Bayesian spatio-temporal modelling for malaria surveillance and residual pockets of transmission identification in Swaziland

Inauguraldissertation

zur

Erlangung der Würde eins Doktors der Philosophie

vorgelegt der

Philosophisch-Naturwissenschaftlichen Fakultät der

Universität Basel

von

Sabelo Nick Dlamini

aus Ezulwini, Swaziland

Basel, 2018

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät auf Antrag von Prof. Dr. Jürg Utzinger, PD Dr. Penelope Vounatsou and Dr. Michael Gebreslasie.

Basel, den 15, November 2016

Prof. Dr. Jörg Schibler

… to my beloved family and to the young minds who hunger and thirst after wisdom

and

to those who are earnestly seeking knowledge in order to break free from the bonds of ignorance.

Summary

Malaria control has been and is in the world spotlight for over 50 years and had been marked by a proliferation of research studies, as interest in finding new control methods increased. Along the full spectrum from malaria control to prevention of reintroduction emphasis on surveillance and response has been made if gains in the fight against the disease are to be realized. International funding for anti-malaria related activities has also been up-scaled and sustained with a significant amount of it focusing on the high burden countries in the sub-Saharan region of continental Africa. Understanding the complex interactions between malaria vectors, parasites and human hosts is key to control and elimination of the disease. Geostatistical methods involving the use of remote sensing (RS) techniques and geographic information system (GIS) tools have proven to be an effective way of estimating spatial and temporal effects of environmental determinants on disease outcomes. They also allow us to produce model-based maps which could be used to predict the disease at explicit geographic scales thus aiding targeted control.

In the context of surveillance, preparedness and response we explored potential methods and tools that could be used for surveillance by malaria control programmes in very low endemic settings like Swaziland. In this country, malaria has drastically declined and the country is currently in its elimination stage as it entered the critical 3-year phase from 2015 to 2018 where it is anticipated that it will receive certification from the World Health Organization (WHO) as a malaria free country. Spatially explicit maps on micro-epidemiological heterogeneities as well as space and time trends and patterns in malaria transmission are needed to aid the country to target and prioritize interventions in this critical phase as it deals with individual episodic cases. Currently achieving malaria elimination remains operationally challenging due to the ever present threat of imported cases from nearby endemic regions and from uncensored immigration. Also the turnaround time from data collection, processing and use for planning purposes is too long for rapid response actions. Therefore a rapid response surveillance system is needed in order to achieve elimination and prevent reintroduction after elimination.

Chapter 1 presents the overall background informing this study including the rationale for undertaking this PhD work. The role of surveillance in malaria control and elimination as well as the importance of rapid response in malaria elimination were also presented. We showed the progress the country has made from the establishment of the malaria control unit in the

1940s to present time. Our study focused on the use of environmental data for disease surveillance. Therefore we detailed the environmental factors associated with malaria transmission and demonstrated how they were interlinked with disease incidence. Such factors included temperature, precipitation and humidity. The use of earth observation (EO) data derived from RS techniques was also presented. Tools that could be used to support surveillance such as GIS and global positioning systems (GPS) are also discussed. We look at the current malaria situation in Swaziland with emphasis on the latest developments following the scaling up of malaria interventions in that country.

In chapter 2 we emphasised the importance of mapping potential vector breeding sites in Swaziland using high resolution remotely sensed data in conjunction with entomological data to aid larval source management (LSM) strategies. We used larva scooping methods to identify potential breeding sites in the country and those identified were fed into a decision tree induction algorithm and a logistic regression to assess which environmental factors characterised larvae presence or absence. Both approaches reliably distinguished between the two set of scenarios of larvae presence or absence and identified the same environmental predictor related to human activity (subsistence farming) as key determinant of potential vector breeding. Models linking presence of larvae with high resolution land-use variables were found to have good predictive ability. Thus we produced a map of predicted potential breeding sites at explicit geographic scales to assist the malaria control programme in planning its LSM budget.

There are many environmental proxies that have been proposed by ecologists and remote sensing experts which have a potential for use in vector-borne disease mapping. However, their uptake by epidemiologists has remained notoriously slow. Therefore in chapter 3 we investigated the litany of available RS variables that could be used in vector-borne disease mapping studies. We reviewed literature on available sources of remotely sensed data and presented a library of supplier processed variables and those that need to be derived by the end-users and processed at different levels before being incorporated into disease mapping studies. We discussed the reasons and criteria used to select the proxies described and presented in our catalogue. Indices investigated were limited to those related to EO data products with continental or global coverage scales, and were grouped according to meteorology, land use/cover, cartography and urban mapping variables which could be used as proxies for disease suitability mapping. We found numerous indices that have been derived by ecologists and remote sensing experts from the various satellite sensors that have been launched over the years. However, they have remained underutilized in epidemiology partly because of lack of remote sensing skills needed to derive them and partly because they were not high demand variables and therefore not provided by remote sensing agents and suppliers of remotely sensed data.

In chapter 4, we explored different scenarios for malaria incidence risk by investigating the environmental effects of weekly distributed lags in Swaziland. A Bayesian geostatistical model based on polynomial distributed lags function was developed to assess how different environmental and socio economic factors influenced malaria incidence in the country. We then produced model based spatially explicit maps of predicted malaria incidence risk which could be used by the control programme to target their control interventions for high impact.

In chapter 5, we evaluated some of the new and potential indices for epidemiological studies by testing and comparing their use in predicting malaria incidence risk in Swaziland. We discussed the inclusion criteria and choice of the selected variables for malaria incidence risk prediction in the country. This was necessitated by the fact that new satellites have been launched with much improved sensor capabilities than previous first generation sensors. Sensor improvements are noticeable in the number of spectral bands, spatial and temporal resolutions, thus presenting unprecedented good image sources for identification of spatial heterogeneities, trends and patterns in disease mapping by epidemiologists. We ended with emphasising the importance of why this research work was carried out including discussing the key findings and overall message that came from this study. The contributions that had been made by this study are also discussed as well as remaining research work that could be undertaken as follow up.

Acknowledgements

It is inconceivable that time has passed so fast since my first entry into Switzerland. What began as a pessimistic and isolated dream in the forlorn streets of Mbabane gradually grew into optimism and has finally become a reality. There have been many challenges, successes and failures but with sustained trial the probability of success remained high. It is impossible to reach such milestones in one`s life without the help and support of others. Firstly, I would like to thank God for sticking to the original plan He had for my life even when I have had so many distractions along the way. I am deeply indebted to my supervisor PD Dr. Penelope Vounatsou for her energetic effort and patience as she helped to metamorphose me into a statistician. Thank you for the many hours we spent together as we tried to create a masterpiece scientific work. Your enthusiasm was a *sine qua non* to my own dedication to the work at hand. I would also like to expend my cordial gratitude to the incumbent and beloved director of the Swiss Tropical and Public Health Institute (Swiss TPH) Prof. Dr. Jürg Utzinger for always taking the first step in all that was required from me as I registered for my PhD with the Institute. I was always relieved when I saw your signature already in the documents that needed higher level signing even before we met in person. Thank you for believing in me. I also thank Dr. Jonas Franke and Dr. Michael Gebreslasie for their valuable expertise and contribution to my PhD project. In the same vein, I thank the entire MALAREO Consortium members, including Dr. Ides Bauwens, Dr. Julie Deleu, Dr. Inbarani Naidoo, Dr. Clotilde de Montpellier and the University of KwaZulu Natal partners.

To the Swiss TPH administration staff, I say “danke schön” for all your support and assistance with whatever I needed no matter how ridiculous it seemed. I am grateful to Christine Mensch for keeping my student file even before I received the financial support for my studies at the institute. It remained a pledge that I would one day return to the institute as a registered student. Thank you for your optimism and for harnessing the seeds of fate in my favour. “Vielen dank” to Dagmar Batra-Seufert and Laura Innocenti for all the patience and travel assistance during my initial trips to Basel, a city I had never known before. My gratitude also goes to Margrith Slaoui for all the help I received at your office with a smile. I also acknowledge Nora Bauer Ott. “Dankeschön” for the open door at your office and for handling all my emergencies. “Vielen dank” to Doris Stamm for handling all the paperwork on my behalf and to Andrea Hofer for a welcoming smile and for always saying to me I was forever smiling. It kept me motivated. I would also like to express my thanks to the Swiss TPH informatics team for the help with setting up my computer. I would like to thank the beloved head of the then Department of Epidemiology and Public Health (now Education and Training) Prof. Dr. Nino Künzli for levelling the ground for all at the institute and for the friendly open door environment. Here is a place where you can literally walk into any office without second thought. Other academic stalwarts who have directly and indirectly inspired, influenced and motivated me include Dr. Graciana Peter, Prof. Dr. Daniel S. Tevera and Dr. Cliff S. Dlamini.

I would like to also thank my fellow students and "comrades in arms" for all the many hours we spent together both formally and informally. I earnestly acknowledge Dr. Federica Giardina for the willingness to share your knowledge and expertise with me as I stumbled along the world of biostatistics. I also thank Dr. Frederique Chammartin, Dr. Yingsi Lai,Dr. Alex Karagiannis-Voules, Dr. Abbas Adigun, Elizaveta Semenova, Anton Beloconi, Oliver Bärenbold, Betty Nambuusi, Verena Jürgens, Serena Scudella, Isidoros Papaioannou, Christos Kokarialis, Eleni Verykouki and Dr. Guojing Yang for all the beautiful moments we shared together. The same goes to the biostatistics 1 study group which included Dr. Henry Ntuku, Samuel Oppong, Dr. Giovanfrancesco Ferrari, Sammy Khagayi, Emmanuel Ngabire, Dr. Rufin K. Assare, Dr. Koku Awoonor-Williams, Clarisse Houngbedji, Dr. Eveline Hürlimann, and others. Thank you for the discussions we had before the exams, they really helped. My appreciation also goes to Dr. Nerina Vischer and her family for the warm welcome into their home during my first Christmas in Basel. I send also my regards and acknowledgement to Dr. Eric Diboulo for the lengthy discussions we had about everything. Thank you for all the help you forwarded towards me when I needed it the most. My thanks also goes to Dr. Richard Yapi and Dr. Ikenna Enze for the time and jokes we shared together. I would like to also mention Steven Akinola for the assistance you gave me when I first arrived in Basel. Thank you for showing me how to take bus 34 to Bottmingen and then 37 to Bruderholspital. I would also like to appreciate Dr. Fransisco Rogerlàndio Martins-Melo for the friendship and companionship in a land where we were both sojourners. It has been a long road indeed and there were many faces that came my way and it's hard to remember all the names but I would like to acknowledge all those we crossed path with during my stay in Basel. "Vielen Dank" to all of you. I hope we will still run to each other with open arms whenever we meet even if we do not remember each other`s name.

My warmest regards goes to the Crossroads International Church of Basel where I found a safe and rejuvenating haven during my stay in Basel. Thank you for your hospitality and for the warm welcome to the church. I would love to specifically thank my church pastor, David Fortune for the deep teachings on Christian living and also Mrs. Cynthia and Mr. Martin for all the bible study sessions we had at your place. I would also like to thank Mr. Jürg and Mrs. Sharon Richner for the smiles, parental love and guidance. At this juncture I cannot go without mentioning my ever fabulous and super energetic salsa team members led by Mrs. Natividad Gagneux. What a best way to unwind and forget it all on the dance floor!

Support for this work came from many sources which made it possible for the work not only to be implemented but also to be completed. I would like to thank the Federal Commission for Scholarship for Foreign Students for affording me the Swiss Government Excellence Scholarship and for all the support throughout my studies in Basel. I would like to thank the National Malaria Control Programme of Swaziland for your willingness to share your data for the purposes of my study. My deepest appreciation goes to Mr. Simon Kunene for committing to assist me with providing the data even when I had left the Malaria Control Programme as I pursued further study. I would further like to acknowledge Mr. Quinton Dlamini for providing the entomological data I needed for my study. I would like to thank Mr. Sibonakaliso Vilakati, Mr. Sicelo Kunene and Mr. Steven Mthethwa for continually sending up-to-date data for my work. To all the programme staff, I would like to say thank you for all the experience I gained while working with you and for all the support you gave me as I executed my duties. Last but not least, I would like to thank the Swaziland Meteorology Services for providing the weather data for my study. Thanks to Mr. Eric Seyama, Mr. Isaac Manyatsi, Mr. Thulani Mkhatshwa, Mr. Amos Jele and all the meteorology staff for your assistance each time I came knocking on your door looking for data.

To God be the Glory! For His unfailing and steadfast love, and for His multiple tender mercies that endureth forever, without Whom nothing will ever make sense!

Table of contents Page

List of Figures Page

List of Tables   Page

List of Abbreviations

ACTs	Artemisinin-based Combination Therapy
ASTER	Advanced Spaceborne Thermal Emission and Reflection Radiometer
DDT	Dichlorodiphenyltrichloroethane
EA	Enumeration Area
EC	European Commission
EO	Earth Observation
ESA	European Space Agency
FEWS NET	Famine Early Warning System Network
FP7	Seventh Framework Programme
GIS	Geographic Information System
GMES	Global Monitoring for Environment and Security
GPS	Global Positioning System
IRS	Indoor Residual Spraying
ITNs	Insecticide-Treated Nets
LSDI	Lubombo Spatial Development Initiative
LSM	Larval Source Management
LST	Land Surface Temperature
LULC	Land Use Land Cover
MALAREO	Earth Observation in Malaria Vector Control and Management
MIS	Malaria Indicator Survey
MODIS	Moderate Resolution Imaging Spectroradiometer
MSD	Malaria Surveillance Database

NASA	National Aeronautics and Space Administration
NDVI	Normalized Difference Vegetation Index
NMCP	National Malaria Control Programme
PCR	Polymerase Chain Reaction
RDT	Rapid Diagnostic Test
RFE	Rainfall Estimates
RS	Remote Sensing
SADC	Southern African Development Community
SMS	Short Message Service
UN	United Nations
WHO	World Health Organization

Chapter 1

Introduction

1.1 Global malaria burden and progress in Swaziland

According to the World Malaria Report (2015), about 3.2 billion people are at risk of malaria worldwide with an estimated incidence risk of more than one malaria case per 1000 population. In addition, there were 214 million new cases and an estimated 438 000 deaths of which 90% occurred in the African region. Although millions more survive malaria episodes, they still suffer from severe anaemia and permanent neurological damage (Hotez, 2008). The UN Secretary General Ban Ki-Moon in April 2008, called for a universal coverage by 2010 to halt malaria deaths. Whereas international and domestic funding has increased to US$ 2.5 billion, there is still a gap of US$ 6.2 billion to reach the estimated US$ 8.7 billion needed to achieve universal access to malaria interventions per year. According to Hsiang, et al. (2012) global progress in malaria control has led to increased interest in and optimism for elimination. However, such optimism must be guided by evidence based decisions and in this regard optimal ways to disburse the limited funding resources have to be sought in order to ensure that the available resources are targeted to areas where they are mostly needed and where they are likely to have the greatest impact.

Out of a total of 99 countries that remain endemic, 32 are moving towards elimination, including four in sub-Saharan Africa and these include: Swaziland, Botswana, Namibia and South Africa. In order for countries to effectively progress towards elimination and further sustain and maintain zero local transmission there is need to develop strong surveillance methods and tools in order to aid preparedness and response. Bayesian spatio-temporal modelling methods can be used in identification, management and targeting response to malaria incidence areas and in understanding spatial heterogeneities in malaria distributions. Strengthening malaria surveillance systems has been identified as critical in high burden countries; however, the same is critical for low burden countries which are still susceptible to resurgence of the disease as seen in Madagascar and Zanzibar (Curtis, 2002; Romi, 2002). Surveillance is the single malaria intervention that is relevant throughout all the phases of the malaria elimination continuum from control to prevention of re-introduction.

Swaziland has already set an elimination target by the year 2016 and has already halted endemic transmission (Churcher et al., 2014). It is therefore important to develop models for malaria surveillance and response solution strategies in order to steer the country towards its malaria elimination target. In such low endemic settings as in Swaziland it is necessary to develop rapid surveillance systems to quickly identify and eliminate all local residual pockets of transmission foci. Among other interventions, mapping mosquito habitats is crucial in order to identify potential breeding sites and carry out larval source management (LSM) programmes based on space and time changes in the local malaria situation (WHO, 2012). While Swaziland is already in the elimination phase it is still faced with the challenge of dealing with the ever present threat of imported cases (Cohen et al., 2013). There is therefore a need to identify receptive areas in order to prevent the risk of local transmission due to importation and also to understand what strategies and responses are needed in order to deal with importation.

The use of Bayesian spatio-temporal methods in malaria mapping and prediction i.e forecasting is essential for disease monitoring solutions, pockets of transmission identification and for prevention of possible disease outbreaks through proactive efforts. Achieving elimination remains an operationally challenging effort which requires strong evidence-based and targeted interventions (Feachem, 2009). Such evidence can be realized with the use of tools and methods for rapid case-based mapping and case load predictions in space and time. Operational requirements for malaria elimination include: identifying environmental predictors of malaria, mapping mosquito vector habitats and identifying both remaining and emerging pockets of transmission. In Swaziland, these requirements have been limited *inter alia* by the high costs of Earth Observation (EO) data at finer geographic scales and lack of capacity in statistics to analyse the data and inform control programs and policy makers.

The high purchase costs for EO products have resulted in malaria transmission and malaria risk estimation being conducted at coarse continental scales (Kulkarni et al., 2010). This has seen a proliferation of vector distribution prediction models and malaria risk being done at continental scales where smaller countries like Swaziland may appear to be completely endemic as they are covered by a single pixel contrary to the true situation of spatial heterogeneities that exist on the ground. There is need to bridge this spatial resolution gap and guide on-the-ground control efforts with more spatially detailed information. Bayesian geostatistical models relating malaria data with climatic predictors have been applied in many settings to guide control efforts (Rumisha et al., 2012; Gemperli et al., 2006; Briët et al., 2008; Giardina et al., 2012). This study was aimed at developing methods and tools for malaria surveillance and response as well as extending the use of Bayesian spatio-temporal modelling in low endemic settings like Swaziland.

1.2 Role of surveillance in malaria control and elimination

Surveillance is important because of its ability to provide information to malaria control programme managers and planners about areas where malaria incidence is high and about locations of the most vulnerable populations. It allows tracking of changes and shift in malaria incidence over time, identification of clustering, and potential transmission foci. Identification of pockets of malaria transmission remains a challenge especially for low endemic countries struggling to eliminate local transmission. These pockets of transmission have been identified as a source of malaria incidence and surveillance can be used to detect and manage them before they fuel onward transmission to the wider population (Bousema et al., 2012).

Studies suggest that geographic clustering of disease is an indicator of residual pockets of transmission which have often been referred to as hotspots (Clark et al., 2008) but there is still need to develop methods to identify such pockets and to define their spatial and temporal clustering in the different transmission settings. In addition, the remaining pockets of transmission can vary in their spatial dimension including their ability to trigger onward transmission to the rest of the population. For instance, the remaining pockets of transmission could be identified by: high malaria incidence rates, presence or proximity of vector breeding sites, serology prevalence, and parasite density (in individual or groups) including concentration of imported cases which all represent a reservoir of parasites. These pockets of transmission present ideal hotspots that can be identified and managed with effective surveillance strategies, however very few surveillance programmes have the capacity to rapidly map and identify such pockets of transmission in their routine surveillance efforts.

The WHO (2012) manual for malaria control emphasises the importance of absolute numbers of cases and deaths as well as the incidence rates of these per 1000 population as important in the decision on where resources maybe allocated. The manual suggests that it is important to address areas where the risk is the greatest (most affected populations). However, pockets of transmission may not always be evidenced by a high concentration of cases and deaths as they could also be due to other factors such as proximity of breeding sites, serology prevalence including importation of cases. When considering the 20%/80% rule which suggest that only 20% of the population is responsible for 80% of the infections, it means that when addressing only high burden areas we may be dealing with the results of the problem instead of the source of problem.

It has also been observed that as control programmes are scaled up, endemic transmission declines, and imported cases become more responsible for the overall malaria incidence. This situation has already been reported in Swaziland where about 90% of the malaria transmission is attributed to importation. The current surveillance system for malaria in Swaziland relies on the cases presenting themselves at health facility and the timely reporting of confirmed cases via the short message service (SMS) based immediate disease notification system by general practitioners. In addition, confirmed cases reported by passive surveillance are then followed up by the National Malaria Control Programme (NMCP) surveillance agents whereby the case is interviewed and travel history is used to classify the case as either imported or local. This study contributed information that could be used in the determination of factors to be considered when defining and specifying thresholds for active case detection surveillance in Swaziland.

1.2.1 Importance of rapid response in malaria elimination

Surveillance systems must be robust and responsive (Ohrt et al., 2015). Scaling up and strengthening of this component of public health intervention is a prerequisite for effective malaria control and elimination. Targeted response must be an immediate action taken following identification through surveillance of any threat of infection or onward transmission to both populations at risk and those not at risk. Rapid detection of infections and delivery of appropriate response is necessary for elimination. According to Cao et al., (2014) surveillance for malaria elimination must include spatial aspect of reporting and clear timelines of response activities. It must also incorporate mechanisms for detecting infections (asymptomatic cases) before they spread using predetermined protocol and effective strategies of surveillance and response (i.e in a flow chart fashion). Furthermore it is important that information collected from affected communities is analysed and used as the basis for preventive actions. In some surveillance manuals, it is suggested that the data collection and processing for decisions and planning is done quickly in order to prevent onward transmission and reduce vectorial density with vector control measures. In order to move forward towards elimination standard procedures and guidelines detailing which actions need to be taken and by whom with clear time frames still need to be developed in order to unpack this rapidness in surveillance response. Quantification of the amount of transmission risk and how it varies in space and time could be used as a basis for the formulation of such rapid response guidelines.

According to Atkinson et al., (2012) timely identification and containment of pockets of infections with targeted response are important in achieving and maintaining elimination especially when the disease has ceased to be of priority concern to communities. This situation has been observed in Swaziland where endemic transmission has already been halted and the country now remains with very few sporadic local cases and thus public awareness has been dulled. Surveillance and response together with advocacy to maintain awareness and funding must be therefore tailored according to the different spectrum of endemicity in each geographic region (malERA, 2011). Technical guidance supported by rigorous analysis of every cost of intervention measure taken must be part of the elimination initiative (Premaratne et al., 2014). In addition defined roles and responsibilities with time frames must be part and parcel of localized strategies for detection and response to identified individual cases. A strong surveillance and response system need to be supported at all levels of societal cadres in order to achieve elimination (Bridges et al., 2012).

1.3 Environmental factors affecting malaria transmission

Epidemics of malaria are caused by a disturbance of the equilibrium between host, parasite and vector. Najera et al., (1998) define three different types of epidemics. Type one epidemics are caused by meteorological conditions which create temporary epidemics that will eventually revert back to the previous condition. Type two epidemics are caused by landscape changes or colonization of sparsely populated areas that create a new equilibrium level of endemicity. And type three epidemics are caused by interruptions in measures that were controlling malaria. Meteorologically stimulated epidemics normally last only one season of transmission. Many areas experience epidemics caused by meteorological changes that occur in inter-annual cycles. These cycles, which have been well illustrated by ENSO (El Niño Southern Oscillation), have been found in many parts of the world to follow the paraquinquennial cycle, which means epidemics happen every 5 to 7 years, however, in some areas the period of the cycle is longer. Because of the periodicity of cycles caused by meteorological factors, there should be a way to predict epidemics based on the risk factors related to epidemics including: a sudden increase in the number of non-immunes that are exposed to malaria, a rapid increase in vectorial capacity (increased density of vectors or invasion of a more efficient vector), land-use change, and failure of control efforts.

Factors that may precipitate a malaria epidemic fall into two categories: natural (climatic variations, natural disasters), and man-made (agricultural projects, dams, mining, logging, failure of control measures). Most of these factors make the physical environment more suitable for mosquito hatching. Other factors, such as local conflicts or development projects, produce massive population movements that expose non-immune populations to the malaria parasite. There is some evidence that this may already be taking place for example in Democratic Republic of Congo and the Central African Republic. In some areas warmer weather transforms rivers into puddles, while in others, it triggers rain and floods that leave behind stagnant pools. In both cases, the standing water serves as a perfect breeding ground for mosquitoes. Hotter weather also shortens the mosquitoes breeding cycle, speeding up their reproduction rate and it lengthens the season during which mosquitoes are plentiful. In warmer weather, mosquitoes are more active. Hotter temperatures reach inside mosquitoes’ gut and intensify the reproduction rate of disease-causing microbes, thereby increasing the likelihood that a single bite will cause infection. To large extent malaria epidemics are predictable, through environmental factors (i.e vegetation cover, land use and or land cover, temperature, rainfall and humidity) and local epidemiological knowledge. Three principal climatic and environmental factors that should be considered in malaria epidemiology are temperature, rainfall (precipitation) and humidity.

1.3.1 Temperature

The life cycles of Plasmodium as well as at the Anopheles mosquito depend on temperature. The optimal temperature for Plasmodium reapplication within the mosquito is 27°C. Higher temperature increases the number of times female mosquitoes bite and lay eggs. The intersections of the ranges of minimum and maximum temperature for parasite and vector development determine the impact of changes in temperature on malaria transmission. The minimum temperature for mosquito development is between 8-10°C, the minimum temperatures for parasite development are between 16-20°C with *P. vivax* (It can exist in places with an average summer temperature of only 16ºC) surviving at lower temperatures than *Plasmodium falciparum* (*Plasmodium falciparum* needs an average ambient temperature of at least 20ºC). The optimum temperature for mosquitoes is 25-27°C, and the maximum temperature for both vectors and parasites is 40°C. There are some areas where the climate is optimal for malaria and Anopheles mosquitoes are present, but there is no malaria. This is called "Anophelism without malaria" which can be due to the fact that the Anopheles mosquitoes present do not feed primarily on humans or because malaria control techniques have eliminated the parasite. If any changes, whether environmental or otherwise, were to occur to bring another species to the area that does act as a vector for human malaria, then the potential for outbreaks of malaria is very high since there is no immunity in the human population there. Therefore temperature is an important indicator for malaria transmission and was used during analysis in the current study.

1.3.2 Precipitation

Precipitation is another factor which affects the behaviour of Anopheles mosquitoes. There must be a certain level of precipitation in order to provide the stagnant water pools for the female mosquito to lay eggs. Anopheline mosquitoes breed in water habitats, thus requiring just the right amount of precipitation in order for mosquito breeding to occur. It is known that different Anopheline mosquitoes prefer different types of water bodies in which to breed. Too much rainfall, or rainfall accompanied by storm conditions can flush away breeding larvae (Savage et al., 1990). The amount and intensity of precipitation, the time in the year, and whether it is the wet or dry season, malaria survival is affected. Rainfall also affects malaria transmission because it increases relative humidity and modifies temperature, and it also affects where and how much mosquito breeding can take place. Some contend that the amount of rainfall may be secondary in its effects on malaria to the number of rainy days or the degree of wetness that exists after a rain event. Nevertheless, rainfall was analysed in this study to assess its effect on malaria incidence in Swaziland.

1.3.3 Humidity

Anopheles mosquitoes are also affected by relative humidity. While, plasmodium parasites are not affected directly by relative humidity. If the average monthly relative humidity is below 60% (Pampana, 1969), it is believed that the life of the mosquito is so shortened that there is no malaria transmission. These environmental factors influence the pattern of malaria distribution, which varies from region to region. Based on long term climatic data a potential distribution of suitable malaria transmission in Africa was produced using long term climatic data. In Mali (Sogoba et al., 2007), in Ethiopia (Asnakew et al., 2009), in Botswana (Craig et al., 2007), in Kenya (Li et al., 2008) vegetation cover along with temperature and rainfall were used to predict malaria transmission rates fairly well. Anthropogenic factors like deforestation, irrigation, urbanization, movements of populations and economic changes have also shown to influence the malaria distribution. Although not directly a measure of humidity, in the current study the Normalized Difference Vegetation Index (NDVI) was used to test its association with malaria incidence.

1.4 Earth Observation and malaria control

Essentially, malaria is an environmental disease since the vector requires specific habitats with surface water for reproduction, humidity for adult mosquito survival and the development rates of both the vector and parasite populations are influenced by temperature. Earth Observation (EO), Geographical Information Systems (GIS), Global Positioning Systems (GPS), spatial modelling and geostatistics are increasingly being recognized as valuable tools for effective management and planning of malaria vector control programmes. GIS, EO, GPS and geostatistics play a crucial role to monitor and analyse vector distribution. In addition these tools have improved knowledge and understanding of the biodiversity and other environmental factors influencing malaria. Any vector borne disease closely related to environmental conditions can be analysed using EO for surveillance, monitoring and early warning *inter alia*. Remotely sensed images of environmental proxies can be powerful predictors of vector distribution patterns and transmission level of malaria parasites by malaria vectors (Rogers, et al., 2002; Zeilhofer et al.,2007). According to Kalluri et al., (2007) remote sensing has been used to associate land use and land cover types with vector habitats based on simple classification techniques, as well as complex statistical models that link satellite-derived multi-temporal meteorological and earth observation with vector biology and abundance. A plethora of studies on remote sensing and epidemiology have already related EO data to vector-borne diseases. Remote sensing and spatial modelling have significantly improved our knowledge on the distribution of malaria over the last 25 years. These technologies provide an information platform to identify areas at risk and assist malaria vector control managers in directing resources and strategies. Rapid access or acquisition, analysis, and spatial display of data permit efficiently rapid response of management and evaluation, enabling the cost effective use of resources.

Previous, conclusions on satellite system products were that system performance, data costs and long turnaround times for products hampered the use of EO data for epidemiology (Beck et al., 2000). However, new satellite with more rapid/efficient and high resolution sensors have been launched but the uptake and use of such products by the health sector remains low mainly due to lack of expertise at control programme level. For over 30 years of research aimed at creating capabilities for malaria control in using remote sensing, the integration of these tools into operational disease control management has rarely taken place (Ceccato et al., 2005, Beck et al., 2000). Also the extension of remote sensing into operational disease surveillance and control has been slow. This study used a combination of high resolution EO, entomological data and surveillance data to develop monitoring solutions for malaria surveillance, control and elimination in Swaziland. The information and models generated in this study added an operational and direct application to malaria control programmes thus bridging the gap between remote sensing and direct application in malaria control and general vector-borne diseases.

