohlke, and F. Loebich. 1977. “Synthesis and Properties of Praziquantel, a Novel Broad Spectrum Anthelmintic with Excellent Activity against Schistosomes and Cestodes.” *Springer* 15 (8): 1036–37. http://download.springer.com/static/pdf/831/art%253A10.1007%252FBF01945954.pdf?auth66=1386376790_762941092c56cc8012ddcd66c0a882df&ext=.pdf.
Shiff, C J, and B Garnett. 1967. “The Influence of Temperature on the Intrinsic Rate of Natural Increase of the Freshwater Snail B. Pfeifferi.” *Arch Hydrobiol* 62.
Simoonga, C, L N Kazembe, T K Kristensen, A Olsen, C C Appleton, and P Mubita. 2008. “The Epidemiology and Small-Scale Spatial Heterogeneity of Urinary Schistosomiasis in Lusaka Province, Zambia.” *Geospat Health* 3.
Simoonga, C, J Utzinger, Simon J Brooker, P Vounatsou, C C Appleton, and A S Stensgaard. 2009. “Remote Sensing, Geographical Information System and Spatial Analysis for Schistosomiasis Epidemiology and Ecology in Africa.” *Parasitology* 136.
Simoonga, C, J Utzinger, Simon J Brooker, P Vounatsou, C C Appleton, A S Stensgaard, A Olsen, and T K Kristensen. 2009. “Remote Sensing, Geographical Information System and Spatial Analysis for Schistosomiasis Epidemiology and Ecology in Africa.” *Parasitology* 136 (13): 1683–93.

doi:10.1017/S0031182009006222.
Soares Magalhães, Ricardo J, Nana-Kwadwo Biritwum, John O Gyapong, Simon J Brooker, Yaobi Zhang, Lynsey Blair, Alan Fenwick, and Archie C A Clements. 2011. "Mapping Helminth Co-Infection and Co-Intensity: Geostatistical Prediction in Ghana." Edited by María-Gloria Basáñez. *PLoS Neglected Tropical Diseases* 5 (6). Public Library of Science: e1200. doi:10.1371/journal.pntd.0001200.
Steinmann, Peter, Jennifer Keiser, Robert Bos, Marcel Tanner, and Jürg Utzinger. 2006. "Schistosomiasis and Water Resources Development: Systematic Review, Meta-Analysis, and Estimates of People at Risk." *The Lancet Infectious Diseases* 6 (7): 411–25. doi:10.1016/S1473-3099(06)70521-7.
Stensgaard, A, A Jørgensen, N B Kabatereine, J B Malone, and T K Kristensen. 2005. "Modeling the Distribution of Schistosoma Mansoni and Host Snails in Uganda Using Satellite Sensor Data and Geographical Information Systems." *Parassitologia* 47 (1): 115–25.
Thompson, D F, J B Malone, M Harb, R Faris, O K Huh, A A Buck, and B L Cline. 1996. "Bancroftian Filariasis Distribution and Diurnal Temperature Differences in the Southern Nile Delta." *Emerging Infectious Diseases* 2 (3). Centers for Disease Control and Prevention: 234–35. doi:10.3201/eid0203.960313.
Thomson, M., S. Connor, K. O'Neill, and J-P. Meert. 2000. "Environmental Information for Prediction of Epidemics." *Parasitology Today* 16 (4): 137–38. doi:10.1016/S0169-4758(00)01648-3.
Utzinger, J, G Raso, Simon J Brooker, D De Savigny, M Tanner, N Ornbjerg, B H Singer, and E K N'goran. 2009. "Schistosomiasis and Neglected Tropical Diseases: Towards Integrated and Sustainable Control and a Word of Caution." *Parasitology* 136: 1859–74. doi:10.1017/S0031182009991600.
Utzinger, J, G Raso, Simon J Brooker, D Savigny, M Tanner, and N Ørnbjerg. 2009. "Schistosomiasis and Neglected Tropical Diseases: Towards Integrated and Sustainable Control and a Word of Caution." *Parasitology* 136.