1.5 Malaria situation in Swaziland

Swaziland, a country in southern Africa has made significant progress in malaria control, has achieved its lowest-ever recorded malaria burden over the past 11 years and it aims to achieve elimination by 2016. Malaria transmission is unstable occurring in the rainy season between November and May, and is closely related to the amount of rainfall. *Plasmodium falciparum* is responsible for over 99% of malaria cases. Vector monitoring indicates that the main malaria vector is *Anopheles arabiensis* (Swaziland Ministry of Health, 2010). In 2010, rapid diagnostic tests (RDTs) were introduced at all health facilities, allowing for definitive diagnosis of all malaria cases. Since RDTs became available, lab confirmed cases have increased from 73 in 2008 to 196 in 2010. Clinically diagnosed cases reported have decreased by 76% from 2009 to 2010. According to Hsiang et al., (2012) from 1999 to 2010 (Figure 1.1.0 and 1.1.1), annual laboratory-confirmed malaria cases decreased from 4005 to 196 (3.8 to 0.2 per 1000 population), and suspected cases decreased from 29,374 to 3470 (27.7 to 2.9 per 1000 population). Among the key interventions attributed to this decline is the sustained annual indoor residual spraying (IRS), strong surveillance system and a cross-border program known as the Lubombo Spatial Development Initiative (LSDI) that has successfully decreased malaria transmission in neighbouring southern Mozambique and South Africa.

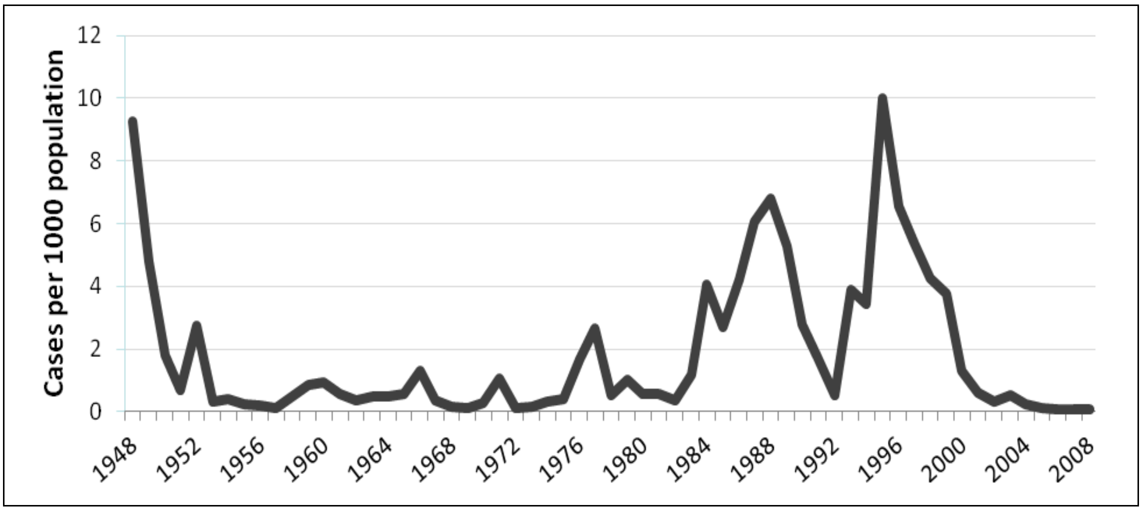

Figure 1.1.0: Malaria Incidence in Swaziland, 1946-2009

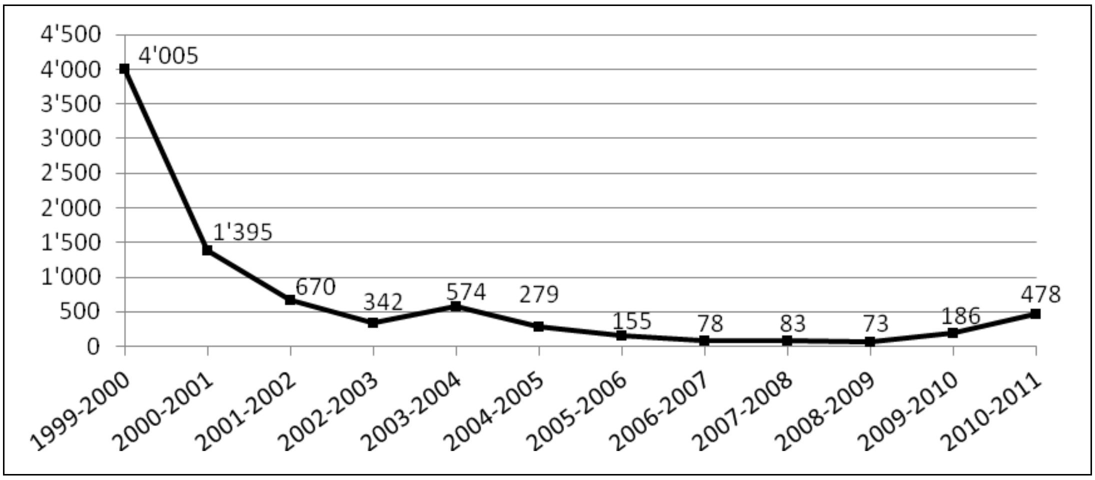

Figure 1.1.1: Parasitological Confirmed Cases, 1999-2011

Based on the increased malaria control achievements and declining disease burden, the country was chosen by the African Union and SADC as one of a few countries in Southern Africa earmarked for malaria elimination. Swaziland then developed a malaria elimination strategy whose overall goal was to: reduce and sustain the locally acquired malaria cases to zero by 2015; reduce and sustain malaria deaths seen at health facilities to zero by 2015; and maintain zero locally acquired malaria cases by prevention of reintroduction for all years following 2015. However, recent data indicates that the country has failed to reach the target of zero local cases by 2015 as in a few months before 2015 local cases doubled from 112 in the July 2012 to June 2013 transmission season to 204 during the July 2013 to June 2014 transmission season. In July 2014 to June 2015 (elimination target year) locally acquired cases increased further to 273 thereby forcing the NMCP to shift its target year from 2015 to 2016. To implement the elimination strategy the NMCP of Swaziland instituted a reactive surveillance system in 2009 in which all cases identified in health facilities are reported via SMS to the NMCP. Surveillance agents then obtain the address of case patients and interview them at their homes and capture the Global Positioning System (GPS) coordinates for the patient's house. Family members and neighbours living within 1 km of the case patient are screened for malaria using rapid diagnostic test (RDTs), and any identified case is referred to the nearest health facility where the patient is treated with artemisinin-based combination therapies (ACTs).

The country therefore, still needs to develop effective surveillance and response tools and methods that are applicable in low transmission settings in order to achieve elimination. WHO certification of elimination requires achieving an absence of all local transmission for three years, as well as a sufficiently strong surveillance system to prove that cases would have been identified had they occurred. According to Cohen et al., (2013) as Swaziland moves towards elimination of malaria, endemic transmission will increasingly occur in residual foci, while the importance of preventing onwards transmission from imported infections will increase. Recently, malaria importation from Mozambique accounts for over 90% of malaria transmission in Swaziland (Koita et al., 2013). As an example, in the recent transmission season from July 2015 to June 2016 imported cases were 230 compared to only 68 local cases. This indicates the ever increasing need for more rapid surveillance and response strategies for pockets of transmission identification, malaria incidence risk mapping and prediction in order to prevent onward transmission from imported cases. However such tools and methods adapted for low endemic settings characterised by diminishing malaria are not yet fully developed.

1.6 Applications of environmental data in malaria epidemiology

There has been an upsurge in the use of environmental data for mapping and predicting malaria transmission and vector distributions. Many of the environmental proxies used as covariates in malaria analysis models are obtained from open source databases and archives of remote sensing agencies. In 2015, Weiss et al., reviewed some of the covariates that had been used to map plasmodium falciparum endemicity. Variables that had been previously used were categorized into those related to climate such as: temperature, rainfall, humidity, vegetation indices, soil moisture, vector breeding site information as well as wind speed. Others included those related to topography such as, land use/cover, elevation, reflectancy, and spatial limits of malaria. In many instances socio economic proxies have also been used to weight the outcome of interest accordingly such as population within defined geographic areas, distances from services among others. The current trend shows that environmental data will continue to be used in disease vector-borne and epidemiological studies especially because the quality of the data has also improved over the years.

1.7 Significance of the proposed research

This study contributed with maps of probability of larva presence at high spatial resolution for all potential vector breeding habitants in Swaziland which could be used to assist the NMCP to deliver cost-effectively larval source management activities. We also contributed new information and demonstrated how multi-scaled remotely sensed data could be used jointly in vector-borne disease mapping and epidemiology to produce geographically explicit and easy to visualize maps. Due to the low uptake of some of the new remote sensing products that have a potential for use in vector-borne disease mapping and epidemiology we created a catalogue of remote sensing indices that have such application by inventorying the products and their sources. Our motivation was from the fact that there are many new and emerging remote sensing products that are available for use from newly launched satellites, however little is known about them since they were not properly documented. This catalogue is important for epidemiologist who might want to incorporate new covariates in their analysis as they evaluate and investigate various environmental factors and their association with certain environmental driven diseases. We also provided for the first time estimates of the space-time patterns and trends of malaria incidence in Swaziland after taking into account the contribution of environmental factors which will assist control programmes to evaluate the efficacy of their control interventions when they are able to better target and prioritise their anti-malaria activities.

The polynomial distributed lags model used in identifying the lag time between environmental factors and malaria incidence can further be improved by developing and incorporating methods and algorithms for malaria forecasting based on the strategy of preventive and cost-effective fashions. In addition, the model could be used as an interactive and real-time tool for malaria early warning as new data is collected and added into the model for update. Overall this study generated knowledge and new information which could be used by the NMCP of Swaziland as its targets malaria elimination by 2016. The information could also be used by policy makers and programme directors as it provides them with scientifically solid and locally responsible knowledge to guide their decision-making process. This study could also contribute in the malaria elimination initiative in the country where statistical and analytical skills are needed in order to build a strong surveillance system in accordance with the elimination phase. Currently, the NMCP of Swaziland collects a lot of surveillance and vector control data but lacks statistical capacity to analyse and make use of the data for planning malaria interventions cost effectively.

1.8 Objectives of the thesis

The main aim of this study was to investigate methods and tools that aid malaria surveillance in low endemic settings like Swaziland. Specific objectives are detailed in the following sub-section.

1.8.1 Specific objectives

Consequently, the overall objective of the study was split into the following specific objectives which were to:

i) Assess potential of high resolution RS images to identify breeding sites in Swaziland.

ii) Review remote sensing products and their sources to assess their potential application in vector-borne disease like malaria mapping and surveillance

iii) Evaluate the predictive performance of new and underutilized remote sensing products to estimate malaria incidence at high geographic resolution

iv) Apply Bayesian modelling based on polynomial distributed lags to map and predict malaria incidence in Swaziland

1.8.2 Sources of data

Data used in this study was obtained from the NMCP in Swaziland and it included Active Case Investigation data (ACI), Active Case Detection data (ACD) and entomological data on potential breeding sites identified through larva scooping field campaign. Remote sensing and other environmental data was obtained from open source data archives such as National Aeronautics and Space Administration (NASA), Moderate Resolution Imaging Spectroradiometer (MODIS) and other Goddard Space Flight Center data archives websites. In addition, a 5 m land cover imagery was purchased from RapidEye (BlackBridge, Germany) through a previous project called MALAREO which preceded this study. The land cover imagery enabled us to map potential breeding sites at high spatial resolution and thus produce geographically explicit maps to aid the NMCP in conducting larval source management activities.

The NMCP data was accessed through the Malaria Surveillance Database (MSD which houses all malaria data such as surveillance, vector control including indoor residual spraying and mosquito net distribution. The data was requested from the control programme manager. This data is also geo-coded which allows spatio-temporal analysis to be performed. This study therefore assembled all the geo-coded surveillance and vector control data from 2009-2015 (monthly aggregated) and it was organized according to each malaria transmission season in Swaziland which is from July-June each year. The data was then applied in Bayesian geostatistics models to conduct various levels of analysis.

Chapter 2 Assessing the relationship between environmental factors and malaria vector breeding sites in Swaziland using multi-scale remotely sensed data

Sabelo Nick Dlamini[1,2], Jonas Franke[3] , Penelope Vounatsou*[1,2]

[1]Swiss Tropical and Public Health Institute, Basel, Switzerland;

[2]University of Basel, Basel, Switzerland;

[3]RSS - Remote Sensing Solutions GmbH, Baierbrunn, Germany

*Corresponding author

This manuscript has been published in Geospatial Health 2015, volume 10(No. 1):302

Abstract

Many entomological studies have analysed remotely sensed data to assess the relationship between malaria vector distribution and the associated environmental factors. However, the high cost of remotely sensed products with high spatial resolution has often resulted in analyses being conducted at coarse scales using open-source, archived remotely sensed data. In the present study, spatial prediction of potential breeding sites based on multi-scale remotely sensed information in conjunction with entomological data with special reference to presence or absence of larvae was realized. Selected water bodies were tested for mosquito larvae using the larva scooping method, and the results were compared with data on land cover, rainfall, land surface temperature (LST) and altitude presented with high spatial resolution. To assess which environmental factors best predict larval presence or absence, Decision Tree methodology and logistic regression techniques were applied. Both approaches showed that some environmental predictors can reliably distinguish between the two alternatives (existence and non-existence of larvae). For example, the results suggest that larvae are mainly present in very small water pools related to human activities, such as subsistence farming that were also found to be the major determinant for vector breeding. Rainfall, LST and altitude, on the other hand, were less useful as a basis for mapping the distribution of breeding sites. In conclusion, we found that models linking presence of larvae with high-resolution land use have good predictive ability of identifying potential breeding sites.

Key words: vector breeding, remote sensing, larva scooping, malaria, Swaziland

2.1 Introduction

Malaria is caused by *Plasmodium* parasites, which are transmitted to people through a bite by an infected female *Anopheles* mosquito. Most mosquito species oviposit in standing waters and pools of varied amounts and sizes depending on the preference of each particular species. Targeting mosquito larvae and pupae with larvicides in standing water or breeding sites is one of the more important intervention measures in the fight against, and elimination of, malaria (Clennon et al., 2010; Dambach et al., 2014). Identification and mapping of all potential vector breeding sites is a prerequisite for successful vector control, especially larval source management (LSM) applied for effective elimination of residual foci. A national inventory of all residual foci is necessary if transmission is to be interrupted and remaining foci cleared (Chanda et al., 2013). Past statistical modelling and mapping efforts have predicted vector distributions at continental scales based on climatic suitability and low-resolution remotely sensed (RS) data (Kulkarni et al., 2010). However, very few studies have used entomological data in conjunction with remotely sensed data to identify, map and predict potential malaria vector breeding sites at explicit geographical locations (Ahmad et al., 2011; Bøgh et al., 2007; Li et al., 2008). The spatial resolution of the satellite-generated imagery is crucial for identifying potential vector habitats, and high spatial resolutions must be applied in order to capture not only larger water bodies, but also smaller ones, which are potentially as important for breeding.

Previous studies using satellite-generated imagery to identify suitable vector habitats, also based their approach on existing knowledge on how factors, such as temperature, humidity and rainfall, influence mosquito population dynamics and distribution (Beck et al., 2000). Dambach et al. (2009) used imagery from the SPOT-5 satellite with supervised classification to identify land cover types known to be suitable as *Anopheles* mosquito breeding sites. Since no field-generated data were used in the analysis, the classification of relative risk was entirely based on the literature on *Anopheles* mosquito presence in different land cover types. Oesterholt et al., (2006) approximated vector breeding sites by assessing malaria incidence in relation to the distance to the nearest water body using geographical information systems (GIS). In their study, 10 houses were mapped and light traps were hung at the end of an occupied bed to catch mosquitoes. Traps were emptied and mosquito species were counted and determined the following morning. The identification of potential breeding sites using adult mosquito entomological data relies on the ability and precision to map distance to the nearest water bodies rather than distance to the actual breeding sites, which can be assessed, for instance, through larval scooping. However adult mosquito dispersal could be influenced by wind speed and wind direction (Bøgh et al., 2007) and therefore approximating breeding sites could be challenging.

Strong progress in the fight against malaria has been made in Swaziland. In 2002, insecticide-treated bed nets (ITNs) were introduced to complement the ongoing indoor residual spraying (IRS) activities. In 2008, after 15 years of progressive reduction of the disease burden (from 4,005 to 369 cases, the country was nominated to spearhead the malaria elimination in the Southern African Development Community (SADC), which is being pursued according to the strategic plan for the period 2008 - 2015 (MIS, 2010). Whereas the country has consistently and annually applied IRS using DDT as its mainstay vector control intervention strategy, studies designed to support these attempts with empirical evidence on its effect on the number of vector breeding sites and their distribution have not been conducted. Identification and elimination of residual foci along with efforts to reduce the number of local malaria cases to zero remains a challenge, especially if there is a lack of geographically explicit supporting maps to target intervention efforts. Following the Stockholm Convention on Persistent Organic Pollutants (http://sites.duke.edu/malaria/the-stockholm-convention/), many countries will soon have very limited supply of DDT, so optimal ways to use this chemical in high priority areas must be sought. This is possible with spatially explicit maps guiding ground IRS activities, thus avoiding the indiscriminate use of DDT that leads to unnecessary waste and environmental damage in addition to potentially increased vector resistance.

The objective of this study was to analyse the relationship between environmental factors and malaria vector breeding sites in Swaziland by linking entomological data with multi-scale RS data and scooping for larvae in selected water bodies and dams, feeding collected information into a statistical regression model and using data mining tools to investigate potential associations. By this approach we aimed to contribute to the existing knowledge about malaria vector breeding habitats in Swaziland and provide high-resolution, spatially explicit maps to assist on-going conventional control efforts as the country targets elimination by 2015.

2.2 Material and methods

2.2.1 Study Area

Swaziland covers an area of 17,363 km^2 and consists of a mountainous Highveld (the wet western part of the country), which has an altitude of about 1,800 m above the sea, and the relatively flat, dry eastern part (between 100-300 m in altitude) called the Lowveld (Fig. 1). The rain that falls on the Highveld flows towards the Lowveld, where it stagnates creating pools of standing water suitable for mosquito breeding because of the flat terrain and high temperatures in this dry zone.

Malaria transmission in Swaziland is prevalent along the eastern part of the country borders with Mozambique and north-east South Africa. Transmission occurs in the rainy season between November and May with a peak in February and March, sometimes extending to April. Transmission is unstable and follows the quantity of rainfall in each particular year and the amount of standing water accumulated during the latest rain episode. The entire population is at risk as it generally lacks acquired immunity and is therefore highly vulnerable. The situation is particularly serious with respect to pregnant women and children under the age of 5 years (MIS, 2010). *Plasmodium falciparum* dominates and about 99% of malaria cases are infected with this species, while infections due to the other malaria species are occasional.

2.2.2 Entomological data

The National Malaria Control Programme (NMCP) of Swaziland recently developed a geo-database of all potential vector breeding sites and on-going larva scooping activities (NMCP 2012 annual report). Data on vector control and entomology were extracted from the NMCP database and used in this study for an analysis including multi-scale RS data. Mosquito larvae were collected from 30 selected water body/wetland sites and peripheral shallow pools of standing water as well as water-filled cattle hoof prints that were identified as potential breeding sites in the Lowveld region. Figure 2.2.1 shows the larvae sampling locations in the different constituencies, locally known as Tinkhundla.

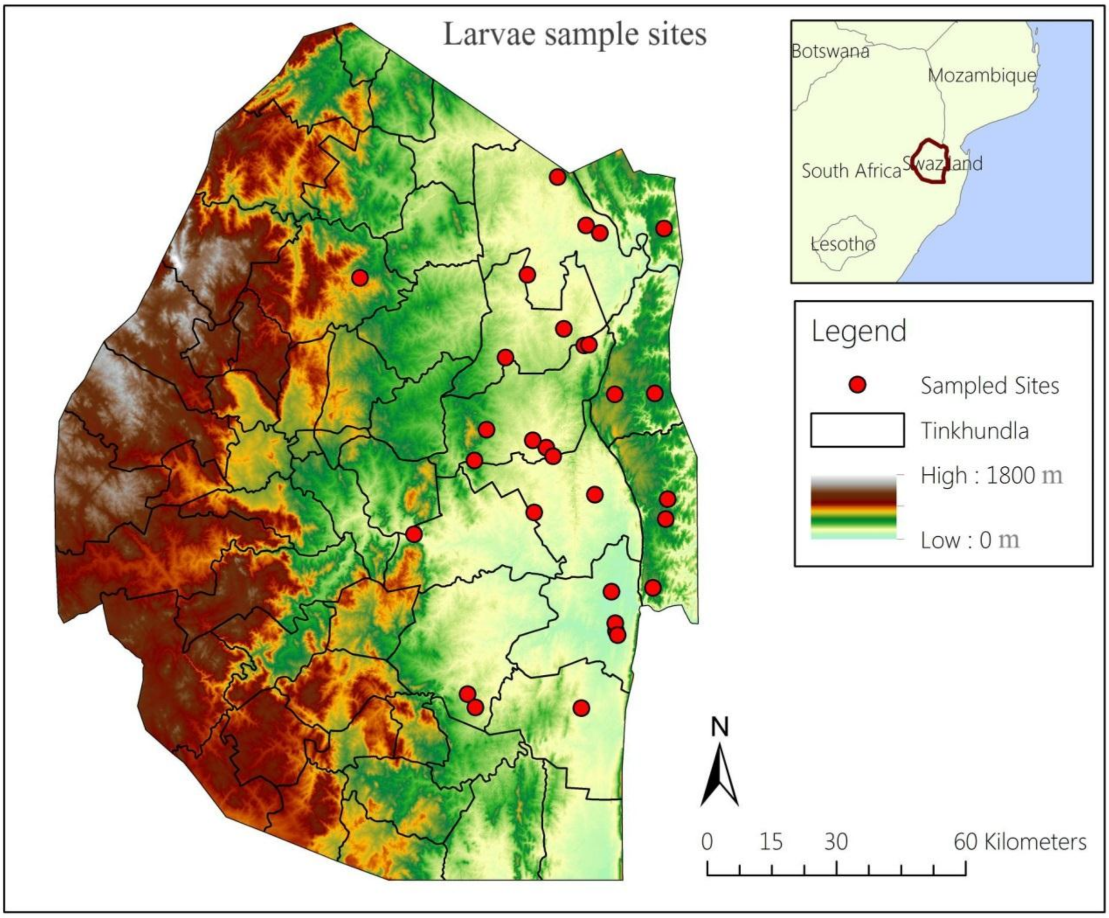

Figure 2.2.1: Altitude map showing locations of larva sampling sites

The morphology of all visible larvae in the water was closely observed in an attempt to collect those larvae known to be anopheline. The larvae were scooped up from high-density larval areas in the water and the number of larvae per scoop counted. Where the larval density was extremely low, or no larvae visible, about 100 scoops were done and the density determined on the basis of the larvae present in each one. Following speciation by a senior entomologist, the mosquitoes belonging to the *Anopheles gambiae* complex and *An. funestus* were kept in Eppendorf tubes, labeled and preserved in iso-propanol for later polymerase chain reaction (PCR) analysis. For identification purposes, the larvae from each sampling location were placed in a plastic bucket leaving a breeding space on top and closed with a lid. Both buckets and lids were both marked to indicate the locality and date for the collection and then transported to NMCP's insectary, where the larvae were transferred by pipette into marked plastic bowls with clean water, which was placed on a cage to secure them from being accidentally tilted and spilled while in the insectary. Each bowl was covered with a net and a cotton wool soaked in 10% sugar solution was placed on top for feeding. A heater was used to ensure maintenance of warm temperatures and to facilitate growth. When the larvae hatched and became adults, they were removed using a sucking tube, placed in marked paper cups and covered with a net. The mosquitoes were still put on the sugar solution which was placed on the side of the cups for daily morphological identification by the entomologist.

2.2.3 Remotely sensed data

Mosquito vector habitat requires specific characteristics, such as sufficient surface water for reproduction, a certain humidity level for adult mosquito survival and a suitable temperature allowing acceptable development rates for both vector and parasite (Ceccato et al., 2005). For that reason, for a period starting 4 weeks prior to the study until the end of the entomological survey, Collection-5 (Land Surface, Temperature & Emissivity data) from the Moderate Resolution Imaging Spectroradiometer (MODIS) (http://modis.gsfc.nasa.gov) was downloaded from the Land Processes Distributed Active Archive Center in USA (https://lpdaac.usgs.gov/) for land surface temperature (LST) data at 1-km spatial resolution. We relied on the MOD11A2 product, which is the average value of clear-sky LSTs during an 8-day period. In addition, a 24-day temperature average was obtained for the three weeks prior to the entomological survey. Rainfall estimates (RFE) at 8-km spatial resolution were downloaded from the Famine Early Warning Systems Network (FEWS NET) Africa (http://earlywarning.usgs.gov/afghan/downloads/index.php?regionID=af& -productID=3&periodID=6). In addition, a digital elevation model (DEM) from the Advanced Space-borne Thermal Emission and Reflection Radiometer (ASTER) (http://asterweb.jpl.nasa.gov) at 30-m spatial resolution was used for altitude data.

High-resolution data from the RapidEye satellite (http://www.satimagingcorp.com/satellite-sensors/other-satellite-sensors/rapideye/) obtained from the ESA Data Warehouse (http://gmesdata.esa.int/web/gsc/news/latest_20110607) covering the malaria-endemic area of Swaziland at 6.5-m spatial resolution was used in order to determine the various land cover types. The RapidEye pixel resolution was resampled to 5-m pixel size during ortho-rectification by the data provider. A standardized atmospheric correction was applied to the 62 RapidEye tiles obtained to ensure consistent image pre-processing and thus a more reliable land cover classification. Land use/land cover (LULC) classification of the high-resolution data was conducted by applying an object-based image analysis with a predefined hierarchical Ruleset (supervised classification) using the software eCognition (Trimble GeoSpatial, Munich, Germany). The overall accuracy of the land cover classification was 80.7% with a Kappa coefficient of 0.78. As separate layers, high-resolution, Euclidean distance-to-land cover layers were generated as input for the statistical modelling of potential mosquito breeding sites. Figure 2.2.2 shows land cover classification while Figure 2.2.3, 2.2.4 and 2.2.5 shows as a set of three examples of distance-to-land cover layers used in the statistical model for prediction. The high spatial resolution of the land cover classes allowed the identification of small water bodies and wetlands that can be potential breeding sites for malaria vectors.

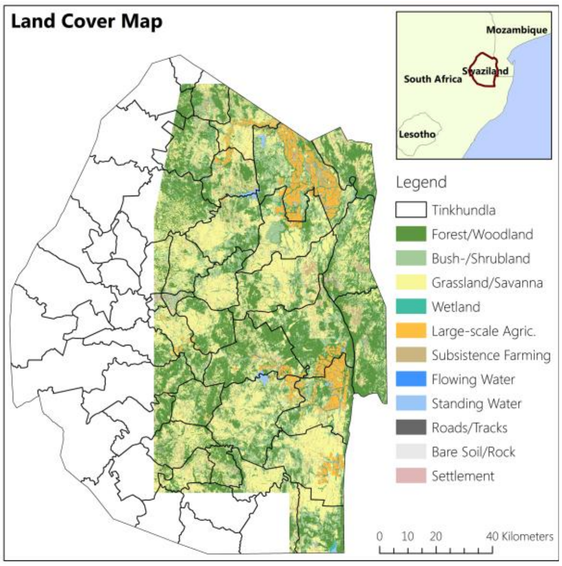

Figure 2.2.2: land cover map covering the malaria-endemic area of Swaziland

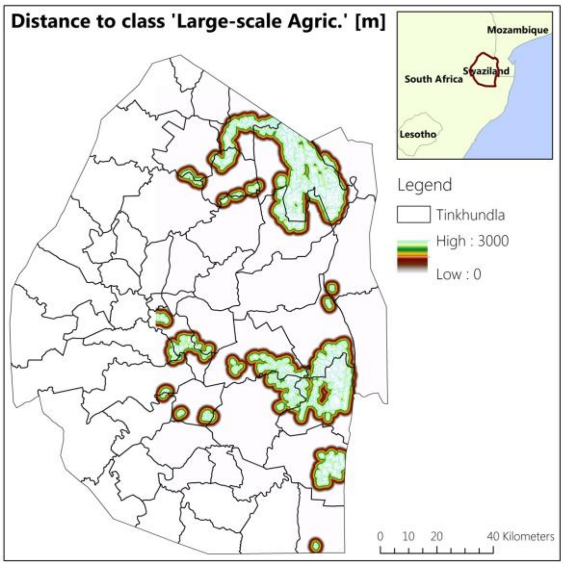

Figure 2.2.3: Distance-to-large-scale agriculture

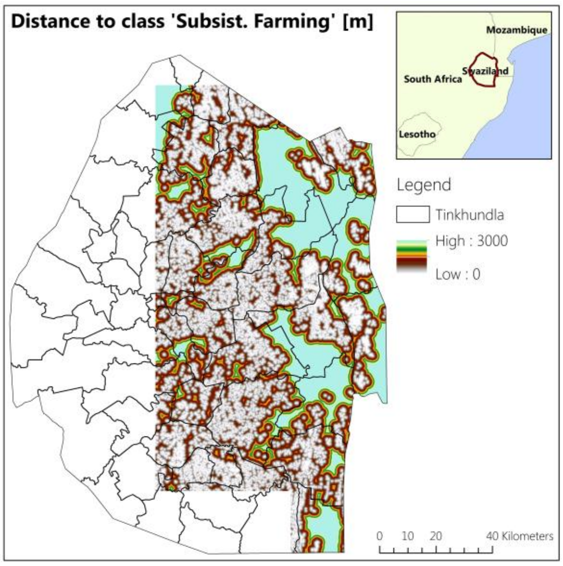

Figure 2.2.4: Distance-to-subsistence farming

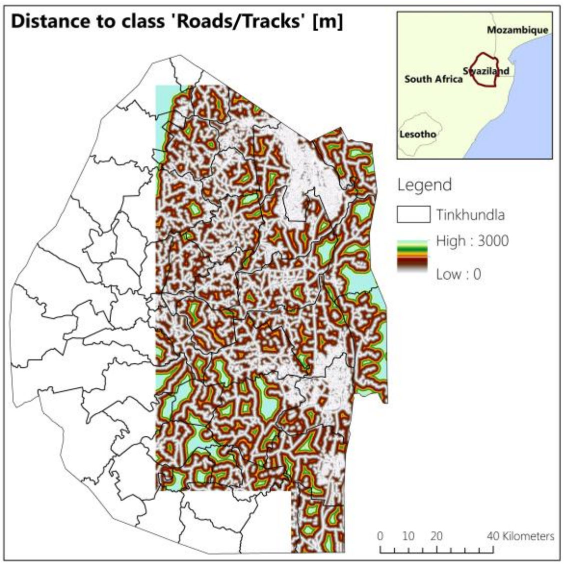

Figure 2.2.5: Distance-to-roads/tracks

An analysis, based on a total of 18 different environmental variables for each larvae sampling point (Table 2.2), was carried out with the aim of identifying which parameters that best describe a vector breeding focus and its environmental requirements (maps in Figures 2.2.6, 2.2.7, 2.2.8 and 2.2.9; Figures 2.3.0, 2.3.1, 2.3.2 and 2.3.3). Eight LULC classes were used: settlements, subsistence farming, large-scale agriculture, savannah, forest, bush, bare soil/rocks and roads. All eight distance-to-land cover variables were categorized into three distance categories which were defined by the 33rd and 66th centiles. This way we were able to estimate the effect of the different distances on the presence or absence of vector breeding sites. Distances from vector breeding sites to each of the distance-to-land cover classes was defined by calculating the Euclidean distance from each sampling point to the centroid of the pixel with the specific land cover type.

Table 2.2: Data sources and properties of the environmental covariates used in predicting potential mosquito breeding sites

Type of data	Source	Date	Temporal resolution	Spatial resolution
Altitude	DEM[1]	2012	-	30 m
Land cover	RapidEye[2]	2011	-	5 m
LST[3]	MODIS data[4]	01.11.2012-05.11.2012	8 days	1 km
Rainfall (RF)	FEWS NET[5]	01.11.2012-10.12.2012	10 days	8 km

[1]Digital Elevation Model (DEM). Available a: http://asterweb.jpl.nasa.gov (accessed: December 2011).

[2]RapidEye satellite imagery. Available at: http://gmesdata.esa.int/web/gsc /news/latest_20110607

[3]Land surface temperature

[4]Processes Distributed Active Archive Center. Available at https://lpdaac.usgs.gov/ (accessed February 2015)

[5]Famine Early Warning Systems Network. Available at http://earlywarning.usgs.gov/afghan/downloads/index.php?regionID=af&productID=3&periodID=6 (accessed February 2015)

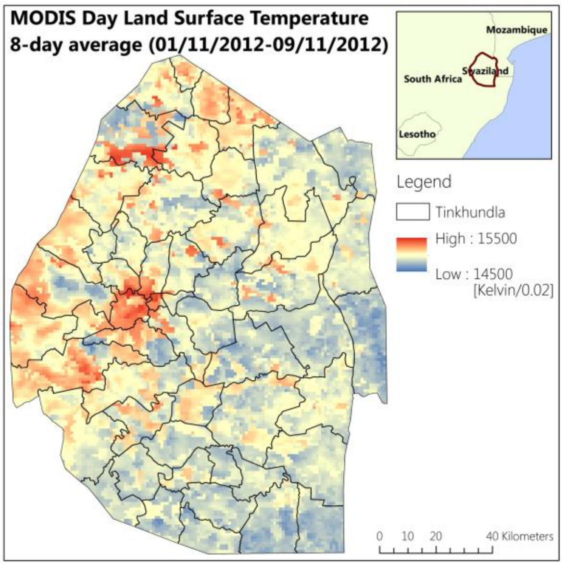

Figure 2.2.6: Land surface temperature (first week

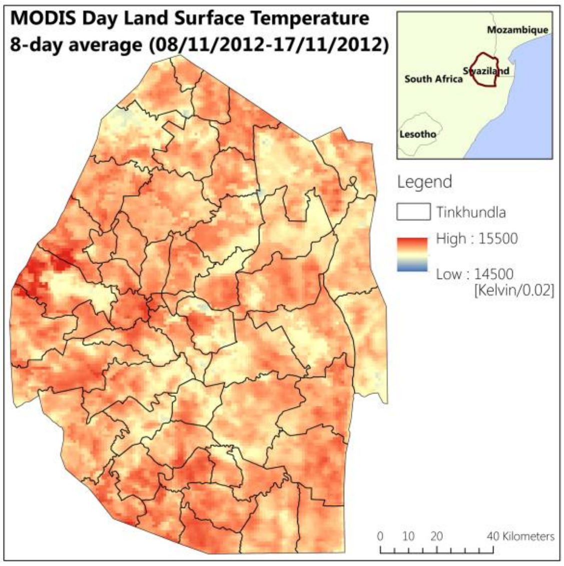

Figure 2.2.7: Land surface temperature (second week)

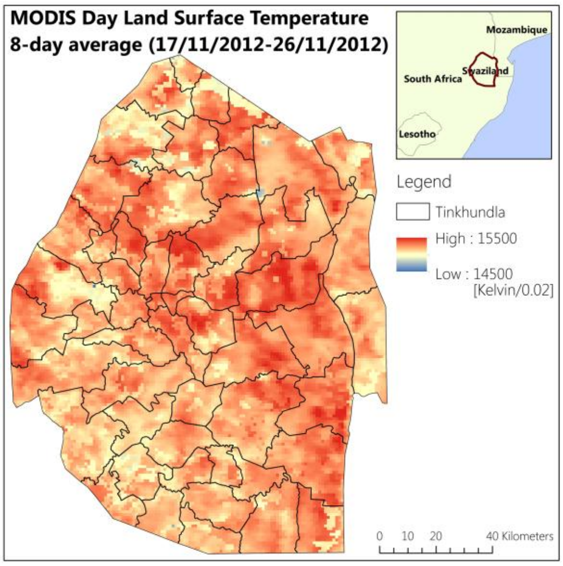

Figure 2.2.8: Land surface temperature (third week)

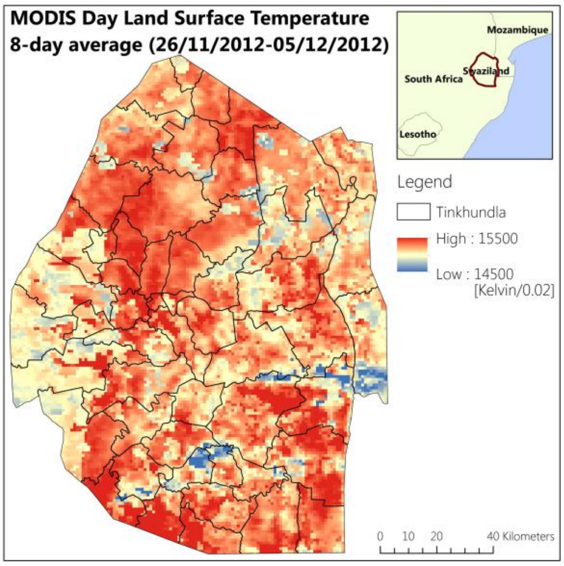

Figure 2.2.9: Land surface temperature (fourth week)

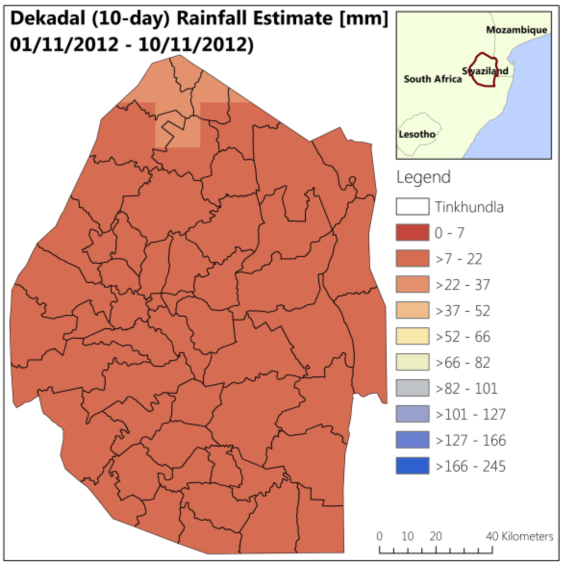

Figure 2.3.0: Ten-day data rainfall estimate (RFE) first dekadal

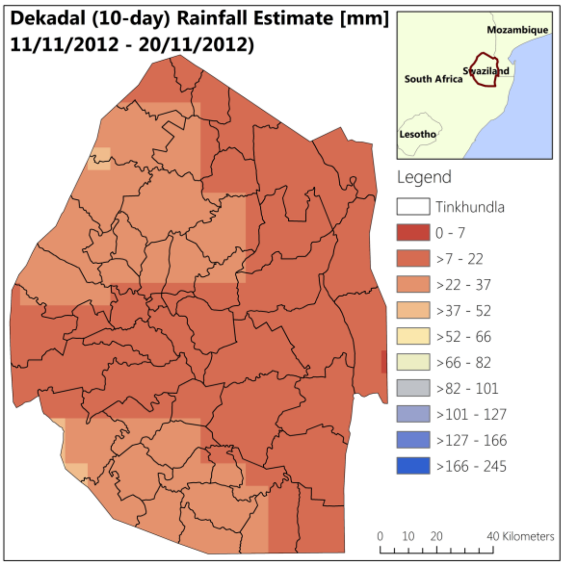

Figure 2.3.1: Ten-day data rainfall estimate (RFE) second dekadal

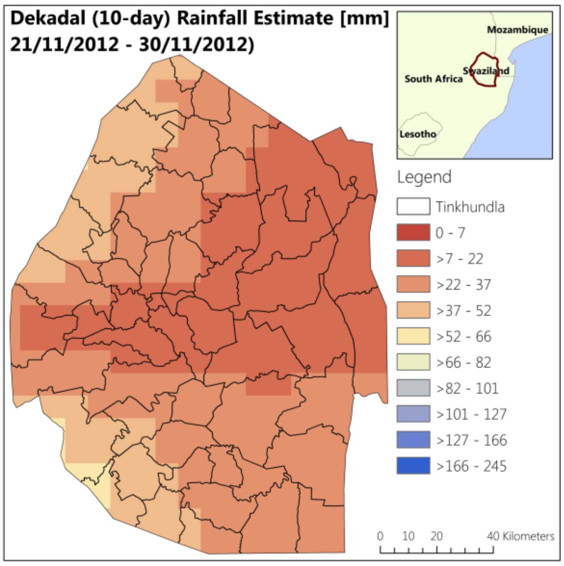

Figure 2.3.2: Ten-day data rainfall estimate (RFE) third dekadal

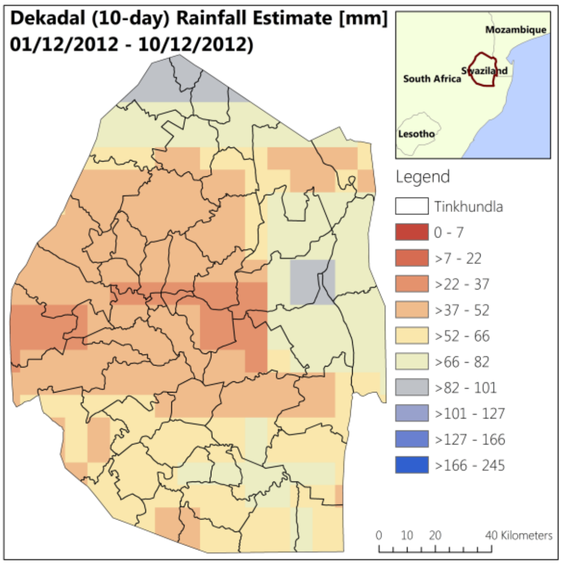

Figure 2.3.3: Ten-day data rainfall estimate (RFE) fourth dekadal

2.2.4 Statistical analysis

The relationship between the above environmental factors and the presence or absence of larvae were analyzed using a data mining approach in the statistical software See5 (RuleQuest Research, St Ives, NSW, Australia), which uses a Decision Tree R3 Ruleset induction engine developed by Quinlan (1993). The algorithm classifies data according to independent variables and a branching method which splits the data to illustrate every possible outcome of a probability-based decision. ArcGIS 10.1 (ESRI, Redlands, CA, USA) was used to conduct zonal statistical analyses at each larval sampling location in order to retrieve the environmental parameters for each location. The Decision Tree Ruleset was thus based both on larval field data from the sampling points and the local environmental variables.

A training dataset was used to select independent variables relevant for the classification of the presence of larvae and to obtain the Decision Tree Ruleset for classification accuracy using the field data provided. Due to the low sample size, the analysis was performed based on the results from all 21 positive sampling points. The classifier construction options in the statistical analysis were set so that it would provide a final Ruleset for classification of unknown water bodies/wetlands. The location of standing water bodies and wetlands from the high-resolution LULC map was afterwards applied to the final Ruleset for analysis. Based on the final classification rule, a map of all potential breeding sites in the malaria-endemic area of Swaziland was produced.

2.2.5 Logistic regression

As a first step, the data were explored for any redundancies between variables via a statistical correlation analysis using Pearson`s correlation coefficient. Bivariate analyses were performed to determine the relationship between each of the environmental covariates and presence of vector breeding sites using logistic regression (Stata Statistical Software, version 13.0). In order to evaluate and compare the results from the Decision Tree analysis a backward-selection, stepwise logistic regression analysis with a 15% significance level for removal was carried out based on the same input data set as used in the Decision Tree approach. Covariates included in the stepwise regression were all those significant at the 15% significance level in the bivariate results. We used a 15% significance level to reduce the chance of excluding important predictors from the stepwise multiple regression. This analysis enabled finding the most effective and parsimonious set of variables predicting the dependent variable (presence or absence of larvae).

2.3 Results and discussion

2.3.1 The field campaign

Larvae were found in 21 out of the total of 30 sample sites (70%) of water bodies. However, these were mostly areas where it had not rained in the past two to three weeks prior to sampling. In areas where it had rained recently, no larvae were found and it was difficult to ascertain whether this was due to the fact that the site was not suitable or the larvae had been washed away due to the rain. In total, 152 mosquito larvae were collected during the field scooping campaign. Upon rearing, about 60% of the captured larvae was found to be *An. pretoriensis* and 15 % *An. gambiae.* The rest of the mosquitoes were C*ulicines* (20%) and *Culex* (5%) species. Sites where no larvae were found were mainly large dams, open turbulent dams and areas where rain was reported less than a week prior to sampling. Most of the larvae were collected from the edges of shallow and sheltered water bodies, but where there was enough sunlight. Larvae were also found in clear pools of water with little aquatic flora and fauna. Geographically, the highest larval densities were found in the southern and northern part of the Lowveld as well as on the eastern plateau bordering Mozambique. These were areas with large-scale irrigated farms such as sugarcane surrounded by subsistence farming. Excess runoff water was common in these areas as most of the fields had sprinklers continually wetting various parts of the field; hence drainage water accumulated in small depressions around each field assuring availability of water also in the dry season.

2.3.2 Decision tree analysis

The analysis of the 21 sample points with larvae using 18 environmental variables showed that the combined use of only 3 variables in 3 rules correctly classified 95.2% of the larvae present sites and they included: distance-to-subsistence farming, 10-day RFE of the second dekadal and distance-to-savannah. The variable with the highest attribute usage was the distance-to-subsistence farming that appeared in 90% of the rulesets in predicting a class. Indeed, distance-to-subsistence farming less than 216 m turned out to be a major determinant for vector breeding sites (Figure 2.3.4).

However, no relation could be found between breeding sites and distance-to-settlements. In addition to the variable distance-to-subsistence farming, the 10-day RFE with 1-week lag had an attribute usage of 38% in the ruleset and the distance-to-grassland/savannah showed an attribute usage of 33%. According to these results, the LST over the 4 weeks before the sampling and altitude had no influence on the occurrence of vector breeding sites. This was also the case for all other LULCs except distance-to-subsistence farming and grassland/savannah and the other three remaining RFEs. As a final step, the classification ruleset derived from the Decision Tree analysis, which used the variables distance-to-subsistence farming, distance-to-grassland/savannah and the 10-day RFE, was applied to the wetland and water body layer with 5-m spatial resolution in order to identify potential vector breeding sites with similar environmental characteristics as those identified through the entomological survey. The resulting map of predicted vector breeding sites for the whole malaria-endemic area in Swaziland explicitly identified potential breeding sites at sizes from 5-m to over 100-m of standing water or dams. This high resolution map could be used as a guide for larval sampling activities. Figure 2.3.5. shows an example of the final potential vector breeding sites classification of the malaria area covered by the 5-m resolution RapidEye data (northern part of the Lowveld) in Swaziland.

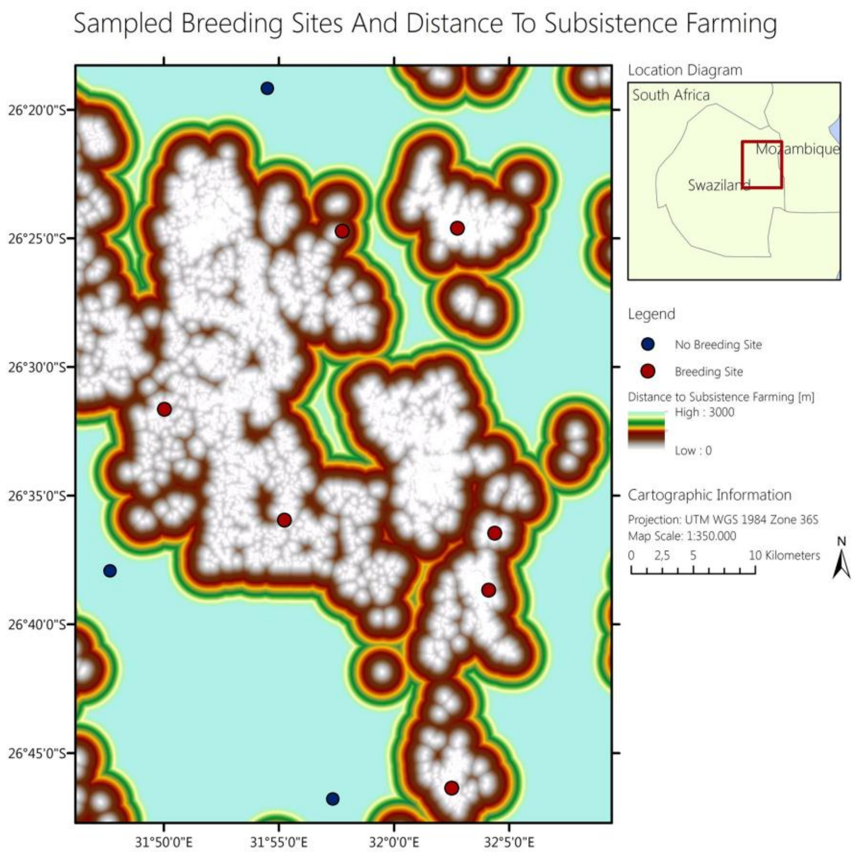

Figure 2.3.4: Example of the spatial distribution of some sampled vector breeding sites and the distance-to-subsistence farming in Swaziland

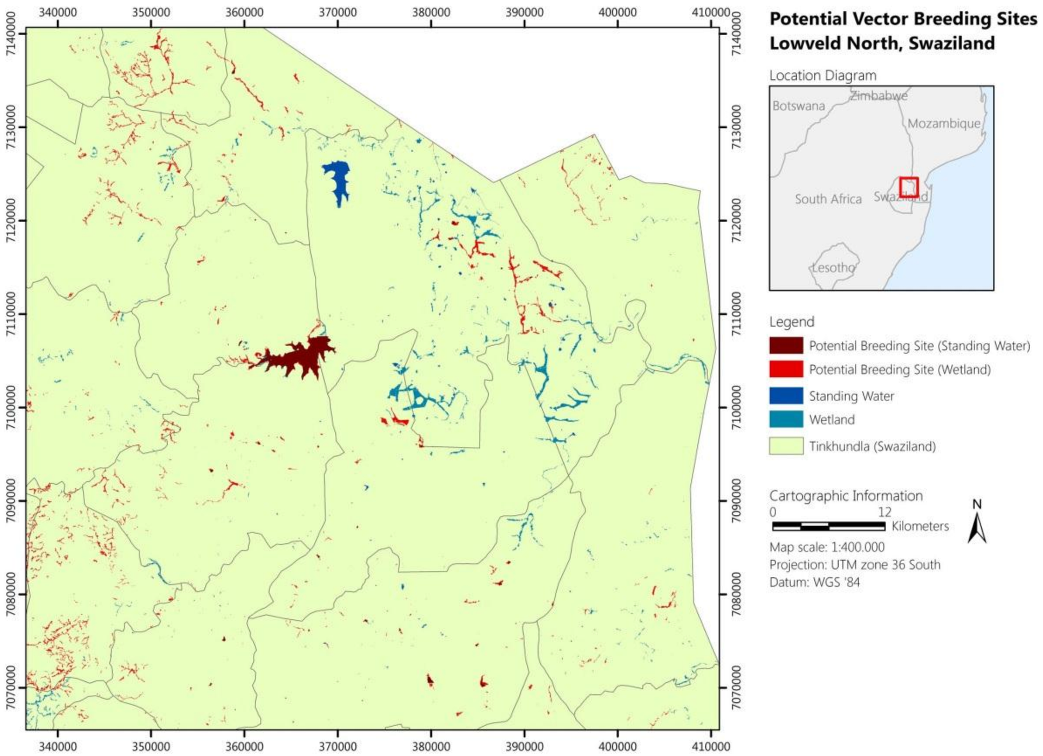

Figure 2.3.5 Example of the final potential vector breeding sites classification of the malaria-endemic area covered by the 5-m resolution RapidEye data (northern part of the Lowveld)

2.3.3 Logistic regression

The exploration for variable redundancies via correlation analysis showed that the average temperature for the 24 days before the larval sampling, which is an aggregate measure of all four temperature periods, was highly correlated with all other temperature variables (Pearson`s $r>0.8$) and was therefore removed from the analysis. The bivariate regression analysis indicated that the following variables were significant at the 15% significance level: Distance-to-subsistence farming, temperature and rainfall of the fourth week prior to larval sampling, altitude and distance-to-settlement. Results from the stepwise logistic regression suggest that the best model was the one including only distance-to-subsistence farming. The higher the distance from subsistence farming the lower the odds of presence of mosquito breeding sites with an odds ratio (OR) 0.21 (95%CI: 0.06-0.71) and a p-value of 0.012 (Table 2.3).

Table 2.3: Analytical results using three different statistical methodologies

Predictor	Bivariate logistic regression		Stepwise logistic regress.		Decision Tree analysis	
	OR (95%CI)	p-value[1]	OR (95%CI)	p-value	Ruleset attribute usage (%)	Cumulative classification accuracy (%)
Subsistence farming*						
Distance (m): <370	1.00					
Distance (m): 370-470	0.05 (0.01; 0.49)	0.003	0.05 (0.01; 0.49)	0.009	90	47.6
Distance (m): ≥470 m	0.21 (0.06; 0.71)	0.003	0.21 (0.06; 0.71)	0.012	90	47.6
Rainfall in 2nd week	1.03 (0.43; 2.48)	0.951			38	76.2†
Distance-to-savannah						
Distance (m): <150	1.00					
Distance (m): 150-620	0.56 (0.90; 3.52)	0.536			33	95.2‡
Distance (m): ≥620	0.93 (0.38; 2.24)	0.864			33	95.2‡
Temperature						
First week	0.64 (0.25; 1.63)	0.337				
Second week	0.76 (0.31; 1.87)	0.549				
Third week	0.93 (0.39; 2.25)	0.873				
Forth week*	3.98 (1.20; 13.23)	0.008				
24-day average temp.	0.75 (0.30; 1.84)	0.520				
Rainfall, first week	0.80 (0.33; 1.95)	0.614				
Rainfall, third week	0.81 (0.34; 1.97)	0.647				
Rainfall, forth week*	2.57 (0.85; 7.79)	0.060				
Altitude*	3.50 (1.08; 11.29)	0.016				
Distance-to-bare soil/rocks						
Distance (m): 120	1.00					
Distance (m): 120-700	1.20 (0.22; 6.68)	0.835				
Distance (m): ≥700	0.75 (0.30; 1.88)	0.536				
Distance-to-bush land						
Distance (m): <420						
Distance (m): 420-460	1.25 (0.22; 7.08)	0.801				
Distance (m): ≥460	0.94 (0.36; 2.41)	0.890				
Distance-to-forest						
Distance (m): < 700	1.00					
Distance (m): 700-720	1.75 (0.31; 10.02)	0.528				
Distance (m): ≥ 720	0.72 (0.26; 2.00)	0.525				
Distance-to-agriculture						
Distance (m): < 200	1.00					
Distance (m): 200-450	2.63 (0.45; 15.31)	0.277				
Distance (m): ≥ 450	1.39 (0.57; 3.43)	0.465				
Distance-to-main road						
Distance (m): <340						
Distance (m): 340-720	1.80 (0.32; 10.20)	0.504				
Distance (m): ≥720	1.16 (0.46; 2.88)	0.757				
Distance-to-settlement*						
Distance (m): <100						
Distance (m): 100-510	0.38 (0.07; 2.22)	0.277				
Distance (m): ≥510	0.39 (0 .12; 1.30)	0.089				

[1]Based on likelihood ratio test; *Variable included in the stepwise regression; †Combined use of Subsistence farming and Rainfall_week_2;‡Combined use of Subsistence Farming, Rainfall in the 2nd week and Savannah

2.4 Conclusions

In this study, we argue that exclusive reliance on RS and knowledge on how climatic factors and other environmental variables influence vector breeding is inadequate for the prediction of potential breeding sites, unless conducted with satellite-generated, high-resolution imagery in conjunction with information on presence or absence of larvae. To effectively guide control activities, it is imperative that actual entomological data on vector breeding and distribution be taken into account, e.g., when constructing regression models, so that previously not fully understood ecological effects and spatial breeding site heterogeneity can be elucidated. Linking entomological field collection with adequate remotely sensed environmental data does not only increase the accuracy of prediction models, but it also assists specification and identification of spatial heterogeneities with regard to vector breeding habitats. Understanding such variations may further facilitate determining the contribution and impact of vector control and other malaria control measures. In addition, as countries move towards malaria elimination, endemic transmission becomes limited to residual foci (Cohen et al., 2013), where prevention of larval breeding will be necessary. Most countries already have geographic databases on water bodies as well as entomological data such as larva scooping. It would be important that this information is incorporated into future databases.

Due to the current availability of only limited number of studies with detailed and explicit spatio-temporal variations on malaria transmission in Swaziland, it is not surprising that the country has not yet reached the goal of zero malaria cases, although current control efforts have been successful to reduce the overall burden. The present work is the first attempt to map potential breeding sites in Swaziland using remotely sensed data in conjunction with information on presence/absence of mosquito larvae. We have demonstrated that distance-to-subsistence farming is the main predictor for presence of active breeding sites. This can be explained by the fact that subsistence farming is a full-time occupation for most rural communities and the proximity of human hosts may explain the strong association between distance-to-subsistence farming and vector breeding sites. This finding is also consistent with a study by Ahmad et al., (2011), which deals with vector breeding habitats located between 100 and 400 m from human settlements. In the rural areas of Swaziland, almost every household is surrounded by hectare-sized subsistence farming fields and this translates to a continuous land use type as fields join at the edges. This situation supports the presence of foci and is thus an important factor for malaria endemicity. Similar result was found in a study by Li et al. (2008), which concluded that houses in great proximity to streams have more abundant mosquitoes than other places and that breeding was high in the nearby valley bottoms.

No relation could be found between breeding sites and distance-to-settlements in the present study. This could further be explained that cattle are not near settlements but exist in abundance at subsistence farming sites. From the field survey results, it was clear that the vector breeds very well in small water bodies down to the size of cattle hoof prints, which provide sheltered conditions. Turbulent water does not favour larvae development, which has already been observed in previous studies (e.g. Ageep et al., 2009), while Dejenie et al. (2011) remark that it is very likely that peripheral water bodies from irrigation channels and small depressions like cattle hoof prints , rather than the larger main dams, become preferred sites for mosquito larvae. Hence, it is very likely that locations characterized by cattle and proximity to water bodies are at risk due to the probable presence of vector breeding sites.

Rainfall was not found to be associated with vector breeding, which is not surprising as the more frequent and heavy the rainfall, the higher the possibility that mosquito larvae are washed away (Savage et al., 1990). As already alluded to, mosquito larvae were not likely to be found in areas reporting rain less than a week prior to the survey. However, as irrigation played a major role in providing excess water, it was not unexpected that rainfall had a lower attribute usage percentage (38%) in the Decision Tree analysis. Indeed, availability of water also in the dry season results in suitable habitats the year round facilitating oviposition compared to other parts where irrigation is not applied. This tells us that rainfall may not always be a useful predictor and that mosquitoes can be plentiful even without rainfall. Based on the discussion above, it would not be unusual that the use of rainfall as a covariate in logistic regression models even indicates a negative correlation, which is not always the case with other studies analysing environmental factors in relation to malaria transmission (Briët et al., 2008; Laneri et al., 2010). Thus, in areas where larval breeding occurs in the absence of rainfall, other environmental proxies will have to be used to identify and map potential breeding sites.

The maps produced in this study are not only useful for targeting residual foci, but could also help reducing the LSM budget. This would be achieved through the implementation of larvicide application guided by risk maps (Dambach et al., 2014).

Acknowledgements

The authors would like to thank the European Commission under the Seventh Framework Programme (FP7) for partially funding the project. Some EO data were provided by the ESA managed GSC-DA, funded under ESA–EC Agreement on the Implementation of the Space Component of Global Monitoring for Environment and Security (GMES). We thank the National Malaria Control Programme (NMCP) of Swaziland for conducting and coordinating the field data collection, the NMCP`s senior entomologist Mr Quinton Dlamini for leading the field team during the entomological survey and NMCP manager Mr Simon Kunene for approval to conduct the study using NMCP equipment. We would also like to acknowledge NMCP officer Mr Sibonakaliso Vilakati for technical assistance during the preliminary planning phase.

Contributions: SD, entomological data collecting, analysing and manuscript writing; JF, SD, remote sensing data processing, analysis and manuscript reviewing; PV, SD, JF, statistical analysis and manuscript reviewing.

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

Funding: the work was supported by the European Commission under the Seventh Framework Programme (FP7)

Chapter 3 Towards a consolidated use of remotely sensed data in epidemiology: a review of existing and potential products for vector-borne disease mapping

Sabelo Nick Dlamini[1,2], Anton Beloconi[1,2] , Penelope Vounatsou*[1,2]

[1]Swiss Tropical and Public Health Institute, Basel, Switzerland;

[2]University of Basel, Basel, Switzerland;

*Corresponding author

This manuscript has been accepted for publication in the International Journal of Remote Sensing

Abstract

There is a sustained interest in the use of remotely sensed (RS) data in vector-borne disease mapping and epidemiology. However, the use of satellite-derived products has been largely ad hoc and based on convenience rather than objectivity. We conducted an online review of literature to find out which RS products were available and how they can be obtained. Although satellite space launches have been meticulously documented, very little is said about the type of data they produce and what variables could be derived from the images captured. Our aim was to document all proxies previously used in epidemiological works and to propose new potential products that could be incorporated in vector-borne disease mapping. Furthermore, in view of the fact that RS products are evolving rapidly, data of higher spatial or temporal resolution than the ones used before were specified. We catalogued indices that have been used in ecological studies and listed potential variables and their sources providing a short description to aid the end-users to assess their relevance to epidemiological application. Our focus was on RS data products which have continental or global coverage, including data related to climate, meteorology, land use/cover, cartography and urban mapping and which can be used as proxies for disease suitability mapping. There remains a lot of work to be done on the evaluation and comparison of some of the proxies presented in this study against the conventionally and commonly used proxies in disease epidemiology; conversely synergies between remote sensing experts and epidemiologists can be useful in the uptake and testing of the novice products presented here.