Utzinger, Jürg, Robert Bergquist, Xiao Shu-Hua, Burton H Singer, and Marcel Tanner. 2003. "Sustainable Schistosomiasis Control--the Way Forward." *Lancet* 362 (9399): 1932–34. doi:10.1016/S0140-6736(03)14968-9.
Venables, W. N. (William N.), and Brian D. Ripley. 1999. *Modern Applied Statistics with S-PLUS*. Springer.
Vounatsou, P., G. Raso, M. Tanner, E. K. N'Goran, and J. Utzinger. 2009. "Bayesian Geostatistical Modelling for Mapping Schistosomiasis Transmission." *Parasitology* 136 (13): 1695. doi:10.1017/S003118200900599X.
Walz, Yvonne, Martin Wegmann, Stefan Dech, Giovanna Raso, Jürg Utzinger, and A Garba. 2015. "Risk Profiling of Schistosomiasis Using Remote Sensing: Approaches, Challenges and Outlook." *Parasites & Vectors* 8 (1). BioMed Central: 163. doi:10.1186/s13071-015-0732-6.
Walz, Yvonne, Martin Wegmann, Benjamin Leutner, Stefan Dech, Penelope Vounatsou, Eliézer K. N'Goran, Giovanna Raso, and Jürg Utzinger. 2015. "Use of an Ecologically Relevant Modelling Approach to Improve Remote Sensing-Based Schistosomiasis Risk Profiling." *Geospatial Health* 10 (2). doi:10.4081/gh.2015.398.
WHO. 1981. "The Schistosomiasis Problem in the World: Results of a WHO Questionnaire Survey." *Bulletin of the World Health Organization* 59 (1): 115–27. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2396036/pdf/bullwho00418-0122.pdf.
———. 1987. "Global Atlas Results for Ghana."
———. 2010. "WHO Country Profile : Ghana, Maps of Infection/disease Endemicity." http://www.who.int/neglected_diseases/preventive_chemotherapy/databank/CP_Ghana.pdf?ua=1.
———. 2011. "CDC - Neglected Tropical Diseases - The Burden of Schistosomiasis." *Centers for Disease Control and Prevention*. https://www.cdc.gov/globalhealth/ntd/diseases/schisto_burden.html.
———. 2012. "Distribution of Schistosomiasis, Worldwide, 2011." *Control of Neglected Tropical Diseases (NTD)*. http://gamapserver.who.int/mapLibrary/Files/Maps/Schistosomiasis_2011_global.png.
———. 2017. "WHO Media Center | Schistosomiasis Fact Sheet." *WHO*. World Health Organization. http://www.who.int/mediacentre/factsheets/fs115/en/.
WHO-AFRO; CDC. 2010. "Integrated Disease Surveillance and Response in the African Region." 2. Brazzaville, Republic of Congo. https://www.cdc.gov/globalhealth/dphswd/idsr/pdf/Technical Guidelines/IDSR Technical Guidelines 2nd Edition_2010_English.pdf.
Woolhouse, M E J. 1998. "Patterns in Parasite Epidemiology: The Peak Shift." *Parasitol Today* 14.
Wright, W H. 1972. "A Consideration of the Economic Impact of Schistosomiasis*." *Bull. Org. Mond. Sante Bull. Wld Hlth Org* 47: 559–66. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2480844/pdf/bullwho00185-0018.pdf.
Yehia, Shaker, Samy Nervana, and Ashour Esmat. 2014. "Hepatobiliary Schistosomiasis." *Journal of Clinical and Translational Hepatology* 2 (3). Xia & He Publishing Inc: 212–16. doi:10.14218/JCTH.2014.00018.
Zinyama-Gutsire, Rutendo B. L., Charles Chasela, Hans O. Madsen, Simbarashe Rusakaniko, Per Kallestrup,

Michael Christiansen, Exnevia Gomo, et al. 2015. “Role of Mannose-Binding Lectin Deficiency in HIV-1 and Schistosoma Infections in a Rural Adult Population in Zimbabwe.” Edited by Cordula M. Stover. *PLOS ONE* 10 (4): e0122659. doi:10.1371/journal.pone.0122659.