Keywords: *Remote sensing, satellite-derived products, vector-borne disease mapping, epidemiology*

3.1 Introduction

There is unwavering interest in the use of remotely sensed (RS) data in vector-borne disease mapping and epidemiology (Machault et al., 2014; Garni et al., 2014). RS products are currently used to predict vector-borne disease distributions such as in malaria (Diboulo et al., 2015; Noor et al., 2014; Amek et al., 2012; Dhimal et al., 2014), soil-transmitted helminths (Karagiannis-Voules et al., 2015) and schistosomiases (Lai et al., 2015; Hu et al., 2015) among others. Previous works include incorporation of RS data in human and animal health studies such as Animal Trypanosomosis and Bluetongue (De La Rocque, 2004); predicting high risk areas for leishmaniasis in Brazil (Almeida and Werneck, 2014; Karagiannis-Voules et al., 2013); developing a national risk map for trachoma in Southern Sudan (Clements et al., 2010) and mapping tsetse fly habitat suitability (Robinson et al., 1997). One of the main advantages of using RS or Earth Observation (EO) products is the good spatial and temporal data resolution coverage, especially for economically-disadvantaged areas which have poor ground measurement station networks. Remote sensing can be particularly useful for monitoring areas where temporal variability in weather conditions, which affect vegetation cover, result in epidemics (Goetz et al., 2000).

From the first generation of ecological studies that demonstrated the capability of RS products in disease mapping (Beck et al., 1994; S. I. Hay, 1997; Thomson et al., 1997), there is a sustained proliferation of studies using RS data in epidemiology and geostatistical mapping, including identification of spatial heterogeneities, pattern and trends in vector-borne disease distributions and forecasting for epidemic preparedness (Giardina et al., 2014; Myers et al., 2000). The idea behind incorporating RS data in disease-mapping is justified by its established association with vector-borne disease distribution (Dlamini et al., 2015; Hassan et al., 1998; Tran et al., 2013). Other studies have demonstrated the association between radiation reflectance as measured by satellites and certain land cover types which have been used as proxies for measurement of presence of a disease and its vectors (Scholte et al., 2012; Innes and Koch, 1998, Curran et al., 2000). However, there has been very little effort made to inventorize existing and potential RS products and proxies that could be used by epidemiologists. Tatem et al. (2004) provided an overview of products relevant to epidemiology and public health, derived from MODIS and ASTER sensors. In most cases past remote sensing products selection criteria have been biased and ad hoc rather than objective and quantitative (Weiss et al., 2015). This has been partly due to data availability and convenience for epidemiologists as it often makes it hard for non-remote sensing experts to locate data related to their needs.

A compendium of civilian and commercial satellites that have been launched with the aim of gathering global Earth Observation (EO) and environmental monitoring data was prepared by (Belward and Skøien, 2015). However, they only documented satellite launches and not the type of data products that could be obtained from the land cover images captured by the satellites. Although a number of indices have been derived from RS images by remote sensing experts and ecologists, very little is known about potential indices or proxies relevant to epidemiological studies that could be derived from current satellite-based products. Furthermore, present documentation of RS data products and environmental proxies estimated from space satellites are ad hoc, incomplete and characterized by duplication and redundancy between access websites. Currently, very little has been done to document all variables and proxies that could be derived from RS images and how they could be used to help improve our vector-borne disease mapping efforts.

Previously, a National Aeronautics and Space Administration *(*NASA) annual bulletin called Spacewarn was launched in 1991 (http://nssdc.gsfc.nasa.gov/spacewarn/) to raise awareness about newly launched satellites and their missions to the general public. Unfortunately, this bulletin was discontinued with the last issue available until July 2011. The initiative headed by the European Commission (EC) in partnership with the European Space Agency (ESA) - Sentinel Scientific Data Hub - a remote sensing website project by Copernicus (http://www.copernicus.eu/) is currently undergoing establishment and promises to be a pool of all RS information around Europe and beyond. Furthermore, most of the websites provide data in a way that is not easily understood by epidemiologists as often coding is used with little elaboration on which products could be derived from the captured RS images. A comprehensive RS data review with the aim to catalogue RS products can assist epidemiologists efficiently access and compare various products.

In this paper we realize the need to make a review of all satellite-based products and proxies that could be used in epidemiological studies. Our review is different in the sense that it also provides a list of new indices that has a potential to be incorporated in vector-borne disease cartography. Although most of the indices have been previously used in ecological and air pollution studies, their uptake in epidemiology has been notoriously slow. A comprehensive catalogue of satellite sensors and specifications of their data products and proxies both those that are already processed by the supplier and those that could be derived by the end-user are presented.

3.2 Methods

3.2.1 Review, collation and inventorization of RS products

We conducted an online web based data triangulation search with the aim to review all existing remote sensing data sources including satellites and spacecrafts currently orbiting the earth. We used the search terms "remote sensing products, earth observation" in Google search engine and had over sixteen million hits in 22 seconds. We then collated the data from internet websites in order to filter duplicated remote sensing data sources after realizing that some hosting organizations were providing data from the same sensors. Both free access and commercial remote sensing products were collated according to the resolution of variables measured and the sensors used to capture them as well as a brief description of the proxies that could be obtained. As a first step we used a satellites list compiled by NASA (http://www.nasa.gov/missions/past/index.html#.VkCoYUYposI) and ESA (http://www.esa.int/ESA/Our_Missions) to find out which satellites were orbiting the earth and had been documented. We then searched for each of the satellites and the responsible organizations to learn more about their capabilities including the number of sensors, spatial and temporal resolutions and to find out which data products could be obtained from the satellites. While searching different online remote sensing websites in order to find out RS data sources, we found various inexhaustive data download ftp sites managed by both remote sensing agencies and private enthusiasts. For instance, such websites included NASA Goddard Space Flight Center, Sentinel Scientific Data Hub (Copernicus, European Space Agency), Observing Systems Capability Analysis and Review Tool (OSCAR), including NORAD Catalog (http://satellitedebris.net/Database/) among others.

In RS, the most solicited technical concerns in the use of EO data products inter alia, is the spatial resolution and the sampling frequency or temporal resolution at which the data is available. Spatial resolution is the maximum separating or discriminating power of a sensor measurement usually referred to as pixel size. Temporal resolution refers to the revisit period or length of time taken by a satellite to complete one orbit cycle. There is consensus that high resolution to very high resolution RS data is the best for vector-borne disease mapping (Franke et al., 2015; Dlamini et al., 2015). For some epidemiological applications, however, the temporal resolution, (e.g. land cover and climate change), plays a bigger role than the spatial resolution (Hugh-Jones, 1989; Riley, 1989). Equally important is the knowledge of the data preprocessing level conducted by the supplier and the data processing expected from the end-user. These technical details were used during products online search and profiling.

3.2.2 Satellites and products selection criteria

Satellites and their products were limited to ones with global or continental coverage and with data related to epidemiology and vector-borne disease mapping. We also considered products with continuous temporal resolution acquisition and coverage as opposed to once-off project based products. We consequently reviewed products that have either been applied in disease modeling or have the potential for such use. This perception was based on our knowledge and evidence from previous studies that incorporated RS variables in disease mapping. Each investigated RS product led to identification of the hosting agency from which the product could be obtained. This way we were able to find the websites where data could be downloaded by end-users. The list of products related to vector-borne disease mapping was divided in the following three categories:

- Meteorological and Climate data
- Land use/ Land cover
- Cartography and urban mapping

3.2.3 Processed vs. derived RS data

The proposed RS products were further split into two groups comprising of processed and derived variables. In our case processed variables refer to ones provided by the data supplying agency after all the necessary preprocessing steps had been conducted. These are products that could directly be fed into models without the necessity to undertake image processing or to derive the estimates of proxies as this has already been conducted before distribution. These constituted of data that could be used even by non-remote sensing experts such as epidemiologists. Derived variables or proxies are those that still require the end-user to estimate them from existing remote sensing products using either novice or established mathematical or physical algorithms. These proxies were investigated from previous studies which have either proposed them as new variables or have tested them in various spatial phenomena of ecological related studies. Thus, we reviewed literature on ecological studies to find out what new proxies or indices were proposed including the formulae used to derive them. We therefore arranged the RS variables and proxies according to the type of environmental condition measured and whether they were derived or already processed. The spatial and temporal resolutions of the products as well as their period of availability were documented.

3.2.4 RS products previously used in disease-mapping and new potential variables

From the review of ecological studies, we presented a library of proxies that have a potential application as variables in epidemiology and vector-borne disease mapping. We found a number of both new and old variables which were proposed by various ecologist and remote sensing experts. Most of the experts also provided the proxy derivation equations for potential end-users as well as presented the strengths and weaknesses of their index in terms of measuring a specific environmental condition in comparison to other known similar indices. We therefore provided a full reference list of all the authors of potential RS variables to aid the end-user to learn more about each index listed in our catalogue. These were indices whose proxy estimation measures the same environmental condition as those variables used in epidemiological studies. We highlighted the proxies which have already been used in epidemiology studies and those that still remained to be explored as potential variables. Thus we created a catalogue of all known and potential remote sensing products that could be used in vector-borne disease mapping with the aim to bring to the attention of end-users other similar indices to the ones already extensively utilized.

The most commonly used remote sensing products in epidemiology applications included proxies of temperature and precipitation i.e Land Surface Temperature (LST) and rainfall estimates (RFE). Vegetation indices such as Normalized Difference Vegetation Index (NDVI) and a host of land use/ land cover (LULC) variables were also widely used. Some new indices like the Temperature Suitability Index (TSI) which converts observed land surface temperatures into predictions of ambient air temperature for malaria distribution in Africa had also been used (Weiss et al., 2014). Other indices were improvement from previous versions, for example a 90 m water resolution database has been recently developed by Yamazaki et al., (2015) while a 15 m water resolution database was developed by Verpoorter et al., (2014). The new potential RS products are therefore important for epidemiology as they provide more explicit spatial detail than previous coarser resolutions of same. This is important in disease mapping especially identification of spatial heterogeneities and understating the underlying courses of spatial variations of certain vector-borne diseases.

3.3 Results

3.3.1 RS data sources

The Goddard Space Flight Center reported that there were about 2,271 satellites currently in orbit (Garner, 2015) while a NORAD Catalog website reports 7142 satellites deployed into space including debris (http://www.satellitedebris.net/Database) since October 1957. These satellites included those deployed for Earth Observation and environmental monitoring as well as global security satellites launched for private use. An example of an online profile of RS data download websites could be found in Exelis Visual Information Solutions (http://www.exelisvis.com/portals/0/pdf/6-14_Geospatial_Imagery_Raster_GIS_Data _ Sources.pdf) which is a remote sensing information site providing links to RS data products. Data access indicates that most RS data is available either free or commercially to end-users. The data products had varied spatial and temporal resolutions making it necessary that caution must be taken whenever they are jointly used during analysis. The results on remotely sensed data products show that those that are available for free have lower resolutions of about 250 m to 1.5 km compared to very high resolution products (from 5 m to 10 m) which are mostly available at high purchase prices.

3.3.2 Data processing

Remote sensing data products were varied according to the level of processing expected from the end-user. Some processing of the data is done by the supplier for most high demand products like NDVI, LST and rainfall. Processing levels vary from level zero which is raw products directly captured from measuring instruments into higher level where data is converted into various parameters of interest. Data preparation and processing includes atmospheric noise cleaning, missing values imputation, alignment correction and rectification of the products; this has usually been a big constraint for epidemiologists being unfamiliar with these issues. Agencies like NASA and ESA process their data into different levels and provide accompanying documentation for each dataset accessed via their ftp sites, making it very useful for the end-users like epidemiologist to *a priori* evaluate the usability of the data for their analysis. Typical data processing levels range from zero (unprocessed data) to four which is modeled output of data or variables derived from multiple measurements. Most epidemiologists use data that is processed up to level 3 and 4 as the basic processing level is meant for end-users with advanced processing skills and are capable of using geometric procedures to correct the images themselves.

3.3.3 Processed RS variables for vector-borne disease modeling

A review of the literature on RS derived proxies showed a litany of available remote sensing processed variables and potential indices that could be used for disease mapping. Most of them were found in ecological and RS studies where an unlimited number of indices that could be derived from remote sensing images are discussed. Numerous vegetation indices that could be used as an alternative to, for instance, NDVI were found. Some of these indices were said to be an improvement of the NDVI and therefore are expected to perform better for example in vector-borne disease distribution models. For example, the soil-adjusted vegetation index (SAVI) was developed in order to improve NDVI estimation by correcting the influence of soil brightness when vegetative cover is sparse.

A majority of the indices were derived by remote sensing experts and have been extensively used in ecological studies, while many of them still remain unknown to epidemiologist and public health experts. The main variables that were already processed by the suppliers included temperature, rainfall and NDVI among others and the same are the ones largely utilized in public health and disease modeling studies. Because they are readily available, their use is mostly out of convenience rather than based on their objective capability to provide better estimates of the distribution of the disease of interest. While epidemiological studies have shown that temperature and rainfall are important factors in the distribution of disease vectors (McMichael et. al., 1996), very few have assessed how modeling and mapping with such data could be affected by their resolutions.

A comparison of the traditionally used variables in epidemiology with the other similar indices which have been proposed in ecological studies is important in order to find out how they could jointly be used in disease modeling. Variables that were processed by the supplier were grouped and are presented in Table 3.1 which shows all data products that could be directly downloaded from remote sensing databases and are ready to be used by end-users. We provided the temporal and spatial resolutions of the data as well as the period at which the data is available. These are variables that are related to vector-borne disease mapping and therefore are of epidemiological importance.

3.3.4 Derived RS variables for vector-borne disease modeling

In Table 3.2 we present a library of indices that could be derived from remote sensing products and have been proposed by the ecologists who have calculated them. We do not imply that one index is more important than the other but we merely provide a description of what each of the indices estimates. We noted that there are a number of vegetation indices that have been calibrated to measure specific plant characteristics in relation to conditions such as moisture content and plant stress. We believe that the description provided can be used by epidemiologist to decide which index is more suitable for the vector-borne disease being mapped. Therefore, it is entirely up to the end-user to decide which indices to combine or to use jointly in their analysis work.

3.4 Discussion

This paper is not meant to be a comprehensive review of all the work that has been done in remote sensing as new satellites for global environmental monitoring purposes are continuously launched into space and new variable potential is presented. This work attempted to compile remote sensing information that is relevant to epidemiology and vector-borne disease mapping. Our motivation arose from the fact that RS coding and abbreviation including nomenclature that is completely different to epidemiology is used to identify data products; this often makes it hard for non-remote sensing experts to locate data related to their needs. Furthermore, remote sensing records and documentation show that launches are escalating each year, yet failed launches are not immediately known and consequently data availability from those missions remains a speculation.

Online websites documenting spacecraft launches and hosting data access sites are ad hoc and tend to present conflicting records as can be seen in the conflicting lists of satellites in orbit as recorded by the Goddard Space Flight Center (2 271) and the NORAD Catalog (7 142) respectively. Although remote sensing information on environmental monitoring can be obtained from agent websites, information from security satellites is not known yet some of their products are available for civilian use like Google Earth. Products from security satellites are usually of the highest quality in terms of both temporal and spatial resolution and they can be the best choice if they can be made available for civilian use when their use is justified such as for public health purposes.

Despite the good resolution commercial data has remained under-utilized mainly because of the high purchase prices associated with it thus limiting access for most end-users. Conversely poor resolution remote sensing products are available for free and a majority of epidemiology and vector-borne disease mapping studies are based on free remotely sensed data with resolutions ranging from 250 m to 1.5 km. In addition, we realized that most epidemiologists do not usually pay attention to the processing levels of the remotely sensed data they use for analysis whereas this is the key to understanding the amount of data preparation efforts needed before the data is ready for use.

There were some indices which could not be found from purported websites and sensor sources and they included humidity from MODIS Atmosphere and the Normalized Difference Water Index (NDWI) another MODIS product. For the processed data, time of availability from hosting websites and spatial resolutions varied across the globe and from region to region. For instance, there were also products that were limited only to national level like the Water resources which is available only for USA. In other cases a description of the data product was well provided while at the same time a direct link to access the same product was not provided. These were products like the Global Interannual Water Extent and Variation from the Special Sensor Microwave Imager; Global Inundation Extent from Multi-Satellites (GIEMS) and Shuttle Radar Topography Mission (SRTM) Water Body.

Some new products like the Water Mask may not be necessarily useful for vector-borne disease mapping since they are only binary proxies showing land and water. Other proxies like Soil moisture had limited spatial coverage as they were not available globally but were limited to national databases such as the US and some parts of Europe which are not accessible to the public. We also presented only the most recent spatial and temporal resolutions for all archived data and indicated when the lowest resolution became available. The resolutions of most of the derived indices depend on the choice of image that was used to derive them and it is possible that with recent improvements in image capturing sensors the proxies would also improve although they were proposed using products from the first generation sensors. The Global Inland Water facility was actually of poor quality for some parts of Southern Africa. This was seen when compared with existing groundtruthed water body map of Swaziland. Some of the existing water bodies were not captured by the Global Inland Water facility product yet the resolution was supposedly comparable at 30 m for both products.

There were many proxies that were proposed by ecologist which have however remained under-utilized partly because they are not fully understood and partly because they are time consuming to derive and interpret. For instance, a list of vegetation indices was found with some of the proxies advanced as an improvement to the commonly used NDVI such as the Atmospherically Adjusted Resistant Vegetation Index (ARVI) and the Soil-Adjusted Vegetation Index (SAVI). However the same indices remain under-utilized in vector-borne disease mapping and epidemiology presumably because no one has used them before. Highly utilized proxies were those of NDVI, rainfall and temperature partly because they are readily available and partly because their association with vector-borne disease distribution has been extensively investigated and established. We present a catalogue of potential and existing remotely sensed indices for end-users like epidemiologists to compare remote sensing products used in their analysis. Whereas some of these indices have been extensively explored in both ecological and remote sensing studies, their use in epidemiology has been limited to variables that are readily available from suppliers of remotely sensed data. This has been partly due to the lack of training in the processing and handling of remote sensing data for epidemiologist and due to lack of knowledge on how these indices can be used to improve our disease mapping efforts.

New vector-borne disease mapping potential is presented as high resolutions products are currently being developed as can be seen in the work of Yamazaki (2015) which shows an improved global water body map of up to 90 m spatial resolution and the GLObal WAter BOdies database project with 15 m spatial resolution. However, some products resolution is not exactly the same as documented from source suppliers. For example after careful examination of the spatial resolution of the G3WBM product using ArcGIS (version 10.2), we found the true spatial resolution to be about 135 m instead of the stated 90 m by the supplier. The present work is aimed at addressing the above issues by increasing the awareness on available satellite-derived products which have the potential to be incorporated in the vector-borne disease mapping field.

3.4.1 Conclusions

Documenting satellite launches into space and providing complete information on their mission and the anticipated data can be useful to interested parties to know what data is expected from these missions. A quick referral guide to remotely sensed data that has been coded or abbreviated on the hosting website can make searching for remote sensing products more efficient when end users can quickly find out what type of variables are available or can be derived from the coded images listed. We have presented a catalogue of potential proxies and currently used variables and indices in vector-borne disease modeling. Our list may not be complete as new indices are being derived from new satellite images but it will help as a guide for what is available to those seeking to use remote sensing products in their analysis. We also compiled a list of proxies that had previously been derived by ecologists but have received very little attention from epidemiologists due to some of the reasons already mentioned above. We noted that documentation of space activities is still ad hoc and uncoordinated with many sites duplicating and often providing conflicting statistics about space missions. This makes it hard for end-users to fully take advantage of the many data sources and products space agencies have to offer. There remains a lot of work on evaluation and comparison of some of the indices and proxies presented in this work against the conventionally and commonly used variables in disease epidemiology. Consequently, synergies between remote sensing experts and public health disease modelers can be useful in the uptake and testing of the novice indices presented in this work.

Acknowledgements

This work has been supported by the Swiss Government Excellence Scholarship program and by the European Research Council (ERC) advanced Grant project no. 323180.

Table 3.1: Supplier processed remote sensing variables

Variable	Source/ Sensor	Temporal resolution	Spatial resolution	Period of data availability	Description & data cost/availability
1) Meteorology and Climate					
Land Surface Temperature (LST) day and night	MOD11L2 MOD11(A1-A2) MOD11B1 MOD11 (C1-C3)	8 days	1 km (since 2000)	1960 - present	Measure of how hot or cold the "surface" of the earth is at a particular location. Data is free
Rainfall Estimates (RFE)	FEWS NET FAO-RFE	-daily -10 days -monthly	8 km	2008 - present	Measures the amount of accumulated rainfall from recent rain episode. Data is free
2) Landuse/Landcover					
Land Surface Water Index	MOD09A1	8 days	500 m	1981 - 2012	Measures the total amount of liquid water in vegetation and its soil background (Chandrasekar, 2010). Data is free
Normalized Difference Vegetation Index (NDVI)	MOD13Q1 MYD13A2	16 days	250 m (since 2000)	1999-present	Indicator used to assess whether the target being observed contains live green vegetation or not. Can be a proxy for water availability. Data is free
Enhanced Vegetation Index (EVI)	MOD13Q1 MYD13A2	16 days	250 m (since 2000)	1999-present	Designed to enhance vegetation sensitivity in high biomass regions and improved vegetation monitoring by correcting for atmospheric influences
Leaf Area Index (LAI)	MYD15A2 MCD15A3 MCD15A2H MCD15A3H	8 days	1 km	1960-present	Characterizes plant canopies and predict photosynthetic primary production, evapotranspiration and is a reference tool for crop growth (Bréda (2003)

Table continues next page

Variable	Source/ Sensor	Temporal resolution	Spatial resolution	Period of data availability	Description & data cost/availability
Global 3 arc-second Water Body Map (G3WBM)	Landsat Global Land Survey (GLS) 1975, GLS1990, GLS2000, GLS2005 and GLS2010	-	+90 m	2015	A high-resolution global water body map with information on the frequency of water body existence (Yamazaki et al., 2015)
Water Mask	MOD44W		250 m	2000	Measures surface water. Data is free
Land/Water mask	-Global Land Cover Facility (GLCF) -MODIS	-	250 m		Measures surface water as improvement from 1km MODIS mask data (Carroll et al., 2009). Data is free
Global Land Cover Facility Inland water (GIW)	-Landsat TM/ETM+	-	30 m	2015	Provides an estimation of regional and global inland water area (Feng et al., 2015). Data is publicly available
Global Lakes and Wetlands Database (GLWD)	Digital Chart of the World (DCW) of ESRI (1993)	-	1 km		Identifies global lakes and wetlands (Lehner and Döll, 2004). Available for free
Soil moisture/Geology maps	- WindSat - Tropical Rainfall Measuring Mission (TRMM) Microwave Imager (TMI)	3 days (archived)	1.5 km	2009	Used for agriculture, ecology, wildlife, and public health and is an important connection between the hydrological cycle and life involving animal, plant, and human (Lakshmi and Lakshmi, 2013). Data is free
Landcover	-MCD12C1 -MCD12Q1 -MCD12Q2	Yearly	500m (since 2001)	1999 - present	Shows how much of a region is covered by forests, wetlands, impervious surfaces, agriculture, etc. Data is free and also for sale

Table continues next page

Variable	Source/ Sensor	Temporal resolution	Spatial resolution	Period of data availability	Description & data cost/availability
Biodiversity/ human impacts maps	-World Atlas Biodiversity -World Map of Human Impacts	-	30 m	2003 - present	Used for understanding the environmental impacts of human populations. Data is free
Lights at night images	-NOAA National centres for environmental information -OLS, SSM/I, SSM/T, SSM/T2, SSJ, SSIES, SSM	-	500m	1992 - 2013	Indicates location and extent of human settlements. Cost about $ 43 per image (subject to scale &shipping costs)
3) Cartography and Urban Mapping					
Altitude/Geo morphology	-NASA -NOAA -ASTER -SRTM30 PLUS -GTOPO30	-	30 m	2011	Height above or below a fixed reference point as well as their topographic characteristics. Data is free
Ecoregions/B iogeographic regions	-Terrestrial Ecoregions Olsen et al. (2001)	-	30 km	2007 - present	Geographical units with characteristic flora, fauna and ecosystems. Data is free
Forest/wildlif e resources	-The world map of intact forest landscapes -World Wilderness Areas -UNEP GEO Data Portal	-	30 m	2000 - present	Shows unbroken natural landscape of a forest ecosystem and its habitat, plant community components, in a current extant forest zone. Data is free

- Details missing or not available

+Actual resolution coarser than the one stated

Table 3.2: Derived remote sensing indices

*Variable/ Index	Reference/ Source	Description
Temperature Suitability Index (TSI)"	(Weiss et al., 2014)	Converts observed land surface temperatures into predictions of ambient air temperature. Derived variable (data unavailable)
Vegetation Change Tracker	Zhao et al., (2015); Vogelmann et al., (2011)	Automated forest change mapping algorithm based on the spectral and temporal properties of forest, disturbance and post-disturbance recovery processes (Li et al., 2009). Calculated (data unavailable)
Soil-Adjusted Vegetation Index (SAVI)	Huete et al., (1988)	Developed as a modification of the NDVI to correct for the influence of soil brightness when vegetative cover is low. Calculated (data unavailable)
Modified Normalized Difference Vegetation Index (MNDV1)	Jurgens (1997)	Index used to determine frost damages in agriculture. Calculated (data unavailable)
Red Edge Normalized Difference Vegetation Index (NDVI705)	Gitelson and Merzlyak, (1994)	Used in precision agriculture, forest monitoring, and vegetation stress detection. Calculated (data unavailable)
Atmospherically Adjusted Resistant Vegetation Index (ARVI)	Kaufman and Tanre (1992)	The index is an enhancement to the NDVI and uses blue reflectance to correct red reflectance for atmospheric scattering. It is most useful in regions of high atmospheric aerosol content, including tropical regions contaminated by soot from slash-and-burn agriculture. Calculated (data unavailable)
Green Near Infrared (G/NIR)	http://www.exelisvis.com/docs/BroadbandGreenness.html	Can be used for soil property analysis. Calculated (data unavailable)
Green Shortwave Infrared (G/SWIR)	http://www.exelisvis.com/docs/BroadbandGreenness.html	Used to study spectral properties of soils and green vegetation. Calculated (data unavailable)
Simple Ratio Index (SR)	http://www.exelisvis.com/docs/BroadbandGreenness.html	It is described as the ratio of light that is scattered in the NIR range to that which is absorbed in the red range. Calculated (data unavailable)

Table continues next page

*Variable/ Index	Reference/ Source	Description
Sum Green Index (SG)	Gamon and Surfus (1999)	Used to detect changes in vegetation greenness and for detecting forest disturbance because it is highly sensitive to small changes in vegetation canopy opening. Calculated (data unavailable)
Modified Red Edge Simple Ration Index (mSR705)	Sims and Gamon (2002)	Used in precision agriculture, forest monitoring, and vegetation stress detection. Calculated (data unavailable)
Modified Red Edge Normalized Difference Vegetation Index (mNDVI705)	Datt (1999)	Used in precision agriculture, forest monitoring, and vegetation stress detection. Calculated (data unavailable)
Vogelmann Red Edge Index 1 (VOG1)	Vogelmann (1993)	Used in vegetation phenology (growth) studies, precision agriculture, and vegetation productivity modeling. Calculated (data unavailable)
Red Edge Position Index (REP)	Curran et al., (1995)	Used in crop monitoring and yield prediction, ecosystem disturbance detection, photosynthesis modeling, and canopy stress caused by climate and other factors. Calculated (data unavailable)
Photochemical Reflectance Index (PRI)	http://www.exelisvis.com	Uses pigments that signify photosynthetic light use efficiency and are useful to quantify vegetation production and stress. Calculated (data unavailable)
Structure Insensitive Pigment Index (SIPI)	http://www.exelisvis.com	Used in areas with high variability in the canopy structure, or leaf area index. Calculated (data unavailable)
Red Green Ratio Index (RGR Ratio)	http://www.exelisvis.com	Used for making foliage development estimations, indicating leaf production and stress. Calculated (data unavailable)
Plant Senescence Reflectance Index (PSRI)	http://www.exelisvis.com	Indicates increased canopy stress (carotenoid pigment), the onset of canopy senescence, and plant fruit ripening. Applications include vegetation health monitoring, plant physiological stress detection and crop production, and yield analysis. Calculated (data unavailable)
Carotenoid Reflectance Index 1 (CRI1)	http://www.exelisvis.com	Measure of stressed vegetation as a consequence of harmful effects of too much light. Calculated (data unavailable)

Table continues next page

*Variable/ Index	Reference/ Source	Description
Carotenoid Reflectance Index 2 (CRI2)	http://www.exelisvis.com	Modified measure of stressed vegetation as a consequence of harmful effects of too much light. Calculated (data unavailable)
Anthocyanin Reflectance Index 1 (ARI1)	Gitelson et al., (2001)	Measure of stressed vegetation and indicates changes in foliage via new growth or death. Calculated (data unavailable)
Anthocyanin Reflectance Index 2 (ARI2)	Gitelson, et al., (2001)	Detects higher concentrations of anthocynanins in vegetation. Measures stressed vegetation and indicates canopy changes in foliage via new growth or death. Calculated (data unavailable)
Water Band Index (WBI)	Penuelas et al., (1995).	Estimates leaf moisture and water content of vegetation. Calculated (data unavailable)
Moisture Stress Index (MSI)	Ceccato et al., (2001).	Indicator of plant water content and water stress. Calculated (data unavailable)
Normalized Difference Infrared Index (NDII)	Hardisky et al., (1983)	Used to detect plant water stress. Calculated (data unavailable)
Normalized Dry Matter Index (NDMI)	Wang et al., (2013)	Estimates the dry matter content in green leaves. Calculated (data unavailable)
Normalized Difference Tillage Index (NDTI)	van Deventer et al., (1997)	Used for differentiation between crop residue and soils. Calculated (data unavailable)
Cellulose Absorption Index (CAI)	Nagler et al., (2000)	Relative depth of cellulose absorption by non-photosynthetic vegetation/biomass cover (Daughtry et al., 2005). Calculated (data unavailable)
Normalized Difference Lignin Index (NDLI)	Serrano et al., (2002)	Estimates Ligno-cellulose content or mass of senesced plant materials. Calculated (data unavailable)
Normalized Difference Nitrogen Index (NDNI)	Serrano et al., (2002)	For foliar nitrogen concentration estimation. Calculated (data unavailable)

Table continues next page

*Variable/ Index	Reference/ Source	Description
Ligno-Cellulose Absorption Index (LCA)	Daughtry et al., (2005)	Used for live and senesced biomass (Numata, 2008). Calculated (data unavailable)
Shortwave Infrared Normalized Difference Residue Index (SINDRI)	Serbin et al., (2009)	Estimates the amount of crop residue cover over multiple locations. Calculated (data unavailable)
Dry Matter Content Index (DMCI)	Romero et al., (2012)	Used for estimation of plant canopy biomass (Wang L. et al. 2013). Calculated (data unavailable)
Reciprocal of Moisture Stress Index (RMSI)	Hunt and Rock (1989)	Measure of the effects of drought and catastrophic plant wetness (https://www.ncdc.noaa.gov/societal-impacts/cmsi/). Calculated (data unavailable)
Simple Ratio Water Index (SRWI)	Zarco-Tejada et al., (2003)	Calculated (data unavailable)
Plant Water Index/ Plant Water Concentration (PWC)	Penäuelas et al., (1995)	Measures amount of water concentration in plants as a proxy for drought assessment. Calculated (data unavailable)
Modified Soil Adjusted Crop Residue Index (MSACRI)	HongRui et al., (2012)	Estimating regional non-photosynthetic biomass. Calculated (data unavailable)
Soil Adjusted Corn Residue Index (SACRI)	HongRui et al., (2012)	Estimating regional non-photosynthetic biomass. Calculated (data unavailable)
Difference Vegetation Index	Tucker (1979)	Distinguishes between soil and vegetation, but it does not account for the difference between reflectance and radiance caused by atmospheric effects or shadows (“Broadband Greenness (Using ENVI) | Exelis VIS Docs Center,” n.d.). Calculated (data unavailable)
Normalized Difference Index (NDI)	HongRui et al., (2012	Estimating regional non-photosynthetic biomass. Calculated (data unavailable)

Table continues next page

*Variable/ Index	Reference/ Source	Description
Normalized Difference Temperature Index (NDTI)	(Peng et al., 2013)	Used for approximating moisture availability. Calculated (data unavailable)
Global Environmental Monitoring Index	Pinty, B., and M. Verstraete (1992)	Used for global environmental monitoring from satellite imagery and attempts to correct for atmospheric effects. Calculated (data unavailable)
Green Atmospherically Resistant Index (GARI)	Gitelson et al., (1996)	This index is more sensitive to a wide range of chlorophyll concentrations and less sensitive to atmospheric effects than NDVI. Calculated (data unavailable)
Green Difference Vegetation Index (GDVI)	Sripada, et al., (2006)	Used to predict nitrogen requirements for corn. Calculated (data unavailable)
Green Normalized Difference Vegetation Index (GDVI)	Gitelson and Merzlyak (1998)	This index is more sensitive to chlorophyll concentration than NDVI. Calculated (data unavailable)
Green Ratio Vegetation Index (GRVI)	Sripada et al., (2006)	Index minimizes the effects of background soil while emphasizing green vegetation and uses global coefficients that weigh the pixel values to generate new transformed bands. Calculated (data unavailable)
Infrared Percentage Vegetation Index (IPVI)	Crippen (1990)	This index is functionally the same as NDVI, but it is computationally faster. Calculated (data unavailable)
Modified Non-Linear Index (MNLI)	Yang et al., (2008)	This index is an enhancement to the Non-Linear Index (NLI) that incorporates the Soil Adjusted Vegetation Index (SAVI) to account for the soil background. Calculated (data unavailable)
Non-Linear Index (NLI)	Goel and Qin (1994)	This index assumes that the relationship between many vegetation indices and surface biophysical parameters is non-linear. Calculated (data unavailable)

Table continues next page

*Variable/ Index	Reference/ Source	Description
Optimized Soil Adjusted Vegetation Index (OSAVI)	Rondeaux et al., (1996)	The index provides greater soil variation than SAVI for low vegetation cover, while demonstrating increased sensitivity to vegetation cover greater than 50% and is best used in areas with relatively sparse vegetation where soil is visible through the canopy. Calculated (data unavailable)
Renormalized Difference Vegetation Index (RDVI)	Roujean and Breon (1995)	Used to highlight healthy vegetation and is insensitive to the effects of soil and sun viewing geometry. Calculated (data unavailable)
Transformed Difference Vegetation Index (TDVI)	Bannari et al., (2002)	Useful for monitoring vegetation cover in urban environments as it does not saturate like NDVI and SAVI. Calculated (data unavailable)
Visible Atmospherically Resistant Index (VARI)	Gitelson et al., (2002)	Used to estimate the fraction of vegetation in a scene with low sensitivity to atmospheric effects. Calculated (data unavailable)
WorldView Improved Vegetative Index (WV-VI)	Wolf (2010)	This index uses WorldView-2 bands to compute NDVI. Calculated (data unavailable)
Topographic Wetness Index (TWI)”	Cohen et al., (2013)	Measure representing the amount of water that should enter a given spatial unit divided by the rate at which the water should flow out of that unit (Sørensen et al., 2006). Calculated (data unavailable)
Urban Index	Kawamura et al., (1996)	Used as an indication for urban built up area intensity. Describes building density. Calculated. (data unavailable)
Normalized Difference Impervious Surface Index (NDISI)	Xu et al., (2010)	Used to investigate the impervious surfaces an area .Calculated (data unavailable)

*Resolution depend on the image that is used during index derivation

“Indicates index already used in epidemiological studies

Chapter 4 Bayesian geostatistical modeling to assess spatio-temporal variations and elapsing time for malaria incidence risk in Swaziland

Sabelo Nick Dlamini[1,2], Federica Giardina[3,4], Penelope Vounatsou*[1,2]

[1]Swiss Tropical and Public Health Institute, Basel, Switzerland;

[2]University of Basel, Basel, Switzerland;

[3]Stockholm University, Stockholm, Sweden

[4]Karolinska Institute, Stockholm, Sweden

*Corresponding author

This is a working manuscript for submission to the Malaria journal

Abstract

Malaria has drastically declined in Swaziland and the country aims for elimination by the year 2016. Spatially explicit maps on micro epidemiological variations in malaria incidence risk are needed to aid the country to target and prioritize interventions in the high risk areas for impact. The study developed a polynomial distributed lags model of up to three months to assess the relative contribution of climatic factors, mosquito breeding sites and human related factors in space and time. A Bayesian geostatistical and temporal negative binomial model was fitted on malaria incidence data to predict the disease at high spatial and temporal resolutions. Thus we produced the country's first model based monthly maps of malaria incidence risk. The results showed that LST day and NDVI lags were associated with malaria after about $1\frac{1}{2}$ months (Lag 4) while LST night was important from the first two weeks up to 1 month (Lags 2 & 4). On the other hand rainfall was negatively associated with malaria on the first month (lag 2) and became important in the second and third month (Lags 5 & 6). Maps of malaria incidence prediction showed high rates in the months of July and September (> 8.2 cases per 1000 people). February and March had the lowest incidence rates at less than 5.3 cases per 1000 people. This could be attributed to the heavy rains that are received around this time, thus inhibiting transmission while the peak in the drier period of July could be a result of previous rainfall episodes. These maps could be useful for timing and targeting of control interventions by the control programme.

Keywords: *malaria elimination, Bayesian modeling, remote sensing, Swaziland*

4.1 Introduction

Recent declines in malaria transmission in Southern Africa shifted the focus of national control programs in the region towards elimination (Cohen *et al*., 2013; Hsiang *et al*., 2012). Swaziland, Namibia, Botswana and South Africa constitute the elimination frontline countries, whereby Swaziland is already on its elimination path which was first targeted to begin in 2015 until certification by the World Health Organization in the year 2018. During the interval from the target year to certification, there should be zero locally acquired malaria cases, however bringing cases down to zero remains operationally challenging (Moonen *et al*., 2010). For instance, the country recorded about 603 malaria cases in the transmission season of July 2014-June 2015 forcing the National Malaria Control Programme (NMCP) to shift its goal of elimination by one year from 2015 to end of 2016. Indeed with recent funding from the Global Fund, the country is geared towards implementing its revised Malaria Elimination Strategic Plan of 2015 to 2020. Consequently, strong surveillance and sustained control with evidenced based intervention strategies is needed in the critical phase of elimination to qualify the country as malaria free.

Swaziland, a landlocked country located in Southern Africa has already scaled up its malaria interventions in an effort to eliminate local transmission by 2016 (Ministry of Health, 2015). According to Churcher *et al.,* (2014) Swaziland has already halted endemic transmission. Nonetheless, recent malaria data show that the country will continue to struggle with emerging cases and importation from neighboring regions resulting in seasonal uncertainties on the national incidence rates. For instance, a study by Koita *et al*., (2013) showed that importation from Mozambique accounts for over 90% of malaria transmission in Swaziland. Furthermore, data from the Swaziland malaria control programme for the transmission season proceeding the then target year of 2015 showed an alarming 82% increase in local cases (112 to 204 malaria cases) for the seasons July 2012 to June 2013 and July 2013 to June 2014. Such seasonal upsurge in cases reemphasizes the need for sustained interventions and watchfulness even when endemic transmission has been halted.

In Swaziland, malaria transmission occurs on the eastern part of the country where it shares a border with Mozambique and on the south eastern part bordered by South Africa (NMCP, 2010). Seasonal low unstable transmission characterizes the malaria situation where its peak is associated with episodes of high rainfall during the summer season between November and May each year (Ministry of Health, 2012). Surveillance is an integral part of the malaria elimination efforts in Swaziland and as such since October, 2009 it has been strengthened to include reactive surveillance through follow up of individual cases reported at health facilities (Cohen *et al*., 2013). This involves collection of demographic information about the patient, travel history which helps in classification of cases as either local or imported based on onset of symptoms. GPS coordinates of the case's residence are also captured.

Studies on spatio-temporal variations in malaria incidence are important not only to assess the problem of malaria in a given region, but also to analyze the effectiveness of preventive strategies (Zacarias and Andersson, 2011). Geostatistical spatio-temporal models are useful for identifying spatial heterogeneities and for assessing those environmental factors associated with disease incidence. These models can be used as tools for assessing both primary and secondary response prevention measures, as well as their effect and impact. In addition spatially explicit model-based maps on micro epidemiological variations are important for malaria elimination as endemic transmission declines to residual foci. These maps are still needed in order to aid surveillance and vector control efforts and to better target and prioritize planned interventions (Giardina *et al*., 2012; Gosoniu *et al*., 2012).

Currently, existing malaria maps in the country are descriptive geolocation of individual cases with little information about the explaining background factors responsible for transmission. Planning and implementation of intervention in a cost effective manner requires more explicit and reliable maps on the geographic distribution of malaria incidence at high spatial and temporal resolution (Kulkarni *et al*., 2010). These maps estimate incidence risk over gridded surfaces of factors driving the micro epidemiology of the disease. A number of factors have been associated with malaria micro-epidemiology such as distance to water body, breeding sites, vegetation cover including population movements (Bousema *et al*., 2012). The contribution of these factors to malaria transmission depends on the local conditions which can be derived from climatic and environmental proxies. In Swaziland studies investigating the contribution of these micro epidemiological factors together with climatic proxies associated with local malaria transmission are limited and have until recently been not available.

Geostatistical models for mapping survey data have been developed to produce high resolution maps of malaria risk (Gosoniu *et al.*, 2010), however very few efforts have been made for modeling incidence data in very low endemic settings characterized by diminishing malaria and episodic individual cases such as observed in Swaziland. Spatial heterogeneities at fine geographic scales can be used to identify patterns in malaria transmission and therefore understand the variations in time and space of the driving factors for better planning, preparedness and response (Amek *et al*. 2012). A combination of Bayesian modeling with Geographic Information Systems (GIS) have led to advancements in the field of spatial epidemiology and disease mapping. Consequently, it is now possible to explore and characterize different sets of spatial disease patterns at a very fine geographical resolution (Banerjee, 2004).

In this study we used Bayesian geostatistical modeling to assess the association between climatic factors and malaria incidence in space and time. A Bayesian geostatistical negative binomial model was fitted on malaria incidence data using a polynomial distributed lag function. We chose distributed lag models (DLMs) because they are useful when the outcome of interest is a result of a cumulative effect from previous time periods. They provide the estimate of the best distributed lag function which describes changes in risk factors for the outcome of interest. This function can be used to assess if the effect of risk factors on the outcome is immediate or rather slowly as a result of a build up from previous conditions. In case of malaria, studies have shown that the elapsing time between climatic factors and onset of malaria depends on past weather conditions Najera *et al.,* (1998) in what is termed epidemic buildup. Understanding the elapsing time and its effect on malaria transmission is important for planning, preparedness and response. Geostatistical models relate the disease data with potential predictors and quantify spatial dependence via the covariance matrix of Gaussian process facilitated by adding random effects at the observed locations (Karagiannis *et al*., 2013). In this study potential predictors included: environmental variables such as rainfall, temperature and normalized difference vegetation index (NDVI) as well as distance to water bodies, altitude, landuse and landcover.

4.2 Methods

4.2.1 Malaria incidence data

Geolocated malaria incidence data for a 5 year period (2010-2015) was obtained from the National Malaria Control Programme (NMCP) of Swaziland. The data comprised of reactively investigated symptomatic cases that have presented at health facility. Cases were already classified into either imported or local based on travel history of patients by investigating officers from the NMCP. Local cases with valid geographic coordinates were aggregated by enumeration area (EA) which is the lowest census unit and analyzed. In addition cases identified via active cases detection in the neighboring households of the index case were also included in the analysis. The population in each of the EA was used as an offset in the negative binomial model in order to take into account the spatially contextual background. Landuse of the EA was also assessed in the model. The data were organized according to the malaria transmission season which is July to June each year. Out of a total of 1229 malaria cases, 43% were local cases and only these were used during analysis.

4.2.2 Environmental data collection and processing

Remotely sensed climatic data were downloaded from the Reverb ECHO system (NASA, 2007). These data included a 1 km resolution normalized difference vegetation indices (NDVI) available biweekly and 1 km day and night land surface temperature (LST) emissivity indices both available weekly. The data are products of the Moderate Resolution Imaging Spectroradiometer (MODIS). Eight km resolution dekadal rainfall was obtained from the Africa Data Dissemination Service (ADDS), a data portal for the Famine Early Warning Systems (FEWS) network. In addition, a 30m resolution Digital Elevation Model (DEM) from ASTER (Advanced Spaceborne Thermal Emission and Reflection Radiometer) and a 1 km resolution landcover image from MODIS was obtained for Swaziland. To understand the association between malaria incidence and environmental conditions, all remotely sensed climatic factors were extracted at each EA centroid with a reported malaria case. A summary of remote sensing data is presented in Table 4.2.

Table 4.2: Variables used in analyzing malaria incidence

Variables	Spatial Resolution	Temporal Resolution	Period	Source
Normalized Difference Vegetation Index (NDVI)	1km	16-days	2009-2014	MODIS
Land Surface Temperature (LST)	1km	8-days	2009-2014	MODIS
Rainfall Estimates	8km	10-days	2009-2014	ADDS
Land Cover	1km	Yearly	2010-2012	MODIS
Altitude	30m	Yearly	-	ASTER

4.2.3 Bayesian geostatistical modeling

A Bayesian model based on distributed lags was formulated and implemented in order to better understand the association between environmental factors and an increased number of malaria reported cases (malaria incidence risk) at distributed lags of up to 12 weeks (3 months). This was done to take into account the possible elapsing time (lag) between the predictive variables which included rainfall, LST and NDVI and the outcome variable (malaria case) and also to determine the best combination of lags that predicted malaria incidence risk. Other factors which consisted of fixed terms included landcover, landuse, altitude and distance to water bodies. Extracted environmental data and malaria incidence data were first processed in STATA version 13.0 and then fitted to a negative binomial model in WinBUGS using a polynomial function constrained to power two as follows: Let Y_i be the average number of malaria cases at a given location s= $i,\ldots.n$ with likelihood Y_i ~dnegbin(P_i, r) where P_i is the proportion of malaria cases in a defined location and r is the dispersion parameter and $\mu_i = r\frac{1-p}{p}$ while $\sigma_i^2 = r(1-p)p^{-2}$. The model was written as: $Logit(\mu_i) = logit(Popu_i) + \beta_0 + \beta_1 X_i \ldots\ldots \beta_{12} X_i + \varepsilon_i$, where μ_i is the number of malaria cases in each location and β are the regression coefficients, X are the model covariates and ε are temporal random effects. The distributed lag model was restricted to a polynomial function of power 2 which was formulated as follows:

$$\beta_i = \sum_{k=0}^{2} a_k \, i^k$$

Where k is the categorical variable for the covariate corresponding to β_i coefficient and a is the intercept for locations $i \ldots . n$. The model describes the relationship between an independent value of X_i and the corresponding dependent mean Y_i. This is summarized as $E(y|x)$. The model gives the expected μ_i of malaria cases given the corresponding value of each categorical variable at location s.

4.2.4 Determining important lags using Bayesian variable selection

We applied Bayesian variable selection to determine the most important lag time between environmental factors and the onset of malaria incidence. We used a Stochastic Search Variable Selection method that tries to find those independent variables that are greatly associated with the outcome variable of interest thus allowing us to fit the model only for those variables that are important in the final model. The set of β_i predictors were fixed into polynomial function describing the distribution of each set of predictors where the third power was selected following first stage testing of the different polynomials. The model was then restricted to power 4 for all the predictors comprising of LST, NDVI and Rainfall. For each of the polynomial functions x_i we introduced a binary indicator with 50% chance of inclusion into the final model by restricting the variable selection to a Bernoulli distribution with probability of inclusion whereby the best set of covariates was indicated by the model with the highest posterior probability ranging from 0 to 1. Any variables with coefficients above 50 % were selected for the final model. An inverse gamma prior was used. To enable prediction we ran the model for 444 parameters using Markov Chain Monte Carlo (MCMC) where prior distributions were assigned to the parameters in order to complete the model formulation. We then applied Bayesian kriging to predict the malaria incidence risk at unsampled locations and produce a parasitaemia risk map at 1 km spatial resolution. We used a randomly selected sample of 150 locations as a training set for fitting the final prediction model.

4.3 Results

4.3.1 Bayesian geostatistical modeling

Spatial variations of malaria incidence with climatic and environmental factors were estimated for each full transmission season in Swaziland. Results were predictions of malaria incidence presented in smoothed surfaces showing monthly variations in malaria incidence making it possible to visualize months of initial and peak transmission in the country. The predictions were based on bi-weeks polynomial lags of up to power 4 which were aggregated up to three months. There was a positive association in increase of temperature, rainfall and NDVI and malaria cases at polynomial lags of up to three months. For instance the current bi-week (LST day[1]) was positively associated with malaria incidence 2.18 (95% CI: 0.98 to 3.19) while the first lag or power 1 (LST day[2]) was negatively associated with malaria cases -2.63 (95% CI: -2.89 to -2.34).

Interestingly, third bi-week or power 2 (LST day[3]) and fourth bi-week or power 3 (LST day[4]) were again positively associated with malaria cases (0.1, 95% CI: 0.05 to 0.17 and 0.22, 95% CI: 0.22 to 0.2357) until the situation start to change in bi-week or power 4 (LST day[5]) which is negatively associated. In LST night, the effects of the lags was different in the sense that the current week and the first lag were not associated with malaria incidence until the second bi-week or power 2 (LST night[3]) which was negatively associated with malaria (-0.12 95% CI: -0.13 to -0.10). The third bi-week or power 4 was also positively associated with malaria which again changes in bi-week of power 4 with negative association. On the other hand NDVI was negatively associated with malaria until the third power (NDVI[4]) which was yet again positively associated (0.06, 95% CI: 0.05 to 0.06). Rain showed negative association in the power 1 (Rain[2]), power 3 rain (Rain[4] and power 4 (Rain[5]) and was only positively associated with malaria in power 2 (Rain[3]) with mean 0.23 (95% CI: 0.22 to 0.25). The results are summarized in Table 4.3 which shows the posterior probabilities of the variables polynomial functions (x^2)

Table 4.3: Posterior estimates of the distributed lags constrained to power four

Variable(x^2)	mean	sd	val2.5pc	median	val97.5pc
LST day[1]	2.18	0.657	0.9857	2.326	3.19
LST day[2]	-2.631	0.1884	-2.89	-2.662	-2.341
LST day[3]	0.1064	0.03389	0.05612	0.09868	0.1741
LST day[4]	0.2284	0.004625	0.2206	0.2295	0.2357
LST day[5]	-0.02924	3.96E-04	-0.03004	-0.02922	-0.02875
LST night[1]	0.006163	0.3635	-0.5658	0.1021	0.5285
LST night[2]	0.2757	0.2006	-0.00967	0.2314	0.5429
LST night[3]	-0.124	0.008557	-0.1372	-0.1264	-0.1077
LST night[4]	0.02868	0.00142	0.02657	0.02925	0.03084
LST night[5]	-0.00271	3.32E-04	-0.00326	-0.00258	-0.00221
NDVI[1]	-0.3156	0.3682	-0.996	-0.2119	0.2182
NDVI[2]	-0.2407	0.1474	-0.4127	-0.2853	0.008079
NDVI[3]	0.001152	0.01623	-0.01974	-0.00238	0.04276
NDVI[4]	0.06359	0.003442	0.05515	0.06441	0.06808
NDVI[5]	-0.00881	2.29E-04	-0.00917	-0.00876	-0.00847
Rain[1]	0.4434	0.4889	-0.9898	0.6221	0.9489
Rain[2]	-0.7833	0.1917	-0.9561	-0.8864	-0.1917
Rain[3]	0.2381	0.01018	0.22	0.2379	0.2587
Rain[4]	-0.01037	0.002428	-0.01754	-0.0102	-0.00579
Rain[5]	-0.00132	2.69E-04	-0.00197	-0.0013	-7.94E-04

The bi-week lags were also fitted into the negative binomial model as fixed covariates in order to assess the effects of each lag on malaria cases. Other fixed non time varying covariates like altitude, waterbody and seasonality, were also added into the model. The earlier lags of the 2nd and 3rd bi-week for LST day were negatively associated with malaria compared to later lags of the 4th, 5th and 6th bi-weeks. The effects changes from a negative association to a positive one after about one month. Interestingly LST night was positively associated with malaria cases in the 2nd and 3rd bi-week lags with means 0.49 (95% CI: 0.17 to 0.91), 1.9 (95% CI: 1.55 to 2.52) and 1.66 (95% CI: 1.39 to 1.62) respectively which jointly are also equivalent to an effect of about one month. Similarly NDVI was negatively associated with malaria in the earlier lags of about $1\frac{1}{2}$ month (bi-week lags 1, 2 & 3) and was positively associated in the latter lags of the 2nd and 3rd months (bi-week lags 4, 5 & 6). Rainfall was negatively associated with malaria case on the 2nd bi-week lag and positively associated in the 5th and 6th lags of the 2nd and 3rd months with means 0.35 (95% CI: 0.09 to 0.93) and 0.36 (CI: 0.16 to 0.67) respectively. The results are summarized in Table 4.4 and the resultant maps are displayed in Figures 4.3.1, 4.3.2 and 4.3.3.

Table 4.4: Posterior probabilities for fixed bi-week lags of environmental factors

variable	Lag	mean	sd	val2.5pc	median	val97.5pc
LST day	**Lag 1**	-0.1451	0.445	-1	-0.05983	0.5578
	Lag 2	-1.297	0.2059	-1.739	-1.261	-0.9551
	Lag 3	-0.9561	0.09027	-1.122	-0.9623	-0.7872
	Lag 4	0.4922	0.2407	0.1771	0.4447	0.9126
	Lag 5	1.963	0.3005	1.555	1.892	2.528
	Lag 6	1.668	0.1928	1.393	1.625	2.091
	Lag 7	-2.883	0.2861	-3.391	-2.884	-2.448
LST night	**Lag 1**	0.1838	0.1798	-0.1447	0.2328	0.4541
	Lag 2	0.2475	0.07481	0.1214	0.2427	0.4208
	Lag 3	0.2718	0.1498	0.03271	0.2404	0.577
	Lag 4	0.2662	0.199	-0.0304	0.1975	0.6114
	Lag 5	0.1751	0.1532	-0.04733	0.1231	0.4599
	Lag 6	-0.1222	0.1056	-0.3501	-0.1102	0.06071
	Lag 7	-0.8112	0.479	-1.7	-0.6842	-0.03606
NDVI	**Lag 1**	-0.5004	0.2157	-0.9203	-0.459	-0.1579
	Lag 2	-0.4247	0.07643	-0.5662	-0.4235	-0.27
	Lag 3	-0.02417	0.112	-0.1951	-0.05566	0.1894
	Lag 4	0.554	0.1873	0.3336	0.4903	0.9159
	Lag 5	0.9512	0.1771	0.7373	0.8958	1.299
	Lag 6	0.5974	0.0683	0.4687	0.595	0.7303
	Lag 7	-1.289	0.3305	-1.922	-1.184	-0.8273
Rain	**Lag 1**	-0.1135	0.3018	-0.9562	-0.02793	0.2342
	Lag 2	-0.2749	0.1329	-0.5975	-0.2561	-0.07677
	Lag 3	-0.1505	0.09798	-0.3069	-0.1598	0.08106
	Lag 4	0.1186	0.1824	-0.1032	0.06273	0.6572
	Lag 5	0.3595	0.2065	0.09986	0.2959	0.9364
	Lag 6	0.3677	0.1306	0.1619	0.3469	0.6715
	Lag 7	-0.0928	0.1959	-0.5978	-0.0567	0.189
Wet season	-	1.532	0.9761	0.04397	1.381	3.896
Dry season	-	-0.07404	9.999	-19.6	-0.07795	19.65
Altitude	-	-0.2638	0.1592	-0.5927	-0.2573	0.03222
Waterbody	-	0.01113	0.09569	-0.1697	0.007787	0.2111

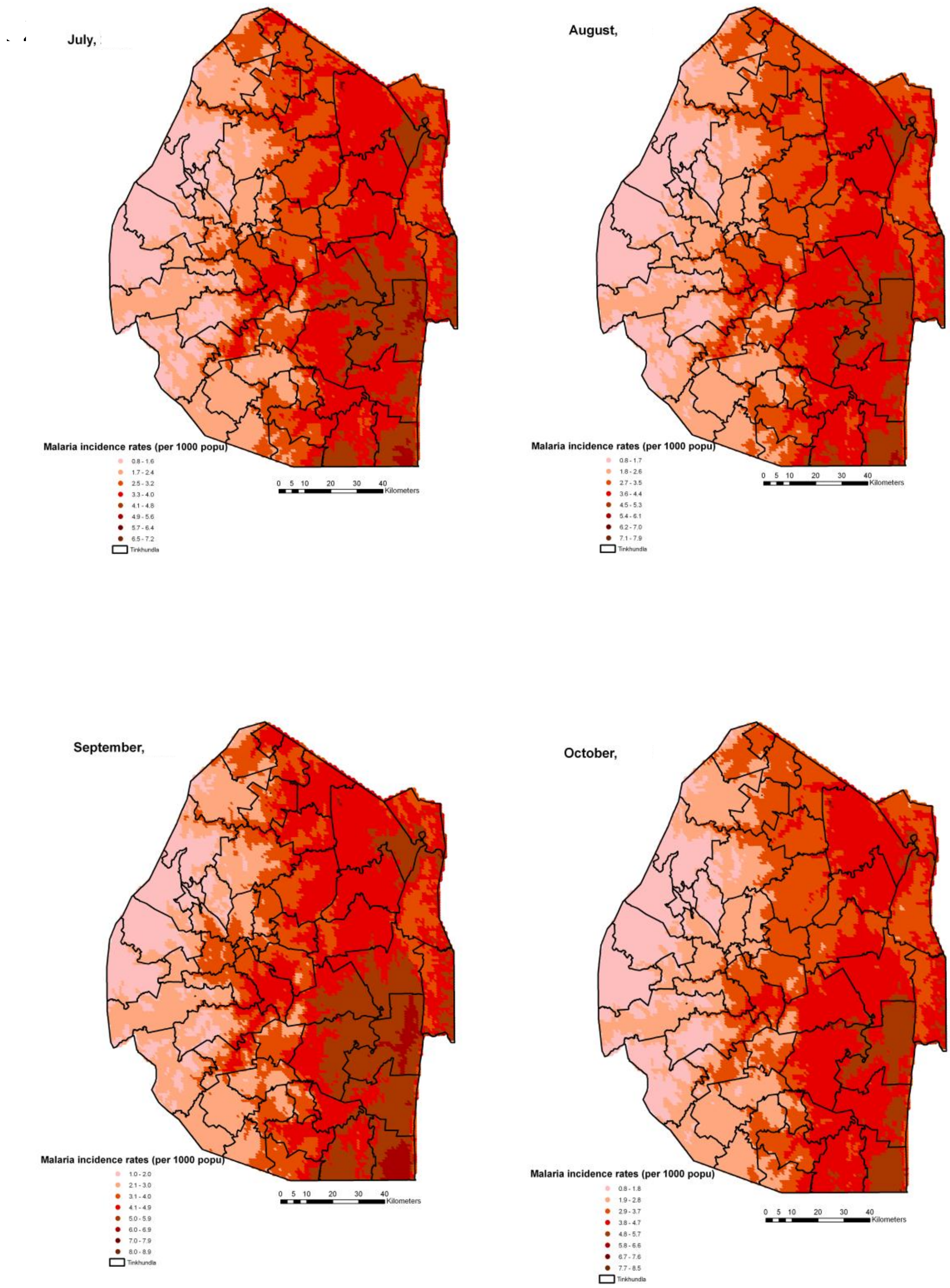

Figure 4.3.1: Predicted malaria incidence for July-October

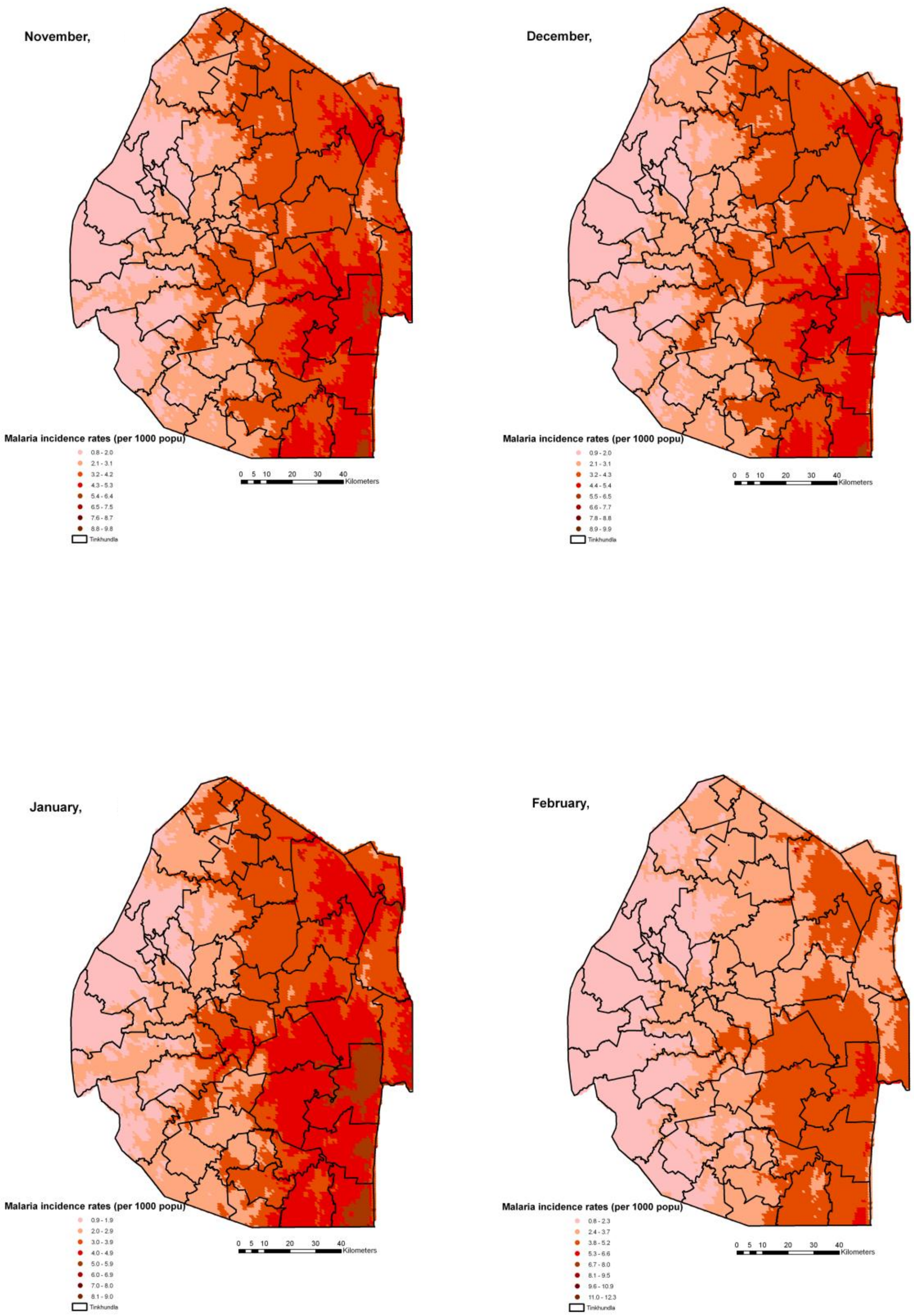

Figure 4.3.2: Predicted malaria incidence for November-February

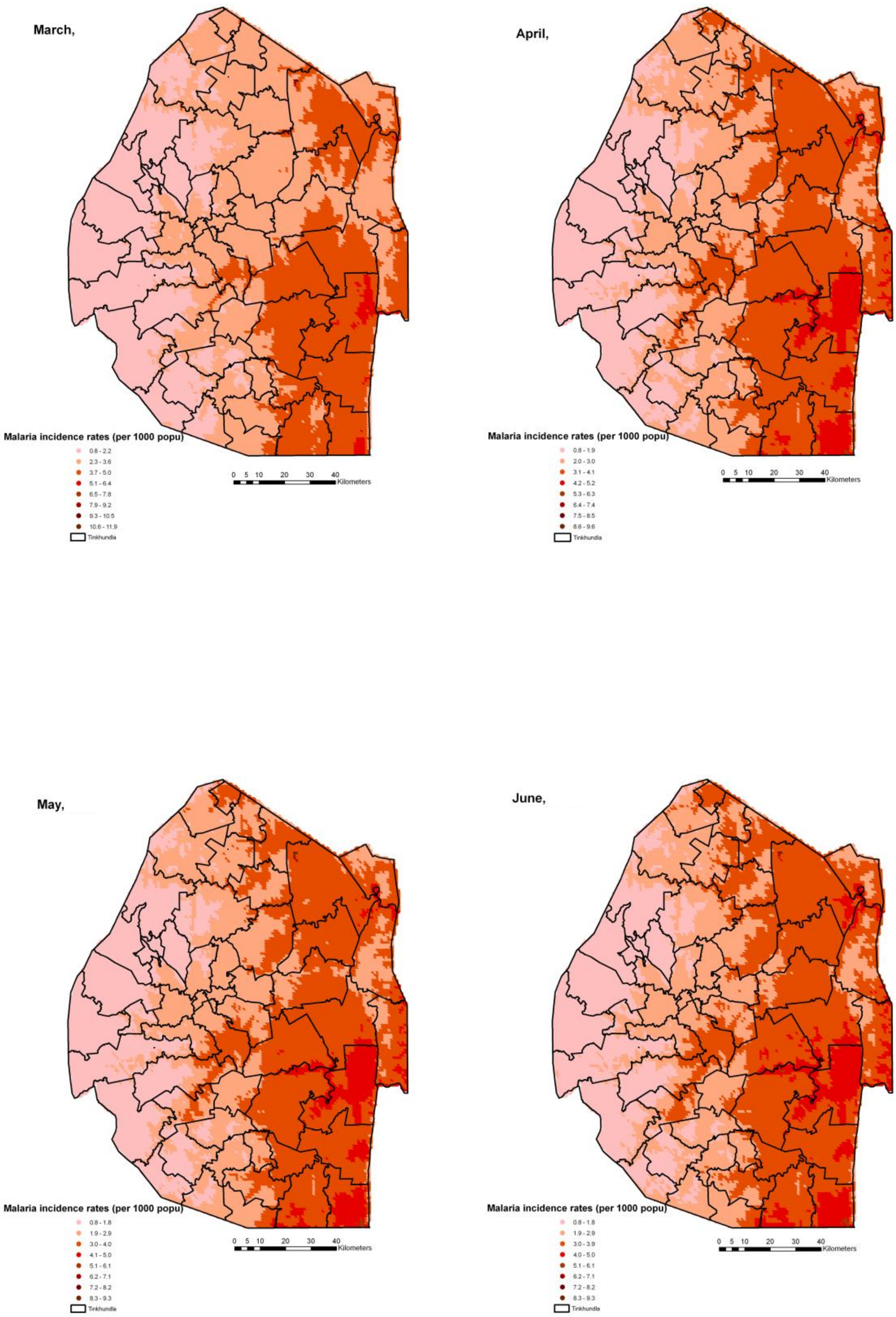

Figure 4.3.3: Predicted malaria incidence for March-June

4.4 Discussion

This study showed a consistent spatio-temporal correlation between climatic factors and malaria incidence risk and as such the maps provide an explicit guide for resource optimization as they show the areas to be targeted with intervention to achieve high impact. While high incidence risk is predominantly in the eastern lowlands of the country its magnitude varies from month to month and the results also suggest that additional non-climatic factors including socio-economic conditions, elevation, etc. also influence malaria transmission. Most of the areas at risk are below an altitude of 400 meters and also dominated by rural settlements. Although the influence of climatic and environmental factors on malaria transmission is known, this study is the first to identify a geostatistically significant correlation between climatic factors and malaria incidence risk in Swaziland. The maps depict a considerable month to month fluctuation in malaria incidence with highest incidence rates occurring in September and January respectively. In sharp contrast, very low incidence rates were observed for the months of February and March. It can be noted that a high proportion of cases that are seen in the months of December to March are usually a result of importation from nearby endemic countries, hence it can be concluded that local transmission essentially begin in the months April to November.

At three months lags, the study was able to quantify, the delay or buildup period for climatic factors and onset of malaria cases. Therefore, the algorithm used in this modeling can be adopted and used by the control programme for short to medium term forecasting of epidemics thus enabling intervention measures to be timely and well placed. Since the study produced the first monthly explicit maps of spatial variations in the country, these can be used as on-the-ground guide for not only timing interventions but also identifying priority areas for high impact as shown in the smoothed maps of monthly predicted malaria incidence in Figures 4.3.1, 4.3.2 and 4.3.3. These predicted maps illustrated spatial variations between months with areas of high predicted malaria incidence risk occurring in the eastern low part of Swaziland and a few locations with low predicted malaria incidence spreading towards the western part of the country. Of note is that the spread towards the west follows areas of low altitude which are channels with the same elevation as the eastern high incidence risk areas.

4.4.1 Conclusions

The Bayesian modeling methods developed and used in this study can be built on and modified for rapid forecasting models to assist countries with very low malaria incidence rates understand the micro epidemiology of the disease over very small areas and therefore optimally deploy malaria interventions in accordance with the severity of the observed malaria episodes. High resolution maps are useful in analyzing factors influencing transmission for very low endemic settings like Swaziland. Results show that the country is on track towards elimination as average predicted malaria cases become less in some months and the area predicted to be at risk demonstrate clearly a case of diminishing malaria.

Chapter 5 An evaluation of potential vegetation indices for predicting malaria incidence risk in Swaziland

Sabelo Nick Dlamini[1,2], Penelope Vounatsou*[1,2]

[1]Swiss Tropical and Public Health Institute, Basel, Switzerland;

[2]University of Basel, Basel, Switzerland;

*Corresponding author

This is a working manuscript to be submitted to Acta Tropica Journal

Abstract

Although there has been much progress in the fight against malaria, threats of resurgence remain. Malaria is driven by multiple factors involving interactions between vectors, parasites and human hosts. The environmental determinants of these factors are what is of interest in epidemiology since understanding them can result in better disease control and even elimination. In this investigation we revisited some of the vegetation indices earlier proposed by ecologists that could be used as predictors for malaria incidence risk. These were indices that were derived using the first generation satellites which however had not been taken up in epidemiological studies. Using the derivation formulae provided by the author of the index we derived some of the indices and assessed how they were associated with malaria incidence risk in Swaziland. All of the twelve vegetation indices that were analysed using bivariate regression were associated with malaria incidence. Following fitting the same indices into a negative binomial regression model the months of May, June, July and August were significantly associated with malaria incidence in seven vegetation indices. The month of November was significant only for the Modified Normalized Difference Vegetation Index (MNDVI) -1.23 (95% CI, -2.42:-0.04) p value 0.042. For the Atmospherically Resistant Vegetation Index (ARVI) in July malaria incidence was estimated at 3.81 cases (95% CI, 1.30:6.32) and p value 0.03 while in August incidence decreased by -3.84 cases (95% CI, -6.95: -0.73) and p value 0.015. Interestingly, all the environmental indices that were significantly associated with malaria cases for the month of May were significant manly in the year 2014, whereby only SARVI showed a significant association also in May 2011. Further research assessing the relationship of these indices with other malaria data could be useful in establishing their potential application in epidemiology as we have already observed in our examination of same using malaria incidence data.

Keywords: *Environmental indices, satellite-derived products, malaria incidence, epidemiology.*

5.1 Introduction

Malaria control and elimination has been a human struggle since ancient times. Although much progress has been made in recent years, threats of resurgence (Pascual et al., 2006; Cohen et al., 2012) remain a challenge and calls for sustained efforts rather than complacency in the fight against malaria. According to the World Health Malaria report of 2015, there were 215 million new cases and 430 000 malaria deaths in 2015 worldwide indicating that more targeted malaria control efforts are still needed to control and eliminate malaria in places with on-going transmission (Protopopoff et al., 2007; Zhou et al., 2010). Malaria transmission is driven by multiple factors that influence the vectors, parasites, human hosts, and the interactions among these. Furthermore, these factors are influenced by meteorological and environmental conditions whose proxies have been widely used as covariates and predictors for malaria transmission (Ageep et al., 2009; Zacarias and Andersson, 2010). Some of these factors could be understood by using remote sensing techniques. Many studies have looked at associations between malaria transmission and meteorological conditions such as temperature and rainfall including investigating related vegetation indices such as the Normalized Difference Vegetation Index (NDVI) among others ((Dantur Juri et al., 2015; Dlamini et al., 2015; Reiner et al., 2015; Srimath-Tirumula-Peddinti et al., 2015). Understanding the environmental conditions driving vector distributions is crucial in developing models for disease prevention, control and elimination (Stensgaard et al., 2016).

Environmental proxies have routinely been obtained from open sources such as Reverb ECHO system by the National Aeronautics and Space Administration (NASA) website since 2007 where a variety of MODIS composites can be obtained. Other environmental data access websites include: United States Geological Survey (USGS) established in 1879, the European Space Agency (ESA) since 2000, Famine Early Warning System (FEWS) since 1985 covering Africa (USAID), Copernicus European Earth monitoring program formerly known as Global Monitoring for Environment and Security (GMES) and other website blogs with links to various remote sensing data sources such as EarthEnable. Remote sensing data is preferred because of its wider horizontal resolution coverage i.e. global or continental scales and near real-time availability. However, there exists an inexhaustible list of environmental indices that have been derived by ecologists and remote sensing experts over the entire duration of satellite launches and space exploration era. Many of these indices were derived during the first generation of remote sensing images and following improvements in sensor capturing capabilities, we believe that indices derivation has also improved. We were therfore interested in evaluating and testing some of the earlier proposed indices using the current remote sensing images for malaria incidence risk estimation in Swaziland. For example, indices like the Soil-Adjusted Vegetation Index (SAVI) were proposed by Heute et al., in 1988 nearly 30 years ago yet they remain unknown in epidemiology. Other indices like the Modified Normalized Difference Vegetation Index MNDVI by Jurgens, (1997), Transformed Vegetation Index (TVI) by Deering et al., (1977), Weighted Difference Vegetation Index (Clevers, 1991) and the Soil Adjusted and Atmospherically Resistant Vegetation Index (SAVRI) by Kaufman and Tanré (1992) have also been derived during the first application of satellite captured images in environmental monitoring. The list of such environmental indices is endless yet they have not received much attention in epidemiology.

Environmental data is one of the main sources for health risk assessment in public health that has been extensively used, for instance in ecological studies on vector densities (Frumkin, 2008). In epidemiology, commonly used environmental proxies include the Normalized Difference Vegetation Index (NDVI), Land Surface Temperature (LST) and rainfall. In addition, a limited list of readily available and easily derivable vegetation indices such as those on landuse and landcover is also used. Weiss et al., (2015) reviewed a list of spatial correlates that have been used in mapping plasmodium malaria mapping and concluded that the new covariates provided improved representation of malaria risk patterns and how the risk was changing over time. Wu (2014) reviewed some of the earlier developed indices for terrestrial ecosystem monitoring and observed that they had sensitivity limitations in dry areas with low vegetation cover. To overcome this limitation a new index called the Generalized Difference Vegetation Index (GDVI) is proposed. Although new indices are being relentlessly explored as new sensors with improved image capture capabilities are launched we believe that there is a need to revisit the earlier proposed indices which were based on the first generation of satellite data as their performance in epidemiological studies was not exhaustively assessed. We therefore embarked on an exploratory analysis to assess if these rarely used and earlier proposed indices were associated with malaria incidence in Swaziland. This was made necessary by the fact that better spatial data resolutions are likely to improve the indices derivation algorithms and thus improve image visualization and interpretation. There is strong consensus among remotely sensed data end-users and experts that high spatial, spectral and temporal resolution data images can improve the way we understand and interpret spatial phenomenon (Hay et al., 1998; Beck et al., 2000; Wulder et al., 2004).

In this current work, we assess the association between some of the vegetation indices that were derived by ecologists using sensor images captured from the first generation of satellites with malaria incidence data in Swaziland. Since most of the earlier images had poor resolutions compared to current images we believe that the earlier derived indices might still be applied in epidemiological studies just like the rest of the commonly used indices. Since their uptake was limited by *inter alia* poor resolutions of the source images that were used to derive them, their potential in epidemiology are therefore not yet fully and exhaustively explored.

5.2 Methods

5.2.1 Study area

Malaria in Swaziland is now near elimination with limited sporadic cases occurring during the malaria transmission season which follows summer rains of around November to May each year. The main parasites are from *Plasmodium falciparum* which is about 99% of all malaria cases (MIS, 2010). Because of the limited number of cases occurring in this pre-elimination phase, identification of the environmental drivers of malaria incidence has become more challenging. The country now aims for total elimination by end of 2016 and has consequently bolstered malaria surveillance and other intervention efforts towards this goal. Over the years of sustained efforts against malaria Swaziland has reduced malaria cases from 9,700 in 1995 to a merely 243 cases in 2013 (Ministry of Health, 2014). However, the cases doubled in the transmission season of July 2014 to June 2015 as reported cases reached 603 demonstrating that active vigilance is required at all times in order to achieve elimination. Indeed, this upsurge in cases shows that even after eliminating malaria, the country could still face resurgence from imported cases of infected people traveling from neighbouring endemic areas. The anti-malaria activities carried out by the National Malaria Control Programme (NMCP) include indoor spraying (IRS) with insecticides, distribution of Long Lasting Insecticide Treated Nets (LLIN) and suspected case rapid diagnosis using Rapid Diagnostic Tests (RDTs) and treatment of all malaria cases with artemisinin-based combination therapy (ACT).

5.2.2 Malaria incidence data

In Swaziland, malaria incidence data is routinely collected during follow up of cases reported from health facilities. The surveillance officers from the NMCP react to a malaria case notification sent via short message service (SMS) by visiting the patient`s place of residence and conducting interview where origin of the case is determined. This surveillance exercise involves capturing the geographic coordinates of the patient`s household with a Global Positioning System which are later entered into a central malaria surveillance database. The malaria incidence data was obtained from the NMCP for a 4 year period from January 2010 to December 2013. The data included the origin of the cases in terms of whether they were imported or locally acquired as well as the corresponding GPS location of the case. The cases were then summed up at Enumeration Area (EA) which is the lowest administrative unit in the country and aggregated by the month and year in which the case was reported. Figure 5.2.0 shows the EA centroids that were sampled for the environmental indices proxies and upon which malaria cases were aggregated.

5.2.3 Vegetation indices data

Following a review of environmental proxies derived by ecologists and remote sensing experts, we selected a list of 12 vegetation indices that had almost the same usage and application as NDVI. We wanted to evaluate how well they performed in predicting malaria incidence in Swaziland by assessing their association with the incidence data. The environmental data was obtained as composites from MODIS and it was made up of four spectral bands which were the: 1 red; 2 blue; 3 near infrared and 4 middle infrared. These bands were used to calculate various vegetation indices using the algorithms and formulae that were provided by those who proposed them. The image composites had a spatial resolution of 250 m and a temporal resolution of 16 days. After processing we aggregated the derived indices over one month in order to take into account the cumulative effect of environmental factors on malaria incidence in Swaziland. Table 5.2 presents the list of vegetation indices that were proposed by ecologists and which were derived using the MODIS composites in this investigation.

Table 5.2: Vegetation indices developed using the first generation remote sensing images and which were used in this analysis

Index Name	Formula	Temporal Resolution	Reference
Simple Ratio Index (SR)	P_{NIR}/P_R	16 days	Birth et al., (1968) Knipling (1970)
Transformed Vegetation Index (TVI)	$(NDVI + 0.5)^{1/2}$	16 days	Deering et al., (1975)
Soil-Adjusted Vegetation Index (SAVI)	$\frac{(1+L)(P_{NIR} - P_R)}{(P_{NIR+}P_R + L)}$	16 days	Huete (1988)
Optimized Soil-Adjusted Vegetation Index (OSAVI)	$(P_{NIR} - P_R)/(P_{NIR} + P_R + 0.16)$	16 days	Rondeaux (1996)
Atmospherically Resistant Vegetation Index (ARVI)	$(P_{NIR} - P_{RB})/(P_{NIR} + P_{RB})$	16 days	Kaufman and Tanré (1992)
Soil Adjusted and Atmospherically Resistant Vegetation Index (SARVI)	$(1+L)(P_{NIR} - P_{RB})/(P_{NIR} + P_{RB} + L)$	16 days	Kaufman and Tanré (1992)
Enhanced Vegetation Index (EVI)	$G * (P_{NIR} - P_R)/(P_{NIR} + C1 * P_R - C2 * P_B + L)$	16 days	Huete et al., (1997), Huete et al., (2002)
Enhanced Vegetation Index (EVI2)	$2.5(P_{NIR} - P_R)/(P_{NIR} + 2.4P_R + L)$	16 days	Jiang et al., (2008)
Non-Linear Vegetation Index (NLI)	$(P_{NIR}^2 - P_R)/(P_{NIR}^2 + P_R)$	16 days	Goel and Qin (1994)
Modified Non-linear Vegetation Index (MNLI)	$(1+L)(P_{NIR}^2 - P_R)/(P_{NIR}^2 + P_R + L)$	16 days	Gong et al., (2003)
Wide Dynamic Range Vegetation Index (WDRVI)	$a * (P_{NIR} - P_R)/(a * P_{NIR} + P_R)$	16 days	Gitelson (2004)
Modified Normalized Difference Vegetation Index (MNDVI)	$\frac{NIR - MIR}{NIR + MIR}$	16 days	Jurgens (1997)

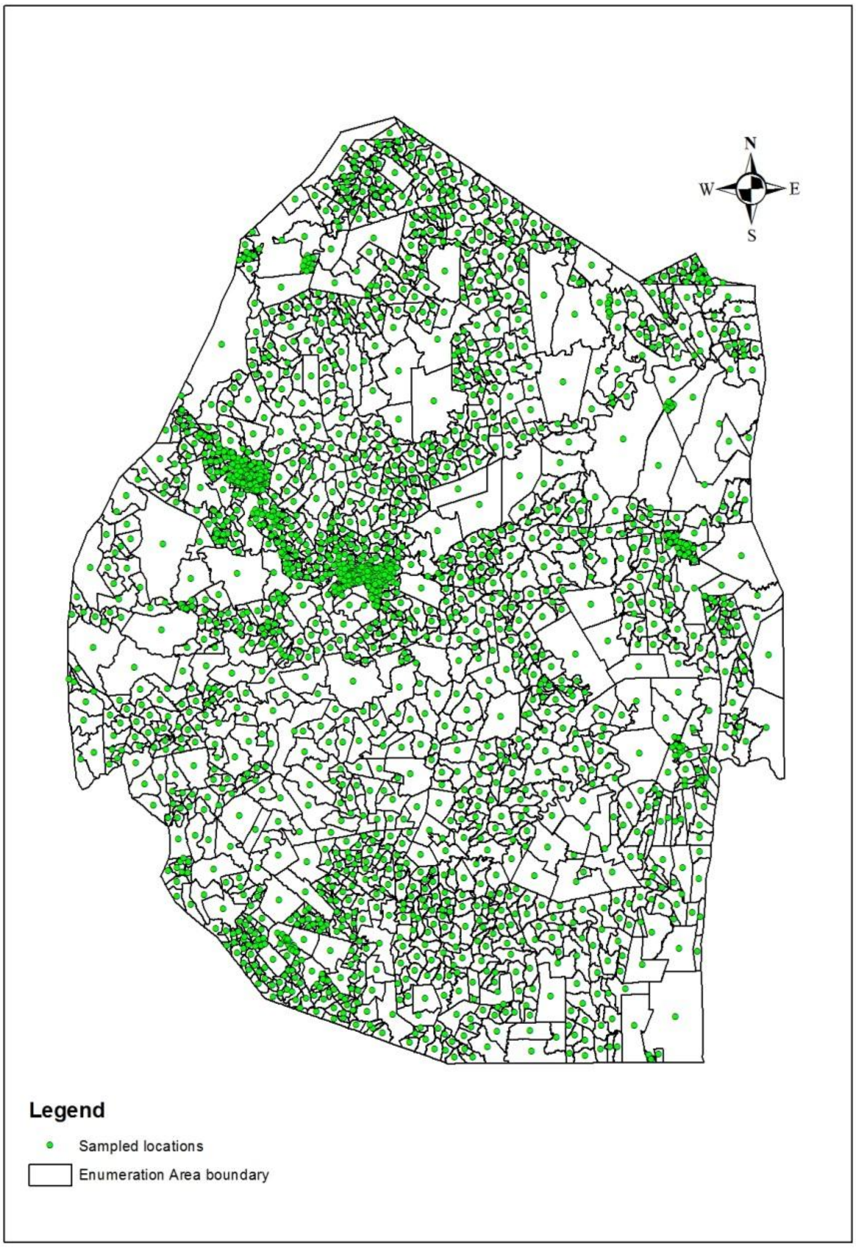

Figure 5.2.0: Sampled enumeration area centroids

5.2.4 Performing principal component analysis on the remote sensing bands

In order to conduct preliminary exploratory analysis, we used Stata to perform a principal component analysis (PCA) on the four remote sensing bands comprising of Red, Blue, Near Infrared and Middle Infrared in order to account for the correlation between them. The PCA converts a set of observations that are correlated into a set of linearly uncorrelated variables known as principal components. The PCA was therefore used to reduce the redundant dimensions of the correlated values of the remotely sensed bands and identify patterns with the highest dimension. By using the correlation matric (corr- PCA) we were able to identify the most contributing component of each of the four remote sensing bands during index derivation and formulation. We used the most contributing component of the remote sensing bands to calculate and derive the environmental indices. The input data was a matrix of rows and columns for each of the individual remote sensing bands used to derive each index. Since the bands were too many for each year (four bands every 16 days), meaning we had 92 variables upon which to perform the PCA. This was too much and could not be handled well in our statistical analysis and we therefore sampled and performed PCA analysis for every third observation of each year.

5.2.5 Statistical analysis

Statistical analysis for the remote sensing indices and malaria incidence data was done in Stata version 14 (StataCorp, 2015). The analysis first involved calculation and derivation of each of the indices used in this investigation using the formula provided by the author of that particular index. This way we calculated each of the indices by utilizing the individual bands extracted from MODIS composites datasets. The 16 day temporal resolution datasets were aggregated into monthly indices whereby dates were generated from the day of the year on which the data was captured in order to identify the month on which the data capture falls. Once the environmental index month was identified it was then used to link with the corresponding month of the malaria case using the Stata command merge. As an exploratory analysis a bivariate regression based on yearly aggregated data was first performed in order to investigate how each of the vegetation indices was associated with malaria incidence. The joint association of vegetation indices with malaria was investigated using a negative binomial regression model which was fitted into the monthly aggregated data in order to identify the set of the monthly lags that better explain malaria incidence in Swaziland. All 12 vegetation indices were used as predictor variables for malaria incidence in each of the Enumeration Areas with reported malaria cases.

5.3 Results

5.3.1 Principal component analysis

The preliminary analysis using PCA allowed us to identify the dominant bands and obtain correlation between the four remote sensing bands. Following PCA calculation, the most significant eigenvalues of the remote sensing bands that were used to derive the different vegetation indices were identified and used in the final model. The eigenvalue which is a list of scalars with a linear system of equation allowed us to select the number of components that were used for indices derivation for the 4 year remotely sensed data. The results showed that only six components contributed eigenvalue of above one for the years 2010-2013, while in the year 2014 five components had an eigenvalue above one. Thus we had to consider only the components that were above the eigenvalue of one for the derivation and calculation of the indices.

5.3.2 Bivariate logistic regression analysis

Following the data processing and aggregation of the 16 day composites by month and merging with malaria incidence data, a bivariate logistic regression model was fitted into the data. All 12 of the vegetation indices that were used in the bivariate regression were significantly associated with malaria incidence and these are presented in Table 5.3. The ARVI and EVI2 which are an enhancement of NDVI were significantly associated with malaria with means 0.187 (95% CI, 0.165:0.99) and 0.667 (95% CI, 0.578:0.756) respectively while the p value was 0.000 for both indices. The incidence rate ratio (IRR) was higher for EVI2 at 1.949 (95% CI, 1.783:2.13) than that of ARVI which was at 1.206 (95% CI, 1.179:1.234). The MNDVI which uses the near infrared and middle infrared bands was also positively associated with malaria incidence with mean 0.289 (95% CI, 0.249:0.329) and the p value was 0.000. The IRR was 1.336 (95% CI, 1.284:1.39). In addition, the MNLI as well as the NLI were also positively associated with malaria incidence whereby MNLI had mean 1.064 (95% CI, 0.855:1.272) and NLI had a mean of 0.068 (95% CI, 0.022:0.113). Their p values were 0.000 and 0.003 while their IRR were 2.898 (95% CI, 2.354:3.569) and 1.07 (95% CI, 1.023:1.12) respectively.

Interestingly, only one of the vegetation indices was negatively associated with malaria incidence and that was the WDRVI which is also another modification of the

NDVI. The mean of the WDRVI was -0.272 (95% CI, -0.312:-0.232) with a p value of 0.000 and IRR of 0.762 (95% CI, 0.762:0.732). The SAVI which is adjusted for areas with sparse vegetation together with its optimised counterpart the OSAVI were also positively associated with malaria incidence. SAVI had a mean of 0.346 (95% CI, 0.300:0.392) and a p value of 0.000 while OSAVI had a mean of 0.321 (95% CI, 0.279:0.363) and p value of 0.000. Similarly their IRR were 1.414 (95% CI, 1.351:1.48) and 1.379 (95% CI, 1.322:1.438) respectively. TVI, SARVI and SR which uses the near infrared and the red bands to identify healthy green vegetation were all positively associated with malaria incidence with p values of 0.000 respectively. In general all 12 vegetation indices were significantly associated with malaria incidence for consideration as determinants in explaining malaria incidence in Swaziland.

Table 5.3: Bivariate analysis of vegetation indices and malaria incidence

Variable	Coefficient	P value	95% CI		IRR	IRR low	IRR high
ARVI	0.187	0.000	0.165	0.099	1.206	1.179	1.234
EVI	-10.612	0.000	-12.002	-9.223	0	0	0
EVI2	0.667	0.000	0.578	0.756	1.949	1.783	2.13
MNDVI	0.289	0.000	0.249	0.329	1.336	1.284	1.39
MNLI	1.064	0.000	0.855	1.272	2.898	2.354	3.569
NLI	0.068	0.003	0.022	0.113	1.07	1.023	1.12
OSAVI	0.321	0.000	0.279	0.363	1.379	1.322	1.438
SAVI	0.346	0.000	0.300	0.392	1.414	1.351	1.48
SARVI	0.257	0.000	0.226	0.289	1.294	1.254	1.335
SR	0.022	0.000	0.019	0.025	1.023	1.02	1.026
TVI	0.138	0.000	0.121	0.154	1.148	1.129	1.167
WDRVI	-0.272	0.000	-0.312	-0.232	0.762	0.732	0.793

5.3.3 Negative binomial regression

Following the data processing and aggregation of the 16 day composites by month and merging with malaria incidence data, a negative binomial regression model was fitted into the data. The vegetation indices that were significantly associated with malaria cases are presented in Table 5.4. The ARVI and EVI which are an enhancement of NDVI were significantly associated with malaria for the months of July and August 2013. For ARVI in July malaria cases were estimated at 3.81 (95% CI, 1.30:6.32) and p value 0.03 while in August cases decreased by -3.84 (95% CI, -6.95: -0.73) and p value 0.015. Furthermore, in the following year 2014 and in the month of March cases also decreased by -2.44 (95% CI, -4.50:-0.38) and p value 0.020. This trend was also the same for NLI, MNLI and EVI2 whereby the months of July and August were also significantly associated with malaria. There was also a negative association for NLI as cases decreased by -1.21 (95% CI, -2.12:0.30) and p value 0.009 and for EVI2 where cases decreased by -4.67 (95% CI, -9.13: 0.21) and p value 0.040, both decreases occurring in the month of June for the year 2013.

Interestingly, the MNDVI which uses the near infrared and middle infrared bands was the only index that was negatively associated with malaria cases for the month of November, -1.23 (95% CI, -2.42:-0.04) and p value 0.042. In addition, six of the environmental indices including EVI, SAVI, SARVI, NLI, MNLI and EV12 were all significantly associated with malaria cases in the month of May. Although SAVI was not significantly associated with malaria cases for all years, it was also positively associated for the month of May in the year 2014 with 2.76 (95% CI, 0.35: 5.16) and p value 0.024. Out of these indices only EVI was negatively associated with malaria cases whereas the rest showed a positive association. SARVI was positively associated with malaria cases for the month of May for both year 2011, 3.64 (95% CI, 0.56: 6.72) p value 0.020 and year 2014, 2.47 (95% CI, 0.29: 4.66) and p value 0.026. Interestingly, all the vegetation indices that were significantly associated with malaria cases for the month of May were significant in the year 2014, whereby only SARVI showed a significant association also in the year 2011.

Vegetation indices that were not associated or important in explaining malaria incidence included the TVI which corrects for negative values of the NDVI by adding a constant of 0.5, the OSAVI which is used for areas with sparse vegetation, the WDRVI which is also another modification of the NDVI and SR which uses the near infrared and the red bands to identify healthy green vegetation. In total out of twelve vegetation indices that were used in the negative binomial regression model, eight of them proved to be important environmental factors for consideration as determinants in explaining malaria incidence in Swaziland.

5.4 Discussion

The bivariate regression analysis of the association between vegetation indices with malaria incidence data showed that they may be important for consideration as covariates in malaria epidemiology and mapping studies. Since some of the indices are an improvement to commonly used indices such as NDVI, there is need to investigate this association further with other malaria data in order to find out if there is consistency in this association. There is also need to investigate the different vegetation indices in order to identify how they change with changes in environmental conditions as well as their ability to predict malaria incidence under different land cover types. For example, it is expected that the different seasons marked by varying amounts of temperature and rainfall and consequently wetness and dryness can help us understand and identify important patterns and spatial heterogeneities in malaria as they will highlight different environmental and land cover conditions.

The negative binomial regression analysis between vegetation indices and malaria incidence data showed that the month of June, July and August were significantly associated with malaria. June was negatively associated with malaria for the NLI and EVI2. This can be assumed to be a result of the fact that June is the beginning of the dry season as winter sets in and rains decrease. Furthermore, most cases that occur around this period are only local cases as it is a period before the travelling season and after the rainy season which is from November to May. In addition, the month of May was also significantly associated with malaria for most of the indices indicating the effect of rainfall around the same period. The SAVI, SARVI, EVI2, NLI and MNLI were positively associated with malaria while only EVI was negatively associated in the month of May. The month of May is also the last month of the rainy season, therefore it is expected that malaria cases still occur during this time especially because temperatures remain high while rains diminish. This therefore creates a suitable and favourable environment for mosquitoes to breed from pools of the previous rain episodes (Dlamini et al., 2015). This also goes to demonstrate that although most of the rains fall around January and February none of the environmental indices were positively associated with malaria cases in this period. Firstly, there are lots of rains that are received in the country which might also result in washing away of mosquito larvae (Savage et al., 1990) or killing of adult mosquitoes from the downpour thus hindering or even delaying transmission.

The environmental indices that are an improvement of NDVI such as ARVI and SAVI were significantly associated with malaria for the months of June, July, and August as well as May respectively indicating a culmination and accumulation from the previous rain season. In August ARVI was positively associated with malaria in July and negatively associated in the following month of August. It is possible that as mosquitoes can only fly when it is not raining and therefore their ability to transmit malaria is hindered during rainy season as they can only fly for short distances. The month of November was also significantly associated with malaria cases for the MNDVI which interestingly coincide with the beginning of the rain season. This is also the time when the mosquitoes that had been breeding during the dry cold months begin to mature into adult and fly as temperatures get warmer. However, the association was negative which might be a result of the heavy rainfall that falls in the beginning of the rain season thus possibly interrupting transmission. The exploratory analysis using the PCA helped us to identify which components of the remote sensing bands were useful for the indices derivation thus avoiding unnecessary redundancy in the final model.

5.4.1 Conclusions

The indices used here are an example of the many potential environmental proxies derived from remote sensing images that could be used in epidemiology. Although the indices are not novel in ecology and remote sensing, they are hardly known in epidemiology. The purpose of this investigation was to assess the potential of these indices in vector-borne disease mapping with intention to bring them into the attention of epidemiologist. The list of indices used here was not all that is out there but our interest was merely to investigate some of the dated indices as potential covariates for disease modelling in epidemiology much as the same way as vegetation indices like NDVI are currently used. Eight of the indices were significantly associated with malaria indicating that they have the potential for use as predictors in vector-borne disease mapping. Further research assessing the relationship of these indices with other vector-borne diseases could be useful in establishing their potential application in epidemiology as we have already observed in our examination of same with malaria incidence data.

Chapter 6

Discussion and outlook

6.1 Significance of the research work

6.1.1 Key messages from the study

This research has contributed new information, tools and methods for consideration in malaria control and surveillance. Although we used the case of Swaziland the methods and tools suggested in this study are applicable to other countries with similar endemic setting as Swaziland. The issues of targeted response are particularly addressed and this work demonstrated the need for high quality spatial data if the distribution of vector-borne diseases is to be understood. The study also showed that it is possible to work with remote sensing data with different spatial resolutions and obtain the best maps that show spatial heterogeneities that would have otherwise not been identified. The maps produced in this work could be used by the control programme of Swaziland as a guiding tool to plan their anti-malaria control interventions as they show both areas of potential vector breeding and areas where malaria cases are likely to occur.

This study contributed entomological data from the field larva scooping exercise that was conducted for the purposes of mapping breeding sites. This was data that would have otherwise not been available to the control programme of Swaziland and the methods used to map breeding sites resulted in the production of the first geographically explicit maps of potential malaria vector breeding sites in the country. The maps and the methods produced and developed in this study can be updated with new geographic data and thus assist the programme with continually tracking the potential breeding sites and monitoring malaria incidence trends in Swaziland.

The study emphasised the importance of high resolution remotely sensed data in areas where human populations and disease causing vectors intersect such as large subsistence farming fields which are not usually targeted with vector control measures like insecticide residual spraying. These are areas which are also frequently visited by people during farming and thus may be sources of contact between humans and disease carrying vectors. This work therefore demonstrated that malaria transmission in Swaziland could essentially be due to occupational behaviour of people located in the Lowveld part of the country with on-going malaria transmission as it is an area with suitable climatic conditions. The study found an association between malaria vector breeding sites and subsistence farming which could be a result of the ever present supply of human host as nearby people frequent the area for cultivation, weeding and harvesting. Therefore our work shows that such areas are also important for consideration for vector control strategies such as conducting larval source management programmes. Current practices in vector control are focused more at indoor household or settlement locations with very limited activities dealing with vectors directly at their outdoor sources and in the environment where they maintain their life cycles.

We have also prepared a first catalogue and library of remote sensing products which has potential application to vector-borne disease mapping. This work has therefore contributed new information relevant to spatial epidemiology and public health by preparing and producing a compendium for environmental data and their products to facilitate access by non-remote sensing experts. The knowledge about existing remote sensing products is important in order to present end-users like epidemiologist with alternatives on the choices of data they use for their disease mapping efforts. This was the first catalogue that attempted to pool together all remote sensing products in order to address public health needs and interests. It had been widely known that epidemiologist used remote sensing data that was already available after being processed and packaged by the supplier; however our catalogue brought to the attention of the public health community, other existing remote sensing indices that could be used as proxies in disease models.

We also highlighted the need to revisit some of the indices proposed earlier by ecologists during the first generation of satellite data products following the fact that better sensors with better resolution image capturing capabilities have been launched in recent years. Thus some of the indices already proposed and derived by ecologists could have potential for use in disease mapping following improved spatial and temporal resolutions as well as in terms of the number of spectral bands resolution definition. Our work was therefore very relevant to current challenges in the ever increasing use of spatial data for disease mapping and public health. We succeeded in this by addressing the challenges of epidemiologists and public health experts which include lack of knowledge and training in the use and handling of remotely sensed data. This knowledge gap had to be addressed as current trends show that remotely sensed data is now an inherent part of the spatial epidemiology and vector-borne disease mapping community.

This study was dedicated to exploring various cumulative effects of environmental factors to malaria incidence in Swaziland by developing and testing various climatic lag effects. This work estimated for the first time in the country the lag time between certain environmental conditions and the probability of occurrence of a local malaria case. The study contributed new information that could be used by the control programme not only to time interventions but also to target those areas identified as drivers of transmission. Monthly maps showing spatial heterogeneities were presented and the same could be used as a guide by the control programme for spatial targeting of interventions.

We developed two weeks lags to assess their effects on malaria cases and to understand how they influence incidence in the country. In addition to this, a polynomial function was fitted into the two weeks lags to estimate the cumulative effect and quantify the time it takes for cases to be made manifest from certain environmental conditions. The methods used in this analysis could also be extended and adapted into algorithms for case load forecasting as this becomes an important component of the surveillance system as the country zeroes down to elimination. This work further emphasised the need for highly robust and effective methods for surveillance and response by demonstrating technical yet applicable algorithms for disease modelling and mapping as an integral part of the surveillance component in very low endemic settings which are already oriented towards elimination.

We also extended the work done from chapter three by testing some of the new indices that have undergone improvement in recent years in terms of both spatial and temporal resolutions. We investigated some of the new products with intention to assess their use in malaria incidence models and disease mapping. Following the availability of new remote sensing data products with better resolutions we felt it was necessary to assess if its incorporation into disease analysis models would improve our understanding of the drivers and determinants of disease causing elements in the environment. This is therefore useful information which could be used to investigate how disease risk is estimated and quantified using remote sensing products with different resolutions. Our analysis proved that high resolution is preferred for disease mapping in order to identify spatial heterogeneities even at small geographic scales.

We therefore, relay the information that it is important for epidemiologist to be constantly looking out for new spatial data sources that have the potential to improve their disease mapping efforts instead of relying on available data sources out of mere convenience. Furthermore we show that epidemiologist can also benefit from synergies with ecologists who are leading the remote sensing community with research of and proposing of new index derivation algorithms, a majority of which are an improvement from existing ones.

6.1.2 Contributions to malaria surveillance and response

This work has contributed technical information to the issues of and considerations for surveillance and response in low endemic settings characterised by diminishing and episodic malaria cases. We addressed the importance of mapping breeding sites and incorporating larval source management as part of the broader strategy of surveillance for elimination. We highlighted the fact that surveillance should not only end with case investigation and detection of asymptomatic cases but should also include foci investigation in order to identify and interrupt transmission chains. These would require that breeding sites are known to the control programme and constantly investigated as possible sources of local transmission.

Furthermore our work contributed to the on-going investigation of which environmental factors are responsible for driving malaria transmission including understating their cumulative effects on malaria incidence. We presented the control programme with spatially explicit maps that could be used as a guide for targeted and timed interventions. Thus our work contributed the first model based country maps not only to aid response interventions but also to assist with proactive and prioritising surveillance strategies.

As Swaziland aims for elimination by 2016 more robust and responsive surveillance systems have to be established and the methods used in this study can be extended and adapted by the control programme in order to meet the demands and sophistication of current surveillance approachess. For example our polynomial lag distribution model can be adapted and extended into methods for early warning system and case load forecasting, an important aspect for malaria elimination and prevention of reintroduction.

6.1.3 Contributions to applications of remote sensing products in epidemiology

The use of remote sensing data products for such research questions as investigating disease outcomes and their determinants, mapping vector-borne diseases and analysing associations between certain environmental factors and disease outcomes has been increasing over the years. Our work contributed to the quest for using remote sensing data in epidemiology by presenting a catalogue on remote sensing that is specifically addressing epidemiologists' challenges in understating and accessing remote sensing information. Evidence from the ever increasing number of epidemiological studies that use remote sensing data suggest that this is an important part of epidemiology and public health in general and thus there is a need to harmonise synergies between remote sensing experts and epidemiologists.

Our work has therefore bridged this skills gap between ecologists and epidemiologists by indicating areas where the two fields intersect and how they could improve and learn from each other by packaging remote sensing information in such a way that it is beneficial and understood by all those interested in it. Therefore our work presented a new platform for collaboration between remote sensing experts and epidemiologists by pointing out areas of common interests. The catalogue could be used by epidemiologist to decide on a criterion for selecting remotely sensed data during disease analysis and mapping.

We further highlighted the need for epidemiologists to routinely check out for new remote sensing proxies instead of relying on what is readily available since remote sensing information is constantly evolving. We showed that other little known and underutilized environmental proxies that had been long derived by ecologists could still be used in disease mapping as some of them are improvements to the commonly used indices. The slow uptake in epidemiology has been merely due to lack of capacity to derive the proposed indices by non-remote sensing experts and not necessarily due to their irrelevance in disease mapping. Therefore our work demonstrated their applicability to disease mapping in the same way as the indices that are used due to their availability and convenience.

6.2 Study limitations

Identifying and mapping potential larval breeding sites is a complex exercise that needs to be repeated in order to ascertain if breeding occurs or not. This would involve repeated observation over entomologically informed larva scooping and sampling intervals. Our study could not have the time and the resources to conduct repeated sampling on identified potential breeding sites and therefore might have missed some of them especially due to the rains that occurred during the sampling period which had a potential to wash away the larva. The results from the field study showed that breeding occurred in very small water bodies down to the size of cattle hooves. Unfortunately it would be impossible to investigate such small water bodies given the limited time and resources. In addition, the remote sensing images used during analysis had a spatial resolution of at least 5 m which is still not enough for identification of potential breeding sites down to the size of cattle hoof.

The inventory and cataloguing of remote sensing products was also limited by the fact that some agencies are private and not open to sharing their information with civilians. This lack of access made it difficult to document all satellite missions and the products that could be obtained, for instance, from military missions. Other agencies were only working on limited geographic coverage scales such as national jurisdiction and therefore their products were not useful for the general global interest and scope. Although some of the products may appear in publications, their access remains limited to national or local end-users. Due to time constraints we also could not test all the new and old indices that were proposed by ecologists, even though some of them were based on the images of the first generation of sensors. We would have liked to test and validate some of the earlier proposed indices using the latest and recent satellite images which have improved resolutions.

6.3 Concluding remarks and extension of this work

Our work on developing surveillance tools for very low endemic settings still needs further development into among others assessing methods that are suitable and can work when local malaria transmission is extremely low or even near non-existent such as in Swaziland. Studies designed to investigate cost effective methods for surveillance in very low endemic settings with episodic malaria cases are still needed. For example, when to continue rolling out or to stop rolling out anti-malaria control interventions such as vector control is still not clear. There are also challenges in keeping enough stock supply of ACTs which are likely not to be used resulting in their expiration and eventual wastage. Nevertheless we have provided new information, tools and methods for surveillance and response that could be used in designing strategies for malaria control and elimination in very low endemic settings.

Bibliography

Ageep TB, Cox J, Hassan MM, Knols BGJ, Benedict MQ, Malcolm CA, Babiker A and El Sayed BB (2009). Spatial and temporal distribution of the malaria mosquito Anopheles arabiensis in northern Sudan: influence of environmental factors and implications for vector control. *Malar J*. 8, 123.

Almeida AS and Werneck GL, (2014). Prediction of high-risk areas for visceral leishmaniasis using socioeconomic indicators and remote sensing data. *Int. J. Health Geogr*. 13, 13

Amek N, Bayoh N, Hamel M, Lindblade KA, Gimnig JE, Odhiambo F, Laserson KF, Slutsker L, Smith T and Vounatsou P (2012). Spatial and temporal dynamics of malaria transmission in rural Western Kenya. *Parasit. Vectors* 5, 86.

Ashton RA, Kefyalew T, Rand A, Sime H, Assefa A, Mekasha A, Edosa W, Tesfaye G, Cano J, Teka H, Reithinger R, Pullan RL, Drakeley CJ and Brooker SJ (2015). Geostatistical modeling of malaria endemicity using serological indicators of exposure collected through school surveys. *Am. J. Trop. Med. Hyg*. 93, 168–177.

Asnakew K., Yeshiwondium, Sucharita Gopal, Afework T. Hailemariam, Dereje O. Degela, and Hrishikesh P Patel, (2009). Spatial analysis of malaria incidence at the village level in areas with unstable transmission in Ethiopia, *Int J. of H. Geogr*, 8:5

Atkinson JA, Johnson M, Wijesinghe R, Bobogare A, Losi L, O'Sullivan M, Yamaguchi Y, Kenilorea G, Vallely A, Cheng Q, Ebringer A, Bain L, Gray K, Harris I, Whittaker M, Reid H, Clements A and Shanks D (2012). Operational research to inform a sub-national surveillance intervention for malaria elimination in Solomon Islands. *Malar J*. 201211:101

Ahmad R, Ali WN, Nor ZM, Ismail Z, Hadi AA, Ibrahim MN and Lim LH (2011). Mapping of mosquito breeding sites in malaria endemic areas in Pos Lenjang, Kuala Lipis, Pahang, Malaysia. *Malar J.* 10, 361.

Baker RHA, Sansford CE, Jarvis CH, Cannon RJC, MacLeod A and Walters KFA (2000). The role of climatic mapping in predicting the potential geographical distribution of non-indigenous pests under current and future climates: *Agri., Ecosy. & Envir.*, 82, pp. 57-71.

Banerjee S, Carlin BP and Gelfand AE (2004). Hierarchical modelling and analysis for spatial data. Chapman and Hall, New York

Bannari A, Asalhi H and Teillet P (2002). Transformed Difference Vegetation Index (TDVI) for vegetation cover mapping. In Proceedings of the Geoscience and Remote Sensing Symposium, IGARSS '02, IEEE International, Volume 5.

Beck LR, Lobitz BM and Wood BL (2000). Remote sensing and human health: new sensors and new opportunities. *Emerg Infect Dis* 6, 217–227.

Beck LR, Rodriguez MH, Dister SW, Rodriguez AD, Rejmankova E, Ulloa A, Meza RA, Roberts DR, Paris JF and Spanner MA (1994). Remote sensing as a landscape epidemiologic tool to identify villages at high risk for malaria transmission. *Am. J. Trop. Med. Hyg.* 51, 271–280.

Belward AS and Skøien JO (2015). Who launched what, when and why; trends in global land-cover observation capacity from civilian earth observation satellites. *ISPRS J. Photogramm. Remote Sens.*, Global Land Cover Mapping and Monitoring 103, 115–128.

Birth GS and McVey GR (1968). Measuring the color of growing turf with a reflectance spectrophotometer. *Agron. J. 60,* 640–643

Bousema T, Griffin JT, Sauerwein RW, Smith DL, Churcher TS, Takken W, Ghani A, Drakeley C and Gosling R (2012): Hitting hotspots: spatial targeting of malaria for control and elimination. *PLoS Med*. 9: e1001165-10.1371/journal.pmed.1001165.

Bøgh C, Lindsay SW, Clarke SE, Dean A, Jawara M, Pinder, M and Thomas CJ (2007). High spatial resolution mapping of malaria transmission risk in the Gambia, west Africa, using LANDSAT TM satellite imagery. *Am J. Trop Med Hyg* 76, 875–881.

Bréda NJJ (2003). Ground-based measurements of leaf area index: a review of methods, instruments and current controversies. *J. Exp. Bot*. 54, 2403–2417.

Bridges DJ, Winters AM, Hamer DH. Malaria elimination: surveillance and response. *Pathogens and Global Health*. 2012;106(4):224-231.

Briët OJ, Vounatsou P, Gunawardena DM, Galappaththy GN and Amerasinghe PH (2008). Temporal correlation between malaria and rainfall in Sri Lanka. *Malar J*. 7, 77.

Briët OJT, Vounatsou P and Amerasinghe PH. (2008). Malaria seasonality and rainfall seasonality are correlated in space. *Geospatial Health* 2, 183-190 .

Broadband Greenness (Using ENVI) | Exelis VIS Docs Center [WWW Document], n.d. URL http://www.exelisvis.com/docs/BroadbandGreenness.html (accessed 9.28.15).

Brooker S, Clements ACA and Bundy DAP, (2006). Global Epidemiology, Ecology and Control of Soil-Transmitted Helminth Infections, in: Simon I. Hay, A.G. and D.J.R. (Ed.), Advances in Parasitology, Global Mapping of Infectious Diseases: Methods, Examples and Emerging Applications. Academic Press, pp. 221–261.

Cao J, Sturrock HJW, Cotter C, Zhou S, Zhou H, Liu Y, Tang L, Gosling RD, Feachem RGA and Gao Q (2014). Communicating and Monitoring Surveillance and Response Activities for Malaria Elimination: China's "1-3-7" Strategy. *PLoS Med* 11(5): 1001642.

Carroll ML, Townshend JR, DiMiceli CM, Noojipady P and Sohlberg RA (2009). A new global raster water mask at 250 m resolution. *Int. J. Digit. Earth* 2, 291–308.

Ceccato P, Connor SJ, Jeanne I and Thomson MC (2005). Application of Geographical Information Systems and Remote Sensing technologies for assessing and monitoring malaria risk. *Parassitologia* 47, 81–96.

Ceccato P, Flasse S, Tarantola S, Jacquemoud S and Grégoire J (2001). Detecting Vegetation Leaf Water Content Using Reflectance in the Optical Domain. *Remote Sens. Environ.* 77 (2001): 22-33

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

Chanda E, Doggale C, Pasquale H, Azairwe R, Baba S and Mnzava A (2013). Addressing malaria vector control challenges in South Sudan: proposed recommendations. *Malar J*. 12, 59.

Chandrasekar (2010). Land Surface Water Index (LSWI) response to rainfall and NDVI using the MODIS Vegetation Index product -_LSWI_MODIS.pdf

Churcher TS, Cohen JM, Novotny J, Ntshalintshali N, Kunene S and Cauchemez S (2014).Measuring the Path Toward Malaria Elimination. *Science (New York, NY)*. 344(6189):1230-1232.

Clark TD, Greenhouse B, Njama-Meya D, Nzarubara B, Maiteki-Sebuguzi C, Staedke SG, Seto E, Kamya MR, Rosenthal PJ and Grant Dorsey (2008). Factors Determining the Heterogeneity of Malaria Incidence in Children in Kampala, *Uganda J Infect Dis*. (2008) 198 (3): 393-400

Clements ACA, Kur LW, Gatpan G, Ngondi JM, Emerson PM, Lado M, Sabasio A and Kolaczinski JH (2010). Targeting Trachoma Control through Risk Mapping: The Example of Southern Sudan. *PLoS Negl Trop Dis* 4, e799.

Clennon JA, Kamanga A, Musapa M, Shiff C and Glass GE (2010). Identifying malaria vector breeding habitats with remote sensing data and terrain-based landscape indices in Zambia. *Int J. Health Geogr* 9, 58.

Clevers JGPW (1991). Application of the WDVI in estimating LAI at the generative stage of barley. *ISPRS J. Photogramm. Remote Sens. 46*, 37–47

Cohen JM, Dlamini S, Novotny JM, Kandula D, Kunene S and Tatem AJ (2013). Rapid case-based mapping of seasonal malaria transmission risk for strategic elimination planning in Swaziland. *Malar J.* 12, 61.

Craig MH, Sharp BL, Mabaso MLH and Kleinschmidt I (2007). Developing a spatial-statistical model and map of historical malaria prevalence in Botswana using a staged variable selection procedure. *International Journal of Health Geographics* 2007**6**:44

Crippen RE (1990). Calculating the vegetation index faster. *Remote Sens. Environ*. 34, 71–73.

Curran PJ, Atkinson PM, Foody GM and Milton EJ (2000). Linking remote sensing, land cover and disease, in: Parasitology, B.-A. in (Ed.), Remote Sensing and Geographical Information Systems in Epidemiology. Academic Press, pp. 37–80.

Curran P, Windham W and Gholz H (1995). Exploring the Relationship Between Reflectance Red Edge and Chlorophyll Concentration in Slash Pine Leaves. Tree Physiology 15 (1995): 203-206

Curtis CF (2002). Should the use of DDT be revived for malaria vector control? *Biomedica*, **22:**455-461.

Dambach P, Sié A, Lacaux J-P, Vignolles C, Machault V and Sauerborn R (2009). Using high spatial resolution remote sensing for risk mapping of malaria occurrence in the Nouna district, Burkina Faso. *Glob Health Action* 2.

Dambach P, Traoré I, Becker N, Kaiser A, Sié A and Sauerborn R (2014). EMIRA: Ecologic Malaria Reduction for Africa--innovative tools for integrated malaria control. *Glob Health Action* 7, 25908.

Dantur Juri MJ, Estallo E, Almirón W, Santana M, Sartor P, Lamfri M, Zaidenberg M (2015). Satellite-derived NDVI, LST, and climatic factors driving the distribution and abundance of Anopheles mosquitoes in a former malarious area in northwest Argentina. J. *Vector Ecol. J.* Soc. Vector Ecol. 40, 36–45.

Datt B (1999). A New Reflectance Index for Remote Sensing of Chlorophyll Content in Higher Plants: Tests Using Eucalyptus Leaves. *Journal of Plant Physiology* 154 (1999): 30-36

Daughtry CST, Hunt ER Jr, Doraiswamy PC and McMurtrey JE (2005). Remote sensing the spatial distribution of crop residues. *Agron. J.* 2005, 97, 864-871

Deering DW, Rouse JW, Haas RH and Schel JA (1975). Measuring Forage Production of Grazing Units from Landsat MSS Data. In Proceedings of the 10th International Symposium of Remote Sensing of Environment, Ann Arbor, MI, USA, 6–10 October 1975; Volume II, pp. 1169–1178

Dejenie T, Yohannes M and Assmelash T (2011). Characterization of mosquito breeding sites in and in the vicinity of tigray microdams. *Ethiop J. Health Sci* 21, 57–66.

De La Rocque S, Michel V, Plazanet D and Pin R (2004). Remote sensing and epidemiology: examples of applications for two vector-borne diseases. Comp. *Immunol. Microbiol. Infect. Dis.*, Advances on some vector-borne diseases 27, 331–341.

Dhimal M, O'Hara RB, Karki R, Thakur GD, Kuch U and Ahrens B (2014). Spatio-temporal distribution of malaria and its association with climatic factors and vector-control interventions in two high-risk districts of Nepal. *Malar. J.* 13, 457.

Diboulo E, Sié A, Diadier DA, Voules DAK, Yé Y and Vounatsou P (2015). Bayesian variable selection in modelling geographical heterogeneity in malaria transmission from sparse data: an application to Nouna Health and Demographic Surveillance System (HDSS) data, Burkina Faso. *Parasit. Vectors* 8, 118.

Dlamini SN, Franke J and Vounatsou P (2015). Assessing the relationship between environmental factors and malaria vector breeding sites in Swaziland using multi-scale remotely sensed data. *Geospatial Health* 10, 302.

Fall VM, Cao Q, and Hong Y (2013). Intercomparison of Vertical Structure of Storms Revealed by Ground-Based (NMQ) and SpaceborneRadars (CloudSat-CPR and TRMM-PR). The Scientific World Journal, Hindawi Publishing Corporation, Volume Article ID 270726, 8 pages

Feachem RG (2009). Shrinking the malaria map. The Global Health Group. San Francisco.USA

Feng M, Sexton JO, Channan S and Townshend JR (2015). A global, high-resolution (30-m) inland water body dataset for 2000: first results of a topographic–spectral classification algorithm. *Int. J. Digit. Earth* 0, 1–21.

Fluet-Chouinard E, Lehner B, Rebelo L.-M, Papa F and Hamilton SK (2015). Development of a global inundation map at high spatial resolution from topographic downscaling of coarse-scale remote sensing data. *Remote Sens. Environ.* 158, 348–361.

Franke J, Gebreslasie M, Bauwens I, Deleu J, Siegert F (2015). Earth observation in support of malaria control and epidemiology: MALAREO monitoring approaches. *Geospatial Health* 10

Frumkin H, Hess J, Luber G, Malilay J, and McGeehin M (2008). Climate Change: The Public Health Response. *American Journal of Public Health*, *98*(3), 435–445.

Gamon JA and Surfus J S (1999). Assessing leaf pigment content and activity with a reflectometer. *New Phytologist* 143, 105–117

Garner R, (2015). NASA’s Goddard Space Flight Center [WWW Document]. NASA. URL http://www.nasa.gov/centers/goddard/home/index.html (accessed 6.15.15).

Garni R, Tran A, Guis H, Baldet T, Benallal K, Boubidi S and Harrat Z (2014). Remote sensing, land cover changes, and vector-borne diseases: Use of high spatial resolution satellite imagery to map the risk of occurrence of cutaneous leishmaniasis in Ghardaia, Algeria. *Infect. Genet. Evol.* 28, 725–734.

Garske T, Ferguson NM and Ghani AC (2013). Estimating Air Temperature and Its Influence on Malaria Transmission across Africa. *PLoS ONE* 8, e56487.

Gemperli A, Sogoba N, Fonjo E, Mabaso M, Bagayoko M, Briet O, Anderegg D, Liebe J, Smith T, and Vounatsou P (2006). Mapping malaria transmission in West and Central Africa. *Tropical Medicine and International Health*, 11, 1032-1046.

Giardina F, Gosoniu L, Konate L, Diouf MB, Perry R, Gaye O, Faye O, and Vounatsou P (2012). Estimating the burden of malaria in Senegal: Bayesian Zero Inflated Binomial geostatistical modelling of the MIS 2008 data. *PLoS One* 7, e32625

Giardina F, Kasasa S, Sié A, Utzinger J, Tanner M and Vounatsou P (2014). Effects of vector-control interventions on changes in risk of malaria parasitaemia in sub-Saharan Africa: a spatial and temporal analysis. Lancet Glob. Health 2, e601–e615.

Gitelson AA (2004). Wide dynamic range vegetation index for remote quantification of crop Biophysical characteristics. *J. Plant Physiol. 161*, 165–173

Gitelson AA, Stark R, Grits U, Rundquist D, Kaufman Y and Derry D (2002). Vegetation and soil lines in visible spectral space: A concept and technique for remote estimation of vegetation fraction. *Int. J. Remote Sens.* 23, 2537–2562.

Gitelson A, Merzlyak M and Chivkunova O (2001). Optical Properties and Nondestructive Estimation of Anthocyanin Content in Plant Leaves. *Photochemistry and Photobiology* 71 (2001): 38-45

Gitelson AA and Merzlyak MN (1994). Spectral reflectance changes associate with autumn senescence of Aesculus hippocastanum L. and Acer platanoides L. leaves. Spectral features and relation to chlorophyll estimation. *Journal of Plant Physiology*, 143, 286–292

Gitelson A and Merzlyak M (1998). Remote sensing of chlorophyll concentration in higher plant leaves. *Advances in Space Research* 22: 689-692

Gitelson A, Kaufman Y and Merzylak M (1996). Use of a green channel in remote sensing of global vegetation from EOS-MODIS. *Remote Sens. Environ.* 58: 289-298

Goel N and Qin W (1994). Influence of canopy architecture on various vegetation indices and LAI and Fpar: Simulation model results. *Remote sensing reviews* 10:309-347

Goetz SJ, Prince SD and Small J (2000). Advances in satellite remote sensing of environmental variables for epidemiological applications, in: Parasitology, B.-A. in (Ed.), Remote Sensing and Geographical Information Systems in Epidemiology. Academic Press, pp. 289–307.

Gong P, Pu R, Biging GS and Larrieu MR (2003). Estimation of forest leaf area index using Vegetation indices derived from Hyperion hyperspectral data. *IEEE Trans. Geosci.Remote Sens. 41*, 1355–1362

Gosoniu L, Msengwa A, Lengeler C and Vounatsou P (2012) Spatially Explicit Burden Estimates of Malaria in Tanzania: Bayesian Geostatistical Modeling of the Malaria Indicator Survey Data. *PLoS ONE* 7(5): e23966.

Gosoniu L, Veta AM and Vounatsou P (2010). Bayesian Geostatistical Modeling of Malaria Indicator Survey Data in Angola. *PLoS ONE* 5(3): e9322.

Hardisky M, Klemas V and Smart R (1983). The influences of soil salinity, growth form, and leaf moisture on the spectral reflectance of spartina alterniflora canopies. Photogrammetric Engineering and Remote Sensing 49 (1983): 77-83

Hassan AN, Beck LR and Dister S (1998). Prediction of villages at risk for filariasis transmission in the Nile Delta using remote sensing and geographic information system technologies. *J. Egypt. Soc. Parasitol.* 28, 75–87.

Hay SI, Omumbo JA, Craig MH and Snow RW (2000). Earth observation, geographic information systems and Plasmodium falciparum malaria in sub-Saharan Africa. *Adv Parasitol.* 47, 173–215.

Hay SI, Snow RW and Rogers DJ (1998). From Predicting Mosquito Habitat to Malaria Seasons Using Remotely Sensed Data: Practice, Problems and Perspectives. Volume 14, Issue 8, Pages 306–313

Hay SI and MJP (1997). The impact of remote sensing on the study and control of invertebrate intermediate hosts and vectors for disease. *Int. J. Remote Sens.* 18, 2899– 2930.

Hongrui R and Guangsheng Z (2012). Estimating senesced biomass of desert steppe in Inner Mongolia using field spectrometric data. Elsevier *Agricultural and Forest Meteorology* 161 (2012) 66–71.

Hotez PJ, Molyneux DH (2008). Tropical Anaemia: One of Africa's Great Killers and rationale for linking malaria and neglected Tropical Disease Control to Achieve a Common Goal. *PLoS Negl Trop DIS* 2(7): e270.

Hsiang MS, Hwang J, Kunene S, Drakeley C, Kandula D, Novotny J, Parizo J, Jensen T, Tong M, Kemere J, Dlamini S, Moonen B, Angov E, Dutta S, Ockenhouse C, Dorsey G and Greenhouse B (2012). Surveillance for Malaria Elimination in Swaziland: A National Cross-Sectional Study Using Pooled PCR and Serology. *PLoS One* 7(1): e29550.

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

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

Huete AR, Liu HQ, Batchily K and van Leeuwen W (1997). A comparison of vegetation indices global set of TM images for EOS-MODIS. *Remote Sens. Environ. 59*, 440–451.

Hu Y, Li R, Bergquist R, Lynn H, Gao F, Wang Q, Zhang S, Sun L, Zhang Z and Jiang Q (2015). Spatio-temporal Transmission and Environmental Determinants of Schistosomiasis Japonica in Anhui Province, China. *PLoS Negl Trop Dis* 9, e0003470.

Hugh-Jones M (1989). Applications of remote sensing to the identification of the habitats of parasites and disease vectors. *Parasitol. Today* 5, 244–251

Hunt ER Jr and Rock BN (1989). Detection of changes in leaf water content usingnear- and middle-infrared reflectance. *Remote Sens. Environ.* 30,43–54

Innes JL and Koch B (1998). Forest biodiversity and its assessment by remote sensing. *Glob. Ecol. Biogeogr. Lett.* 7, 397–419.

Jiang Z, Huete AR, Didan K and Miura T (2008). Development of a two-band enhanced vegetation index without a blue band. *Remote Sens. Environ. 112*, 3833–3845.

Jurgens C (1997). The modified normalized difference vegetation index (nNDVI)—a new index to determine frost damages in agriculture based on Landsat TM data. *Int. J. Remote Sens.*18, 3583-3594

Kalluri S, Gilruth P, Rogers D and Szczur M (2007). Surveillance of Arthropod Vector-Borne Infectious Diseases Using Remote Sensing Techniques: A Review. *PLoS Pathog.*, 26, pp. 1361–1371

Karagiannis-Voules D-A, Biedermann P, Ekpo UF, Garba A, Langer E, Mathieu E, Midzi N, Mwinzi P, Polderman AM, Raso G, Sacko M, Talla I, Tchuenté L-AT, Touré S, Winkler MS, Utzinger J and Vounatsou P (2015). Spatial and temporal distribution of soil-transmitted helminth infection in sub-Saharan Africa: a systematic review and geostatistical meta-analysis. *Lancet Infect. Dis.* 15, 74–84.

Karagiannis-Voules DA, Scholte RGC, Guimaraes LH, Utzinger J and Vounatsou P (2013). Bayesian Geostatistical Modeling of Leishmaniasis Incidence in Brazil. PLoS *Negl Trop Dis* 7(5): e2213.

Kaufman YJ and Tanré D (1992). Atmospherically resistant vegetation index (ARVI) for EOS-MODIS. *IEEE Trans. Geosci. Remote Sens. 30*, 261−270.

Kawamura K, Sempéré Y, Imai Y, Fujii M and Hayashi (1996). Water soluble dicarboxylic acids and related compounds in Antarctic aerosols, *J. Geophys. Res.* 101 (D13), 18,721–18, 728

Keevallika S, Soomereb T, Pärg R and Žukova V (2007). Outlook for wind measurement at Estonian automatic weather stations. *Estonian Acad. Sci. Eng.*, 2007, **13**, 3, 234–251

Knipling EB (1970). Physical and physiological bases for the reference of visible and near infrared radiation from vegetation. *Remote Sens. Environ. 1*, 155–159

Koita K, Novotny J, Kunene S, Zulu Z, Nyasatu N, Ghandhi M and Gosling R(2013). Targeting imported malaria through social networks: a potential strategy for malaria elimination in Swaziland. *Malar J.* 12:219.

Kulkarni MA, Desrochers RE and Kerr JT (2010). High resolution niche models of malaria vectors in northern Tanzania: a new capacity to predict malaria risk? *PloS One* 5, e9396.

Kumar V, Mangal A, Panesar S, Yadav G, Talwar R, Raut D, Singh S, Kumar V, Mangal A, Panesar S, Yadav G, Talwar R, Raut D and Singh S (2014). Forecasting Malaria Cases Using Climatic Factors in Delhi, India: A Time Series Analysis, Forecasting Malaria Cases Using Climatic Factors in Delhi, India: A Time Series Analysis. *Malar. Res. Treat. Malar. Res. Treat.* 2014, 2014, e482851.

Kunene S, Phillips AA, Gosling RD, Kandula D and Novotny JM (2011). A national policy for malaria elimination in Swaziland: a first for sub-Saharan Africa**.** *Malar J* 2011, 10:313.

Lai Y-S, Biedermann P, Ekpo UF, Garba A, Mathieu E, Midzi N, Mwinzi P, N'Goran EK, Raso G, Assaré RK, Sacko M, Schur N, Talla I, Tchuenté L-AT, Touré S, Winkler MS, Utzinger J andVounatsou P (2015). Spatial distribution of schistosomiasis and treatment needs in sub-Saharan Africa: a systematic review and geostatistical analysis. Lancet Infect. Dis. 15, 927–940.

Lakshmi V and Lakshmi V (2013). Remote Sensing of Soil Moisture, Remote Sensing of Soil Moisture. *Int. Sch. Res. Not. Int. Sch. Res. Not.* 2013, 2013, e424178.

Laneri K, Bhadra A, Ionides EL, Bouma M, Dhiman RC, Yadav RS and Pascual M (2010). Forcing Versus Feedback: Epidemic Malaria and Monsoon Rains in Northwest India. *PLoS Comput Biol* 6.

Lehner B and Döll P (2004). Development and validation of a global database of lakes, reservoirs and wetlands. *J. Hydrol.* 296, 1–22.

Li L, Bian L and Yan G (2008). A study of the distribution and abundance of the adult malaria vector in western Kenya highlands. *Int J Health Geogr* 7, 50.

Li M, Huang C, Zhu Z, Wen W, Xu D and Liu A (2009). Use of remote sensing coupled with a vegetation change tracker model to assess rates of forest change and fragmentation in Mississippi, USA. *Int. J. Remote Sens.* 30, 6559–6574.

Machault V, Yebakima A, Etienne M, Vignolles C, Palany P, Tourre YM, Guerecheau M and Lacaux J-P (2014). Mapping Entomological Dengue Risk Levels in Martinique Using High-Resolution Remote-Sensing Environmental Data. *Isprs Int. J. Geo-Inf.* 3, 1352–1371.

Mass CF, Ovens D, Westrick K and Colle BA (2002). Does increasing horizontal resolution produce more skilful forecasts? *Bull. Am. Meteorol. Soc.*, 83, 407–430.

McMichael AJ, Haines A, Slooff R and Kovats S (1996). Climate change and human health: an assessment prepared by a task group on behalf of the World Health Organization, the World Meteorological Organization and the United Nations Environmental Programme. Geneva, Switzerland, World Health Organization.

Ministry of Health (2012). National Malaria Control Programme annual report. Mbabane

Ministry of health (2010). Swaziland Malaria Elimination Strategy. Mbabane

Moonen B, Cohen JM, Snow RW, Slutsker L, Drakeley C, Smith DL, Abeyasinghe RR, Rodriguez MH, Maharaj R, Tanner M and Targett GA (2010). Operational strategies to achieve and maintain malaria elimination. *Lancet* 2010, 376:1592–1603

Myers MF, Rogers DJ, Cox J, Flahault A and Hay SI (2000). Forecasting Disease Risk for Increased Epidemic Preparedness in Public Health. *Adv. Parasitol.*

47, 309–330

Nagler PL, Daughtry CST and Goward SN (2000). Plant litter and soil reflectance. *Remote Sens. Environ.* 2000, 71, 207-215

NASA, National Aeronautics and Space Administration (2007). Reverb ECHO System. In: http://reverb.echo.nasa.gov; accessed: December, 2014.

Nájera JA, Kouznetsov RL and Delacollette C (1998). Malaria epidemics: detection and control, forecasting and prevention. Geneva: World Health Organization

Noor AM, Kinyoki DK, Mundia CW, Kabaria CW, Mutua JW, Alegana VA, Fall IS and Snow RW (2014). The changing risk of Plasmodium falciparum malaria infection in Africa: 2000–10: a spatial and temporal analysis of transmission intensity. Lancet 383, 1739–1747.

Numata I, Roberts DA, Chadwick OA, Schimel JP, Galvão LS and Soares JV (2008). Evaluation of hyperspectral data for pasture estimate in the Brazilian Amazon using field and imaging spectrometers. *Remote Sens. Environ.*, Remote Sensing Data Assimilation Special Issue 112, 1569–1583.

Oesterholt MJAM, Bousema JT, Mwerinde OK, Harris C, Lushino P, Masokoto A, Mwerinde H, Mosha FW and Drakeley CJ (2006). Spatial and temporal variation in malaria transmission in a low endemicity area in northern Tanzania. *Malar J.* 5, 98.

Ohrt C, Roberts KW, Sturrock HJW, Wegbreit J, Lee BY and Gosling RD (2015). Information Systems to Support Surveillance for Malaria Elimination. *The American Journal of Tropical Medicine and Hygiene*. 2015;93(1):145-152.

Oluwole AS, Ekpo UF, Karagiannis-Voules D-A, Abe EM, Olamiju FO, Isiyaku S, Okoronkwo C, Saka Y, Nebe OJ, Braide EI, Mafiana CF, Utzinger J and Vounatsou P (2015). Bayesian geostatistical model-based estimates of soil-transmitted helminth infection in Nigeria, including annual deworming requirements. *PLoS Negl. Trop.Dis.* 9, e0003740.

Pampana E (1969). A Textbook of Malaria Eradication London: Oxford University Press.

Papa F, Prigent C, Aires F, Jimenez C, Rossow WB and Matthews E (2010). Interannual variability of surface water extent at the global scale, 1993–2004. *J. Geophys. Res. Atmospheres* 115, D12111.

Pascual M , Ahumada JA, Chaves LF, Rodó X, and Bouma M (2006).Malaria resurgence in the East African highlands: Temperature trends revisited. vol.

103 no. 15, 5829–5834

Peng J, Liu Y and Loew A (2013). Uncertainties in Estimating Normalized Difference Temperature Index From TOA Radiances. IEEE *Trans. Geosci. Remote Sens.* 51, 2487–2497.

Penuelas J, Filella I, Biel C, Serrano L and Save R (1995). The Reflectance at the 950-970nm region as an indicator of plant water status. *Int. J. Remote Sens.* 14 (1995): 1887-1905

Pinty B and Verstraete M (1992). GEMI: a non-linear index to monitor global vegetation from satellites. Vegetation 101 15-20.

Premaratne R, Ortega L, Navaratnasinghe J, Mendis KN. Malaria elimination from Sri Lanka: What it would take to reach the goal. WHO South-East Asia *J Public Health*. 2014;3:85–9

Protopopoff N, Bortel W, Marcotty T, Herp M, Maes P, Baza D, D'Alessandro U and Coosemans M (2007). Spatial targeted vector control in the highlands of Burundi and its impact on malaria transmission. *Malar. J.* 6, 158.

Quinlan JR (1993). C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers Inc, San Francisco, CA, USA.

Reiner RC, Geary M, Atkinson PM, Smith DL and Gething PW (2015). Seasonality of Plasmodium falciparum transmission: a systematic review. *Malar. J.* 14, 343.

Riley JR (1989). Remote Sensing in Entomology. *Annu. Rev. Entomol.* 34, 247–271.Robinson T, Rogers D and Brian W (1997). Mapping tsetse habitat suitability in the common fly belt of Southern Africa using multivariate analysis of climate and remotely sensed vegetation data. *Med Vet Entomol* 1997; 11:235 – 45.

Rogers DJ, Randolph SE, Snow RW and Hay SI (2002). Satellite imagery in the study and forecast of malaria. *Nature* **415**, 710-715

Romi R, Razaiarimanga MC, Raharimanga R, Rakotondraibe EM, Ranaivo LH, Pietra V, Raveloson A and Majori G (2002). Impact of the Malaria Control Campaign (1993-1998) in the Highlands of Madagascar: Parasitological and Entomological Data**.** *Am J Trop Med*, **66:**2-6.

Romero J, Giovannini D, Franke-Arnold S, Barnett SM and Padgett MJ (2012). Increasing the dimension in high-dimensional two-photon orbital angular momentum entanglement. *Phys. Rev.* A 86, 012334.

Rondeaux G, Steven M and Baret F (1996). Optimization of Soil-Adjusted Vegetation Indices." *Remote Sens. Environ.* 55: 95-107

Roujean J and Breon F (1995). Estimating PAR absorbed by vegetation from bidirectional reflectance measurements. *Remote Sens. Environ.*51: 375-384

Rue H, Martino S, Chopin N (2009). Approximate Bayesian inference forlatent Gaussian models by using integrated nested Laplace approximations. *J R Stat Soc Ser B Stat Methodol* 71: 319–392.

Rumisha SF, Smith T, Abdulla S, Masanja H and Vounatsou P (2014). Modeling heterogeneity in malaria transmission using large sparse spatio-temporal entomological data. *Global Health Action*, [S.l.], v. 7, jun. 2014. ISSN 1654-9880 Savage HM, Rejmankova E, Arredondo-Jim'enez JI, Roberts DR and Rodr'iguez MH (1990). Limnological and botanical characterization of larval habitats for two primary malarial vectors, Anopheles albimanus and Anopheles pseudopunctipennis, in coastal areas of Chiapas State, Mexico. *J. Am Mosq Control Assoc* 6, 612–620.

Scholte RGC, Carvalho OS, Malone JB, Utzinger J and Vounatsou P (2012). Spatial distribution of Biomphalaria spp., the intermediate host snails of Schistosoma mansoni, in Brazil. Geospatial Health 6, S95–S101.

Serbin G, Daughtry CST, Hunt ER Jr, Reeves JB and Brown DJ (2009). Effects of soil composition and mineralogy on remote sensing of crop residue cover. *Remote Sens. Environ.* 2009, 113, 224-238

Serrano L, Peñuelas J and Ustin SL (2002). Remote sensing of nitrogen and lignin in Mediterranean vegetation from AVIRIS data: Decomposing biochemical from structural signals. *Remote Sens. Environ.* 81 (2002) 355–364

Sims D and Gamon J (2002). Relationships Between Leaf Pigment Content and Spectral Reflectance Across a Wide Range of Species, Leaf Structures and Developmental Stages. *Remote Sens. Environ.* 81 (2002): 337-354

Sirje Keevallika, Tarmo Soomereb, Riina Pärg and Veera Žukova (2007). Outlook for wind measurement at Estonian automatic weather stations. *Estonian Acad. Sci. Eng.*, 2007, 13, 3, 234–251

Sogoba N, Doumbia SO, Vounatsou P, Bagayoko MM, Dolo G, Traoré SF, Maïga HM, Touré YT and Smith T (2007). Malaria transmission dynamics in Niono, Mali: the effect of the irrigation systems. Acta Tropica. 101(3):232–240

Sørensen R, Zinko U and Seibert J (2006). On the calculation of the topographic wetness index: evaluation of different methods based on field observations. Hydrology and Earth System Sciences, 10, 101–112, 2006. www.copernicus.org/EGU/hess/hess/10/101/ SRef-ID: 1607-7938/hess/2006-10-101 European Geosciences Union

Srimath-Tirumula-Peddinti RCPK, Neelapu NRR and Sidagam N (2015). Association of Climatic Variability, Vector Population and Malarial Disease in District of Visakhapatnam, India: A Modeling and Prediction Analysis. PloS One 10, e0128377.

Sripada RP, Heiniger RW, White JG and Meijer AD (2006). Aerial color infrared photography for determining early in-season nitrogen requirements in corn. *Agronomy* 2006, 98, 968–977

StataCorp. 2015. *Stata Statistical Software: Release 14*. College Station, TX: StataCorp LP

Stensgaard A-S, Vounatsou P, Onapa AW, Utzinger J, Pedersen EM, Kristensen TK, Simonsen PE (2016). Ecological Drivers of Mansonella perstans Infection in Uganda and Patterns of Co-endemicity with Lymphatic Filariasis and Malaria. PLoS Negl Trop Dis 10, e0004319.

Stensgaard AS, Vounatsou P, Onapa AW, Simonsen PE, Pedersen EM, Rahbek C and Kristensen TK (2011). Bayesian geostatistical modelling of malaria and lymphatic filariasis infections in Uganda: predictors of risk and geographical patterns of co-endemicity. *Malar J* 2011, 10:298

Swaziland Malaria Indicator Survey (2010). Available at: http://www.rollbackmalaria.org/countryaction/docs/sarn/Swaziland2010MISreport.pdf (Acessed on March 2014)

Swaziland Malaria Report (2014). Mbabane: Ministry of Health

Tatem AJ and Hay SI (2004). Measuring Urbanization Pattern and Extent for Malaria Research: A Review of Remote Sensing Approaches. Journal of Urban Health: Bulletin of the New York Academy of Medicine. Vol. 81, NO. 3. 2004

The malERA Consultative Group on Drugs (2011) A Research Agenda for Malaria Eradication: Drugs. *PLoS Med* 8(1): e1000402

Thomson MC, Connor SJ, Milligan P and Flasse SP (1997). Mapping malaria risk in Africa: What can satellite data contribute? *Parasitol. Today Pers.* Ed 13, 313–318.

Tran A, Ippoliti C, Balenghien T, Conte A, Gely M, Calistri P, Goffredo M, Baldet T and Chevalier V (2013). A geographical information system-based multicriteria evaluation to map areas at risk for Rift Valley fever vector-borne transmission in Italy. Transbound. *Emerg. Dis.* 60 Suppl 2, 14–23.

Tucker CJ (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sens. Environ.* 8, 27−150.

van Deventer AP, Ward AP, Gowda PH and Lyon JG (1997). Using Thematic Mapper data to identify contrasting soil plains to tillage practices. *Photogramm. Eng. Remote Sens.* 63, 87-93.

Verpoorter C, Kutser T, Seekell DA and Tranvik LJ (2014). A global inventory of lakes based on high-resolution satellite imagery. *Geophys. Res. Lett.* 41, 2014GL060641.

Vogelmann J, Rock B and Moss D (1993). Red Edge Spectral Measurements from Sugar Maple Leaves. *Int. J. Remote Sens.* 14 (1993): 1563-1575.

Vogelmann JE, Kost JR, Tolk B, Howard S, Short K, Chen X, Associate Member, IEEE, Huang C, Pabst K and Rollins MG (2011). Monitoring Landscape Change for LANDFIRE Using Multi-Temporal Satellite Imagery and Ancillary Data. IEEE *Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 4, no. 2, June 2011.

Wang L, Hunt E, Raymond Jr, Qu, John J, Hao, Xianjun, Daughtry and Craig ST (2013). Remote sensing of fuel moisture content from ratios of narrow-band vegetation water and dry-matter indices. Publications from USDA-ARS / UNL Faculty. Paper 1175

World Health Organization (2015). World malaria report. Geneva, Switzerland: World Health Organization.

World Health Organization (2012). Disease Surveillance for Malaria Elimination: An Operational Manual. 1211 Geneva 27, Switzerland

Weiss DJ, Bhatt S, Mappin B, Boeckel TPV, Smith DL, Hay SI and Gething PW (2014). Air temperature suitability for Plasmodium falciparum malaria transmission in Africa 2000-2012: a high-resolution spatiotemporal prediction. *Malar. J.* 13, 171.

Weiss DJ, Mappin B, Dalrymple U, Bhatt S, Cameron E, Hay SI and Gething PW (2015). Re-examining environmental correlates of *Plasmodium falciparum* malaria endemicity: a data-intensive variable selection approach. *Malar J*. 201514:68

Wilschut LI, Addink EA, Heesterbeek JaP, Dubyanskiy VM, Davis SA, Laudisoit A, Begon M, Burdelov LA, Atshabar BB and de Jong SM (2013). Mapping the distribution of the main host for plague in a complex landscape in Kazakhstan: An object-based approach using SPOT-5 XS, Landsat 7 ETM+, SRTM and multiple Random Forests. *Int. J. Appl. Earth Obs. Geoinformation ITC J.* 23, 81–94.

Wolf A (2010). Using WorldView 2 Vis-NIR MSI Imagery to Support Land Mapping and Feature Extraction Using Normalized Difference Index Ratios. Unpublished report, Longmont, CO: DigitalGlobe.

Wu W (2014). The Generalized Difference Vegetation Index (GDVI) for Dryland Characterization. ISSN 2072-4292. *Remote. Sens.* 2014, *6*, 1211-1233

Wulder MA, Hall RJ, Coops NC, and Franklin SE (2004). High Spatial Resolution Remotely Sensed Data for Ecosystem Characterization. *BioScience,* 54 (6): 511-521 [0511:HSRRSD]2.0.CO;2

Xia X, Chen H, Goloub P, Zong X, Zhang W and Wang P (2013). Climatological aspects of aerosol optical properties in North China Plain based on ground and satellite remote-sensing data. *J. Quant. Spectrosc. Radiat. Transf.* 127, 12–23.

Xu H (2010). Analysis of impervious surface and its impact on urban heat environment using the normalized difference impervious surface index (NDISI) *Photogram Engin Remote Sens.* 76:557–565

Yamazaki D, Trigg MA and Ikeshima D (2015). Development of a global ~ 90 m water body map using multi-temporal Landsat images. *Remote Sens. Environ.* 10.014

Yang Z, Willis P and Mueller R (2008). Impact of band-ratio enhanced AWIFS image to crop classification accuracy. In: Proc. Pecora 17. The Future of Land Imaging Going Operational. November 18–20, Denver, Colorado, USA

Zacarias OP and Andersson M (2010). Mapping malaria incidence distribution that accounts for environmental factors in Maputo Province - Mozambique. *Malar. J.* 9, 79.

Zacarias O and Andersson M (2011). Spatial and temporal patterns of malaria incidence in Mozambique. *Malar J* 10:189.

Zarco-Tejada PJ, Rueda CA, Ustin SL (2003). Water content estimation in vegetation with MODIS reflectance data and model inversion methods. *Remote Sens. Environ.* 85, 109–124

Zeeshan M and Kim Oanh NT (2014). Assessment of the relationship between satellite AOD and ground PM_{10} measurement data considering synoptic meteorological patterns and Lidar data. *Sci. Total Environ.* 473–474, 609–618.

Zeilhofer P, Soares dos Santos E, Ribeiro ALM, Miyazaki RD and Atanaka dos Santos M, (2007). Habitat suitability mapping of *Anopheles darlingi* in the surroundings of the Manso hydropower plant reservoir, Mato Grosso, Central Brazil. *International Journal of Health Geographics*2007**6**:7

Zhao F, Huang C, Member IEEE and Zhu Z (2015). Use of Vegetation Change Tracker and Support Vector Machine to Map Disturbance Typesin Greater Yellowstone Ecosystems in a 1984–2010 Landsat Time Series. IEEE *Geosci. Remote Sens lett*. 2418159

Zhou G, Githeko AK, Minakawa N and Yan G (2010). Community-wide benefits of targeted indoor residual spray for malaria control in the Western Kenya Highland. *Malar. J.* 9, 67.

Zhou XN, Malone JB, Kristensen TK and Bergquist NR (2001). Application of geographic information systems and remote sensing to schistosomiasis control in China. Acta Trop., A Global Network on Schistosomiasis Information Systems and Control of Snail-Borne Diseases 79, 97–106.

Curriculum vitae

Name and Surname:	Sabelo Nick Dlamini
Profession:	Geographic Information Systems/Spatial Epidemiology
Nationality:	Swazi
Permanent Home Address:	P. O. Box 8070, Mbabane, Swaziland
Email :	sabelodlamini2004@yahoo.com

Academic resume

Institution (Date from-Date to)	**Degree(s) or Certificate(s) obtained**
Swiss TPH/University of Basel, Switzerland- (2013-2016)	**Doctor of Philosophy in Epidemiology (Ph.D.)** Major: Biostatistics
University of Swaziland -(2009-2012)	**Master of Science in Environmental Resources Management (MSc. ERM)** (Major: Land and Water Resources Management)
University of Swaziland -(1999-2003)	**Bachelor of Arts (BA)** (Major: Geography, Environmental Science and Planning)
University of Swaziland -(2003-2004)	**Post Graduate Certificate in Education (PGCE)** (Major: Curriculum Studies in Geography)

Professional Qualification

Institution - Date	Certificates Obtained
Environmental Systems Research Institute, South Africa, Pretoria - 2012	Desktop III: Workflows and Analysis in ArcGIS (GIS)
Martsinovsky Institute of Medical Parasitology and Tropical Medicine, Moscow, Russia - 2010	Malaria Surveillance, Monitoring and Evaluation (Malaria Control & Elimination)
Environmental Systems Research Institute, South Africa, Pretoria - 2010	Working with Spatial Analyst (GIS)
Environmental Systems Research Institute, South Africa, Durban, South Africa - 2010	Performing Analysis with ArcGIS Desktop (GIS)
Technoserve, Swaziland - 2007	Business Management Studies (Entrepreneurship skills)

Work Resume:

Designation	Institution	Period
GIS Consultant (GIS application in Malaria Control and Elimination)	World Health Organization and Southern African Development Community	Aug 2010- Feb 2013
GIS Analyst (Malaria Elimination)	National Malaria Control Programme; Government of Swaziland	Sep 2009-Sep 2013
Research and Planning Assistant (Town Planning and Community Development)	Municipal Council of Mbabane, Swaziland	Mar 2007- Aug 2009
Teacher (Geography and Development Studies)	Setsembiso Sebunye High School, Swaziland	Jan 2004- Dec 2005

Publications

Published books

1. **Dlamini Sabelo Nick** and Peter G., (2013). **Policy Integration and Stakeholder Participation in Swaziland**. Lambert Academic Publishing, Germany - ISBN: 978-3-659-37765-5.

2. Bauwens Ides, Franke Jonas, Gebraslasie Michael, **Dlamini Sabelo**, Ahmed Fethi and Vounatsou Penelope (2012). **Let`s Embrace Space; vol. 2**: Chapter 20- MALAREO: Earth Observation in Malaria Vector Control and Management. Luxemburg, France; Pages 234-240, ISBN: 978-92-79-22207-8.

Journal publications

3. **Sabelo Nick Dlamini**, Jonas Franke, Penelope Vounastou (2015). Assessing the relationship between environmental factors and malaria vector breeding sites in Swaziland using multi-scale remotely sensed data. Geospatial Health J., 10. 302

4. Sturrock HJW, Novotny JM, Kunene S, **Dlamini S**, Zulu Z, *et al*., (2013). Reactive case detection for malaria elimination: real-life experience from an ongoing program in Swaziland. Plos One 8(5): e63830.

5. Justin Cohen, **Sabelo Dlamini**, Joe Novotny, Deepika Kandula, Simon Kunene and Andrew J. Tatem (2013). Rapid case-based mapping of seasonal malaria transmission risk for strategic elimination planning in Swaziland**.** Malaria Journal. **12**:61.

6. Hsiang MS, Hwang J, Kunene S, Drakeley C, Kandula D, **Dlamini S**, *et al.*, (2012). Surveillance for Malaria Elimination in Swaziland: A National Cross-Sectional Study Using Pooled PCR and Serology. PLoS One 7(1): e29550.

Published technical reports

7. Franke, J.; Bauwens, I.; Deleu J.; de Montpellier C.; **Dlamini, S**.; Gebreslasie M.; Giardina, F.; Sierget F., and Vounatsou P. (2013). MALAREO MapAtlas- Exploring the spatial dimension of malaria and its explaining environmental factors in Southern Africa by Earth Observation. RSS- Remote Sensing Solutions, GmbH

National and International Conference/Workshop Presentations;

1. **Dlamini S.N.** (2012) Malareo end user needs assessment and GIS training workshop**.** University of Kwazulu Natal, Durban, South Africa, January 2012-Septemeber, 2013

2. **Dlamini, S.N.** (2011). Malaria Epidemiology and Key Interventions in Swaziland. Micro Planning Workshop held at Harare, Zimbabwe, 27-29 September, 2011.

3. **Dlamini, S.N.** (2010). Meteorological Observation for the Health Sector; A case of Swaziland's Malaria Elimination Campaign**:** St George Hotel, Pretoria, South Africa, November, 2010.

4. **Dlamini, S.N.** (2010). Malaria Epidemiological Surveillance: Interventions in pursuit of Elimination, held at The Martsinovsky Institute of Parasitology and Tropical Medicine, Moscow, Russia from 21 Sept- 12 Oct 2010.

Computer Skills:

Mapping/Geodatabase: ArcGIS, GRASS GIS, QGIS, ILWIS, EpiInfo

Statistical Modeling: WinBUGS, STATA

Spread Sheets/Databases: Microsoft Access, Microsoft Excel